

Lecture Notes: PH279

Fourth Semester Introductory Physics

R. Todd Lines
Brigham Young University-Idaho

Created in Scientific Workplace™

Copyright ©2020 by Author

Preface: How PH279 works

It seems that professors always have an agenda for each class. This class is no exception. Not only will we learn about the universe, we will also try to get you ready for your junior and senior level physics classes. Here are my goals for this class:

1. To help you understand relativity and quantum mechanics
2. To complete the picture of physics that you started with PH121.
3. To toughen up your math skills up to prepare you for junior level classes
4. To toughen up your reading skills so to prepare you for the rest of physics and life in STEM

Relativity and Quantum mechanics are not intuitive like PH121.

- Trying to do as little as possible won't meet that goal
- If you are over-scheduled, you will have to change how you work to become more efficient.
- Not doing the reading, the examples, and the homework is likely to leave you not understanding the physics. (and to bad test scores)
- How PH279 works

We are now in a class that is just for physics majors (and minors). PH121 through PH220 were designed to teach physics, but often were designed to help our fellow scientists and engineers along the way to their majors. Now it is just us. The author of our text book and your professor don't have to care about whether the students can build bridges after the class is done. We get to care about if you better understand the universe and how it works. Understanding is now the goal. This means your professor may act a little differently, and your professor may ask you to act a little differently. Let's look at some of those potential changes.

Problem Solving

We will still use the physics problem solving techniques. Some of your professors will have called this the problem solving method, some D3BSNUB, and some might not have given it a name. Here are the basic parts.

Problem Solving Process

Process Step	Purpose	Value if Present	Value if Absent
Label the problem with chapter and problem number	This is essential if I am to figure out what problem to grade	0	0 to -25
Restate the problem in your own words. One line may do! List any assumptions you are making. You may wish to classify the problem in your problem statement	Most major mistakes come from misinterpreting the problem. This step asks you to slow down and determine what the problem really is asking	1	-1
Identify the type of problem. Pick a strategy, is it a Newton's Second Law problem? Is it a rotational problem with constant rotational acceleration?	By identifying the type of problem it is, you are more likely to be able to find the right equations to get started and to successfully complete the problem. The idea is to pause and see what approach will likely be involved in forming the solution. This is not as hard as it sounds. If the word "force" appears, you can reasonably assume Newton's Second Law will be involved, for example.	2	-2
Draw a picture , label items, define coordinate systems, etc. <i>This picture should be a visual restatement of the problem. View this as a graphical restatement of the problem.</i>	Many mistakes happen because we do not have a clear picture of the problem. This step may save hours of grief. Also, many physics problems will have different symbolic answers because of the freedom to choose coordinate systems, etc. Drawing a diagram gives the reader the ability to understand your vision of the problem.	5	-5
Define variables used, Identify known and unknown quantities	Choose reasonable names for physical quantities, and let me know what they are. Don't forget to include units.	2	-2
List basic equations that apply to the problem	This step gives you a firm starting place.	2	-2
Solve the problem algebraically starting from the basic equations,	This is the heart of the solution. The symbolic answer tells you the relationships between physical quantities.	10	-10
Determine numerical answer	The specific numerical answer is not the point of doing the problem in this class, but is a great indicator that you have succeeded in understanding the physics.	1	-1
Check units. If you have not done the algebra on the units earlier, do it here.	Many mistakes are evident in a units analysis. It is a good habit to always check units.	1	-1
Determine if the numerical answer is reasonable. Indicate if you are comfortable with the result, if you have little experience with the result and can't tell if it is reasonable, or if it is not reasonable, but you don't know why (or else you would fix it).	From your understanding of the physics, state whether the answer is reasonable. For example, if you are calculating the mass of a ping pong ball, and get an answer that is many times the mass of the earth, you should note that there may be a problem even if you do not know where you went wrong.	1	-1 to -25
Total Possible		25	

The change you might see is that we will become ever more unconcerned about the number at the end. In PH332 and PH333 often there won't be a number at the end. And we will see less numbers and more equations as answers. This is just as it should be. Equations tell us about how the universe works. Numbers usually don't. So if our goal is to understand the universe, equations are likely to be the goal of a problem rather than a numeric answer. For some of the problems in our class, drawing a picture will be harder. But we should draw the situation if we possibly can.

Reading

I was told by a former student who is now a professor (I won't give a name, but his

initials are Matt Zachreson) that he got through my PH123 class without reading the text book at all. This won't work from here on out. The definition of "read" that I will use in this and our higher level physics classes is as follows.

1. Turn to the assigned section of the book.
2. Read the words on the page until you get to a derivation or an example
3. If it is a derivation, work through the derivation, filling in all the missing parts that the author didn't give you
4. If it is an example, work through the example and make sure you get the same answer as the author (or have a good reason why you are right and why the author might be wrong).
5. Repeat this process until you are done with the assigned sections.

I didn't discover that this is what my professors meant by "read" until I had spent a year or so struggling in physics classes. They never told us. They just expected this of us. I am telling you. And I will motivate you as well. The examples in the book and the derivations will show up as homework problems. Let's look at our first chapter to see how this will go.

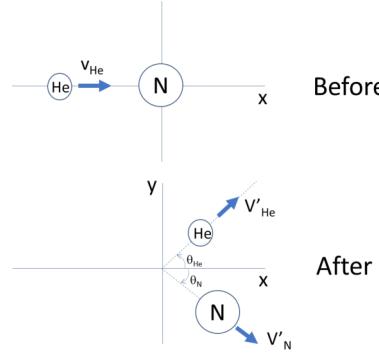
In the first chapter there is a snazzy diagram about the Cassini Interplanetary Trajectory that the author understands and thinks is cool. I looked at this (and thought it was cool). I read the caption in the blue box below the figure. It told me that I needed to go beyond Newtonian physics to make cool space probe trajectories which intrigued me. On the next page there was a brief history of what we call "classical physics" and what we call "modern physics." I read them. On the next page I read section 1.1 and was relieved to find equations that I knew. And on the next page was an example. I worked the example, and here is what I got.

Example 1.1: A helium atom ($m = 6.6456 \times 10^{-27} \text{ kg}$) moving at a speed of $v_{He} = 1.518 \times 10^6 \text{ m/s}$ collides with an atom of nitrogen ($m = 2.3253 \times 10^{-26} \text{ kg}$) at rest. After the collision, the helium atom is found to be moving with a velocity of $v'_{He} = 1.99 \times 10^6 \text{ m/s}$ at an angle of $\theta_{He} = 78.75^\circ$ relative to the direction of the original motion of the helium atom.

- a) Find the velocity (magnitude and direction) of the nitrogen atom after the collision.
- b) Compare the kinetic energy before the collision with the total kinetic energy of the atoms after the collision.

(Subsubsubsubsection head:)Solution

a) The word “collision” is a dead give away that this is a conservation of momentum type problem. I always draw before and after pictures for conservation of momentum. That is, before and after the collision.



We know the following:

$$\begin{aligned}
 m_{He} &= 6.6465 \times 10^{-27} \text{ kg} \\
 m_N &= 2.3253 \times 10^{-26} \text{ kg} \\
 v_{He} &= 1.518 \times 10^6 \text{ m/s} \\
 v_N &= 0 \\
 v'_{He} &= 1.199 \times 10^6 \text{ m/s} \\
 \theta_{He} &= 78.75^\circ
 \end{aligned}$$

Our basic equations are

$$\begin{aligned}
 P_i &= P_f \\
 p &= mv
 \end{aligned}$$

and we can write this out as

$$\begin{aligned}
 \vec{P}_i &= \vec{P}_f \\
 \vec{p}_{iHe} + \vec{p}_{iN} &= \vec{p}_{fHe} + \vec{p}_{fN}
 \end{aligned}$$

and of course we split this into components

$$\begin{aligned}
 p_{x,iHe} + p_{x,iN} &= p_{x,fHe} + p_{x,fN} \\
 p_{y,iHe} + p_{y,iN} &= p_{y,fHe} + p_{y,fN}
 \end{aligned}$$

then

$$\begin{aligned}
 m_{He}v_{He} + 0 &= m_{He}v'_{He} \cos \theta_{He} + m_Nv'_N \cos \theta_N \\
 0 + 0 &= m_{He}v'_{He} \sin \theta_{He} + m_Nv'_N \sin \theta_N
 \end{aligned}$$

It is worth noting what we know

$$\begin{aligned}\underline{m_{He}v_{He}} &= \underline{m_{He}v'_{He} \cos \theta_{He}} + \underline{m_N v'_N \cos \theta_N} \\ 0 &= \underline{m_{He}v'_{He} \sin \theta_{He}} + \underline{m_N v'_N \sin \theta_N}\end{aligned}$$

This looks like two equations and two unknowns. We should be able to solve for this. Of course there is more than one way to do this. Let's solve for $v'_N \cos \theta_N$ and $m_N v'_N \sin \theta_N$

$$\begin{aligned}\underline{m_{He}v_{He}} - \underline{m_{He}v'_{He} \cos \theta_{He}} &= \underline{m_N v'_N \cos \theta_N} \\ -\underline{m_{He}v'_{He} \sin \theta_{He}} &= \underline{m_N v'_N \sin \theta_N}\end{aligned}$$

Let's try dividing the two equations

$$\frac{-\underline{m_{He}v'_{He} \sin \theta_{He}}}{\underline{m_{He}v_{He}} - \underline{m_{He}v'_{He} \cos \theta_{He}}} = \frac{\underline{m_N v'_N \sin \theta_N}}{\underline{m_N v'_N \cos \theta_N}}$$

this gives

$$\frac{-\underline{m_{He}v'_{He} \sin \theta_{He}}}{\underline{m_{He}v_{He}} - \underline{m_{He}v'_{He} \cos \theta_{He}}} = \frac{\sin \theta_N}{\cos \theta_N} = \tan \theta_N$$

which is a mess, but we know all the parts except θ_N . To fit this on the page I am going to have to break it up a bit.

$$\begin{aligned}-\underline{m_{He}v'_{He} \sin \theta_{He}} &= -(6.6465 \times 10^{-27} \text{ kg}) (1.199 \times 10^6 \text{ m/s}) \sin(78.75^\circ) \\ &= -7.8161 \times 10^{-21} \frac{\text{m}}{\text{s}} \text{ kg} \\ \underline{m_{He}v_{He}} - \underline{m_{He}v'_{He} \cos \theta_{He}} &= (6.6465 \times 10^{-27} \text{ kg}) \left(1.518 \times 10^6 \frac{\text{m}}{\text{s}} \right) \\ &\quad - (6.6465 \times 10^{-27} \text{ kg}) \left(1.199 \times 10^6 \frac{\text{m}}{\text{s}} \right) \cos(78.75^\circ) \\ &= 8.5347 \times 10^{-21} \frac{\text{m}}{\text{s}} \text{ kg}\end{aligned}$$

so then

$$\begin{aligned}\theta_N &= \tan^{-1} \left(\frac{-\underline{m_{He}v'_{He} \sin \theta_{He}}}{\underline{m_{He}v_{He}} - \underline{m_{He}v'_{He} \cos \theta_{He}}} \right) \\ &= \tan^{-1} \left(\frac{-7.8161 \times 10^{-21} \frac{\text{m}}{\text{s}} \text{ kg}}{8.5347 \times 10^{-21} \frac{\text{m}}{\text{s}} \text{ kg}} \right) \\ &\quad (6.6465 \times 10^{-27} \text{ kg}) \left(1.518 \times 10^6 \frac{\text{m}}{\text{s}} \right) - (6.6465 \times 10^{-27} \text{ kg}) \left(1.199 \times 10^6 \frac{\text{m}}{\text{s}} \right) \cos(78.75^\circ) \\ &= -0.74148 \text{ rad} \\ &= -42.5^\circ\end{aligned}$$

Now we can use one of the two equations to solve for v'_N . I want the shorter of the two

$$\begin{aligned} \frac{-m_{He}v'_{He} \sin \theta_{He}}{m_N \sin \theta_N} &= m_N v'_N \sin \theta_N \\ \frac{-m_{He}v'_{He} \sin \theta_{He}}{m_N \sin \theta_N} &= v'_N \end{aligned}$$

so

$$\begin{aligned} v'_N &= \frac{- (6.6465 \times 10^{-27} \text{ kg}) (1.199 \times 10^6 \text{ m/s}) \sin (78.75^\circ)}{(2.3253 \times 10^{-26} \text{ kg}) \sin (-42.48333^\circ)} \\ &= 497693.22 \frac{\text{m}}{\text{s}} \end{aligned}$$

which is what they asked for

$$\begin{aligned} v'_N &= 497693.22 \frac{\text{m}}{\text{s}} \frac{\text{m}}{\text{s}} \\ \theta_N &= -42.48333^\circ \end{aligned}$$

Note the annoying inconsistent use of the primes!

b) We know the kinetic energy is given by

$$k = \frac{1}{2}mv^2$$

for the before case we have

$$\begin{aligned} k_{total,i} &= \frac{1}{2}m_{He}v_{He_i}^2 + \frac{1}{2}m_Nv_{N_i}^2 \\ &= \frac{1}{2}m_{He}v_{He_i}^2 + 0 \\ &= \frac{1}{2}(6.6465 \times 10^{-27} \text{ kg})(1.518 \times 10^6 \text{ m/s})^2 \\ &= 7.6578447 \times 10^{-15} \frac{\text{m}^2}{\text{s}^2} \text{ kg} \\ &= 7.658 \times 10^{-15} \text{ J} \end{aligned}$$

For the after case

$$\begin{aligned} k_{total,f} &= \frac{1}{2}m_{He}v_{He_f}^2 + \frac{1}{2}m_Nv_{N_f}^2 \\ &= \frac{1}{2}m_{He}v'_{He_f}^2 + \frac{1}{2}m_Nv'_{N_f}^2 \\ &= \frac{1}{2}(6.6465 \times 10^{-27} \text{ kg})(1.199 \times 10^6 \text{ m/s})^2 \\ &\quad + \frac{1}{2}(2.3253 \times 10^{-26} \text{ kg})\left(4.977 \times 10^5 \frac{\text{m}}{\text{s}}\right)^2 \\ &= 7.6573746 \times 10^{-15} \frac{\text{m}^2}{\text{s}^2} \text{ kg} \\ &= 7.657 \times 10^{-15} \frac{\text{m}^2}{\text{s}^2} \text{ kg} \end{aligned}$$

To within the accuracy of our numbers, we seem to have conservation of kinetic energy as well. But note that the book got another decimal place further than we did. I

don't know how. I used Scientific Workplace, a TI calculator, and Excel to check my numbers. All agreed with what I have here. Now it is time to check with my homework group to see if they got what the book got!

There weren't any words between this example and the next so I did the next as well. Both show up as homework problems in our first homework assignment.

After Example 1.2 there were words with equations, but these words and equations were intended as a review, they didn't come up with anything new.

On the next page there is a carefully hidden example on the bottom of the page. It is a transform of the initial velocities of Example 1.1 into another reference frame. I did this as well.

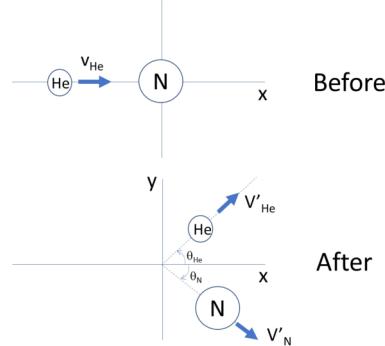
a) Find the center of mass velocity for the initial state of the particles in Example 1.1.

b) Transform the initial velocities of the helium atom and the Nitrogen atom into the reference frame of the center of mass of the system.

ANSWER:

This is a relative velocity problem

From Example 1.1 we know



and

$$\begin{aligned}
 m_{He} &= 6.6465 \times 10^{-27} \text{ kg} \\
 m_N &= 2.3253 \times 10^{-26} \text{ kg} \\
 v_{He} &= 1.518 \times 10^6 \text{ m/s} \\
 v'_{He} &= 1.199 \times 10^6 \text{ m/s} \\
 \theta_{He} &= 78.75^\circ \\
 v_N &= 0
 \end{aligned}$$

For part a) we want the center of mass velocity for the initial state of the particles

We “remember” that the equation for center of mass is

$$x_{CL} = \frac{x_{He}m_{He} + x_Nm_N}{(m_{He} + m_N)}$$

We will find that the word “remember” means we look it up again, but hopefully you have your old physics text book and can just look it up. Once we have our center of mass equation, we can move on. We know that

$$v = \frac{dx}{dt}$$

so we can find the center of mass speed with respect to the laboratory by taking a derivative

$$\frac{dx_{CL}}{dt} = \frac{\frac{dx_{He}}{dt}m_{He} + \frac{dx_N}{dt}m_N}{(m_{He} + m_N)}$$

because the masses are not changing. This is simply

$$v_{CL} = \frac{v_{He}m_{He} + v_Nm_N}{(m_{He} + m_N)}$$

Notice that our author left it for us to figure this out. The numeric value is

$$\begin{aligned}
 v_{CL} &= \frac{(1.518 \times 10^6 \text{ m/s}) (6.6465 \times 10^{-27} \text{ kg}) + 0}{(6.6465 \times 10^{-27} \text{ kg} + 2.3253 \times 10^{-26} \text{ kg})} \\
 &= 337443.34 \frac{\text{m}}{\text{s}}
 \end{aligned}$$

For part b) we want to transform the initial velocities of the helium atom and the Nitrogen atom into the reference frame of the center of mass of the system.

We can use the standard Galilean reference frame transform that we “remember”

$$v_{bA} = v_{bB} + v_{BA}$$

For the helium atom

$$\begin{aligned}
 v_{HeC} &= v_{HeL} - v_{LC} \\
 &= 1.518 \times 10^6 \text{ m/s} - 337443.34 \frac{\text{m}}{\text{s}} \\
 &= 1.1805567 \times 10^6 \frac{\text{m}}{\text{s}}
 \end{aligned}$$

$$\begin{aligned}
v_{NC} &= v_{NL} - v_{LC} \\
&= 0 - 337443.34 \frac{\text{m}}{\text{s}} \\
&= 3.3744334 \times 10^5 \frac{\text{m}}{\text{s}} \\
&= 0.33744334 \times 10^6 \frac{\text{m}}{\text{s}}
\end{aligned}$$

U check

R check

There are pages of review with some new suggested notations that follow. I read these. Section 1.3 was more difficult. It starts with a result that we will get in PH412

$$N(E) = \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_B T}} \quad (0.1)$$

You have to love a mysterious equation from a class you haven't taken that has $\sqrt{\pi}$ in it. And since this is in the "review" chapter we learn that our author has a different definition of "review" than we are used to. The author tells us this equation is the number of molecules as a function of energy such that $N(E)dE$ is the number of molecules that have energies that are within a small delta energy (dE) centered about the energy E . We can write this as

$$dN = N(E) dE$$

and then the total number of molecules, N would have to be the integral of this quantity over all energies from zero energy to infinite energy

$$N = \int_0^\infty dN = \int_0^\infty \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE$$

and if you read carefully you will notice that the author totally expects us to do this nasty integral because the author hints about how to do it. He says we should end up looking up the form

$$\int_0^\infty x^{\frac{1}{2}} e^{-x} dx = \frac{1}{2} \sqrt{\pi}$$

Often we will look up integrals in tables of integrals. You can solve them yourself, if you want to. But more often we will work to get our integral to look like one in a table, and just use the table result. Apparently we better do this now. We need to make our integral look like this last one. We can do that

$$\begin{aligned}
N &= \int_0^\infty \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} \frac{E^{\frac{1}{2}}}{\sqrt{k_B T}} e^{-\frac{E}{k_B T}} dE \\
&= \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right) \int_0^\infty \left(\frac{E}{k_B T} \right)^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE
\end{aligned}$$

and define

$$\begin{aligned}x &= \frac{E}{k_B T} \\dx &= \frac{1}{k_B T} dE\end{aligned}$$

$$\begin{aligned}N &= \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right) k_B T \int_0^{\infty} \left(\frac{E}{k_B T} \right)^{\frac{1}{2}} e^{-\frac{E}{k_B T}} \frac{dE}{k_B T} \\&= \frac{2N}{\sqrt{\pi}} \int_{x=0}^{x=\infty} (x)^{\frac{1}{2}} e^{-x} dx \\&= \frac{2N}{\sqrt{\pi}} \left(\frac{1}{2} \sqrt{\pi} \right) \\&= N\end{aligned}$$

Which doesn't seem to be super instructive, but it sets us up for the next derivation, showing that

$$E_{av} = \frac{3}{2} k_B T$$

and the next one that let's us find the number of molecules that have energies in the interval between E_1 and E_2 .

You might be saying, "Whoa there hoss, this kind of reading is a *lot more work* than just reading the words." And you would be right. But this kind of reading gives you the practice with the equations and the practice with the ideas that you will need to be successful in understanding how the universe works.

Homework

You have been doing homework for a long time. You know how to do homework. But even this is going to change. You should form homework groups. You should start early. You should try the problems (many you will try as you read) and then collaborate to finish the assignment. This is what real physicists do. Where your group is stuck, you should find your instructor or one of the other professors to get help. The walk-in tutoring lab is likely not going to be a help anymore. Normally you would get a code to the Major's room. There are tables there (and into the labs across the hall) that you can use for your groups to meet. This semester we will have a discussion board to help you collaborate, but you might build your own group and your own way to communicate. It is increasingly important to do the homework. Relativity and quantum mechanics are often counter intuitive. Banding together is the way to make sure you are thinking right. This is what physicists do.

- Work efficiently.
- Start early—start a homework assignment the day it is given, not the day it is due.
- Use professors and colleagues for help (the walk-in lab is not much use from now on).

Exams

Exams will still be exams. Normally they would in the testing center. Generally this is one of the last physics classes that uses the testing center. In the upper division classes exams are less likely to be short and are more likely to be take home exams. This semester we will likely have mostly take home exams.

What will we learn?

A little about everything! We will build the universe (using the standard model).

Feedback

This class will be more collaborative than many past classes might have been. Our job is to work together to make sure we all learn this material

Acknowledgments

A big thanks to all the students of my Modern Physics class Fall 2019 that suffered through the initial draft of these notes, and the brave students of 2020 than not only survived my class, but the COVID-19 pandemic. Thanks for all the feedback and suggestions.

BYU-I

R. Todd Lines.

Contents

Preface: How PH279 works	iii
Acknowledgments	xiii
1 Mechanics and Thermo: Where we left off in PH121 and PH123	3
Fundamental Concepts in the Lecture	3
Galilean Relativity	3
Molecular Model of an Ideal Gas	8
What to think of this “review”	22
2 Classical Relativity	23
Fundamental Concepts in the Lecture	23
Relative Motion	23
3 The Postulates of Special Relativity	35
Fundamental Concepts in the Lecture	35
The Two (3) Postulates	35
A more detailed look at simultaneity	40
4 Special Relativity	47
Fundamental Concepts in the Lecture	47
Time dilation	47
Length Contraction	54

The Invariant Interval	58
Lorentz Transformation	60
5 Using the Lorentz Transformation (and Time Dilation and Length Contraction)	71
Fundamental Concepts in the Lecture	71
How much do leading clocks lag trailing clocks?	71
Putting it all together	74
6 Special Relativity Dynamics	83
Fundamental Concepts in the Lecture	83
Spacetime Diagrams and the Twin Paradox	83
Velocity Transforms	96
7 Relativistic Conservation Laws	105
Fundamental Concepts in the Lecture	105
Relativistic momentum	108
Relativistic Kinetic Energy	113
Universal Speed Limit	118
8 Quantization: The Beginnings of Modern Physics	121
Fundamental Concepts in the Lecture	121
Waves and Fields Review	121
Quantum Waves: Photoelectric Effect	127
Blackbody Radiation	131
9 Photons and Matter	141
Fundamental Concepts in the Lecture	141
The discovery of X-rays	141

Compton Effect	144
Pair Production	148
Wave-Particle Duality of Light	149
Wave-particle Duality of Matter	152
10 De Broglie Waves and Uncertainties	155
Fundamental Concepts in the Lecture	155
Particle Interference	155
Uncertainty Relationships	158
11 Working with Wave Packets	167
Fundamental Concepts in the Lecture	167
Building the mathematical concepts behind wave packets	167
Motion of the wave packet	175
12 Wave Packets and the Schrödinger Equation	185
Fundamental Concepts in the Lecture	185
Wave Packets and Boundaries	185
Particle in a box	196
the Schrödinger equation	199
13 Using the Schrödinger Equation	207
Fundamental Concepts in the Lecture	207
Constant Potential Energy	207
Free Particle	209
Infinite potential well revisited	212
Multi-dimensional infinite potential wells, and finite potential wells	222

14 Barriers, Oscillators, and Early Hints about the Structure of the Atom 225

Fundamental Concepts in the Lecture	225
Barriers and Tunneling	225
Harmonic Oscillator	233
Ancient thoughts on atomic structure	238

15 The Bohr Model of the Atom 243

Fundamental Concepts in the Lecture	243
Atomic Spectra give a clue	243
Bohr Misses the Clue, but hits on quantization	246
Extension to “Hydrogen-Like” Atoms	252
Bohr Model - What doesn’t work	254
De Broigne Discovers the Clue	255

16 Quantum Model for the Atom 257

Fundamental Concepts in the Lecture	257
One Dimensional first look at Quantum Hydrogen	257
Fixing the angular momentum	261
Hydrogen wave functions	265

17 Details of the Quantum Model for the Atom 275

Fundamental Concepts in the Lecture	275
Radial probability densities	275
Angular probability densities	278
Angular momentum and space quantization	280

18 Hydrogen Atoms in Magnetic Fields 285

Fundamental Concepts in the Lecture	285
Stern-Gerlach Experiment	286
19 Many Electron Atoms	295
Fundamental Concepts in the Lecture	295
Pauli exclusion principle	295
Screening in Many-Electron Atoms	297
Optical Transitions	300
The Periodic Table of the Elements	300
20 Electron Inner Structure	307
Fundamental Concepts in the Lecture	307
Addition of angular momentum	307
Transitions involving inner electrons	314
Moseley's Law	315
Stimulated Emission	318
21 Molecular Bonds and their Structure	325
Fundamental Concepts in the Lecture	325
Molecular Wave Functions and the H_2^+ Molecule	325
The H_2 Molecule	331
Covalent bonds	332
Ionic Bonds	340
22 Molecular Spectra, Vibrations and Rotations	343
Fundamental Concepts in the Lecture	343
Molecular oscillators	343
Rotational molecular states	349

Vibration and Rotation 352

23 Statistical approach for many atoms 357

Fundamental Concepts in the Lecture 357

More than two atoms 357

Classical Statistics 357

Beginnings of Quantum Statistics 368

Distribution Functions and Density of States Functions 371

Gas of Particles Density of States 377

Photon gas density of states 381

24 Probability Distribution Functions and Quantum Statistics 383

Fundamental Concepts in the Lecture 383

Distribution Functions 384

Classical Maxwell-Boltzmann distribution function 388

How do we know when to use quantum statistics? 396

Examples of finding energy distribution functions 397

25 More Probability Distribution Functions and Quantum Statistics 403

Fundamental Concepts in the Lecture 403

26 Solids and Solid State Physics 415

Fundamental Concepts in the Lecture 415

Structure and bonding of elemental solids 415

Heat Capacity of Solids 421

Quantum Conduction 426

Superconductivity 433

27 Examples of Solid State Physics, and Introduction to the Nucleus 437

- Fundamental Concepts in the Lecture 437
- Semiconductors 437
- Quantum Magnetic Materials 450
- Matter and Energy 456
- Nature of the Nucleus 457

28 Nuclear Binding Energy and Radioactivity 461

- Fundamental Concepts in the Lecture 461
- Binding Energy 461
- Nuclear Stability 465
- Nuclear energy states 467
- Separation energy 469

29 Radioactive Decay 473

- Fundamental Concepts in the Lecture 473
- Radioactive decay basics 473
- Details of the decay processes 478
- Mathematically describing radioactive decay 486

30 Artificial Nuclear Reactions 491

- Fundamental Concepts in the Lecture 491
- Producing nuclear reactions 491
- Cross Sections 495
- Radioisotope production 501

31 Nuclear Energy Reactors 507

- Fundamental Concepts in the Lecture 507

Nuclear Fission	507
Fusion	516
32 Nucleosynthesis and the Particle Zoo	523
Fundamental Concepts in the Lecture	523
Making the atoms	523
Applications of Nuclear Physics	526
Fundamental forces	530
Antimatter	533
Classification Schemes	534
33 Particle Conservation Laws and particle detectors	537
Fundamental Concepts in the Lecture	537
Conservation laws	537
Detecting particles	541
Detection strategies	546
Energy-momentum in particle experiments	550
34 Quantum Chromo Dynamics in a Nut Shell	553
Fundamental Concepts in the Lecture	553
Momentum and Energy in particle interactions	553
Quarks	556
The Standard Model	565
35 Cosmology	569
Fundamental Concepts in the Lecture	569
Expansion of the universe and Hubble's law	569
Dark Matter	582

36 General relativity 585

Fundamental Concepts in the Lecture 585

General relativity 585

Stellar Evolution 597

37 General Relativity and Cosmology 603

Fundamental Concepts in the Lecture 603

Size and Shape of the Universe 603

A very brief history of time 607

38 Current Cosmology 613

The universe today 613

Curvature of the Universe 616

Does it all work? 617

A Fourier Series Review 619

Music and Non-sinusoidal waves 619

Fourier Series: Mathematics of Non-Sinusoidal Waves 624

Fourier Analysis 625

Back to our Square Wave 628

B Tables 631

1 Mechanics and Thermo: Where we left off in PH121 and PH123

Sections 1.1 through 1.2

Fundamental Concepts in the Lecture

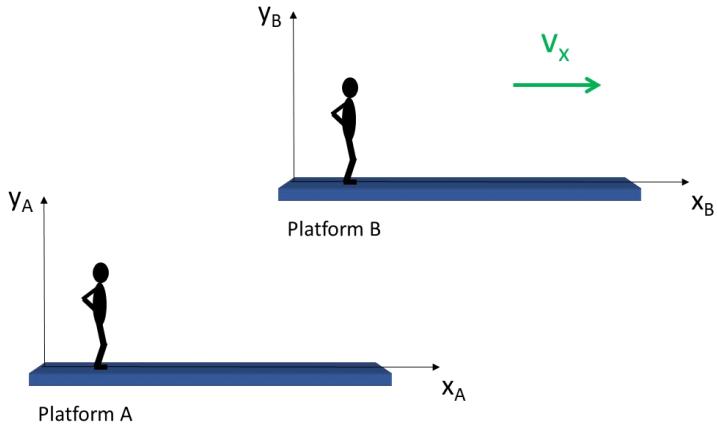
- Galilean Newtonian Relativity
- Add in the idea of forces
- Tie the idea forces to potential energy as a way of solving problems
- Added in electric and magnetic forces and their potential energies
- Breakdown of the concept of space
- Breakdown of the concept of time.
- Classical Molecular Energies
- Equapartition of energy in molecules
- Heat capacities and ideal gasses
- Theory, experiment, and law

Galilean Relativity

Way back in your introductory mechanics class (PH121) we learned about reference systems and relative motion. If you took PH121 from me, you used the Tutorials in Physics and relative motion kept coming back in the later Tutorials. If you are normal, you didn't like that. But there was a purpose for all that relative motion stuff. Relative motion is the underlying concept behind Einstein's theories of relativity. Let's review relative motion and the Galilean transformation before we dive into Einstein's work.

Galilean Transformation

The Galilean Transformation is a name we give to a set of equations that allow us to consider motion on one reference frame from the viewpoint of someone on a different reference frame. Back in PH121 we used simple notation and in that notation our situation might look like this



Hopefully you used a set of subscripts to keep things strait. For example, we could write v_x as v_{BA} , the speed of frame B as viewed from frame A . Then the person on platform A might be observing the motion of the person on platform B and find the position of person b to be.

$$x_A = x_B + v_{BA}t$$

$$y_A = y_B$$

$$z_A = z_B$$

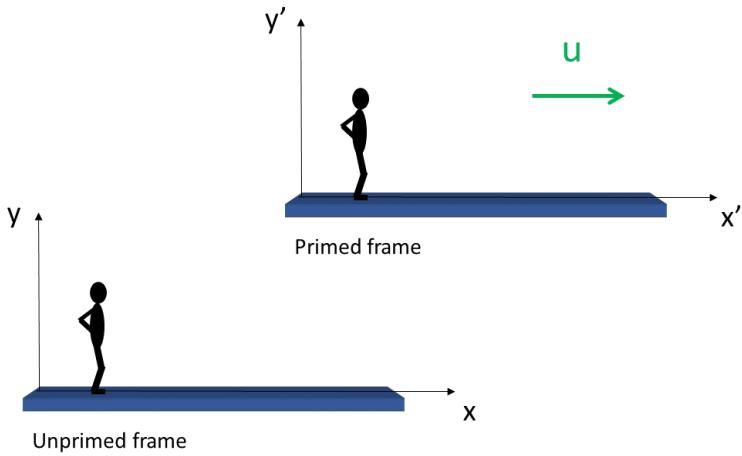
where the subscript A indicates a measured position in the A -frame and the subscript B indicates a measured position in the B -frame. Note that the y and z components don't change because the relative motion is all in the x -direction. The velocities also transform. Suppose something is moving in frame B as viewed by a person in frame B with speed v_{bB} , but we view the object from frame A . The speed the person in frame A sees for this object is v_{bA} . And we know from PH121 that we just add the velocities in the direction of motion.

$$v_{bAx} = v_{bBx} + v_{BA}$$

$$v_{bAxy} = v_{bBy}$$

$$v_{bAz} = v_{bBz}$$

Note how nicely the subscripts tell us what is being viewed from what frame. But in Einstein's relativity there are some traditional ways we write this situation. You will likely see this in other textbooks. Let's see the same situation again, but with a new notation.



We can see that our relative speed v_{BA} is now written as u . And we can see that instead of A and B labels, one reference frame has prime marks on the axis labels. We call this reference frame the “primed frame” and the other reference frame the “unprimed frame.” A location in the unprimed frame also has a location in the primed frame. The difference will be how far the primed frame has moved from where we started the experiment to some later time t when we wish to know the location. In the primed frame, the location (x, y, z) would be given by

$$x' = x - ut$$

$$y' = y$$

$$z' = z$$

And the speed of an object moving in the primed frame would have speed seen from the unprimed frame given by.

6 Chapter 1 Mechanics and Thermo: Where we left off in PH121 and PH123

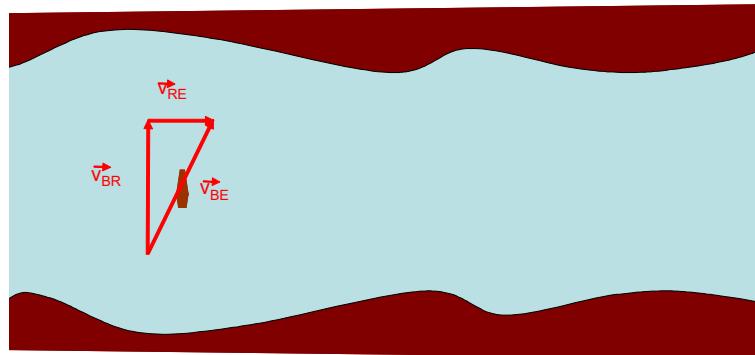
$$\begin{aligned} v_x &= v'_x + u \\ v_y &= v'_y \\ v_z &= v'_z \end{aligned}$$

Of course just like before, in these equations we have assumed the velocity u is all in the x -direction. This system of notation requires us to remember much more in our heads. I will try to use subscripts, but what you read in other books will often use the prime notation. I suppose you could use both, if you want to keep the primes so you can follow along with other books.

$$\begin{aligned} v_{bAx} &= v'_{bBx} + v_{BA} \\ v_{bAxy} &= v'_{bBy} \\ v_{bAz} &= v'_{bBz} \end{aligned}$$

Let's do an example from PH121.

Problem Statement: You have herd the fishing is great in Idaho and so you rent a boat and take your FHE group out fishing on the river. Your heading is due north and your speed with respect to the water is $v_{BR} = 10.0 \text{ km/h}$. The river goes by to the east with a speed with respect to the Earth (shore) is $v_{RE} = 5.00 \text{ km/h}$. How fast would a forest ranger see the boat if the ranger is standing on the shore?



Let's choose the y axis to be positive in the northern direction. The x axis will be positive in the eastern direction.

Variables:

$\vec{v}_{BR} = 10.0 \text{ km/h} \hat{j}$	Velocity of the boat with respect to the river
$\vec{v}_{RE} = 5.00 \text{ km/h} \hat{i}$	Velocity of the river with respect to the shore (Earth)
\vec{v}_{BE}	The velocity of the boat with respect to the shore

Basic Equations:

We will use our transformation equations

$$\begin{aligned} v_{bAx} &= v_{bBx} + v_{BA} & v'_x &= v_x - u \\ v_{bAxy} &= v_{bBy} & v'_y &= v_y \\ v_{bAz} &= v_{bBz} & v'_z &= v_z \end{aligned}$$

but we will also need our vector recombination set

$$\begin{aligned} v &= \sqrt{v_x^2 + v_y^2} \\ \theta &= \tan^{-1} \left(\frac{v_y}{v_x} \right) \end{aligned}$$

We identify

$$\begin{aligned} \mathbf{v}_y &= \vec{v}_{BR} = 10.0 \text{ km/h} \hat{j} & \text{Velocity of the boat with respect to the river} \\ \mathbf{u} &= \vec{v}_{RE} = 5.00 \text{ km/h} \hat{i} & \text{Velocity of the river with respect to the shore (Earth)} \\ \mathbf{v}' &= \vec{v}_{BE} & \text{The velocity of the boat with respect to the shore (Earth)} \end{aligned}$$

Note that we have to pick one of the frames to be the prime frame if we use the prime notation, then stick with that choice. I have picked the A frame to be the prime frame in our basic equations. We can put any zeros into our basic equations and to switch to the right subscripts for our particular problem. We then have

$v_{bAx} = 0 + v_{BA}$	$v'_x = 0 + u$
$v_{bAxy} = v_{BR} = v_{bBy}$	$v'_y = v_{BR} = v_y$
$v_{bAz} = v_{bBz} = 0$	$v'_z = v_z = 0$

Symbolic Solution:

We can see that our Galilean transformation becomes (using both notation systems)

$v_{BE_x} = 0 + u$	$v'_x = 0 + u$
$v_{BE_y} = v_{BR} = v_y$	$v'_y = v_{BR} = v_y$
$v_{BE_z} = 0$	$v'_z = 0$

And using our vector component equations we can find $v_{BE} = v'$

$\begin{aligned} v_{BE} &= \sqrt{(v_{BE_x})^2 + (v_{BE_y})^2} \\ &= \sqrt{(u)^2 + (v_{RE})^2} \\ &= \sqrt{(5.00 \frac{\text{km}}{\text{h}})^2 + (10.0 \frac{\text{km}}{\text{h}})^2} \\ &= 11.18 \frac{\text{km}}{\text{h}} \end{aligned}$	$\begin{aligned} v' &= \sqrt{(v'_x)^2 + (v'_y)^2} \\ &= \sqrt{(-u)^2 + (v_{RE})^2} \\ &= \sqrt{(5.00 \frac{\text{km}}{\text{h}})^2 + (10.0 \frac{\text{km}}{\text{h}})^2} \\ &= 11.18 \frac{\text{km}}{\text{h}} \end{aligned}$
--	---

8 Chapter 1 Mechanics and Thermo: Where we left off in PH121 and PH123

and of course we need a direction

$$\begin{aligned}\phi' &= \tan^{-1} \left(\frac{v_{BEy}}{v_{BEx}} \right) & \phi' &= \tan^{-1} \left(\frac{v'_y}{v'_x} \right) \\ &= \tan^{-1} \left(\frac{(5.00 \frac{\text{km}}{\text{h}})}{(10.0 \frac{\text{km}}{\text{h}})} \right) & &= \tan^{-1} \left(\frac{(5.00 \frac{\text{km}}{\text{h}})}{(10.0 \frac{\text{km}}{\text{h}})} \right) \\ &= 26.565^\circ & &= 26.565^\circ\end{aligned}$$

The math is the same, the answer is the same, but with the prime notation you have to keep in your head that the prime frame is the shore or earth frame and that the velocity of the boat with respect to the shore is v' . The subscript method makes everything clear without remembering as much.

Molecular Model of an Ideal Gas

We know from our second semester physics class (PH123) that the internal energy of a system must be the energy associated with the atoms and molecules that make the system. We used the Idea Gas Model in PH123, and we will continue to use it in PH279. Here is a reminder of what the Ideal Gas Model contains.

Ideal Gas Model
The number of molecules in the gas is very large,
The average separation between molecules is large compared the their dimensions
The molecules obey Newton's laws of motion,
On the whole the molecules move randomly
The molecules interact only by short-range forces during elastic collisions
The molecules make elastic collisions with the walls
The gas under consideration is a pure substance; that is; all molecules are identical.

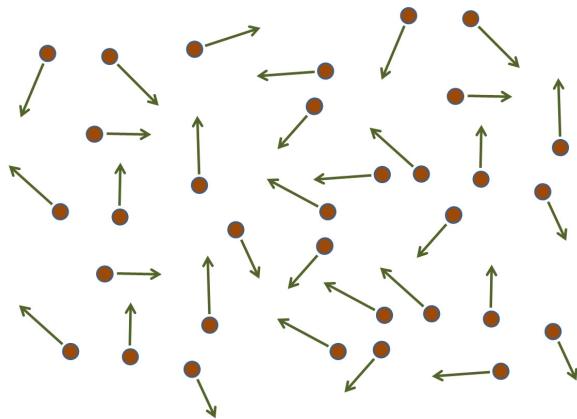
We will often say “molecule” but for an ideal gas atoms and molecules act alike.

Speed of molecules

We know from PH123 that thermal energy is associated with the motion of molecules, but we don't mean bulk motion of the molecules as a whole.



We mean the motion of the individual molecules in our sample of gas



The molecule motions have random directions. So the average velocity is zero, but that really doesn't describe the situation well. All gasses would have an average molecular velocity of zero over time. Instead we used the rms speed.

$$v_{rms} = \sqrt{\bar{v^2}}$$

This quantity is like an average, and gives a representative value that is near the most probable speeds.

Using this, we can get an expression that relates the temperature of our gas to the speed of the molecules. But it would be better if we knew something about how many molecules have what speed. Having the “average” is not enough. To find the *distribution* of speeds, let's remember the idea of number density

$$n_V = \frac{\# \text{ of molecules}}{V}$$

We expressed the number of molecules this way assuming we knew the volume. From this we worked to get the speed of the molecules (well some PH123 classes did). But along the way we found a sort of “energy density,” that is, the density of molecules that have energy between two amounts of energy, say, E_1 and E_2 . We want E_1 and E_2 to be quite close together. So let's let $E_1 = E$ and $E_2 = E + \Delta E$ where ΔE is a small amount of energy. Then the number of molecules with a particular energy between E_1 and E_2 could be written as

$$n_V(E) dE = \frac{\# \text{ of molecules with energy between } E \text{ and } E + \Delta E}{V} \quad (1.1)$$

This is called a *distribution function*. This might not be familiar to you. If not, don't panic. We will do much more with this later in our course. But for now, let's use this

10 Chapter 1 Mechanics and Thermo: Where we left off in PH121 and PH123

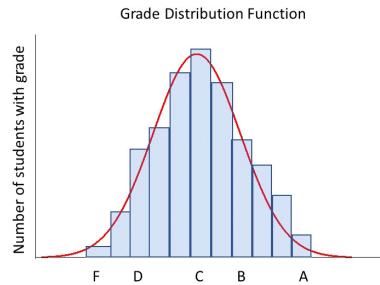


Figure 1.1.

notation $n_V(E) dE$ to give the number of molecules with a particular energy between E_1 and E_2 .

We find distribution functions in statistics. They are associated with probabilities. The standard “bell curve” used sometimes in grading is a distribution function. It tells the total number of students that got a particular number of points in a class.

What we need is the probability that the molecules will have a particular energy (or speed, since for molecules in an ideal gas, the molecular speed associated with the kinetic energy). A function that gives the amount of molecules that have a particular amount of energy also is called a *distribution* function. The distribution function that we will use is written symbolically in this cryptic fashion, $N(E) dE$. It is the number of molecules with a particular energy divided by the total number of molecules.

We will derive this function later in our class and in PH412, but for now let’s borrow the result.

$$N(E) dE = \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE$$

This distribution function is called the *Maxwell-Boltzmann distribution law*. It tells us that the probability of finding the molecules in a particular energy state varies exponentially as the negative of the energy divided by $k_B T$.

Distribution of Molecular Speeds

You might recall (but probably not) that we can use our energy distribution function to find the distribution of molecular speeds assuming the gas was ideal. An ideal gas only

has kinetic energy, so

$$E = \frac{1}{2}mv^2$$

Since there is a distribution of energies, we expect our gas molecules to have a distribution of speeds. That is, the molecules in the gas do not all go the same speed. We can make the transition from E to v by substituting in $mv^2/2$ for E . The number of molecules with energy E that corresponds with the speed v is just that, a number. So

$$N(v)dv = N(E)dE$$

Then we can say that

$$N(v) = N(E) \frac{dE}{dv}$$

or

$$\begin{aligned} N(v) &= \frac{2N}{\sqrt{\pi}(k_B T)^{\frac{3}{2}}} \sqrt{E} e^{-E/(k_B T)} \frac{dE}{dv} \\ &= \frac{2N}{\sqrt{\pi}(k_B T)^{\frac{3}{2}}} \sqrt{\frac{1}{2}mv^2} e^{-(\frac{1}{2}mv^2)/(k_B T)} (mv) \\ &= N \sqrt{\frac{2}{\pi}} \left(\frac{m}{k_B T}\right)^{\frac{3}{2}} v^2 e^{-(mv^2)/(2k_B T)} \\ &= N \frac{2^{\frac{1}{2}} \pi^{\frac{3}{2}} 2^{\frac{3}{2}}}{\pi^{\frac{1}{2}} \pi^{\frac{3}{2}} 2^{\frac{3}{2}}} \left(\frac{m}{k_B T}\right)^{\frac{3}{2}} v^2 e^{-(mv^2)/(2k_B T)} \\ &= 4N\pi \frac{1}{\pi^{\frac{3}{2}} 2^{\frac{3}{2}}} \left(\frac{m}{k_B T}\right)^{\frac{3}{2}} v^2 e^{-(mv^2)/(2k_B T)} \\ &= 4N\pi \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} v^2 e^{-(mv^2)/(2k_B T)} \end{aligned}$$

We will take this up again later. For now the important thing is that the distribution should depend on temperature, T , since we know the internal energy is tied to temperature. The distribution is as follows:

$$N_v = 4\pi N \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}} \quad (1.2)$$

where m is the mass of the molecule.

If there are a small number dN molecules with speeds between v and $v + dv$ then

$$dN = N_v dv \quad (1.3)$$

where N_v is the number with speed v , so there should be

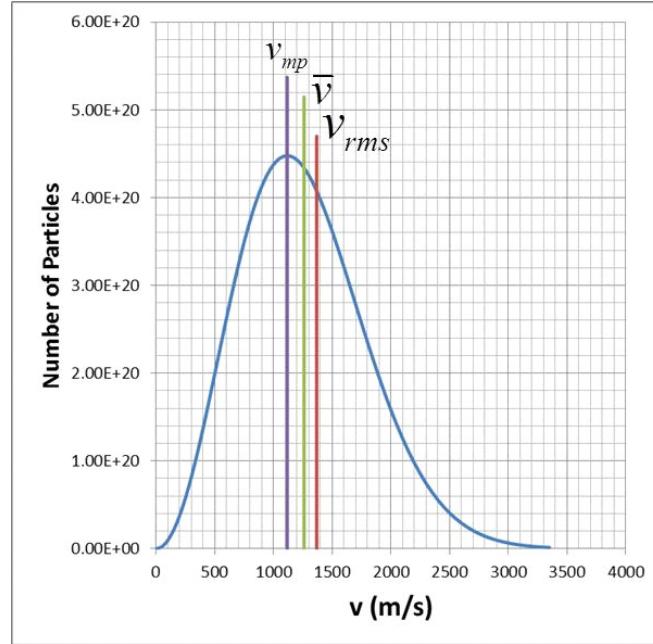
$$N = \int_0^\infty N_v dv \quad (1.4)$$

12 Chapter 1 Mechanics and Thermo: Where we left off in PH121 and PH123

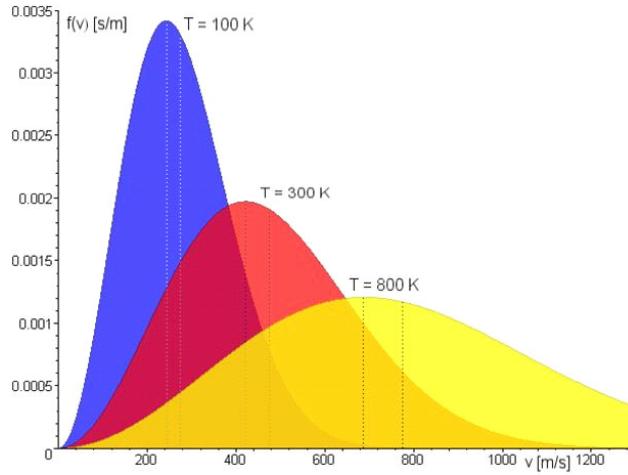
total molecules.

$$N = \int_0^\infty 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}} dv \quad (1.5)$$

If we plot N_v vs. v we get the figure below. The number of molecules with speeds between v and $v + dv$ is the area under the blue curve. The peak of the curve tells us the most probable speed, that is, the speed the most molecules have, v_{mp} . The curve is not symmetric, so the most probable speed is not the average speed, \bar{v} . There is also our new speed estimate marked v_{rms} .



If we plot N_v for different temperatures, we observe that the peak shifts, and the curve broadens



Temperature dependence of the Maxwell-Boltzmann distribution (Image in the Public Domain courtesy Fred Stober)

A motivated student could now find the most probable speed by finding the maximum of N_v . To do this, we take a derivative

$$\begin{aligned}
 \frac{dN_v}{dv} &= \frac{d}{dv} \left(4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}} \right) \\
 &= 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \frac{d}{dv} \left(v^2 e^{-\frac{mv^2}{2k_B T}} \right) \\
 &= 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \left(-2 \frac{m}{2k_B T} v^3 e^{-\frac{mv^2}{2k_B T}} + 2v e^{-\frac{mv^2}{2k_B T}} \right) \\
 &= -2v e^{-\frac{mv^2}{2k_B T}} \left(\frac{mv^2}{2k_B T} - 1 \right) 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}}
 \end{aligned}$$

set this equal to zero

$$\begin{aligned}
 0 &= -2v e^{-\frac{mv^2}{2k_B T}} \left(\frac{mv^2}{2k_B T} - 1 \right) 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \\
 \frac{mv^2}{2k_B T} &= 1 \\
 v_{mp} &= \sqrt{\frac{2k_B T}{m}}
 \end{aligned}$$

$$v_{mp} = \sqrt{\frac{2k_B T}{m}} \quad (1.6)$$

and this is great! We have related the temperature, T , to the most probable speed of the molecules. But it is more convenient to use v_{rms} , so let's see if we can modify this

14 Chapter 1 Mechanics and Thermo: Where we left off in PH121 and PH123

expression to be in terms of v_{rms} .

The average value of v^n is given by

$$\overline{v^n} = \frac{1}{N} \int_0^\infty v^n N_v dv$$

The motivated student could also use this to find the average speed (not the average velocity, which is zero). He or she will want the average value of v^1

$$\begin{aligned}\overline{v^1} &= \frac{1}{N} \int_0^\infty v^1 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}} dv \\ &= 4\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \int_0^\infty v^3 e^{-\frac{mv^2}{2k_B T}} dv \\ &= 4\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \left(-\frac{1}{2 \left(\frac{m}{2k_B T} \right)^2} \left(e^{-\frac{mv^2}{2k_B T}} + \frac{m}{2k_B T} v^2 e^{-\frac{mv^2}{2k_B T}} \right) \right) \Big|_0^\infty \\ &= 4\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \left(-\frac{1}{2 \left(\frac{m}{2k_B T} \right)^2} \right) (0 - (1 + 0)) \\ &= 2\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \left(\frac{1}{\left(\frac{\pi m}{2\pi k_B T} \right)^2} \right) \\ &= 2\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \left(\left(\frac{m}{2\pi k_B T} \right)^{-\frac{4}{2}} \pi^{-2} \right) \\ &= 2\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{1}{2}} (\pi^{-2}) \\ &= 2 \left(\frac{2\pi k_B T}{m} \right)^{\frac{1}{2}} (\pi^{-1}) \\ &= \sqrt{\frac{8k_B T}{\pi m}}\end{aligned}$$

so

$$\bar{v} = \sqrt{\frac{8k_B T}{\pi m}}$$

which is also great, but not what we wanted. But it is close. This time let's find the average value of v^2 . We hinted that the root mean squared value would be useful. that is, $\overline{v^2} = v_{rms}$. We can find this like we found the average velocity

$$\begin{aligned}\overline{v^2} &= \frac{1}{N} \int_0^\infty v^2 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}} dv \\ &= 4\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \int_0^\infty v^4 e^{-\frac{mv^2}{2k_B T}} dv\end{aligned}$$

But the math is a bit more cumbersome. Let

$$a = \frac{m}{2k_B T}$$

and note (from a table of integrals)

$$\begin{aligned}\int x^4 e^{-ax^2} dx &= -\frac{1}{8a^{\frac{5}{2}}} \left(6\sqrt{a}xe^{-ax^2} - 3\sqrt{\pi} \operatorname{erf}(\sqrt{a}x) + 4a^{\frac{3}{2}}x^3 e^{-ax^2} \right) \Big|_0^\infty \\ &= -\frac{1}{8a^{\frac{5}{2}}} (-3\sqrt{\pi})\end{aligned}$$

The quantity $\operatorname{erf}(\sqrt{a}x)$ is called the “error function.” If you study this function in a good mathematical book on integration you will find that

$$\operatorname{erf}(\sqrt{a}\infty) = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt = 1$$

and we used that to get the last result. For the other terms, we had to look at limits. At any rate we are left with

$$\begin{aligned}\overline{v^2} &= 4\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \left(-\frac{1}{8 \left(\frac{m}{2k_B T} \right)^{\frac{5}{2}}} (-3\sqrt{\pi}) \right) \\ &\quad 4 \left(\pi^{\frac{2}{2}} \right) \left(\frac{1}{\pi} \right)^{\frac{3}{2}} \pi^{\frac{1}{2}} \left(\frac{1}{8 \left(\frac{m}{2k_B T} \right)^{\frac{5}{2}}} (3) \right)\end{aligned}\tag{1.7}$$

or simply

$$\overline{v^2} = 3 \frac{T}{m} k_B$$

now recognize that $v_{rms} = \sqrt{\overline{v^2}}$ so

$$v_{rms} = \sqrt{\frac{3k_B T}{m}}\tag{1.8}$$

and this is what we wanted. We have an expression that relates the temperature of the gas to the *rms* speed of the molecules of the gas. Returning to v_{rms} we see that, indeed, v_{rms} is close to the average and most probable speeds. It will show up in the next topic. So we will need to recognize it.

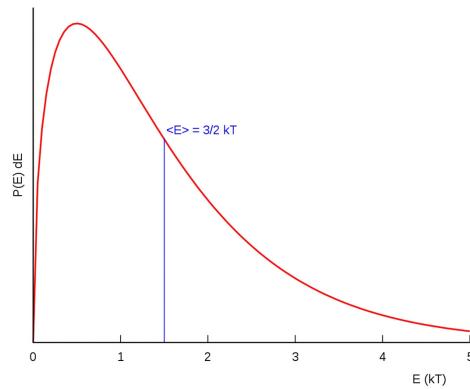
That we have some speeds higher than others explains why the molecules in a liquid

16 Chapter 1 Mechanics and Thermo: Where we left off in PH121 and PH123

don't all evaporate or all boil away at once.

Gas	Molar Mass (kg/mol)	v_{rms} at 20 °C (m/s)
H ₂	2.02×10^{-3}	1902
He	4.0×10^{-3}	1352
H ₂ O	18×10^{-3}	637
N ₂	28×10^{-3}	511
O ₂	28×10^{-3}	478
CO ₂	44×10^{-3}	408

This gives us some insight into the energy distribution because for an ideal gas, we have just kinetic energy. But for more realistic gasses, E might not be restricted to just kinetic energy. In PH279 we will go back to our distribution of energies



Maxwell-Boltzmann Energy Distribution

$$N(E) dE = \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE$$

and work with this equation directly.

We should realize that an integration of $dN = N(E) dE$ over all energies must give the number of molecules, since every molecule must have some energy. We can easily do this math by integrating from zero energy to infinite energy.

$$N = \int_0^\infty dN = \int_0^\infty \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE$$

We can do this by looking up the form

$$\int_0^\infty x^{\frac{1}{2}} e^{-x} dx = \frac{1}{2} \sqrt{\pi}$$

We need to make our integral look like this last one from the table. We can do that

$$\begin{aligned} N &= \int_0^\infty \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} \frac{E^{\frac{1}{2}}}{\sqrt{k_B T}} e^{-\frac{E}{k_B T}} dE \\ &= \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right) \int_0^\infty \left(\frac{E}{k_B T} \right)^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE \end{aligned}$$

We can make this look like the standard integral form by defining

$$\begin{aligned} x &= \frac{E}{k_B T} \\ dx &= \frac{1}{k_B T} dE \end{aligned}$$

Then

$$\begin{aligned} N &= \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right) k_B T \int_0^\infty \left(\frac{E}{k_B T} \right)^{\frac{1}{2}} e^{-\frac{E}{k_B T}} \frac{dE}{k_B T} \\ &= \frac{2N}{\sqrt{\pi}} \int_{x=0}^{x=\infty} (x)^{\frac{1}{2}} e^{-x} dx \\ &= \frac{2N}{\sqrt{\pi}} \left(\frac{1}{2} \sqrt{\pi} \right) \\ &= N \end{aligned}$$

which is what we said should happen.

Just like we did with molecular speeds, we can find the average molecular energy. To do this, we weigh each energy that the molecules can have by the number of molecules that have that energy.

$$E_{av} = \frac{1}{N} \int_0^\infty (E) N(E) dE$$

18 Chapter 1 Mechanics and Thermo: Where we left off in PH121 and PH123

Then

$$\begin{aligned}
 E_{av} &= \frac{1}{N} \int_0^\infty (E) \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE \\
 &= \frac{1}{N} \int_0^\infty \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{3}{2}} E^{\frac{1}{2}} e^{-\frac{E}{k_B T}} dE \\
 &= \frac{1}{N} \int_0^\infty \frac{2N}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{\frac{3}{2}} E^{\frac{3}{2}} e^{-\frac{E}{k_B T}} dE \\
 &= \frac{1}{N} \int_0^\infty \frac{2N}{\sqrt{\pi}} \left(\frac{E}{k_B T} \right)^{\frac{3}{2}} e^{-\frac{E}{k_B T}} dE \\
 &= \frac{2}{\sqrt{\pi}} \int_0^\infty \left(\frac{E}{k_B T} \right)^{\frac{3}{2}} e^{-\frac{E}{k_B T}} dE \\
 &= \frac{2k_B T}{\sqrt{\pi}} \int_0^\infty \left(\frac{E}{k_B T} \right)^{\frac{3}{2}} e^{-\frac{E}{k_B T}} \frac{dE}{k_B T} \\
 &= \frac{2k_B T}{\sqrt{\pi}} \int_0^\infty (x)^{\frac{3}{2}} e^{-x} dx
 \end{aligned}$$

Once again define

$$\begin{aligned}
 x &= \frac{E}{k_B T} \\
 dx &= \frac{1}{k_B T} dE
 \end{aligned}$$

From your friendly neighborhood table of integrals we find another standard form

$$\int_0^\infty x^{\frac{3}{2}} e^{-x} dx = \frac{3}{4} \sqrt{\pi}$$

Using this we find

$$\begin{aligned}
 E_{av} &= \frac{2k_B T}{\sqrt{\pi}} \left(\frac{3}{4} \sqrt{\pi} \right) \\
 &= \frac{3}{2} k_B T
 \end{aligned}$$

The k_B (written as just k in our textbook) is Boltzmann's constant.

$$k_B = 1.3806568 \times 10^{-23} \frac{\text{J}}{\text{K}}$$

and if the temperature in your room is about $T = 293 \text{ K}$ we would have

$$\begin{aligned}
 \bar{v} &= \sqrt{\frac{8k_B T}{\pi m}} = \sqrt{\frac{8(1.3806568 \times 10^{-23} \frac{\text{J}}{\text{K}})(293 \text{ K})}{\pi 28 \text{ u}}} \\
 &= 470.70 \frac{\text{m}}{\text{s}}
 \end{aligned}$$

and

$$\begin{aligned} E_{av} &= \frac{3}{2} \left(1.3806568 \times 10^{-23} \frac{\text{J}}{\text{K}} \right) (293 \text{ K}) \\ &= 6.0680 \times 10^{-21} \text{ J} \end{aligned}$$

The Equipartition of Energy

For ideal gasses we found the molar heat capacities

$$\begin{aligned} C_V &= \frac{3}{2} R \\ C_P &= \frac{5}{2} R \end{aligned}$$

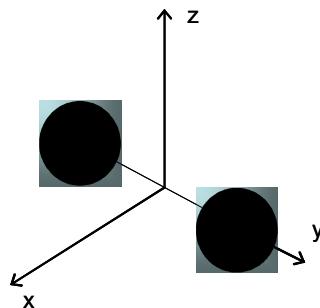
These equations work well for monotonic gasses, but fail badly for more complex gasses. Our ideal gas formulation starts to break down with more complex gasses at this point. To go farther, we would need to include the rotational and vibrational energy of the molecules. These rotational and vibrational states are beyond the particle model of classical physics and it turns out they are firmly in the region of quantum mechanics. In PH123 we used a little quantum to extend our ideal gas model a little by using what we know about degrees of freedom. Remember that we found that for each degree of freedom the internal energy was

$$E_i = \frac{1}{2} k_B T \quad (1.9)$$

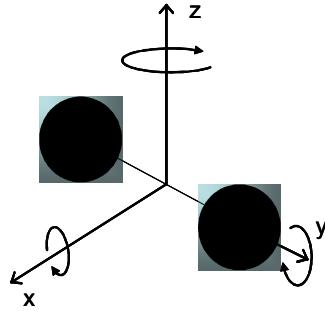
we found for an ideal monotonic gas that the internal energy was

$$E_{int} = 3E_i = \frac{3}{2} k_B T \quad (1.10)$$

where each E_i came from a translational degree of freedom. But a diatomic molecule has several more degrees of freedom.



It can rotate about any of the axes.



Here rotation about the y axis does not contribute significantly because the moment of inertia of a sphere (we will take the atom to be roughly spherical) about its axis is

$$\mathbb{I} = \frac{2}{5}mr^2 \quad (1.11)$$

where m is the mass of the atom. Most of the mass is centered in the nucleus (proton mass = 1.67×10^{-27} kg, electron mass = 9.11×10^{-31} kg), which has a radius of about $r = 1.7 \times 10^{-5}$ Å is the radius of the atom. The moment of inertia for rotation about the center of the two mass system is

$$\mathbb{I} = \sum_i m_i R_i^2 \quad (1.12)$$

where R is the distance from the center of mass. For diatomic hydrogen, $R = \frac{1}{2}1.06$ Å so we can see that the rotation about the y -axis is not very important, so we are left with three translational and two rotational degrees of freedom. This gives

$$E_{int} = \left(\frac{3}{2}k_B T \right)_{trans} + \left(\frac{2}{2}k_B T \right)_{rot} = \frac{5}{2}k_B T \quad (1.13)$$

Writing this in molar terms

$$E_{int} = \frac{5}{2}nRT \quad (1.14)$$

then

$$C_V = \frac{1}{n} \frac{dE_{int}}{dT} = \frac{5}{2}R \quad (1.15)$$

and

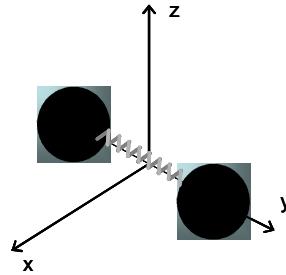
$$C_P = C_V + R \quad (1.16)$$

gives

$$C_P = \frac{7}{2}R \quad (1.17)$$

BUT WAIT, we did not include vibration! The atoms are bond together with an electrical attraction that acts quite like a spring force. So vibration along the axis is possible and we need to add in one more degree of freedom. We also have potential energy involved for a spring force, so we expect an additional degree of freedom for

vibration.



When we add all these up, we get

$$E_{int} = \left(\frac{3}{2}k_B T\right)_{trans} + \left(\frac{2}{2}k_B T\right)_{rot} + \left(\frac{2}{2}k_B T\right)_{vib} = \frac{7}{2}k_B T \quad (1.18)$$

which gives

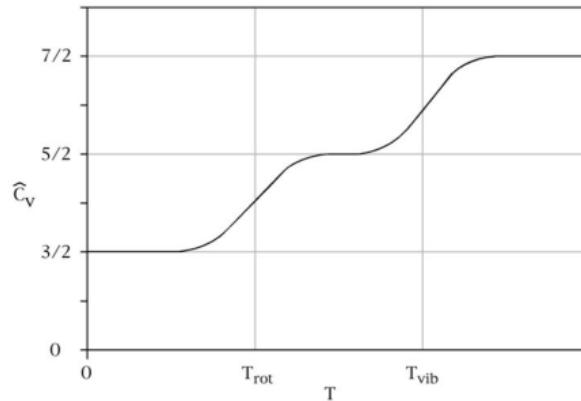
$$C_V = \frac{7}{2}R \quad (1.19)$$

and

$$C_P = \frac{9}{2}R \quad (1.20)$$

We should pause to ask, what values do we use? for diatomic gasses, is $C_V = \frac{7}{2}R$ all the time?

It turns out that when energy is added to a collection of molecules, it does not pick randomly from the degrees of freedom. We start at low temperatures with $C_V = \frac{3}{2}R$, then with an increase of temperature we get $C_v = \frac{5}{2}R$ and with a higher temperature we get $C_V = \frac{7}{2}R$



But why? Quantum mechanics will have to come to our rescue to explain why the molecules act as they do.

What to think of this “review”

You might be saying at this point, “I have never seen most of this stuff! and it is supposed to be a review.” And you might be right. Depending on your introductory physics professors, you might have done all of this or you might have missed large parts. Not to worry. We will go through relative motion and the thermal stuff in this class, so it should eventually make sense.

But what if you are saying that you have never done so much math in a problem before and are worried? Well this will be a theme for this semester. Expect problems to be more mathematical and to be longer than in the first three semesters of physics. It’s really OK, it just is not what you may be used to. You just need to get used to longer problems. And as a physics major, you knew that was coming. So it is all OK. Once you are used to the longer math, it will become second nature.

2 Classical Relativity

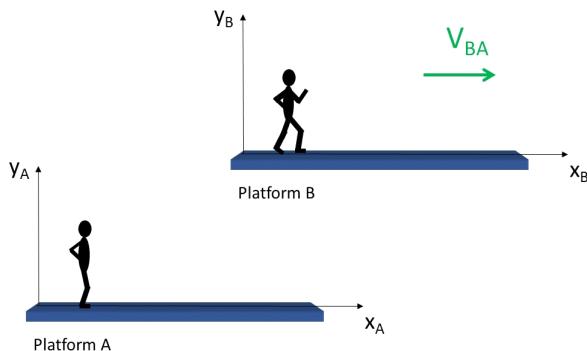
Section 2.1

Fundamental Concepts in the Lecture

- Galilean Newtonian Relativity
- The Michelson-Morley Experiment

Relative Motion

Back in PH121 we studied relative motion. We ended up with a system of equations that describe the velocity and position of an object moving in one reference frame from the point of view of another reference frame.



Let's assume a reference frame B in which a person, Bob, walks around. And further let's assume that frame B is moving with a speed v_{BA} to the right (positive x -direction) with respect to reference frame A . We could write the position of the person, Bob, in reference frame B as x_B and the position of Bob as viewed from frame A as x_A . Then

the position of Bob as viewed from the A frame would be

$$x_A = x_B + v_{BA}\Delta t$$

$$y_A = y_B$$

$$z_A = z_B$$

and the velocity of Bob as viewed from frame A would be.

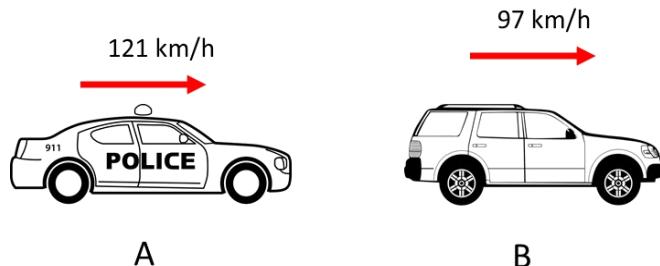
$$v_{bA} = v_{bB} + v_{BA}$$

where the little b is for Bob. This set of equations is called the *Galilean Transformation*. And you will remember from PH121 that there is not “right” reference frame. Both the A perspective and the B perspective are valid.

Let's review relative motion with a numeric example.

Relative motion example:

Two automobiles are traveling in the same direction toward Idaho Falls. One automobile, A , is traveling in the left lane of a highway at $121.0 \frac{\text{km}}{\text{h}}$ with respect to the highway frame of reference and the other car, B , is traveling in the right lane at $97 \frac{\text{km}}{\text{h}}$ with respect to the highway frame of reference. Using Galilean relativity, find the velocity of car B in the reference frame of car A .



Variables:

In the highway frame,

$$\begin{aligned}v_{BH} &= 121 \frac{\text{km}}{\text{h}} \\v_{AH} &= 97 \frac{\text{km}}{\text{h}}\end{aligned}$$

where here H stands for highway. Notice that we are using two subscripts. The first is the object that is moving (car A or car B in this case) and the second is the reference frame (like H for the highway reference frame).

In the frame of car A

$$v_{AA} = 0 \frac{\text{km}}{\text{h}}$$

That is, car A is not moving in the A reference frame.

And v_{BA} or the velocity of car B as viewed from the reference frame A is what we want.

Equations:

The Galilean transformation is just

$$\vec{v}_{AB} = \vec{v}_{AE} - \vec{v}_{BE}$$

Solution

We want to find the relative velocity between the frames of car A and car B . We can write that from our PH121 experience as

$$v_{AB} = v_{AE} - v_{BE}$$

or, since we are using the highway as a reference frame,

$$v_{AB} = v_{AH} - v_{BH}$$

We find that

$$\begin{aligned}v_{AB} &= 121 \frac{\text{km}}{\text{h}} - 97 \frac{\text{km}}{\text{h}} \\&= 24.0 \frac{\text{km}}{\text{h}}\end{aligned}$$

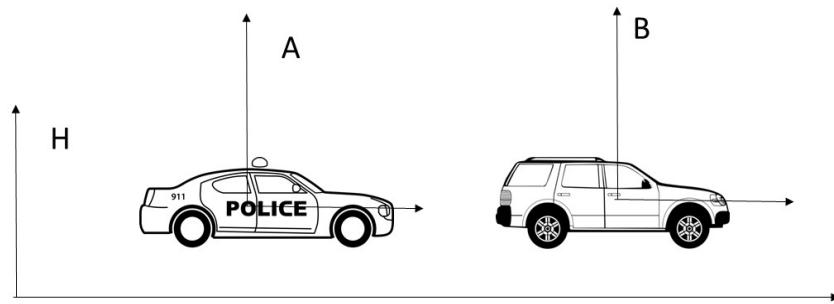
This is the speed of police car (frame) A as viewed from SUV (frame) B , but we wanted the speed of SUV (frame) B as viewed from police car (frame) A . We recognize that in frame A car A has a velocity of $v_{AA} = 0 \frac{\text{km}}{\text{h}}$. Since we know that car A is traveling with speed v_{AB} relative to frame B . So we can use our transformation again to find

$$\begin{aligned}v_{BA} &= v_{AA} - v_{AB} \\&= 0 - 24.0 \frac{\text{km}}{\text{h}}\end{aligned}$$

or in other words, if we were riding in car *A*, we would see car *B* approaching us at $-24.0 \frac{\text{km}}{\text{h}}$ relative to our speed. Car *B* is coming toward us in our reference frame.

Reference frames

We have done this example by assuming that we could assume that each car carried its own reference system or set of coordinate axes. There are three sets of coordinate axes in this example. one in each car, and one attached to the highway (or to the earth)



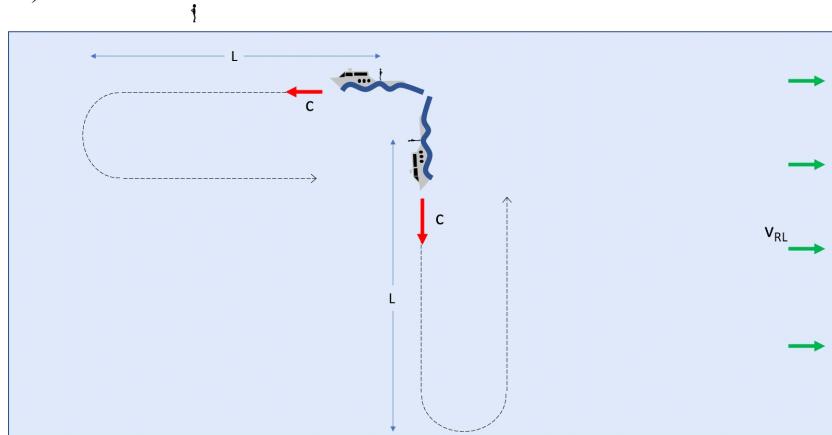
We view each reference frame as a perfectly good set of coordinates for doing physics. We expect the laws of physics to be the same no matter which frame we are in. This is like saying that we expect the laws of physics to be the same no matter which car we are in or if we are standing on the highway watching both cars go by. Einstein asked

I stand at the window of a railway carriage which is travelling uniformly, and drop a stone on the embankment, without throwing it. Then, disregarding the influence of the air resistance, I see the stone descend in a straight line. A pedestrian who observes the misdeed from the footpath notices that the stone falls to earth in a parabolic curve. I now ask: Do the “positions” traversed by the stone lie “in reality” on a straight line or on a parabola?

The path we see depends on whether we are in the rail car, or whether we are standing by the tracks, but are the laws of physics the same? Do we both agree that there is gravity acting on the rock? Would the free-body diagram look the same? How can both be paths be true at the same time? Curiously, the answer requires us to look more deeply into what we mean by “the same time.”

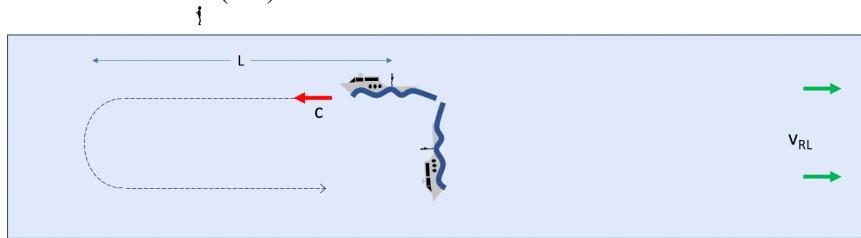
Galilean Relativity and boating on a River

Let's take a new specific example that we will use later to describe Einstein's theory of Special Relativity. Let's take a boat moving on a river. And let's take two cases. One in which the boat goes directly upstream, and then goes back downstream to where it started. We'll call this case *UD* (for up-down). And a second case where the boat crosses the river and comes back to where it started. We can call this case *CS* (for cross stream).



In both cases the boat travels the distance $2L$. Let's start with the case where the boat moves up stream and back. Let's find the time it takes to make this trip for both cases.

Upstream downstream (UD) case



Let's start analyzing the *UD* case by looking at basic motion. If the boat moves with a constant velocity compared to the land we can write

$$v_{BL} = \frac{\Delta x}{\Delta t}$$

where B is for the boat, and L is for an observer on the land watching the boat move. Using Galilean relativity we find

$$v_{BL} = v_{BR} + v_{RL}$$

where R is for the reference frame of the river water flowing by. You might picture this reference frame by considering an observer on a log floating by. This observer would be at rest with respect to the river water. A person on the land would observe the water of the river going by at a speed of v_{RL} . Then as the boat goes upstream for the first leg of its trip we would have

$$v_{BL_1} = -v_{BR_1} + v_{RL}$$

(the negative sign on v_{BR_1} is because the boat is going to the left). The time it takes the boat to make this first leg of the trip would be

$$\Delta t_1 = \frac{\Delta x_1}{v_{BL_1}} = \frac{-L}{v_{BL_1}}$$

and we know v_{BL_1}

$$\begin{aligned}\Delta t_1 &= \frac{-L}{-v_{BR_1} + v_{RL}} \\ &= \frac{L}{v_{BR_1} - v_{RL}}\end{aligned}$$

But this is just one leg of the trip. Let's find the time for the second leg back to where the boat started. Now

$$v_{BL_2} = +v_{BR_2} + v_{RL}$$

so

$$\Delta t_2 = \frac{\Delta x_2}{v_{BL_2}} = \frac{L}{v_{BR_2} + v_{RL}}$$

Since the boat is going with the current, we expect $\Delta t_2 < \Delta t_1$. If we compare our expressions for Δt_1 and Δt_2 we see that this is true.

We also know that the boat speed with respect to the water is constant so

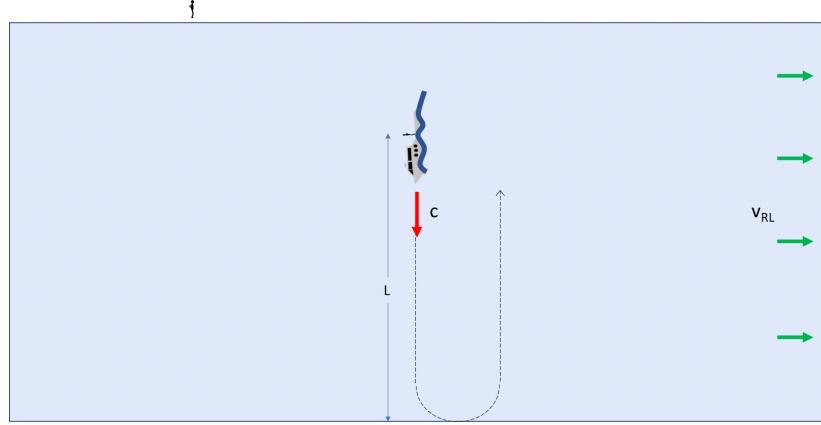
$v_{BR_1} = v_{BR_2} \equiv v_{BR}$. The total time of the trip would then be

$$\begin{aligned}\Delta t_{UDtrip} &= \Delta t_1 + \Delta t_2 \\ &= \frac{L}{v_{BR_1} - v_{RL}} + \frac{L}{v_{BR_2} + v_{RL}} \\ &= L \left(\frac{v_{BR} + v_{RL}}{(v_{BR} - v_{RL})(v_{BR} + v_{RL})} + \frac{v_{BR} - v_{RL}}{(v_{BR} + v_{RL})(v_{BR} - v_{RL})} \right) \\ &= L \left(\frac{v_{BR} + v_{RL}}{v_{BR}^2 - v_{RL}^2} + \frac{v_{BR} - v_{RL}}{v_{BR}^2 - v_{RL}^2} \right) \\ &= L \frac{2v_{BR}}{v_{BR}^2 - v_{RL}^2} \\ &= \frac{2L}{v_{BR}} \frac{1}{\left(1 - \frac{v_{RL}^2}{v_{BR}^2}\right)}\end{aligned}$$

This is the trip time for the upstream-downstream trip.

Cross Stream (CS) case

Now let's compare this to the cross stream trip.



Now we have to be more careful because we have a two dimensional problem. We must aim the boat upstream in order to cross the river directly. We don't want to get washed downstream. So we need to take x and y -components. Crossing to the other side we have

$$\begin{aligned} v_{BL_{1x}} &= v_{BR_{1x}} + v_{RL_x} \\ v_{BL_{1y}} &= v_{BR_{1y}} + v_{RL_y} \end{aligned}$$

We want $v_{BL_{1x}} = 0$. So

$$0 = v_{BR_{1x}} + v_{RL_x}$$

or

$$v_{BR_{1x}} = -v_{RL_x}$$

Now consider the y -part of the motion. We know

$$v_{BL_1} = \sqrt{v_{BR_{1x}}^2 + v_{BR_{1y}}^2}$$

we want $v_{BL_{1y}}$

$$\begin{aligned} v_{BR_1}^2 &= v_{BR_{1x}}^2 + v_{BR_{1y}}^2 \\ v_{BR_1}^2 - v_{BR_{1x}}^2 &= v_{BR_{1y}}^2 \\ v_{BR_{1y}} &= \sqrt{v_{BR_1}^2 - v_{BR_{1x}}^2} \end{aligned}$$

but we know from above that $v_{BR_{1x}} = -v_{RL_x}$ so

$$v_{BR_{1y}} = \sqrt{v_{BR_1}^2 - (-v_{RL_x})^2}$$

And the velocity of the river with respect to the land is all in the x -direction so

$$v_{BR_{1y}} = \sqrt{v_{BR_1}^2 - (-v_{RL})^2}$$

As in our first case, we can find the time it takes for the boat to cross the river. This is leg one

$$\Delta t_1 = \frac{\Delta y_1}{v_{BL_{1y}}} = \frac{-L}{-\sqrt{v_{BR_1}^2 - v_{RL}^2}}$$

and going back is similar, and of course $v_{BR_1} = v_{BR_2} = v_{BR}$ still

$$\Delta t_1 = \frac{-L}{-\sqrt{v_{BR}^2 - v_{RL}^2}}$$

$$\Delta t_2 = \frac{\Delta y_2}{v_{BL_{2y}}} = \frac{L}{\sqrt{v_{BR}^2 - v_{RL}^2}}$$

so the total trip time would be

$$\begin{aligned} \Delta t_{CStrip} &= \Delta t_1 + \Delta t_2 \\ &= \frac{2L}{\sqrt{v_{BR}^2 - v_{RL}^2}} \\ &= \frac{2L}{v_{BR}} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}} \end{aligned}$$

Notice that the trip time for the two cases is very different. Even the form of the equation for the total trip time is very different.

case UD up-stream	case CS cross stream
$\Delta t_{UD} = \frac{2L}{v_{BR}} \frac{1}{\left(1 - \frac{v_{RL}^2}{v_{BR}^2}\right)}$	$\Delta t_{CS} = \frac{2L}{v_{BR}} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}}$

We could compare the two by writing out Δt_{UD}

$$\begin{aligned} \Delta t_{UD} &= \frac{2L}{v_{BR}} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}} \sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}} \\ &= \left(\frac{2L}{v_{BR}} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}} \right) \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}} \\ &= \Delta t_{CS} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}} \end{aligned}$$

And we can think that if $v_{RL} = 0$ the relative speed of our land and river reference frames, the two times will be exactly the same. But if v_{RL} is not zero, the v_{RL}/v_{BR} is

less than one, and $1 - v_{RL}^2/v_{BR}^2$ is slightly less than one so $1/\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}$ is slightly greater than one. That is the up-down stream trip takes a little longer than the cross stream trip.

$$\Delta t_{UD} = \Delta t_{CS} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}}$$

Since we have assumed that the two trips cover the same distance, $2L$ then the two boats must travel at different average speeds

$$v_{UD} = \frac{2L}{\Delta t_{UD}} = \frac{2L}{\Delta t_{CS} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}}} = \frac{2L}{\Delta t_{CS}} \sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}$$

and

$$v_{CS} = \frac{2L}{\Delta t_{CS}}$$

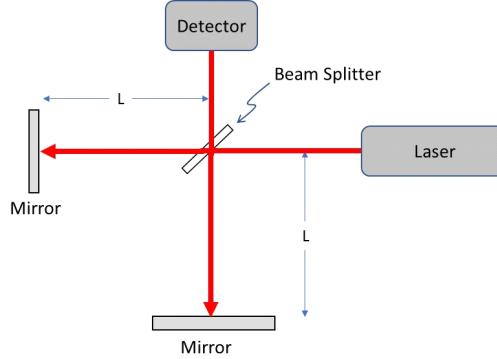
so that

$$v_{UD} = v_{CS} \sqrt{1 - \frac{v_{RL}^2}{v_{BR}^2}}$$

This is just what we expect from relative motion.

We can envision this same experiment, except instead of launching boats, we could replace the boats with lasers and launch pulses of light. We would need to install mirrors at the turning points to reflect the light back. Using the Galilean transformation equations we would expect to see the same result, just with c being a much higher speed than a boat could achieve. This would represent the state of physics in the late 1800's. It was assumed that there was a material wave medium for light and that the wave medium flowed past the Earth like a river. The name for this hypothesized material wave medium was the *ether*. The Earth would be like an island in that river with the ether flowing by.

Two researchers devised a clever way to find the speed of the Earth through the ether using a device we met back in PH123, the Michelson Interferometer.



Let's review the interferometer. It has a coherent light source, a beam splitter, two mirrors, and a detector. The beam is split and the resulting beams travel different paths to the mirrors. After reflection the beams recombine at the beam splitter and are sent off to a detector. This is just two wave mixing where the waves are electromagnetic field waves. We expect to see constructive and destructive interference. The math is familiar to us

$$\begin{aligned} E_1 &= E_{\max} \sin(k_1 \mathbf{r}_1 - \omega_1 t + \phi_1) \\ E_2 &= E_{\max} \sin(k_2 \mathbf{r}_2 - \omega_2 t + \phi_2) \end{aligned}$$

so the combined wave would be

$$E_r = E_{\max} \sin(k_1 \mathbf{r}_1 - \omega_1 t + \phi_1) + E_{\max} \sin(k_2 \mathbf{r}_2 - \omega_2 t + \phi_2)$$

Using a trig identity, we can write this as

$$\begin{aligned} E_r &= 2E_{\max} \cos\left(\frac{(k_2 \mathbf{r}_2 - \omega_2 t_1 + \phi_2) - (k_1 \mathbf{r}_1 - \omega_1 t_2 + \phi_1)}{2}\right) \\ &\quad + \sin\left(\frac{(k_2 \mathbf{r}_2 - \omega_2 t_1 + \phi_2) + (k_1 \mathbf{r}_1 - \omega_1 t_2 + \phi_1)}{2}\right) \end{aligned}$$

If the waves travel through air or a vacuum, then $k_2 = k_1 = k$ and $\omega_2 = \omega_1 = \omega$ so that

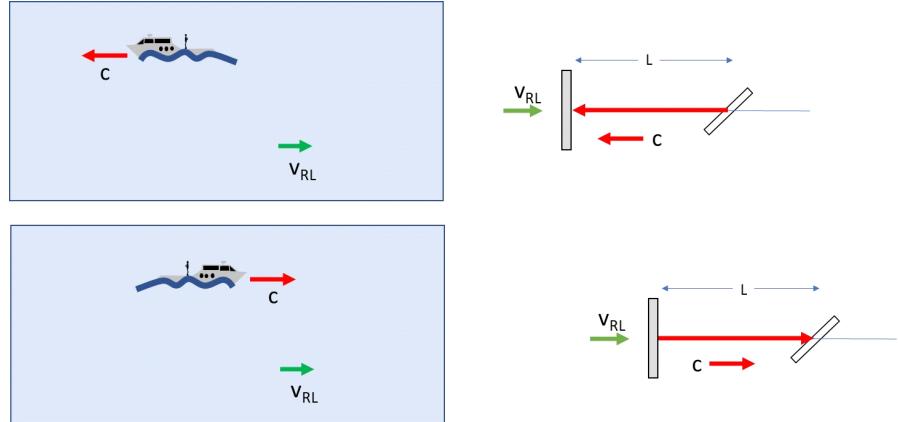
$$\begin{aligned} E_r &= 2E_{\max} \cos\left(\frac{(k(\mathbf{r}_2 - \mathbf{r}_1) - \omega(t_1 - t_2)) + (\phi_2 - \phi_1)}{2}\right) \\ &\quad + \sin\left(k \frac{(\mathbf{r}_2 + \mathbf{r}_1)}{2} - \frac{\omega(t_1 + t_2)}{2} + \frac{\phi_2 + \phi_1}{2}\right) \end{aligned}$$

The sine part is a wave. The cosine part is the amplitude. And if the amplitude is zero we have total destructive interference and if it is $2E_{\max}$ we have total constructive interference. If you follow the reflections of the two waves through the interferometer

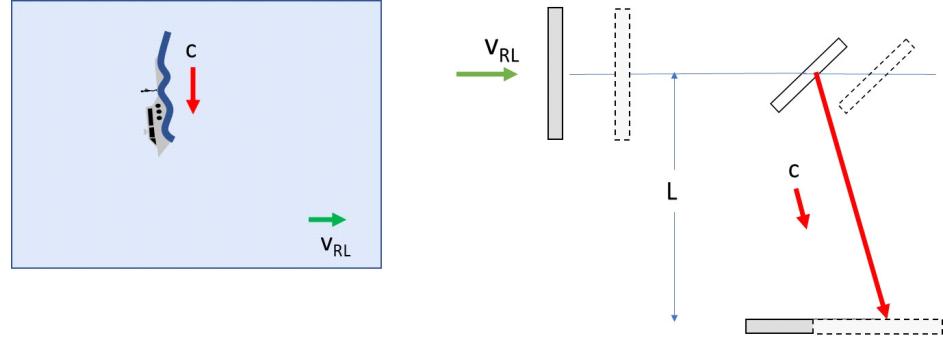
you will find that $\phi_2 = \phi_1 = \pi$, so

$$E_r = 2E_{\max} \cos \left(\frac{\left(\frac{2\pi}{\lambda} (\mathbf{r}_2 - \mathbf{r}_1) \right) - \frac{2\pi}{T} (t_1 - t_2)}{2} \right) + \sin \left(k \frac{(\mathbf{r}_2 + \mathbf{r}_1)}{2} - \frac{\omega (t_1 + t_2)}{2} + \pi \right)$$

Because the interference depends on the Δr and Δt we expect that a change in the path due to the motion of the device through the ether would have an effect on the interference pattern. You could take the device and align one arm with the flow through the fluid much like we did for our boat.



The other arm would cross the ether flow. This is just our up-down stream case and our cross stream case.



So we expect there to be an amplitude pattern shift because t_{UD} will not be equal to t_{CS} . By measuring the pattern once, and then turning the device 90° and measuring the pattern again we should be able to tell if the device is moving though the ether by observing a pattern change (even if we are not perfectly precise on our measurements of the distance between the mirror and the beam splitter L). In fact, we can tell how different the two paths are by counting the reversals of the interference pattern as

we turn the device. A reversal in the pattern gives a time difference of one quarter a period (because the waves travel from the beam splitter to the mirrors and back). The wavelength of light is about 500 nm, so the frequency would be

$$f = \frac{v}{\lambda} = \frac{3 \times 10^8 \frac{\text{m}}{\text{s}}}{500 \text{ nm}} = 6.0 \times 10^{14} \frac{1}{\text{s}}$$

and therefore the period would be

$$T = \frac{1}{f} = \frac{1}{6.0 \times 10^{14} \text{ Hz}} = 1.6667 \times 10^{-15} \text{ s}$$

and therefore very small time differences should have been measurable.

Michelson and Morley didn't find any shift in the pattern. None at all. And this was the end of ether theory!

The result was startling. The Earth is clearly moving. So from the Galilean transformation there should have been a shift. This failed result matched others that hinted a very strange thing. The speed of light seems to be the same in any reference frame. The physics we learned in PH121 doesn't seem to work for light! We need a new way of looking at motion that can include light (and other fast moving things).

3 The Postulates of Special Relativity

2.1, 2.2 and 2.3

Fundamental Concepts in the Lecture

- Postulate 1: Galilean Newtonian Relativity
- Postulate 2: Invariance of the speed of light
- Postulate 3: spacetime consists of points with x , y , z , and t values and these values are not invariant

In our last lecture, we found that Michelson and Morley did not find a difference in the time for the two light paths in their instrument. The implication was that the speed of the apparatus through the universe didn't change the speed of light. The speed of light is not relative! Let's pause for a moment. It's easy to not catch the full implication of this. Many people do miss what this means (mostly news reporters and people who are not physics students). Galileo told us that motion is relative. Michelson and Morley showed that one motion is not relative. The speed of light is the same no matter the reference frame. This is a very strange thing! Sometimes sociologists quote Einstein as telling us that everything is relative. But this is *not* what special relativity tells us. Galileo told us that speeds are relative. Einstein tells us that at least one speed is *not* relative. That one speed is the speed of light. It is probably not proper to extrapolate this to social science anyway, but social scientists who say everything is relative in their field and blame it on physics are actually using a 370 plus year old theory of motion due to Galileo. That makes it sound less cool. What we now know is that something is *not* relative and thus we have a theory of *special relativity*.

The Two (3) Postulates

Einstein made two postulates to solve the problem of the speed of light in vacuum being

invariant (not relative).

1. Galilean relativity is still true
2. The speed of light is invariant

Okay this doesn't seem too profound. This is just what we got through saying in our last lecture. But what Einstein did was to come up with a set of equations (and ideas) that could make this set of seemingly contradictory postulates make sense. The Galilean transformation couldn't be exactly right because it does not work for light. Einstein's theory fixes this. But to do this he really made another postulate, and that postulate is very profound.

The third Postulate

Let's go back to our Galilean relativity study of the motion of two boats. We assumed that our boats traveled at the same speed, v_{BR} . But let's say that instead of boats, we have light beams with $v_{BR} = c$. Then we would have

$$\Delta t_{UD} = \Delta t_{CS} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{c^2}}}$$

This is just due to normal Galilean relative motion. We could find a sort of average speed of our two boats, for our two cases, We know the times,

case UD	case CS
$\Delta t_{UD} = \frac{2L}{c} \frac{1}{1 - \frac{v_{RL}^2}{c^2}}$	$\Delta t_{CS} = \frac{2L}{c} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{c^2}}}$

and so far we think that in both cases $v = \frac{d}{\Delta t}$ should work, where d is still the round trip distance. For case UD

$$v_{UD} = \frac{d_{UD}}{\Delta t_{UD}} = \frac{d_{UD}}{\frac{2L}{c} \frac{1}{1 - \frac{v_{RL}^2}{c^2}}}$$

and for case CS

$$v_{CS} = \frac{d_{CS}}{\Delta t_{CS}} = \frac{d_{CS}}{\frac{2L}{c} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{c^2}}}}$$

if we demand this crazy idea that for light $v_{UD} = v_{CS}$ then we find

$$\begin{aligned}
 \frac{v_{UD}}{\frac{d_{UD}}{\frac{2L}{c} \frac{1}{\left(1 - \frac{v_{RL}^2}{c^2}\right)}}} &= \frac{v_{CS}}{\frac{d_{CS}}{\frac{2L}{c} \sqrt{1 - \frac{v_{RL}^2}{c^2}}}} \\
 \frac{d_{UD}}{\frac{1}{\left(1 - \frac{u^2}{c^2}\right)}} &= \frac{d_{CS}}{\frac{1}{\sqrt{1 - \frac{v_{RL}^2}{c^2}}}} \\
 d_{UD} \left(1 - \frac{v_{RL}^2}{c^2}\right) &= d_{CS} \sqrt{1 - \frac{v_{RL}^2}{c^2}} \\
 d_{UD} \sqrt{1 - \frac{v_{RL}^2}{c^2}} &= d_{CS} \\
 d_{UD} &= d_{CS} \frac{1}{\sqrt{1 - \frac{v_{RL}^2}{c^2}}}
 \end{aligned}$$

What we have found is that we can make the speeds match, *if we allow the distances d_{UD} and d_{CS} to be different and the times t_{UD} and t_{CS} to be different*. That is, if we can say that the distances are really somehow different in the cross motion and up-down motion directions just because of the motion of our reference frame (apparatus moving thought the Ether, or river flow) we could adjust our theory of motion to account for the result of Michelson and Morley experiment (and many other experiments that confirmed their result). This idea is the heart of Einstein's theory of Special Relativity.

Let's summarize what we have just said. We have said that in any reference frame the speed of light is the same. Since speed is

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

we conclude that either Δt is different, based on the motion of our object (which is true by Galilean relativity, see last lecture), or d is different to preserve the invariance of the speed of light, or both such that v for light never changes. What we are saying is that space and time change rather than the speed of light, and they change such that they make the speed of light always constant in any reference frame.

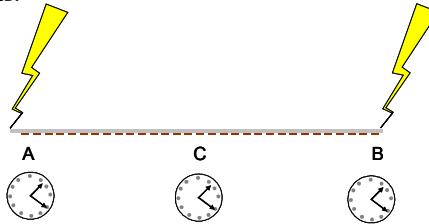
The third postulate is that spatial position and spatial time are not invariant.

Simultaneity

The third postulate is a little hard to imagine. To adjust the speeds so they are the same for light, we can either adjust the distance, d , or the time Δt or both. It remains to be seen which adjustments we need to make. Before we proceed to find out if d or Δt need

adjustment. Let's look at our concept of time more carefully. We should ask ourselves how we know if something happens "at the same time" as something else.

Einstein suggested that to answer this question we think of the example of a train track (he likes trains) that is struck by lightning in two places, *A* and *B* at the same time. He asks us what that means.



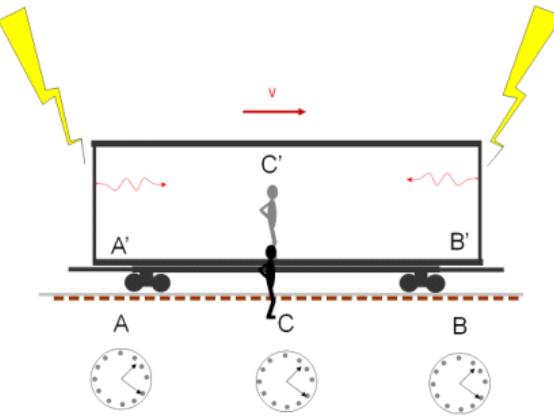
It turns out that this is a little bit hard to explain. How could you tell if the lightning bolts struck "at the same time?"

How about this, we measure along the rails to a midpoint *C*. We place a person at the midpoint. We give the observer a set of mirrors inclined at 90° which allows him or her to visually observe both places *A* and *B* at the same time. If the observer perceives the two flashes of lightning at the same time, then they are simultaneous.

This seems to work, provided light travels at the same rate along *AC* and *BC*. And we know that it does, but even if it does not, we really won't know the difference. We will use this definition of being "at the same time" as seeing the flashes in the mirrors together. Now we set up a series of clocks at *A* and *B* and *C*. The clocks are identical (all work at the same rate). And let's set up our experiment so we can see the clocks when we see the lightning. We can compare the clocks and say we see things happening "at the same time."

Whenever we do an experiment to try to see if things are simultaneous, we have a situation analogous to this.

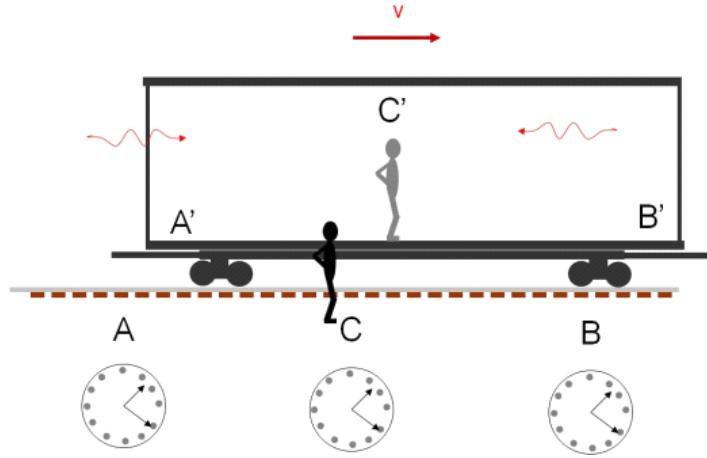
Now let's place an observer on the train as well.



The train is moving. The observer on the train also sees the lightning flashes through windows in the ends of the train car. But does the observer on the train see them “at the same time?”

Following Einstein, we say that the lightning strikes at A and B are simultaneous with respect to the observer standing by the tracks. What we mean by this is that the rays of light emitted at the places A and B , where the lightning occurs, meet each other at the mid-point C of the length AB along the tracks. But the events A and B also correspond to positions A' and B' on the train. Let C' be the mid-point of the distance $A'B'$ on the travelling train. Just when the flashes (as judged from the embankment) of lightning occur, this point C' happens to be passing the point C but it moves towards the right in the diagram with the velocity v of the train. If an observer sitting in the position C' in the train did not possess this velocity, then he would remain permanently at C (until he collided with the back of the train car!), and the light rays emitted by the flashes of lightning A and B would reach him just where he is situated.

Now more probably the passenger is moving towards the beam of light coming from B , while he is riding on ahead of the beam of light coming from A . So he will see the beam of light emitted from B earlier than he will see that emitted from A .



Observers who take the railway train as their reference-body must therefore come to the conclusion that the lightning flash B took place earlier than the lightning flash A . We thus arrive at the important result:

Events which are simultaneous with reference to the embankment are not simultaneous with respect to the train, and vice versa. This is known as the relativity of simultaneity. Every reference-frame has its own particular time; unless we are told the reference-frame to which the statement of time refers, there is no meaning in a statement of the time of an event.¹

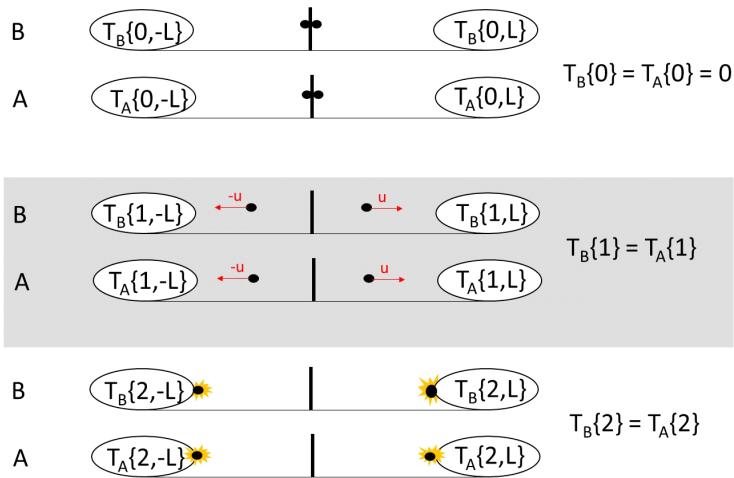
This is true no matter what we use to measure the two events. It turns out that in all of PH121 through PH220 we have assumed that the speeds of our reference frames were so slow that the simultaneity problem was never noticed. We could just assume things in our problems happened “at the same time” and not worry about it. But now we see that this is not enough. With this new expanded version of time, let’s go back to our speed-of-light problem and consider what might happen if we let Δt change with speed.

A more detailed look at simultaneity

We used Einstein’s words to describe the relativity of simultaneity, but I want to go beyond Einstein’s example. Let’s start with just the first of Einstein’s postulates, and look at Galilean relativity. And let’s envision a new observation device that can tell you its position (like a GPS positioning device) and the time of an event at that location. I

¹ Einstein, Albert, *Relativity: The Special and General Theory*, <http://www.gutenberg.org/files/5001/5001-h/5001-h.htm>

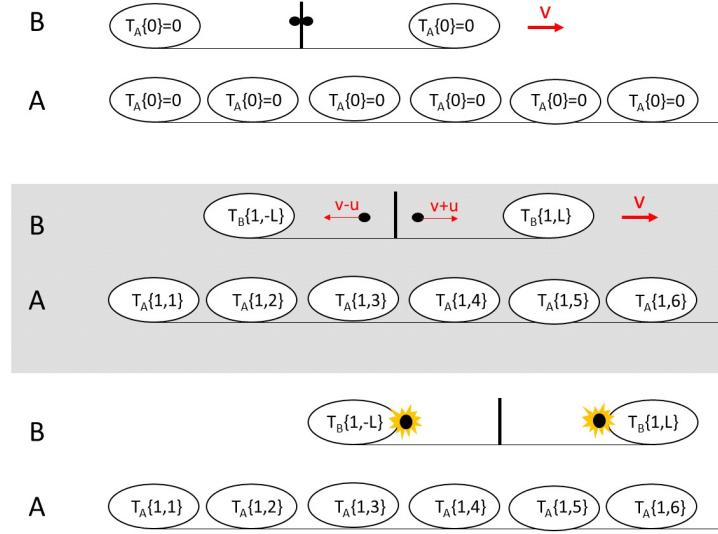
will call it a geoclock. We really envisioned geoclocks in our train example. But let's take our study of relativity into an imaginary laboratory and build a device that can launch balls with the same velocity in opposite directions. This device can then be used to launch balls into two geoclocks which would record the impact, recording the geoclocks location and the time of the impact. Here is a depiction of two such setups



where the subscript system used is complicated. The subscripts on the times gives which system (A or B) we are using. And the parameters in the curly braces give the time and then the location of the geoclock that makes the measurement. These could also be subscripts, but the number of subscripts started to get too big to easily write. So I have moved the rest of the subscripts up and placed them in curly brackets.

In this initial experiment there are no surprises. Both systems A and B act exactly the same. But now let's let system B move. And let's put a geoclock at every point in reference frame A . And let's synchronize all the frame A clocks so they all read the same time (so long as we view them from reference frame A).

42 Chapter 3 The Postulates of Special Relativity



Note what happens with our balls. The one going to the left will travel to the left geoclock at $x_B = -L$ with a speed $v - u$ where U is the speed of the ball. The ball going to the right will travel with a speed of $v + u$ toward a geoclock at $x_B = L$. And so the balls will hit “at the same time” as viewed from both reference frame A and reference frame B. This is just what the Galilean relativity equations say should happen. No surprise.

But now suppose we replace the apparatus that throws balls with one that shoots pulses of light. And now let's look at this new case from the B reference frame. Let's introduce a person riding along in the B reference frame. We'll call him Bob. Bob would see the light pulses hit the geoclocks and be detected at the same time. To Bob there is no surprise. Everything works just the same as with the balls.

We will designate Bob's clocks with a small b because “Bob's clock” is too long to be a subscript and we will let the details in the curly braces $\{\}$ be the details of which of Bob's clocks we are talking about. So

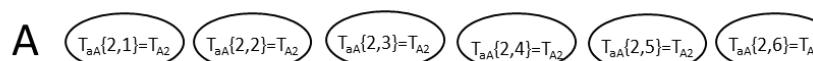
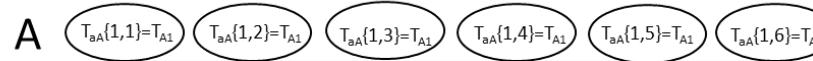
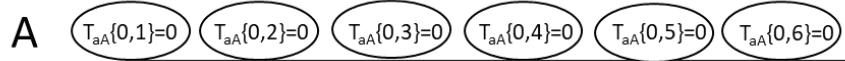
$$t_{bB}\{0, -L\}$$

means the zeroth measurement of Bob's (b) clock that is at location $-L$ as viewed *from the B reference frame*. We added an additional subscript to tell us our view point. Here is our experiment as viewed from the B reference frame. Bob, of course, has synchronized all the clocks (b) in his reference frame (B).



Bob's view of the experiment from the *B* reference frame.

But now let's consider a person standing in the *A* reference frame. Let's call her Alice. Alice sees all the clocks in the *A* frame as synchronized. Alice observes the *B* reference frame moving by with speed v_{BA} . So the first postulate of Galilean relativity applies. But we are dealing with light, so the second postulate also applies.



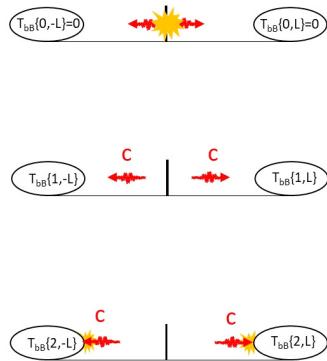
Alice's view of Bob's experiment from the *A* reference frame.

44 Chapter 3 The Postulates of Special Relativity

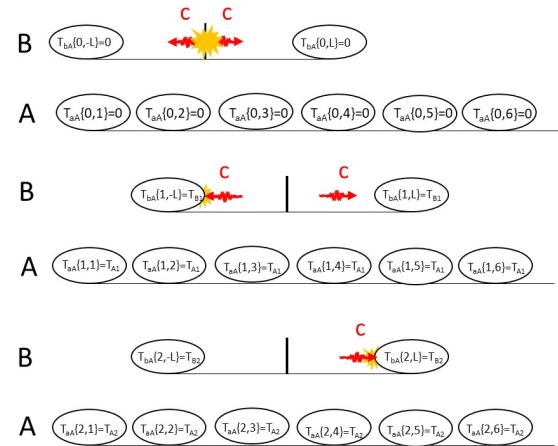
We see that Alice sees the (b) geoclocks moving, and the speed of the light pulses is not relative, so it is always the speed of light (c). Then Alice sees the left light pulse hit Bob's left geoclock ($t_{bA}\{1, -L\}$) when her geoclock at that position reads $t_{aA}\{1, 2\} = T_{A1}$ and she sees Bob's geoclock at that position read $t_{bA}\{1, -L\} = T_{B1}$. Then at a later time, $t_{aA}\{2, 5\} = T_{A2}$ (and a different location, location 5) Alice sees the right light pulse hit Bob's geoclock at $t_{bA}\{2, L\} = T_{B2}$ and she sees her own geoclock at that position reads $t_{aA}\{2, 5\} = T_{A2}$.

Note that Alice does not see the flashes hit Bob's geoclocks at the same time! In fact, She sees Bob's left clock running faster than his right clock. This must be the case because Bob's geoclocks record the time. And Alice can know that time by, say, taking a photograph when the light hits. That photograph will be the same no matter who looks at. They can even gather the photographs and look at them together later when Bob get's back to Alice's spot in the lab. So the photographic geoclock evidence must agree no matter what perspective we use to look at it. Here are the A and B frame situations side by side.

Bob's clocks as viewed in the B frame (bB)
(Bob sees his reference frame as not moving)



Bob's clocks are viewed from Alice's reference frame (bA)
(Alice sees Bob's reference frame as moving and her own as not moving)



So then Bob's left clock as he sees it must record the same time for the left pulse to hit as Alice sees on Bob's clock.

$$t_{bB}\{2, -L\} = t_{bA}\{1, -L\} = T_{B1}$$

That is, t_{bB} is the time on Bob's clock as viewed by Bob and t_{bA} is the time on Bob's clock as viewed by Alice and these two readings must be the same for a single event (like the left pulse striking the Bob's geoclock). So we really could write this time coordinate as just

$$t_b\{2, -L\} = t_b\{1, -L\} = T_{B1}$$

that is, the t_b coordinates at location $x_b = -L$ are the same regardless of the frame from which we view them. Now let's look at the right pulse hitting the right geoclock.

$$t_{bB}\{2, L\} = t_{bA}\{1, L\} = T_{B2}$$

So it is also true that the t_b coordinates at location $x_b = L$ are the same.

$$t_b\{2, L\} = t_b\{1, L\} = T_{B2}$$

But I find it helpful to think about our reference frame view point as we do problems.

Bob's clocks didn't stop at any point, so since the light hit the left clock at $x_B = -L$ first, and kept running, then from Alice's point of view

$$T_{B1} > T_{B2}$$

so we can write

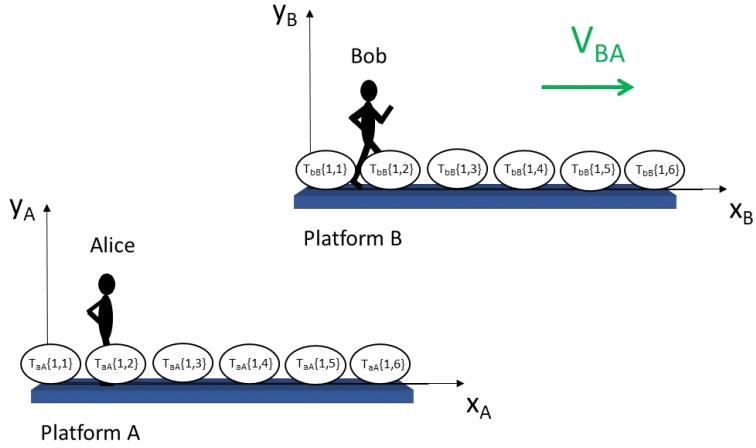
$$t_{bA}\{2, -L\} > t_{bA}\{2, L\}$$

That is, Alice doesn't see the light pulses hit at the same time and the trailing end of Bob's apparatus will read a longer time than the leading edge clock Bob's apparatus moves. You might find this described as "leading clocks lag."

It's important to ask the question, "does Bob see his clocks as synchronized?" And the answer is "yes!" Remember from Bob's point of view, the light pulses hit Bob's left and right geoclocks "at the same time." But Alice does not see the light pulses hit Bob's clocks at the same time. Who is right? And that is the nature of relativity, both are right. Simultaneity is relative.

Third Postulate again

Let's add a mental image to our third postulate We can envision reference frames as having a "lattice" of geoclocks that describe the positions and times for those positions.



In relativity we will consider a location in spacetime as having four coordinates, x , y , z , and t and we acknowledge that as we view one reference frame from another, we might see different times at each position in the lattice of geoclocks from different reference frames. We expect the value of x , y , and z to change when viewed from another reference frame. Now we know that t changes as well. Of course we could envision our lattice to be as finely spaced as we need. In my notation I will use subscripts and extended subscripts in curly braces to designate what is being viewed

$$t_{bA}\{1, 3\}$$

in this case the b says we are viewing something in the B reference frame (i.e. Bob's B frame geoclocks, or a flash of light, etc.). The second subscript tells us from what frame we are viewing. In this case we are viewing the b event from the A reference frame. The curly braces tell us the details of our event. Usually we will put the time details first (in this case the first t_{bA1}) and the position second (in this case the x_{bA3}). But think of the curly brace stuff like subscripts. We might at times leave off one or more types of subscripts or even add more. So it will be important as we go to give descriptions of the subscripts.

Because we can only have one clock reading for an event, we can write this as just

$$t_b\{1, 3\}$$

Whether we use subscripts or primes (as most books do), this is a very different view of space and time than we used in the first three classes of introductory physics!

4 Special Relativity

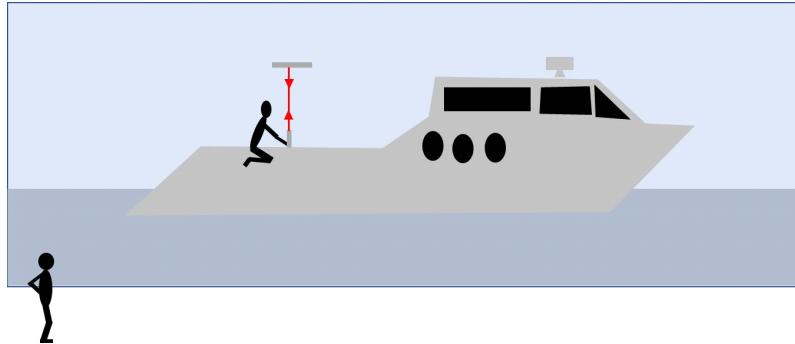
Sections 2.4 and 2.5

Fundamental Concepts in the Lecture

- Time Dilation
- Length Contraction
- Lorentz Transformation

Time dilation

To see how this time-change-with-speed works, let's start with our boats again. Consider our river and land reference frames. We found that the time of travel was different if we considered the cross and up-down stream paths. But how about if we switch from the land reference frame to the river reference frame? Let's picture a person (Bob) in the boat that has a very special clock. The clock is made from a pulsed laser and a mirror. The mirror is placed on a stand on the back of the boat (stand not shown in the figure below). The boat reference frame, B , is fixed to the boat. So in the boat reference frame there is no motion except for the light bouncing off of the mirror. Further suppose that our boat is just sitting on the water (engines are not on). The person points the laser straight up at the mirror.

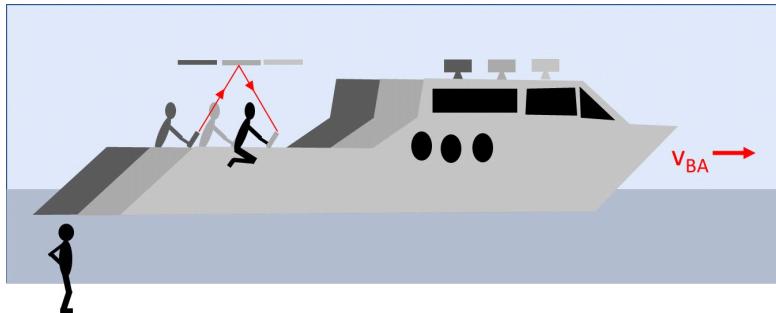


Inside the boat (that is, in the boat's reference frame, B) we know that if we aim the laser at the mirror that the pulse will bounce back. The path will be a straight line. The time of flight (time it takes to go to the mirror and back) is

$$\delta t_{bB} = \frac{2d}{c} \quad (4.1)$$

where now c is the speed of light and d is the distance between the source and the mirror. Let's say that δt_{bB} is the time it takes for our new clock to "tick" once. So we will call this a "tick time." We do need two subscripts, because δt , a duration or interval, does depend on which reference frame we use.

Now let's view this from outside the boat standing on the shore. Let's put another person on the shore called Alice. From our shore reference frame (now called A for Alice) the boat is moving with speed v_{BA} (the speed of the river with respect to the land).

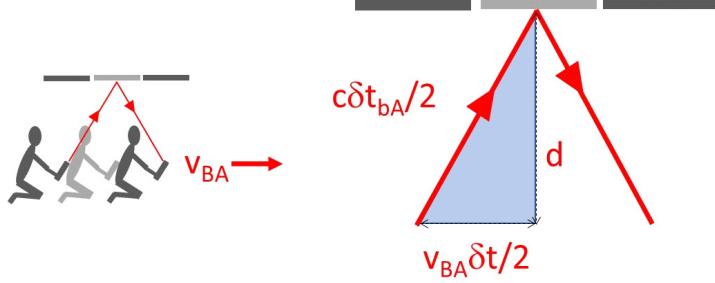


The observer near the river bank sees the light travel in a triangle.

Before we even start any math, we can see the change in time that we experienced before. The light path as viewed from the shore (A reference frame) is a longer path than as viewed from the boat (B reference frame). But from postulate 2, the speed of the light is the same, (c). Then the time of flight as viewed from the shore must be longer

than the time of flight as viewed from the boat in order to keep c constant! And this time δt_{bA} is still the tick duration. So the tick time got bigger when we view the clock from the shore. This is the effect that is known as *time dilation*.

Let's calculate how different the times will be. If we draw just the light path it is easier to see the triangle path as seen from the A reference frame. For our calculations, let's look at just half that triangle (the part marked in blue in the next figure).



we can see that the sides are of length

$$\frac{c\delta t_{bA}}{2}$$

where recall that Δt_{bA} is the time it takes on Bob's clock as viewed by Alice, and

$$\frac{v_{BA}\delta t_{bA}}{2}$$

where $\frac{v_{BA}\delta t_{bA}}{2} < \frac{c\delta t_{bA}}{2}$ because the boat floating along with the river current is much slower than the light. We can use the Pythagorean theorem to form an expression using these lengths

$$\left(\frac{c\delta t_{bA}}{2}\right)^2 = \left(\frac{v_{BA}\delta t_{bA}}{2}\right)^2 + d^2 \quad (4.2)$$

Now let's solve for δt_{bA}

$$\frac{c^2\delta t_{bA}^2}{4} = \frac{v_{BA}^2\delta t_{bA}^2}{4} + d^2 \quad (4.3)$$

$$\frac{c^2\delta t_{bA}^2}{4} - \frac{v_{BA}^2\delta t_{bA}^2}{4} = d^2 \quad (4.4)$$

$$\delta t_{bA}^2 \left(\frac{c^2}{4} - \frac{v_{BA}^2}{4} \right) = d^2 \quad (4.5)$$

$$\delta t_{bA}^2 = \frac{4d^2}{(c^2 - v_{BA}^2)} \quad (4.6)$$

so

$$\delta t_{bA} = \frac{2d}{\sqrt{(c^2 - v_{BA}^2)}} \quad (4.7)$$

or

$$\delta t_{bA} = \frac{2d}{c\sqrt{\left(1 - \frac{v_{BA}^2}{c^2}\right)}} \quad (4.8)$$

I want to break this into two pieces

$$\delta t_{bA} = \left(\frac{2d}{c} \right) \left(\frac{1}{\sqrt{\left(1 - \frac{v_{BA}^2}{c^2} \right)}} \right) \quad (4.9)$$

The first piece is just the time it took in the reference frame on the boat, δt_{bB} .

Sometimes this is called the *proper time*

$$\delta t_{bB} = \delta t_p = \frac{2d}{c}$$

although it is not a good name because there is nothing more “proper” about this time than any other time. It is the time we would get by measuring the clock tick from any reference frame in which the clock is sitting still in that reference frame. So it is also true that

$$\delta t_{aA} = \delta t_{bB} = \delta t_p$$

Notice that in my notation the b indicates a clock in the B reference frame and the B subscript tells us we are viewing the clock from the same, B , reference frame. Proper times are found in a reference frame where the clock is stationary with respect to that reference frame. If we moved the clock to the A reference frame and let it be stationary there then we would measure δt_{aA} and we would find that this would also be a proper time because it is a clock tick time measured in its own reference frame.

Then our b clock tick time as viewed by A is

$$\delta t_{bA} = \delta t_{bB} \left(\frac{1}{\sqrt{\left(1 - \frac{v_{BA}^2}{c^2} \right)}} \right) \quad (4.10)$$

The second term is long so I also want to give it its own symbol, γ_{BA} .

$$\gamma_{BA} = \frac{1}{\sqrt{\left(1 - \frac{v_{BA}^2}{c^2} \right)}}$$

Then the time as viewed from the shore would be

$$\delta t_{bA} = \delta t_{bB} \gamma_{BA} \quad (4.11)$$

or maybe

$$\delta t_{bA} = \delta t_p \gamma_{BA} \quad (4.12)$$

This is just what we expected from thinking about our boat trips. We now expect the motion of the river to cause there to be a time difference when we switch reference frames, and this is just what we see. On the shore, we see the clock tick time δt_{bA} that is bigger by a factor γ than the clock tick time $\delta t_p = \delta t_{bB}$ that people on the boat saw.

But what is this γ_{BA} ? Suppose that $v_{BA} \ll c$, then

$$\gamma_{BA} = \left(\frac{1}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}} \right) \quad (4.13)$$

$$\approx \left(\frac{1}{\sqrt{(1 - 0)}} \right) \quad (4.14)$$

$$= 1 \quad (4.15)$$

so for small river velocities we just have the same time as we would see if we were in the boat frame. This is why we never noticed the time difference before.

Suppose v_{BA} is large, say $\frac{c}{2}$

$$\gamma = \left(\frac{1}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}} \right) \quad (4.16)$$

$$= \left(\frac{1}{\sqrt{1 - \frac{\left(\frac{c^2}{4}\right)}{c^2}}} \right) \quad (4.17)$$

$$= \left(\frac{1}{\sqrt{1 - \frac{1}{4}}} \right) \quad (4.18)$$

$$= \left(\frac{1}{\sqrt{\left(\frac{3}{4}\right)}} \right) \quad (4.19)$$

$$= \left(\frac{2}{\sqrt{3}} \right) \quad (4.20)$$

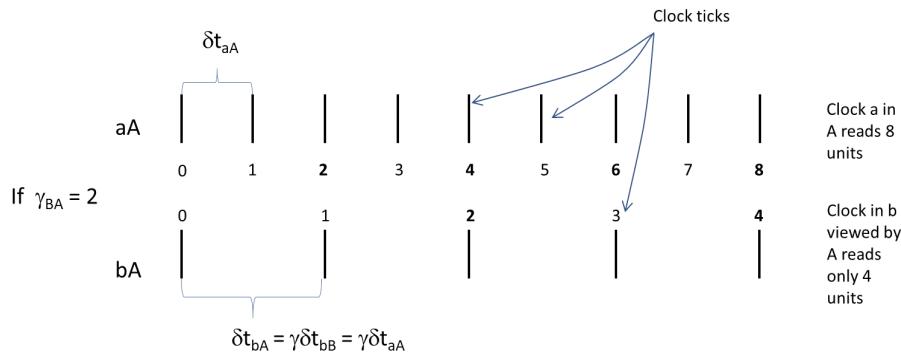
$$= 1.1547 \quad (4.21)$$

This is greater than 1! So what this tells us is that it seems to take longer for the light to travel and make the clock tick in the river bank frame than in the boat frame. As we said before, we call the time Δt_p the *proper time*. **It is always the time difference between two events for a person who sees the object (clock) experiencing the events at the same position in their reference frame (not moving).** This is like our person in the boat, Bob. He sees the strange optical clock as not moving in his reference frame. A person on the shore (Alice) would see the clock measurements taken at different positions in the shore reference frame.

Note that the time we measured from the observer on the bank is a longer tick time than the proper time. It is as though the time units got bigger. This will be true no matter the sign of v_{BA} and c because of the squares in γ . Because the measured b clock tick duration is longer when measured from the A frame than the proper time, this effect is known as *time dilation*. It is like the tick times got wider.

$$\delta t_{bA} = \delta t_p \gamma_{BA} \quad (4.22)$$

But so far we have just looked at the clock tick times. Now let's consider what the clocks on the boat and on the land read. The elapsed clock time $\Delta t_{aA} = t_{af} - t_{ai}$ for a clock on the shore (clock set a , viewed from frame A) is shown in the top of the next figure as a series of ticks. This is a clock that is not moving in its reference frame. So remember that $\delta t_{aA} = \delta t_{bB} = \delta t_p$. The clock time on the boat as observed from the shore is also shown (bottom set of ticks) for the case of $v_{BA} = 0.86603c$ which gives us $\gamma_{BA} = 2$. As you can see, the δt_{bA} are twice as fat as the δt_{aA} (or δt_{bB}) ticks.



This means that the b clock on the boat as observed from the A frame (shore) will read only 4 units when the clock a in the A frame (shore frame) as viewed from the shore frame will read 8 units. Of course a relative speed of $v_{BA} = 0.86603c$ would be hard to achieve with a floating boat, or even with a space craft. But we can achieve such speeds for small particles with particle accelerators.

We could summarize this by saying *moving clocks run slow*. And we could write the elapsed time as

$$\Delta t_{bA} = \frac{\Delta t_{aA}}{\gamma_{BA}}$$

Now you might say that this seems wrong. How do we know that it was the boat that was moving, It might be the boat is stationary and the Earth is moving under it! From

the B frame it looks like the A frame goes to the left with speed $v_{AB} = -v_{BA}$. So

$$\delta t_{aB} = \gamma_{AB} \delta t_{aA} \quad (4.23)$$

and

$$\Delta t_{aB} = \frac{\Delta t_{bB}}{\gamma_{AB}} \quad (4.24)$$

That is, would a person on the boat see a clock sitting on the shore as running slow? And the answer is yes! But for now, we will stick with the boat being the object that is moving.

Nice gamma values

We used our gamma factor in a nice unit of $\gamma_{BA} = 2$. It will prove useful to have a few more nice values of γ_{BA} . We can easily solve for v_{BA} for nice γ_{BA} values.

$$\begin{aligned} \gamma_{BA} &= \left(\frac{1}{\sqrt{\left(1 - \frac{v_{BA}^2}{c^2}\right)}} \right) \\ \gamma_{BA}^2 &= \frac{1}{\left(1 - \frac{v_{BA}^2}{c^2}\right)} \\ \left(1 - \frac{v_{BA}^2}{c^2}\right) &= \frac{1}{\gamma_{BA}^2} \\ 1 - \frac{1}{\gamma_{BA}^2} &= \frac{v_{BA}^2}{c^2} \\ c^2 \left(1 - \frac{1}{\gamma_{BA}^2}\right) &= v_{BA}^2 \\ c \sqrt{1 - \frac{1}{\gamma_{BA}^2}} &= v_{BA} \end{aligned}$$

We can get

v_{BA}	γ_{BA}
0	1
$0.06c$	1.25
$0.86603c$	2
$0.94281c$	3
$0.96825c$	4
$0.97980c$	5
$0.99381c$	9
$0.99875c$	20

Length Contraction

In studying the Galilean transformation we said that we would need to choose either time or length as changing from one reference frame to another, or maybe both. Let's look at length now.

Pause and think of what you might do to measure the moving boat.

Suppose that I give you a 4 meter stick and ask you to measure a moving boat. How would you do it?

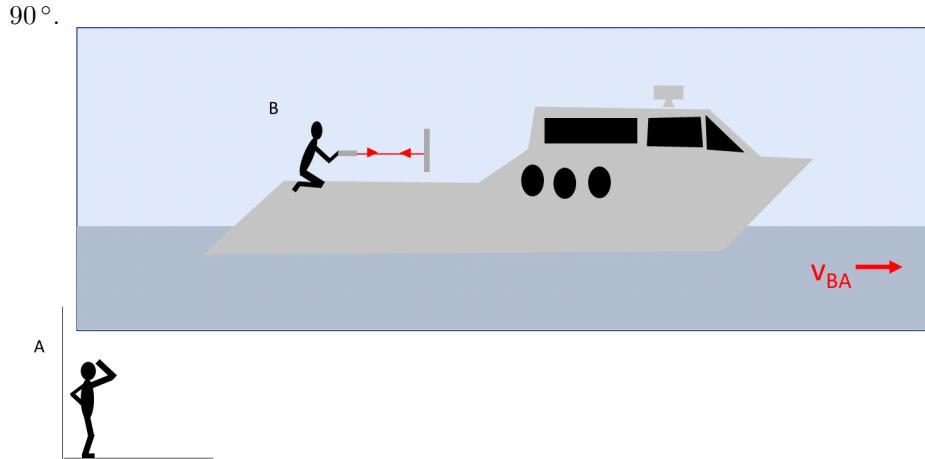
Most of the time we measure things that are standing still in our own reference frame. But now I want to measure the length of a moving object. This is hard to do. I can place one end of the meter stick on the front of the boat, but before I can align the end of the 4 meter stick with the back of the boat, the boat has moved!

To measure the boat length, we must take both the front position and the back position “at the same time.” But oh no! we have learned that “at the same time” depends on which reference frame you are in! We should expect some trouble with this.

Suppose our observer is in the boat holding a rod horizontally so it is parallel to the boat x -axis. Our boat person can measure the length of the rod as $\Delta x_{bB} = L_p$ by simply marking both ends of the rod along the bottom of the boat “simultaneously,” That will give the distance. Observers along the river bank do the same, they mark the ends of the rod “simultaneously” on their coordinate system along the bank as the boat goes by. This measurement will be Δx_{bA} .

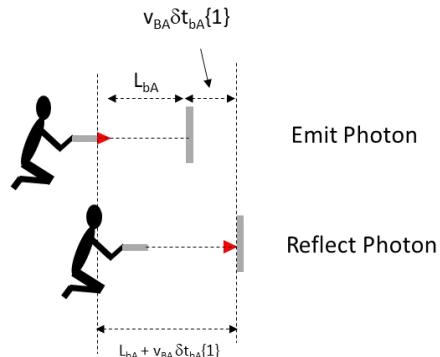
But now think of our lightning experiment with the train. The boat is like the train car. The boat observer would see a flash of light from the right ahead of a flash of light from the left because the boat is moving. So as the boat person aligns the rod with a mark on the deck, the light travels toward the person. But the person on the bank sees the flashes at the different times. They don't agree on what “at the same time” means. The person on the boat thinks the person on the ground marked the right hand side of the rod before he marked the left hand side! And in that time difference the boat moves. So the person on the shore would measure a shorter distance because she didn't measure the beginning of the rod and the end of the rod “at the same time.” Of course, she thinks she *did* measure the ends of the rod at the same time, and she thinks the boat person measured at different times. We can see that observers in the two reference frames won't agree on the length of the boat.

To make this computational, let's envision modifying our clock by turning the system



Now the light is traveling in the x -direction and the boat might move in that direction if we view it from the shore reference frame. Again the engines are not going, but the boat will float downstream with speed v_{BA} . The light travels at speed c .

If we let the clock pulse go, it will hit the mirror a time $\delta t_{bA}\{1\}$ later, which is about half a “tick” later



During that time, The boat would float a distance $v_{BA}\delta t_{bA}\{1\}$

$$\Delta x_{bA}\{1\} = v_{BA}\delta t_{bA}\{1\}$$

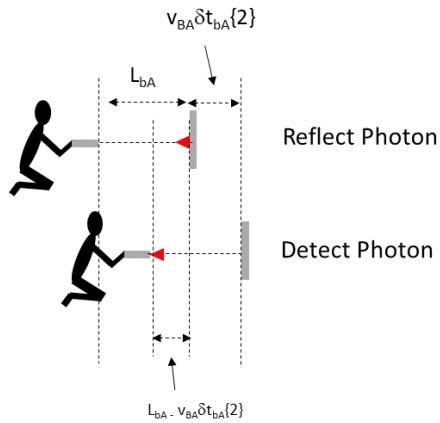
Let's call the distance from the source to the mirror $\Delta x_{bB} = L_{bB}$. In the last section, we said the clock had height d . But let's now relabel this distance L_{bA} , the length of the clock device on the boat as viewed from the shore (A) reference frame so we can be sure we understand our viewpoint. The light must go the distance L_{bA} , but because boat moves due to the moving water, it must go a little farther.

$$\Delta x_{bA}\{1\} = c\delta t_{bA}\{1\} = L_{bA} + v_{BA}\delta t_{bA}\{1\}$$

We could be tempted again to substitute in d , the length of the device in the boat frame, for L_{bA} but let's not. Let's just stick with L_{bA} as the source-mirror distance as viewed from the shore.

On the trip back, we have the same sort of thing, but this time the water and light are going opposite directions.

The boat still floats a distance $x_{bA}\{2\} = v_{BA}\delta t_{bA}\{2\}$ in the time the photon goes from the mirror to the detector.



The light must travel a distance

$$x_{ba}\{2\} = c\delta t_{bA}\{2\} = L_{bA} - v_{BA}\delta t_{bA}\{2\}$$

We can solve for the time it takes for both legs of the trip by solving for $\delta t_{bA}\{1\}$ and $\delta t_{bA}\{2\}$ and adding them together.

$$\begin{aligned} c\delta t_{bA}\{1\} - v_{BA}\delta t_{bA}\{1\} &= L_{bA} \\ \delta t_{bA}\{1\} (c - v_{BA}) &= L_{bA} \\ \delta t_{bA}\{1\} &= \frac{L_{bA}}{(c - v_{BA})} \end{aligned}$$

and likewise

$$c\delta t_{bA}\{2\} + v_{BA}\delta t_{bA}\{2\} = L_{bA}$$

gives

$$\delta t_{bA}\{2\} = \frac{L_{bA}}{(c + v_{BA})}$$

then the total tick time should be

$$\begin{aligned}
\delta t_{UD} &= \delta t_{bA}\{1\} + \delta \Delta t_{bA}\{2\} \\
&= \frac{L_{bA}}{(c - v_{BA})} + \frac{L_{bA}}{(c + v_{BA})} \\
&= \frac{L_{bA}(c + v_{BA})}{(c - v_{BA})(c + v_{BA})} + \frac{L_{bA}(c - v_{BA})}{(c + v_{BA})(c - v_{BA})} \\
&= \frac{2L_{bA}c}{c^2 - v_{BA}^2} \\
&= \frac{2L_{bA}c}{c^2 \left(1 - \frac{v_{BA}^2}{c^2}\right)} \\
&= \frac{2L_{bA}}{c \left(1 - \frac{v_{BA}^2}{c^2}\right)} \\
&= \frac{2L_{bA}}{c} \gamma_{BA}^2
\end{aligned}$$

as viewed from the shore.

Now here is the thing. The person on the boat in their reference frame, B , and the person on the shore in their reference frame, B , will see different durations for this same light reflection and detection “tick.” We know that from our previous calculation that lead to time dilation. In the boat frame, where the beginning and ending interval happen with the clock at the same boat frame coordinate, we have

$$\delta t_{bB} = \delta t_p = \frac{2L_{bB}}{c}$$

where L_{bB} is the length of the clock as seen on the boat. That is, in the boat frame there doesn’t seem to have been any motion of the clock apparatus, so the light just travels the path $2L_{bB}$. This tick time δt_{bB} must be related to our tick time as viewed from the shore, δt_{bA} , using the time dilation equation

$$\delta t_{bA} = \gamma_{BA} \delta t_{aA} = \gamma_{BA} \delta t_{bB}$$

and we called our total tick time δt_{bA} just δt_{UD} so

$$\delta t_{UD} = \gamma_{BA} \delta t_{bB} \quad (4.25)$$

or

$$\delta t_{UD} = \frac{2L_{bB}}{c} \gamma_{BA} \quad (4.26)$$

We have two expressions for the duration of the light pulse flight as viewed from the shore, δt_{UD} . Setting them equal gives

$$\begin{aligned}
\frac{2L_{bB}}{c} \gamma_{BA} &= \frac{2L_{bA}}{c} \gamma_{BA}^2 \\
L_{bB} &= L_{bA} \gamma_{BA} \\
L_{bA} &= \frac{L_{bB}}{\gamma_{BA}}
\end{aligned}$$

We remember that L_{bB} is the length of the device in its own rest frame (on the boat), and L_{bA} is the length of the device as viewed from the shore.

$$L_{bA} = L_{bB} \sqrt{1 - \frac{v_{BA}^2}{c^2}}$$

A person on the shore sees the clock with a length different than a person on the boat by a factor of $1/\gamma$!

And this is wonderful, it tells us how the perception of the length of our clock changed due to the motion of the clock apparatus. Notice that γ is generally greater than 1, so $\frac{1}{\gamma}$ is generally smaller than one. The length of the clock (and everything on the boat) measured from the river bank is shorter than the length of the clock (and everything on the boat) measured by a person on the bank. This effect is called *length contraction*.

Notice that when $v_{BA} \rightarrow 0$ that $L_{bA} = L_{bB} = d$. When it is not moving, a person on the shore and a person on the boat agree on the size of the clock.

We can now see that the way to reconcile the Michelson and Morley results is to allow both δt and Δx (or L as we have been calling it in the last equation) to change. We can expect both time dilation and length contraction. Let's define the proper length L_p as the length that would be measured in a reference frame **where the object is at rest in the reference frame**. This is called the *proper length*. That would be the boat reference frame. So

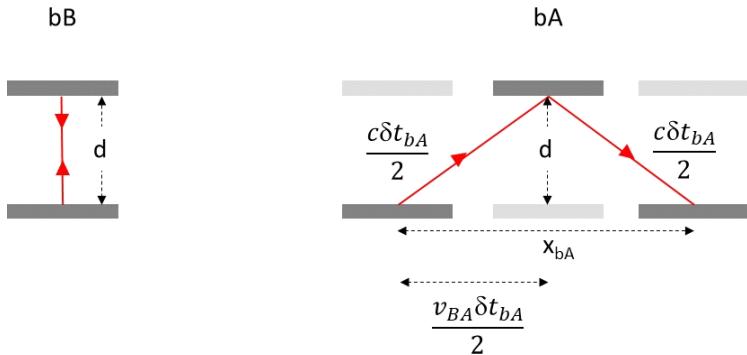
$$L_{bA} = L_p \sqrt{1 - \frac{v_{BA}^2}{c^2}}$$

In our class notation, this would again be when the subscripts match

$$L_p = L_{bB}$$

The Invariant Interval

So far we have said that the speed of light is invariant under transformation. This means that we see the same speed of light (in a vacuum) is always the same. But is there anything else that is invariant? Let's go back to our light clock.



For the light clock that is in frame B as observed in that frame the light travels a distance of $2d$ up and down. And the distance d can be written as

$$d = \frac{c\Delta t_{bB}}{2}$$

So the total distance the light travels is

$$2d = c\Delta t_{bB}$$

which can be written (though it is not apparent why yet) as

$$4d^2 = c^2 \Delta t_{bB}^2 + 0$$

Now let's look at the light clock b in B from frame A . We see that the light travels a path that looks like a triangle. The distance traveled from the source to the first mirror is

$$d_{bA}\{1\} = \frac{c\Delta t_{bA}}{2}$$

and likewise from the mirror to the detector

$$d_{bA}\{2\} = \frac{c\Delta t_{bA}}{2}$$

We can form a right triangle using d and using the Pythagorean theorem

$$\left(\frac{c\Delta t_{bA}}{2}\right)^2 = (d)^2 + \left(\frac{\Delta x_{bA}}{2}\right)^2$$

so then

$$\frac{c^2 \Delta t_{bA}^2}{4} = d^2 + \frac{\Delta x_{bA}^2}{4}$$

$$c^2 \Delta t_{bA}^2 = 4d^2 + \Delta x_{bA}^2$$

$$4d^2 = c^2 \Delta t_{bA}^2 - \Delta x_{bA}^2$$

We once again have $4d^2$ and once again we have a term like $c^2 \Delta t^2$ and now we have an x^2 term. If we think about what x_{bA} means, it is how far the light clock moved in Δt_{bA} . In the bB frame the light clock moved $x_{bB} = 0$ so we actually see the same for

for our bB view and our BA view. It turns out that the quantity

$$c^2 \Delta t_{bB}^2 - \Delta x_{bB}^2 = c^2 \Delta t_{bA}^2 - \Delta x_{bA}^2 = \text{constant}$$

for any object, b , in any reference frame (A, B, Q or what have you) viewed from any reference frame (A, B, Q or what have you). This is a second quantity that is invariant in special relativity like the speed of light in vacuum. And it works for any reference frame.

Lorentz Transformation

What we have done was to find a change in position and a change in time and how the displacement and duration are different when observed from a different reference frame. But a displacement is a difference between two positions, and a duration is a difference between two times. What of the positions and times themselves? Are they invariant? A quick thought back to the Galilean transformation tells us they must not be invariant. If we have our person in reference frame A observing a person in reference frame B we could describe the location of that person

$$x_a = x_b + v_{BA}t$$

and this is profound. The position of person in our coordinate system in frame A depends on when we ask the question. This makes sense, because reference frame B is moving. We need something like this for special relativity, a special relativity transformation.

Lorentz Transformation method 1

As we just said, let's think again about our Galilean relativity. If we have our person in reference frame A observing a person in reference frame B we could describe the location of that person

$$x_{bA} = x_{bB} + v_{BA}\Delta t$$

and this is profound. The position of person b in our coordinate system in frame A depends on when we ask the question. This makes sense, because reference frame B is moving. We could now write this as

$$\begin{aligned} x_a &= x_b + v_{BA}\Delta t \\ y_a &= y_b \\ z_a &= z_b \end{aligned}$$

in our subscript system. We only need one subscript. These are locations measured

within the A frame or the B frame. They don't really depend on an object. But we do need to modify this. We know that lengths contract, and a length is defined using two positions. We expect that the length represented by the position x_b and the origin will also be contracted. Let's write this as

$$x_a = G(x_b + v_{BA}t_b)$$

where we assume that G is a number that is dimensionless but depends on v_{BA} . This G is the fraction that the position changes due to our high speed motion so that the speed of light can stay constant. This number must go to one as v_{BA} approaches c , the speed of light. Note also that we wrote Δt as t_b . We know in special relativity that time coordinates are not invariant, so we need to be specific about which time we are using. This t_b is the time recorded at the x_b geoclock in the B reference frame. The reference frames B and A are equivalent so this transformation should work the other way around

$$x_b = G(x_a - v_{BA}t_a)$$

We want to find this G so we know how to accurately transform positions. There is a G in both of the last two equations, let's try to use them to solve for G . Substituting the previous equation into the last equation gives.

$$x_a = G(G(x_a - v_{BA}t_a) + v_{BA}t_b)$$

and now let's solve for t_b

$$\begin{aligned} x_a &= G^2(x_a - v_{BA}t_a) + Gv_{BA}t_b \\ x_a &= G^2x_a - G^2v_{BA}t_a + Gv_{BA}t_b \\ x_a - G^2x_a + G^2v_{BA}t_a &= Gv_{BA}t_b \\ (1 - G^2)x_a + G^2v_{BA}t_a &= Gv_{BA}t_b \\ \frac{(1 - G^2)}{G}x_a + Gv_{BA}t_a &= v_{BA}t_b \\ \frac{(1 - G^2)}{G}\frac{x_a}{v_{BA}} + Gt_a &= t_b \\ t_b &= \frac{(1 - G^2)}{G}\frac{x_a}{v_{BA}} + Gt_a \end{aligned}$$

This is not illuminating, but we will need it later. It is not immediately clear how to proceed. But it might help to take the derivatives

$$dx_b = G(dx_a - v_{BA}dt_a)$$

and let's take a derivative of t_b

$$dt_a = \frac{(1 - G^2)}{G}\frac{dx_a}{v_{BA}} + Gdt_a$$

and we can form $v_{bx} = dx_b/dt_b$

$$v_b = \frac{dx_b}{dt_b} = \frac{G(dx_a - v_{BA}dt_a)}{\frac{(1 - G^2)}{G}\frac{dx_a}{v_{BA}} + Gdt_a}$$

$$v_b = \frac{dt_a G \left(\frac{dx_a}{dt_a} - v_{BA} \right)}{dt_a \frac{(1-G^2)}{Gv_{BA}} \frac{dx_a}{dt_a} + G}$$

$$v_b = \frac{G \left(\frac{dx_a}{dt_a} - v_{BA} \right)}{\frac{(1-G^2)}{Gv_{BA}} \frac{dx_a}{dt_a} + G}$$

$$v_b = \frac{G(v_a - v_{BA})}{\frac{(1-G^2)}{Gv_{BA}} v_a + G}$$

Now from postulate 2 we require that for light $v_{ax} = v_{bx} = c$

$$c = \frac{G(c - v_{BA})}{\frac{(1-G^2)}{Gv_{BA}} c + G}$$

and solve for G

$$c \left(\frac{(1-G^2)}{Gv_{BA}} c + G \right) = G(c - v_{BA})$$

$$\frac{(1-G^2)}{Gv_{BA}} c + G = G - \frac{v_{BA}}{c} G$$

$$\frac{(1-G^2)}{Gv_{BA}} c = -\frac{v_{BA}}{c} G$$

$$\frac{(1-G^2)}{G^2} = -\frac{v_{BA}^2}{c^2}$$

$$\frac{1}{G^2} - 1 = -\frac{v_{BA}^2}{c^2}$$

$$\frac{1}{G^2} = 1 - \frac{v_{BA}^2}{c^2}$$

$$G^2 = \frac{1}{1 - \frac{v_{BA}^2}{c^2}}$$

$$G = \frac{1}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}}$$

Which is great! G is just γ ! So our coordinate transformation should look like

$$x_b = \frac{(x_a - v_{BA} t_a)}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}} = \gamma_{BA} (x_a - v_{BA} t_a)$$

$$x_a = \frac{(x_b + v_{BA} t_b)}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}} = \gamma_{BA} (x_b + v_{BA} t_b)$$

and of course

$$y_a = y_b$$

$$z_a = z_b$$

This makes a lot of sense. If our coordinate x_b is part of a length, we expect the length to contract. We should see length contraction in our transformation. And we do!

$$x_b = \frac{x_a}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}} - \frac{v_{BA} t_a}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}}$$

The first term is just our expression for length contraction provided we have the special case where in both the A and B reference frames we have the other end of the length at $x_a\{1\} = x_b\{1\} = 0$. But of course such a special case could only last a moment because the x_b frame is moving relative to the x_a frame. Just like with our Galilean transformation, the x_b coordinate must depend on t_a . The quantity $v_{BA}t_a$ is a length as well, and it must be contracted as well. So this transformation for x_b is just what we should expect.

But in special relativity, we know that time changes as well. We need to also transform t . We can do this with our equation for t_a , but now we know G is γ_{BA}

$$t_a = \gamma_{BA} \left(\left(\frac{1}{\gamma_{BA}^2} - 1 \right) \frac{x_b}{v_{BA}} + t_b \right)$$

and putting in what γ_{BA} is yields

$$\begin{aligned} t_a &= \gamma_{BA} \left(\left(\frac{1}{1 - \frac{v_{BA}^2}{c^2}} - 1 \right) \frac{x_b}{v_{BA}} + t_b \right) \\ t_a &= \gamma_{BA} \left(\left(1 - \frac{v_{BA}^2}{c^2} - 1 \right) \frac{x_b}{v_{BA}} + t_b \right) \\ t_a &= \gamma_{BA} \left(\left(\frac{v_{BA}^2}{c^2} \right) \frac{x_b}{v_{BA}} + t_b \right) \\ t_a &= \gamma_{BA} \left(\left(\frac{v_{BA}}{c} \right) x_b + t_b \right) \\ &= \gamma_{BA} \left(t_b + \frac{v_{BA}}{c^2} x_b \right) \end{aligned}$$

This is also something we should expect

$$t_a = \gamma_{BA}t_b + \gamma_{BA} \frac{v_{BA}}{c^2} x_b$$

Again the first part looks like time dilation. But the second term might be a surprise.

Notice it's units.

$$\frac{\text{m} \frac{\text{m}}{\text{s}}}{\text{m}^2} = \text{s}$$

It is an amount of time measured in the A frame. But it is an amount of time that depends on where the position of the geoclock is in the B frame. Like our position in the A frame depends on time in the B frame, our time in the A frame depends on our position of our object in the B frame.

We found x_b in terms of x_a and t_a but we found t_a in terms of x_b and t_b . We would like t_b in terms of t_a and x_a . We can find this by solving for t_b

$$\begin{aligned}
t_a &= \gamma_{BA} t_b + \gamma_{BA} \frac{v_{BA}}{c^2} x_b \\
t_a - \gamma_{BA} \frac{v_{BA}}{c^2} x_b &= \gamma_{BA} t_b \\
t_b &= \frac{1}{\gamma_{BA}} \left(t_a - \gamma_{BA} \frac{v_{BA}}{c^2} x_b \right)
\end{aligned}$$

but we need to turn our x_b into a x_a . We know

$$x_b = \gamma_{BA} (x_a - v_{BA} t_a)$$

so then

$$\begin{aligned}
t_b &= \frac{1}{\gamma_{BA}} \left(t_a - \gamma_{BA} \frac{v_{BA}}{c^2} x_b \right) \\
&= \frac{1}{\gamma_{BA}} \left(t_a - \gamma_{BA} \frac{v_{BA}}{c^2} (\gamma_{BA} (x_a - v_{BA} t_a)) \right) \\
&= \frac{1}{\gamma_{BA}} \left(t_a - \gamma_{BA} \left(\gamma_{BA} \frac{v_{BA}}{c^2} x_a - \gamma_{BA} \frac{v_{BA}}{c^2} v_{BA} t_a \right) \right) \\
&= \frac{1}{\gamma_{BA}} \left(t_a - \gamma_{BA}^2 \frac{v_{BA}}{c^2} x_a + \gamma_{BA}^2 \frac{v_{BA}^2}{c^2} t_a \right) \\
&= \frac{1}{\gamma_{BA}} \left(t_a + \gamma_{BA}^2 \frac{v_{BA}^2}{c^2} t_a - \gamma_{BA}^2 \frac{v_{BA}}{c^2} x_a \right) \\
&= \frac{1}{\gamma_{BA}} \left(\left(1 + \frac{1}{1 - \frac{v_{BA}^2}{c^2}} \frac{v_{BA}^2}{c^2} \right) t_a - \gamma_{BA}^2 \frac{v_{BA}}{c^2} x_a \right) \\
&= \frac{1}{\gamma_{BA}} \left(\left(\frac{c^2 \left(1 - \frac{v_{BA}^2}{c^2} \right)}{c^2 \left(1 - \frac{v_{BA}^2}{c^2} \right)} + \frac{1}{1 - \frac{v_{BA}^2}{c^2}} \frac{v_{BA}^2}{c^2} \right) t_a - \gamma_{BA}^2 \frac{v_{BA}}{c^2} x_a \right) \\
&= \frac{1}{\gamma_{BA}} \left(\left(\frac{1}{\left(1 - \frac{v_{BA}^2}{c^2} \right)} \right) t_a - \gamma_{BA}^2 \frac{v_{BA}}{c^2} x_a \right) \\
&= \frac{1}{\gamma_{BA}} \left(\gamma_{BA}^2 t_a - \gamma_{BA}^2 \frac{v_{BA}}{c^2} x_a \right) \\
&= \gamma_{BA} \left(t_a - \frac{v_{BA}}{c^2} x_a \right)
\end{aligned}$$

and after all that we see that we just changed the minus sign in the second term. Our transform is

$$\begin{aligned}
x_b &= \gamma_{BA} (x_a - v_{BA} t_a) \\
y_b &= y_a \\
z_b &= z_a \\
t_b &= \gamma_{BA} \left(t_a - \frac{v_{BA}}{c^2} x_a \right)
\end{aligned}$$

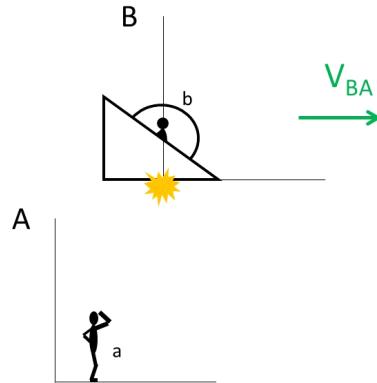
and as we saw, we can make the inverse transform

$$\begin{aligned}x_a &= \gamma_{BA} (x_b + v_{BA} t_b) \\y_a &= y_b \\z_a &= z_b \\t_a &= \gamma_{BA} \left(t_b + \frac{v_{BA}}{c^2} x_b \right)\end{aligned}$$

Lorentz Transformation method 2

We have the Lorentz transformation, and a reasonable person would let this be good enough. But we also learned about the invariant interval. We can find the Lorentz transformation using the invariant interval. Let's try it.

Suppose we have two reference frames, A and B . And suppose frame B is moving to the right with speed v_{BA} . We can put Bob in frame B in a space ship, for example. And we could have Alice observing Bob's ship go by.



We could imagine a special case where Bob's space ship makes a flash at position $x_B\{f\} = 0$. In Galilean relativity we would find the position of this flash in the A coordinates to be

$$\begin{aligned}x_a\{f\} &= x_b\{f\} + v_{BA} t_a\{f\} \\&= v_{BA} t_a\{f\}\end{aligned}$$

and we expect something similar in the special relativity. Since we are trying to find a coordinate in one reference frame and transform it into a coordinate in another reference frame we won't need to keep our subscripts that tell us from which frame we are viewing. So we will drop the capital subscripts for now. But we know that in special relativity we have to be careful about time. Our time coordinates are elapsed times. So

we start with

$$t_{bA}\{f\} = \frac{t_{aA}\{f\}}{\gamma_{BA}}$$

and removing the view point subscript (because we are doing that)

$$t_b\{f\} = \frac{t_a\{f\}}{\gamma_{BA}}$$

This tells us that a time on a geoclock in reference frame A will transform to a different time on a geoclock in frame B . We could write this as

$$t_a\{f\} = t_b\{f\}\gamma_{BA}$$

So from our transformation equation we expect

$$x_a\{f\} = v_{bA}t_b\{f\}\gamma_{BA}$$

But this was for the special case of a flash right at $x_b\{f\} = 0$. What if the flash is ahead of the space ship at $x_b\{f : x_{b2}\}$? We want any x_b to transform into an x_a and t_b to transform into a t_a .

Let's say that we start our experiment again just as Bob passes Alice and that we synchronized our clocks and choose our reference frame origins such that just as they pass

$$\begin{aligned} t_b\{1\} &= 0 \\ t_a\{1\} &= 0 \\ x_a\{1\} &= 0 \\ x_b\{1\} &= 0 \end{aligned}$$

which doesn't sound less restrictive, but the generality of this derivation comes from assuming a form for our transform. Let's assume that our relativistic transformation should be something like our Galilean transformation. It should not have x^2 or t^2 terms or the units would be difficult. Let's try

$$\begin{aligned} x_a &= Mx_b + Nt_b \\ t_a &= Gx_b + Ht_b \end{aligned}$$

where G , H , M , and N are functions to be found.

We can use what we did before for the special $x_b\{1\} = 0$ case to help find H and N

$$\begin{aligned} x_a\{1\} &= M(0) + Nt_b\{1\} \\ t_a\{1\} &= G(0) + Ht_b\{1\} \end{aligned}$$

then

$$\begin{aligned} x_a\{1\} &= Nt_b\{1\} \\ t_a\{1\} &= Ht_b\{1\} \end{aligned}$$

and from what we did before we know

$$t_a = t_b \gamma_{BA}$$

$$x_a = v_{BA} t_b \gamma_{BA}$$

and let's say we have two geoevents and we will mark them as {1} and {2}, then, using our time equation from our special case (because it still applies so far)

$$\Delta t_a = t_a\{2\} - t_a\{1\} = \gamma_{BA} (t_b\{2\} - t_b\{1\})$$

or

$$\Delta t_a = t_a\{2\} - 0 = \gamma_{BA} (t_b\{2\} - 0)$$

$$t_a\{2\} = \gamma_{BA} t_b\{2\}$$

where our second point could be any t_a (we didn't specify exactly when {2} is so long as it is not {1}), so we can drop the {2}

$$t_a = \gamma_{BA} t_b$$

Matching with our transform equations above we can see that

$$H = \gamma_{BA}$$

We can do this for position as well to see that

$$x_a = v_{BA} t_b \gamma_{BA}$$

leads to

$$\begin{aligned} x_a\{2\} &= v_{BA} \gamma_{BA} (t_b\{2\} - t_b\{1\}) \\ &= v_{BA} \gamma_{BA} t_b\{2\} - 0 \\ &= v_{BA} \gamma_{BA} t_b\{2\} \end{aligned}$$

or for any second value

$$x_a = v_{BA} \gamma_{BA} t_b$$

which implies that

$$N = v_{BA} \gamma_{BA}$$

Our transform so far is

$$\begin{aligned} x_a &= M x_b + v_{BA} \gamma_{BA} t_b \\ t_a &= G x_b + \gamma_{BA} t_b \end{aligned}$$

But we still need to find M and G . There are many ways to do this, but let's do it a simple way by utilizing our invariant interval.

$$c^2 \Delta t_b^2 - \Delta x_b^2 = c^2 \Delta t_a^2 - \Delta x_a^2 = \text{constant}$$

and for our case

$$c^2 t_b^2\{2\} - \Delta x_b^2\{2\} = c^2 t_a^2\{2\} - \Delta x_a^2\{2\} = \text{constant}$$

and since we are doing this for any {2} we can drop the {2} designation.

$$c^2 t_b^2 - \Delta x_b^2 = c^2 \Delta t_a^2 - \Delta x_a^2 = \text{constant}$$

And let's plug in our x_a and t_a .

$$c^2 (Gx_b + \gamma_{BA} t_b)^2 - (Mx_b + v_{BA} \gamma_{BA} t_b)^2 = c^2 \Delta t_b^2 - x_b^2$$

and hopefully we can use this to solve for M and G (this is not obvious, but it will work, keep reading). Let's expand out the squares

$$c^2 (G^2 x_b^2 + 2Gx_b \gamma_{BA} t_b + \gamma_{BA}^2 t_b^2) - (M^2 x_b^2 + 2Mx_b v_{BA} \gamma_{BA} t_b + v_{BA}^2 \gamma_{BA}^2 t_b^2) = c^2 \Delta t_b^2 - x_b^2$$

and rearranging

$$c^2 G^2 x_b^2 + 2c^2 Gx_b \gamma_{BA} t_b + c^2 \gamma_{BA}^2 t_b^2 - M^2 x_b^2 - 2Mx_b v_{BA} \gamma_{BA} t_b - v_{BA}^2 \gamma_{BA}^2 t_b^2 = c^2 \Delta t_b^2 - x_b^2$$

and combining like terms

$$c^2 G^2 x_b^2 + x_b^2 - M^2 x_b^2 + c^2 \gamma_{BA}^2 t_b^2 - c^2 \Delta t_b^2 - v_{BA}^2 \gamma_{BA}^2 t_b^2 + 2c^2 Gx_b \gamma_{BA} t_b - 2Mx_b v_{BA} \gamma_{BA} t_b = 0$$

or

$$(c^2 G^2 + 1 - M^2) x_b^2 + (c^2 \gamma_{BA}^2 - c^2 - v_{BA}^2 \gamma_{BA}^2) t_b^2 + (2c^2 G \gamma_{BA} - 2Mv_{BA} \gamma_{BA}) x_b t_b = 0$$

and here is our first time in this class that we will use a clever idea. For the whole equation to be equal to zero, each part must be also, so we can say

$$\begin{aligned} (c^2 G^2 + 1 - M^2) x_b^2 &= 0 \\ (c^2 \gamma_{BA}^2 - c^2 - v_{BA}^2 \gamma_{BA}^2) t_b^2 &= 0 \\ (2c^2 G \gamma_{BA} - 2Mv_{BA} \gamma_{BA}) x_b t_b &= 0 \end{aligned}$$

and if x_{bB} and t_{bB} are not zero, then the stuff in parenthesis must be. Then, taking the first

$$\begin{aligned} c^2 G^2 + 1 - M^2 &= 0 \\ c^2 G^2 + 1 &= M^2 \end{aligned}$$

so that

$$M = \sqrt{c^2 G^2 + 1}$$

Now let's take the second

$$\begin{aligned}
 c^2\gamma_{BA}^2 - c^2 - v_{BA}^2\gamma_{BA}^2 &= 0 \\
 c^2\gamma_{BA}^2 - v_{BA}^2\gamma_{BA}^2 &= c^2 \\
 (c^2 - v_{BA}^2)\gamma_{BA}^2 &= c^2 \\
 \gamma_{BA}^2 &= \frac{c^2}{(c^2 - v_{BA}^2)} \\
 \gamma_{BA} &= \frac{1}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}}
 \end{aligned}$$

which is nice, but not helpful.

$$\begin{aligned}
 2c^2G\gamma_{BA} - 2Mv_{BA}\gamma_{BA} &= 0 \\
 2c^2G\gamma_{BA} &= 2Mv_{BA}\gamma_{BA} \\
 c^2G &= Mv_{BA} \\
 G &= \frac{Mv_{BA}}{c^2}
 \end{aligned}$$

Now let's substitute this into our equation for M

$$\begin{aligned}
 M &= \sqrt{c^2G^2 + 1} \\
 M &= \sqrt{c^2 \left(\frac{Mv_{BA}}{c^2} \right)^2 + 1} \\
 M &= \sqrt{c^2 \frac{M^2v_{BA}^2}{c^4} + 1} \\
 M &= \sqrt{\frac{M^2v_{BA}^2}{c^2} + 1} \\
 M^2 &= \frac{M^2v_{BA}^2}{c^2} + 1 \\
 M^2 - \frac{M^2v_{BA}^2}{c^2} &= 1 \\
 M^2 \left(1 - \frac{v_{BA}^2}{c^2} \right) &= 1 \\
 M^2 &= \frac{1}{\left(1 - \frac{v_{BA}^2}{c^2} \right)} \\
 M &= \frac{1}{\sqrt{1 - \frac{v_{BA}^2}{c^2}}} = \gamma_{BA}
 \end{aligned}$$

so then

$$G = \frac{\gamma_{BA}v_{BA}}{c^2}$$

So our transformation.

$$\begin{aligned}x_a &= \gamma_{BA}x_b + v_{BA}\gamma_{BA}t_b \\t_a &= \frac{\gamma_{BA}v_{BA}}{c^2}x_b + \gamma_{BA}t_b\end{aligned}$$

or

$$\begin{aligned}x_a &= \gamma_{BA}(x_b + v_{BA}t_b) \\t_a &= \gamma_{BA}\left(t_b + \frac{v_{BA}}{c^2}x_b\right)\end{aligned}$$

We can add in to this a transform for y and z parts which don't change, so

$$\begin{aligned}x_a &= \gamma_{BA}(x_b + v_{BA}t_b) \\y_a &= y_b \\z_a &= z_b \\t_a &= \gamma_{BA}\left(t_b + \frac{v_{BA}}{c^2}x_b\right)\end{aligned}$$

We called this the special relativity transform, but it has a name. It is called the *Lorentz Transformation* after a scientist that actually derived this from electrodynamics before Einstein figured out why it worked. Of course we could do all this from Bob's perspective, Bob sees Alice going to the left with speed $v_{AB} = -v_{BA}$ so

$$\begin{aligned}x_b &= \gamma_{BA}(x_a - v_{BA}t_a) \\y_b &= y_a \\z_b &= z_a \\t_b &= \gamma_{BA}\left(t_a - \frac{v_{BA}}{c^2}x_a\right)\end{aligned}$$

And of course this is just the same thing we got with our first method.

We should try an example to see how this works.

5 Using the Lorentz Transformation (and Time Dilation and Length Contraction)

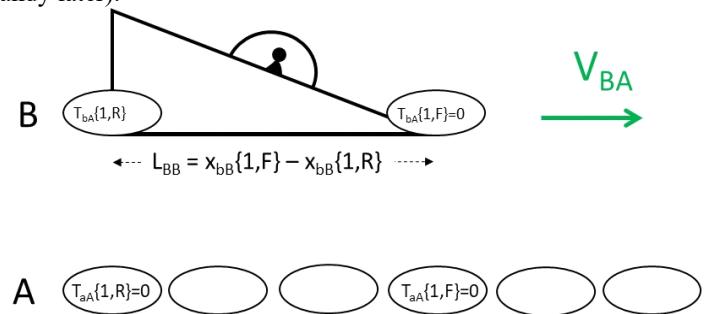
No new sections

Fundamental Concepts in the Lecture

- Time Dilation
- Length Contraction
- Lorentz Transformation

How much do leading clocks lag trailing clocks?

We found earlier that when geoclocks are observed in a reference frame that is seen to move from the perspective of another reference frame, the leading clocks trail or run behind the clocks that follow them. We could use the Lorentz transformation to discover what the observed time difference is between leading and trailing clocks (it will come in handy later).



Let's start with a spaceship being piloted by Bob with his B reference frame moving to the right past frame A where Alice is looking at Bob pass. Bob measures his ship which is at rest in his reference frame to have a length

$$\Delta x_{bB} = L_{bB} = x_b\{1, F\} - x_b\{1, R\}$$

where F and R stand for the front and rear of the ship. And this is a proper time because the measurement is made with the b ship stationary in the B reference frame. Bob also sees all his geoclocks as synchronized

$$t_{bB}\{1, R\} = t_{bB}\{1, F\} = 0$$

Alice, of course, sees a different length for Bob's ship.

$$\Delta x_{bA} = L_{bA} = x_a\{1, F\} - x_a\{1, R\}$$

And what we want to know is what times does Alice see on Bob's clocks?

Knowing the time on the front and rear clock will allow us to calculate the time difference

$$\Delta t_{bA}\{FR\} = t_{bA}\{R\} - t_{bA}\{F\}$$

Let's make things easier by choosing the situation where as Bob passes Alice his front clock reads zero

$$t_{bB}\{1, F\} = 0$$

and as they pass for that split second Alice's clocks also read zero

$$t_{aA}\{1, i\} = 0$$

Note we can always synchronize one clock in each reference frame with one clock in the other reference frame by carefully choosing how we set up our experiment. But we can never synchronize more than one clock across reference frames. That is because leading clocks will lag trailing clocks. We will see all the clocks in the moving reference frame as not synchronized with each other.

Recall that we want to know what Alice sees on Bob's clocks. We can start by finding out what Alice's clock's values would be in Alice's frame. Then, this will be a job for the Lorentz transformation to transform Alice's times to Bob's frame. Here is the basic set of transform equations:

$$x_b = \gamma_{BA} (x_a - v_{BA} t_a)$$

$$y_b = y_a$$

$$z_b = z_a$$

$$t_b = \gamma_{BA} \left(t_a - \frac{v_{BA}}{c^2} x_a \right)$$

and we fill in the details, say, for the front of the ship.

$$\begin{aligned}
 x_b\{1, F\} &= \gamma_{BA} (x_a\{1, F\} - v_{BA} t_a\{1, F\}) \\
 y_b\{1, F\} &= y_a\{1, F\} \\
 z_b\{1, F\} &= z_a\{1, F\} \\
 t_b\{1, F\} &= \gamma_{BA} \left(t_a\{1, F\} - \frac{v_{BA}}{c^2} x_a\{1, F\} \right)
 \end{aligned}$$

where it bears repeating that we are looking at the geoclock event time from one of Alice's geoclocks (t_a) and working to find what that geoclock event time it would correspond with in the reference Frame B (t_b).

We just want the time difference between the front and the rear of the ship, so let's start with the last of the set of Lorentz equations

$$t_b\{1, F\} = \gamma_{BA} \left(t_a\{1, F\} - \frac{v_{BA}}{c^2} x_a\{1, F\} \right)$$

and what we want is the time difference from the front to the rear. So we also need the time on the rear clock.

$$t_b\{1, R\} = \gamma_{BA} \left(t_a\{1, R\} - \frac{v_{BA}}{c^2} x_a\{1, R\} \right)$$

Remember that we expect our leading clock to lag behind the trailing clock. So

$$t_b\{1, F\} < t_b\{1, R\}$$

so let's define the difference in time on Bob's clocks as seen by Alice as

$$\Delta t_{bA}\{FR\} = t_b\{1, R\} - t_b\{1, F\}$$

to keep things positive. Let's use the "subscript" FR to tell us that the Δt_{bA} is the change in time from the front to the rear. Writing it all out gives

$$\begin{aligned}
 \Delta t_{bA}\{FR\} &= \gamma_{BA} \left(t_a\{1, R\} - \frac{v_{BA}}{c^2} x_a\{1, R\} \right) - \gamma_{BA} \left(t_a\{1, F\} - \frac{v_{BA}}{c^2} x_a\{1, F\} \right) \\
 &= \gamma_{BA} t_a\{1, R\} - \gamma_{BA} \frac{v_{BA}}{c^2} x_a\{1, R\} - \gamma_{BA} t_a\{1, F\} + \gamma_{BA} \frac{v_{BA}}{c^2} x_a\{1, F\} \\
 &= \gamma_{BA} t_a\{1, R\} - \gamma_{BA} t_a\{1, F\} + \gamma_{BA} \frac{v_{BA}}{c^2} x_a\{1, F\} - \gamma_{BA} \frac{v_{BA}}{c^2} x_a\{1, R\}
 \end{aligned}$$

But of course Alice sees all her clocks as synchronized so Alice's clock events

$t_a\{1, F\} = t_a\{1, R\}$ so the first two terms cancel

$$\begin{aligned}
 \Delta t_{bA}\{FR\} &= \gamma_{BA} \frac{v_{BA}}{c^2} x_a\{1, F\} - \gamma_{BA} \frac{v_{BA}}{c^2} x_a\{1, R\} \\
 \Delta t_{bA}\{FR\} &= \gamma_{BA} \frac{v_{BA}}{c^2} (-x_a\{1, F\} - x_a\{1, R\}) \\
 \Delta t_{bA}\{FR\} &= \gamma_{BA} \frac{v_{BA}}{c^2} (x_a\{1, F\} - x_a\{1, R\})
 \end{aligned}$$

Looking at the part in parenthesis, we can say that this must be the length of Bob's ship

as measured by Alice

$$\Delta x_{bA} = L_{bA} = x_a\{1, F\} - x_a\{1, R\}$$

then

$$\Delta t_{bA}\{FR\} = \gamma_{BA} \frac{v_{BA}}{c^2} (L_{bA})$$

When Bob's ship is stationary, it has a length $L_{aA} = L_{bB}$ because we would measure the ship to be the same length in any stationary reference frame (that is, when the ship is sitting still in that frame). This is a proper length. But we know that lengths are not invariant. So the length of the ship measured by Alice, L_{bA} is shorter than the proper length, L_{bB} .

$$L_{bA} = \frac{L_{bB}}{\gamma_{BA}}$$

It would be more convenient to measure Bob's ship when it isn't moving. So let's say we do that. Then we know $L_{bB} = L_{aA} = L_p$. Then our time difference is

$$\begin{aligned} \Delta t_{bA}\{FR\} &= \gamma_{BA} \frac{v_{BA}}{c^2} \left(\frac{L_{bB}}{\gamma_{BA}} \right) \\ &= \frac{v_{BA} L_{bB}}{c^2} \end{aligned}$$

So Alice sees Bob's trailing clock as

$$\Delta t_{bA}\{FR\} = \frac{v_{BA} L_{bB}}{c^2}$$

seconds ahead of the leading clock. Or said another way, if Alice takes a picture of Bob's ship as it goes by, in the picture the front clock will be read $\Delta t_{bA}\{FR\}$ seconds behind the rear clock (so long as Bob is flying forward). The "leading" and "trailing" only correspond to "front" and "back" if the ship is flying forward.

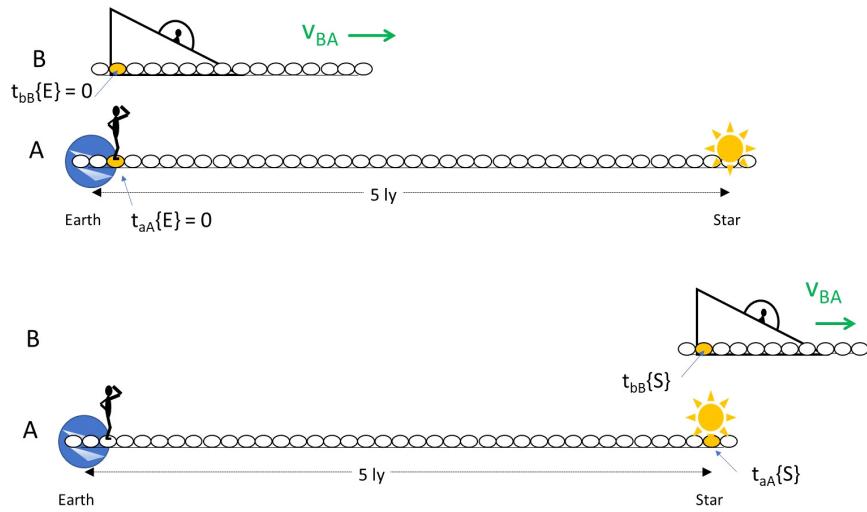
Of course Bob sees his clocks as synchronized. But Alice sees this time difference between his clocks. The subscripts are still a little hard. Think of $\Delta t_{bA}\{FR\}$ as the difference of the clock times on Bob's clocks as observed by Alice in the A frame.

Putting it all together

We should do a more difficult example.

Suppose we have Bob in a spaceship again and Alice on the Earth. Bob's space ship passes Alice with a relative speed of $v_{BA} = 0.9428c$ (so that $\gamma_{BA} = 3$). Bob is to travel to a distant star that Alice sees as 5 light years ($5ly$) from the Earth. Further suppose that one of Bob's clocks and one of Alice's clocks are zero when Bob passes the Earth ($t_{bB}\{1, E\} = 0$ and $t_{aA}\{1, E\} = 0$). The particular clocks are shown on the figure by making them orange.

- a) from Alice's perspective, how long does Bob's trip to the star take?
- b) Bob sends back data from his geoclock when he arrives at the star. What does Alice see on Bob's geoclock when the photo evidence from Bob's geoclock arrives?
- c) Bob sees himself as stationary in the *B* reference frame, and he sees the star moving at $-v_{BA}$ toward him. From Bob's perspective, how far away is the star as the Earth begins to move away from Bob? That is, what is the Earth-Star distance as viewed by Bob?
- d) What is the time on Alice's *A* frame clock at the star location when the star passes Bob?
- e) Alice sees Bob's trip take 5.3 y. Bob sees the star's trip take 0.589 y. But we use geoclocks and their photo evidence as observations, so both Bob and Alice must agree on what the clocks read. Use the results of our analysis of leading clocks to explain how these trip times are consistent.



- a) From Alice's perspective, how long does Bob's trip take?

We know

$$\begin{aligned} v_{BA} &= 0.9428c \\ \gamma_{BA} &= 3 \\ \Delta x_{aA}\{ES\} &= 5ly \\ t_b\{1, E\} &= 0 \\ t_a\{1, E\} &= 0 \end{aligned}$$

It is important to note any distances and time intervals by what their reference frame is and from which reference frame we view them. So $\Delta x_{aA}\{ES\}$ is marked with an *A* because we are viewing this distance from the *A* frame, and it is marked with an *a* because the objects (*a* geoclocks) are stationary in the *A* frame. So we immediately see this as a proper distance.

Within a reference frame, velocity is easy. It is just $\Delta x/\Delta t$ in that frame. If you are using my subscripts, you just need to check that the subscripts match. If you are using primes, you need to check that the primes match and mentally think if the view reference frames also match. For our case, if we view the space ship frame as an object moving in the *A* frame we can say write

$$v_{BA} = \frac{\Delta x_{aA}\{ES\}}{\Delta t_{aA}\{ES\}}$$

where $\Delta x_{aA}\{S\}$ is the distance from Alice to the star as viewed from the *A* frame and $t_{aA}\{S\}$ is the trip time to the star also viewed from the *A* reference frame. These subscripts match. So we have just $v = \Delta x/\Delta t$ because it is all in one frame. So the trip time can be written as

$$\Delta t_{aA}\{ES\} = \frac{\Delta x_{aA}\{ES\}}{v_{BA}}$$

and we know all the parts

$$\Delta t_{aA}\{ES\} = \frac{5ly}{0.9428c}$$

And now we can choose our units for *c*. Let's use some convenient units. Let's say that

$$c = 1 \frac{ly}{y}$$

A light year is how far light travels in a year (it is a distance) and so light will travel one light year in a year's time. Then *c* has a value of 1 ly/y and we have

$$\Delta t_{aA}\{ES\} = \frac{5ly}{0.9428 \left(1 \frac{ly}{y}\right)} = 5.3034 \text{ y}$$

and since we said $t_{aA}\{1, E\} = 0$ we expect the elapsed time to be just

$$\begin{aligned}\Delta t_{aA}\{ES\} &= t_a\{2, S\} - t_a\{1, E\} \\ \Delta t_{aA}\{ES\} &= t_a\{2, S\} - 0 \\ t_a\{2, S\} &= \Delta t_{aA}\{ES\}\end{aligned}$$

so Alice's star clock should read

$$t_a\{2, S\} = 5.3034 \text{ y}$$

as Bob flies by the star.

- b) Bob sends back data from his geoclock when he arrives at the star. What does Alice see on Bob's geoclock when the photo evidence from Bob's geoclock arrives?

When Bob arrives, we know that Alice will see Bob's clock as slow. Elapsed time changes due to time dilation (tick size expansion). We can write the elapsed time as

$$\Delta t_b\{2, S\} = \frac{1}{\gamma_{BA}} \Delta t_a\{2, S\}$$

and since $t_b\{1, E\} = 0$ we can say that

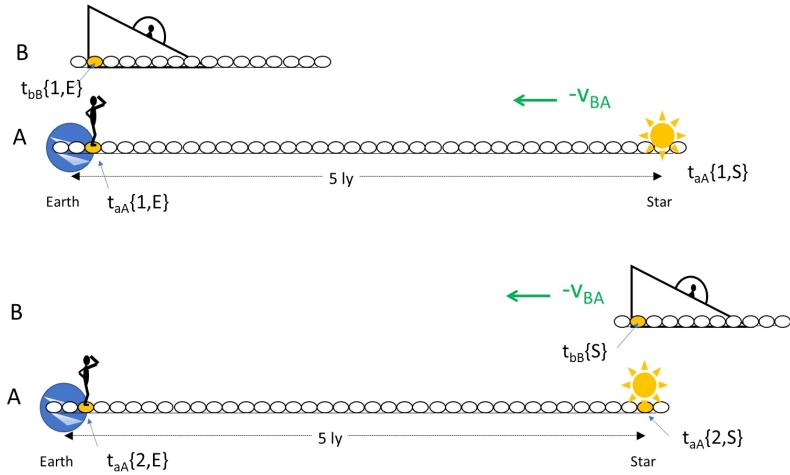
$$t_b\{2, S\} = \Delta t_b\{2, S\} - t_b\{1, E\} = \Delta t_b\{2, S\} - 0$$

must the time on Bob's clock as he arrives at the star. We have these pieces so let's write

$$t_b\{2, S\} = \frac{1}{3} (5.3034 \text{ y}) = 1.7678 \text{ y}$$

But you might worry about this. This is what Alice says Bob's clock should read. But the clock reads what it reads. Alice and Bob can't disagree on this. They can disagree on why the clock reads what it reads, but not on the reading, itself. So the reading on Bob's clock as he passes the star has to be 1.7678 y as viewed by Bob as well. And this seems a little strange. But we will resolve the strangeness in the next part of the problem.

- c) Bob sees himself as stationary in the B reference frame, and he sees the star moving at $-v_{BA}$ toward him. From Bob's perspective, how far away is the star as the Earth begins to move away from Bob? That is, what is the Earth-Star distance as viewed by Bob?



We know that objects in moving reference frames as viewed from other frames will experience length contraction. Bob sees the Earth - Star distance of Alice's frame to be moving relative to him with a speed $v_{AB} = -v_{BA}$. So Bob will see the distance contracted.

$$\begin{aligned}\Delta x_{AB}\{ES\} &= \frac{1}{\gamma_{BA}} \Delta x_{aA}\{ES\} \\ &= \frac{1}{3} (5ly) \\ &= 1.6667ly\end{aligned}$$

And now we see why Bob sees his geoclock read only $t_{bB}\{2,S\} = 1.7678y$. Bob sees the star travel only $1.6667ly$ and he sees the star come toward him at $v_{AB} = -0.9428c$. So he sees the time of the trip to be (note the matching subscripts).

$$\Delta t_{bB}\{ES\} = \frac{\Delta x_{AB}\{ES\}}{v_{AB}} = \frac{-1.6667ly}{-0.9428 \left(1\frac{ly}{y}\right)} = 1.7678y$$

So indeed Bob and Alice agree on the time reading of Bob's clock. But Bob would say this reading is due to a smaller distance traveled and Alice would say it is because Bob's clocks run slow.

d) How much time has passed on Alice's A frame clock at the star location from the start of our experiment until when the star passes Bob?

Bob's geoclock took a photograph with Bob next to the star $1.7678y$ after the trip started. Both Alice and Bob must agree on this. The photo can't be wrong or read two different things. So we know Bob sees the star travel for just $1.7678y$ to get to him.

And this is possible because he sees the star as only 1.6667 ly away. So it works.

$$v_{AB} = \frac{\Delta x_{aB}\{ES\}}{\Delta t_{bB}\{2, S\}} = \frac{1.6667 \text{ ly}}{1.7678 \text{ y}} = 0.94281 \frac{\text{ly}}{\text{y}}$$

which is just right in our new units for c . We can use time dilation again to find out what Alice's A frame clock at the star location will read when it passes Bob. This is an elapsed time, so

$$\begin{aligned}\Delta t_{aB}\{2, S\} &= \frac{1}{\gamma_{BA}} \Delta t_{bB}\{2, S\} \\ &= \frac{1}{3} (1.7678 \text{ y}) \\ &= 0.58927 \text{ y}\end{aligned}$$

So the A frame geoclock at the star will have an elapsed time of 0.58927 y *past its start time* since the star's voyage to Bob. Of course, Bob doesn't see Alice's clocks as synchronized. So we should probably worry about what that start time was for Alice's star location clock.

- e) Alice sees Bob's trip take 5.3 y. Bob sees the star's trip take 0.589 y. But we use geoclocks and their photo evidence as observations, so both Bob and Alice must agree on what the clocks read. Bob must agree that Alice's clock reads 5.3 y when the star reaches him. Use the results of our analysis of leading clocks to explain how these trip times are consistent.

Our notation and postulate 3 can help us here. First note that

$$t_{aA}\{2, S\} = 5.3034 \text{ y}$$

This is the time in the A frame at the star. But for Alice in the A frame, all A frame clocks are synchronized, so she sees the clock at the star and the clock on the Earth to both read 5.3034 y when Bob passes the star. But Bob won't see the clocks in the A frame as synchronized. This seems like a job for the Lorentz transformation. But we did this Lorentz problem in our last example so we know that the clocks will be off by

$$\Delta t_{bA}\{FR\} = \frac{v_{BA} L_{bB}}{c^2}$$

where L_{bB} was a proper distance. In our case, we want the proper distance between the Earth and the star. That would be the distance measured in the A frame because in that frame the Earth and the star are not moving.

$$L_{bB} = L_{aA} = \Delta x_{aA}\{ES\}$$

so we expect there to be a time difference between the star's clock and the Earth's clock

of

$$\begin{aligned}\Delta t_{bA}\{FR\} &= \frac{v_{BA}\Delta x_{aA}\{ES\}}{c^2} \\ &= \frac{\left(0.9428\left(1\frac{ly}{y}\right)\right)(5ly)}{\left(1\frac{ly}{y}\right)^2} \\ &= 4.714 \text{ y}\end{aligned}$$

but we need to think about how to apply this $\Delta t_{bA}\{FR\}$. From Bob's perspective, the Earth's clock will be the leading clock because everything in the A frame is moving to the left. The star location clock will be the trailing clock. We said that at the Earth, Alice's clock starts at zero $t_{aA}\{1, E\} = 0$. But since it is leading clock from Bob's perspective, then the Star's clock will already be ahead of the Earth's clock by an amount $\Delta t_{bA}\{FR\}$. The star's clock didn't start at zero according to Bob. So from the B frame perspective, the a star clock reading when our experiment started must have been.

$$t_{aB}\{1, S\} = 4.714 \text{ y}$$

Then the reading on the start's clock as it passes Bob would be

$$\begin{aligned}t_{aB}\{2, S\} &= t_{aB}\{1, S\} + t_{aB}\{2, S\} \\ &= \Delta t_{bA}\{FR\} + \Delta t_{aB}\{2, S\} \\ &= 4.714 \text{ y} + 0.58927 \text{ y} \\ &= 5.3033 \text{ y}\end{aligned}$$

So Bob does see the star's geoclock with 5.3033 y on it as he passes. All is consistent.

But Alice and Bob disagree on *why* the clocks read what they do. Alice sees her clocks as synchronized and sees Bob travel $5ly$ at $0.9428c$ so her clock at the Earth and her clock at the star will read 5.3 y. Bob sees the star travel just $1.6667ly$ at $0.9428c$ so he says the trip takes 0.58927 y, but he sees Alice's clocks as not synchronized. He sees Alice's clock at the star location as running ahead so when our experiment starts her star location clock already reads 4.7 y. So at the end of the experiment Bob sees Alice's star location clock read 5.3 y, but only because it was running ahead.

The observations are the same, the interpretations different, and all just because Alice and Bob use different reference frames. This is another cherished idea we give up in special relativity. The invariance of explanation for events! But really this is just a special form of the idea of the non-invariance of simultaneity.

Technically we used the Lorentz transformation in part e, but we did it borrowing a result from a previous problem.

As you can see, special relativity problems are tricky and take some careful thinking through.

6 Special Relativity Dynamics

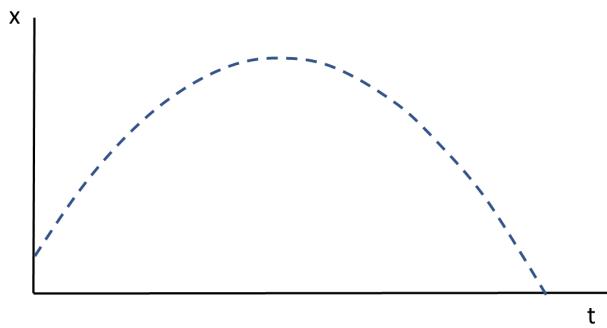
Sections 2.6 and 2.7

Fundamental Concepts in the Lecture

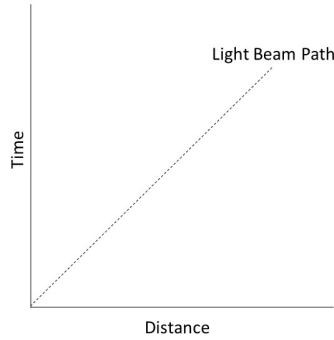
- Time Dilation and the Twin Paradox
- Spacetime Diagrams
- Velocity addition

Spacetime Diagrams and the Twin Paradox

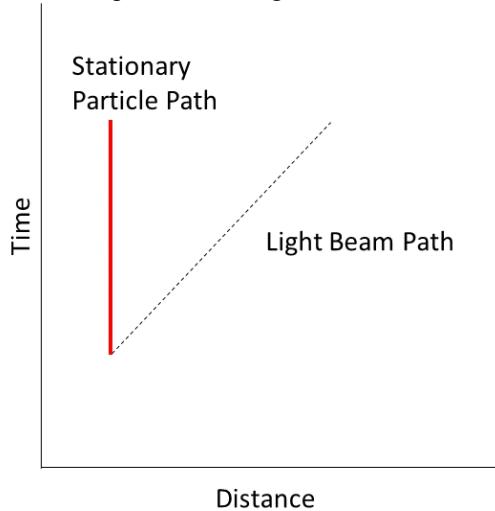
As we have derived the Lorentz transformations, you may have noticed that we placed the direction of motion along the x -axis for convenience. And that the transformation equations were different for the x -direction than for the y and z -directions. Often we reduce relativistic problems to one-dimensional problems with our one dimension in the direction of the motion because it is the direction that is effected most by the motion. We are used to doing this, we have been doing this since PH121.



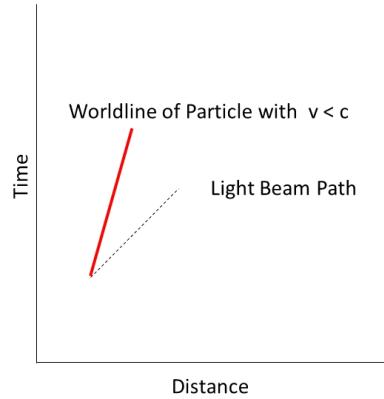
where we put the x axis as vertical and the time axis as horizontal. It turns out that this kind of graph will be helpful in special relativity. But we will switch the position of the graphs (because it is traditional). It is called a spacetime diagram and it looks like this.



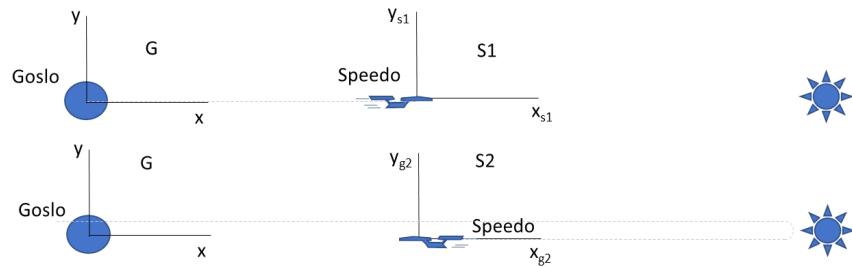
Time is on the vertical access and distance (along the direction of motion) is on the horizontal access. And let's set the units for this graph to be distance in light years and time in years. Then the path of a light beam would be a horizontal path at 45° . A stationary object would have a path on the diagram that looks like this.



The distance doesn't change with time. A path on a spacetime diagram is called a world line. The red line is the world line for a stationary particle. If the object has a speed v that is less than c .

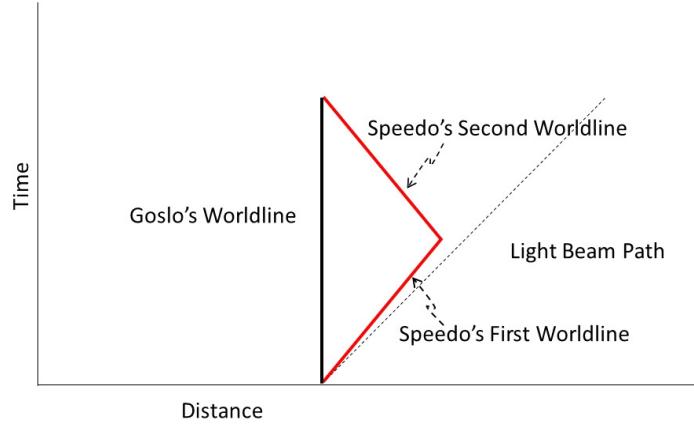


Let's analyze an example using spacetime diagrams. Let's take a specific example.



Two twins are involved in an experiment. One stays on Earth. The other travels in a space ship at a high speed, say, $0.6c$ to a star $3ly$ away. The Earthbound twin we call Goslo and the traveler we call Speedo (no swim suit jokes please). We would expect that Goslo will see Speedo's clocks as running slow. And from what we have done before we know that Speedo will see Goslo's clocks as running slow. But when the experiment is run, we find that when Speedo and Goslo meet again, Speedo seems to have aged less than Goslo. How do we resolve this apparent contradiction.

We can use such a diagram to show an asymmetry for our twins Speedo and Goslo.



Say Speedo goes $3ly$ and then comes back. Goslo would watch Speedo travel with a powerful telescope. He would conclude that the time of the trip to the star was

$$\begin{aligned}\Delta t_{sG}\{E, S\} &= \frac{2\Delta x_{gG}\{E, S\}}{v_{GS}} \\ &= \frac{3ly}{0.6 \left(1 \frac{ly}{y}\right)} \\ &= 5.0 \text{ y}\end{aligned}$$

But to Speedo the Earth and Star are what are moving. Speedo sees himself as stationary in his reference frame. And we know that moving lengths shorten. So the Earth-star distance is shorter for Speedo.

$$\begin{aligned}\Delta x_{gS}\{E, S\} &= \frac{1}{\gamma_{SG}} \Delta x_{gG}\{E, S\} \\ &= \frac{1}{(1.25)} (3ly) \\ &= 2.4ly\end{aligned}$$

So Speedo doesn't think the distance between the Earth and the star is all that far. Let's

put this into our time equation for the Earth-Star movement for Speedo.

$$\begin{aligned}
 \Delta t_{gS}\{S, E\} &= \frac{\Delta x_{gS}\{S, E\}}{v_{SG}} \\
 &= \frac{\frac{1}{\gamma_{SG}} \Delta x_{gG}\{E, S\}}{v_{SG}} \\
 &= \frac{1}{\gamma_{SG}} \frac{\Delta x_{gG}\{E, S\}}{v_{SG}} \\
 &= \frac{1}{(1.25)} \frac{-(3ly)}{-0.6 \left(1 \frac{ly}{y}\right)} \\
 &= 4.0 \text{ y}
 \end{aligned}$$

Thus, Speedo thinks the star comes to meet him in just 4 y. And since we started Speedo's clocks with $t_s\{1, E\} = 0$ this last value is just what should show up on Speedo's clock when he gets to the star.

Now if half the trip takes 5 y we might guess that the whole trip would take 10 y as viewed by Goslo and if half the trip takes 4 y then the whole trip should take 8 y as viewed by Speedo. And so Speedo will be younger than his brother when they get back together. And this is what we see on the spacetime diagram. The diagram we drew was Goslo's reference frame. From Goslo's perspective Speedo's moving clock should run slow.

But when Speedo observes Goslo's clocks he sees Goslo as moving. So Speedo sees Goslo's clocks as running slower. So

$$\begin{aligned}
 \Delta t_{gS}\{E, S\} &= \frac{1}{\gamma_{GS}} \Delta t_{sS}\{ES\} \\
 &= \frac{1}{1.25} (4.0 \text{ y}) \\
 &= 3.2 \text{ y}
 \end{aligned}$$

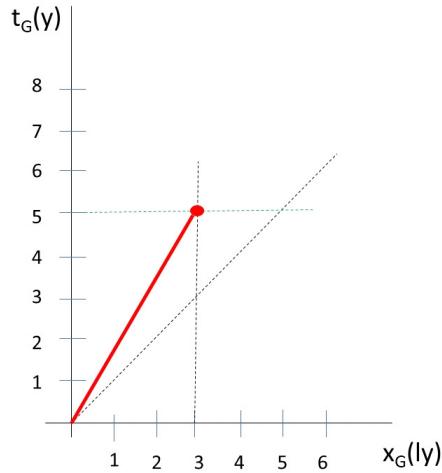
So Speedo thinks only 3.2 y have passed on Goslo's star geoclock.

So Goslo thinks Speedo's clocks are slow, and Speedo thinks Goslo's clocks are slow. Special relativity says both are right. But when Speedo gets back to Earth, it is clear that he is younger (and so would be any living things or any objects, like radioactive elements, etc.—we can do this kind of experiments, we just don't really do them on people).

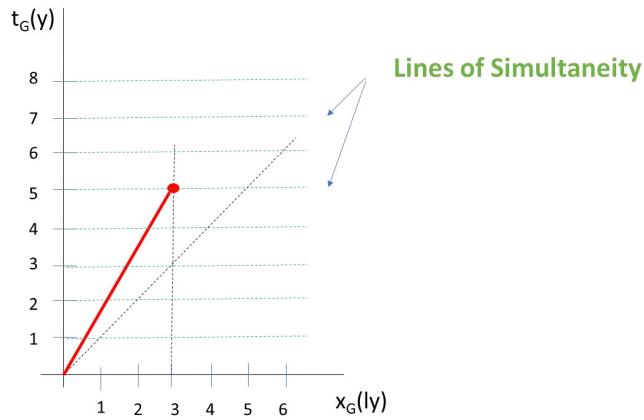
So how can this work?

Let's use the power of spacetime diagrams to illustrate what is happening. We can start with a spacetime diagram for the outbound trip. From Goslo's reference frame we

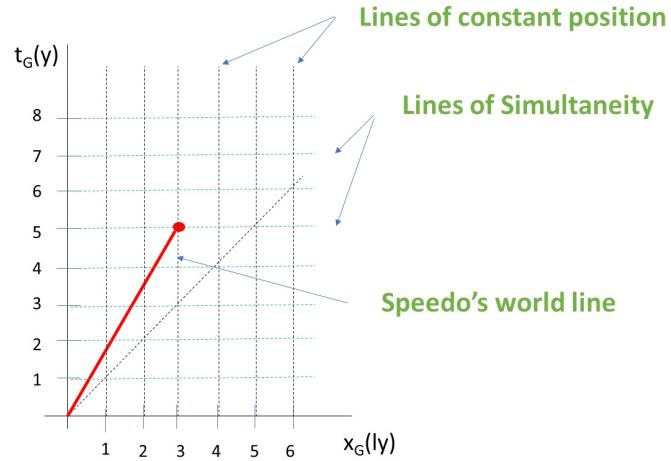
have



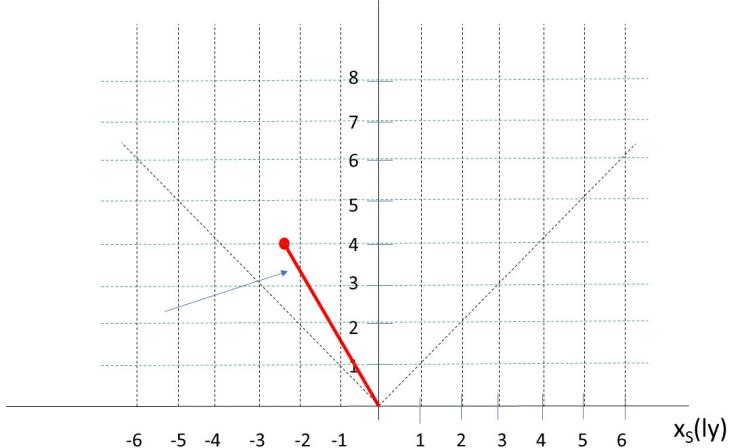
where the red thick line is Speedo's world line as viewed from the G frame. Note that in the G frame horizontal lines are lines of constant time. All events on this line are “at the same time.” We call these lines of simultaneity.



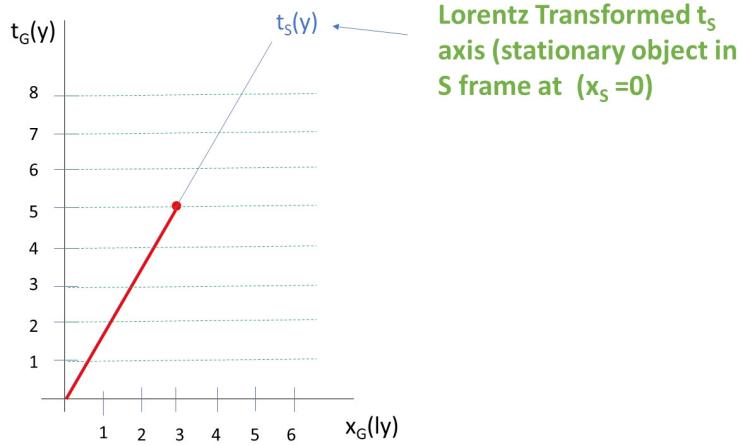
We could also add lines of constant position.



but this graph is all from Goslo's view point. Speedo would see the star moving. For the star's "outbound" trip Speedo's reference frame diagram would look like this.



It is a little hard to compare the two reference frames. So let's do something very strange. Let's use the Lorentz Transformation to transform Speedo's coordinate axes onto Goslo's spacetime diagram. This would let us see what both viewpoints observe with just one diagram. We can start with Speedo's time axes. Speedo is at rest in his reference frame. So his time axis in on his graph will have to be the world line on Goslo's graph.



We should transform a few points to mark whole year values on the line. We know for the t_s axis this is where $x_s = 0$. We can use our Lorentz Transformation

$$x_g = \gamma_{SG} (x_s - v_{SG} t_s)$$

$$y_g = y_s$$

$$z_g = z_s$$

$$t_g = \gamma_{SG} \left(t_s - \frac{v_{SG}}{c^2} x_s \right)$$

Then, using the last of the set of equations

$$\begin{aligned} t_g &= \gamma_{SG} \left(t_s - \frac{v_{SG}}{c^2} (0) \right) \\ &= (1.25) (t_s) \end{aligned}$$

and we can transform a few values

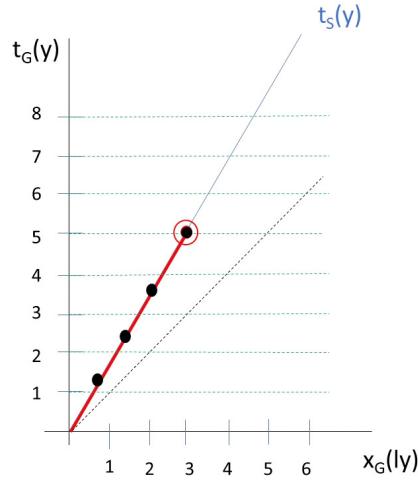
$$t_g = 1.25 (1) = 1.25$$

$$t_g = 1.25 (2) = 2.5$$

$$t_g = 1.25 (3) = 3.75$$

$$t_g = 1.25 (4) = 5.0$$

We can mark these points on our line for $t_s \{1\}$ through $t_s \{4\}$



The time axis was easy. Let's try the position axis. Let's write our Lorentz transformation for position

$$x_g = \gamma_{SG} (x_s - v_{SG} t_s)$$

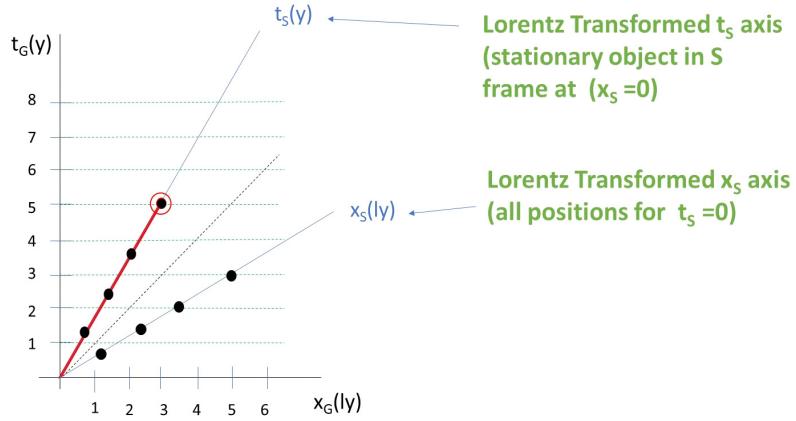
and since we know that the x_S axis is at $t_S = 0$ we can put this in

$$\begin{aligned} x_g &= \gamma_{SG} (x_s - v_{SG} t(0)) \\ &= \gamma_{SG} x_s \\ &= 1.25 x_s \end{aligned}$$

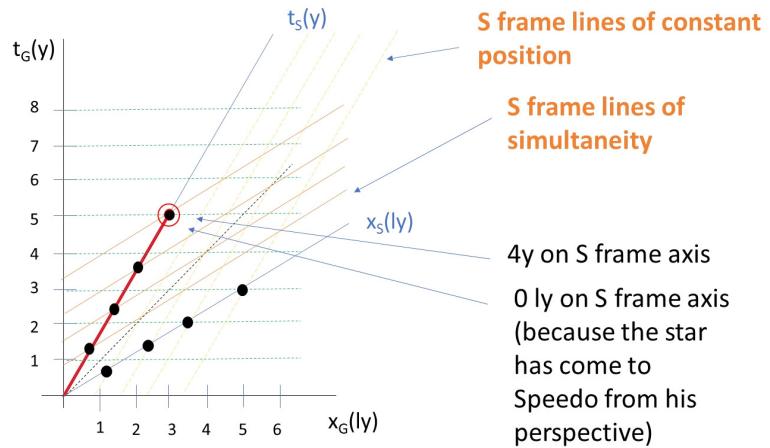
which is a line with slope of 1.25 and horizontal intercept of 0. Calculating a few values gives the picture

$$\begin{aligned} x_g &= 1.25(1) = 1.25 \\ x_g &= 1.25(2) = 2.5 \\ x_g &= 1.25(3) = 3.75 \\ x_g &= 1.25(4) = 5.0 \end{aligned}$$

Here is the graph



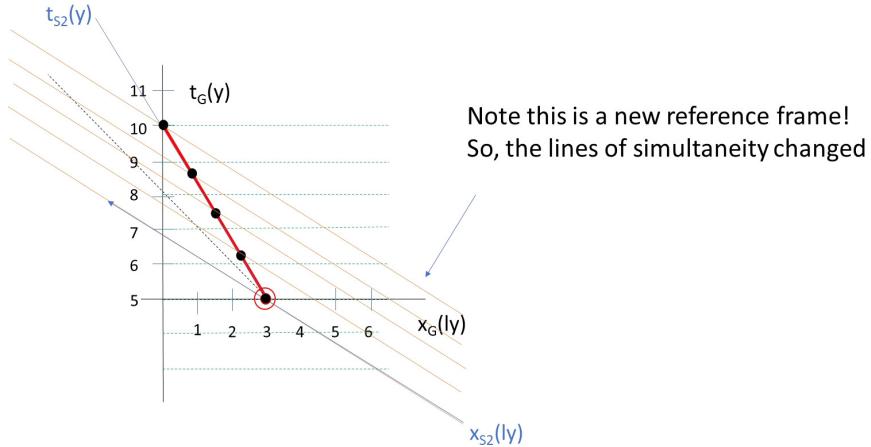
Note that the transformed x_s and t_s axes are symmetric about the worldline for light. We can add in S frame lines of simultaneity and constant position. the lines of simultaneity will be parallel to the x_s axis and the lines of constant position will be parallel to the t_s axis.



And look at the point with the red circle around it. This is when and where Goslo sees Speedo reach the star. It is as $x_g = 3ly$ and $t_g = 5y$. But if we use the skewed S frame axes we can see that it is at $x_s = 0$ (because the star has come to Speedo, who is at $x_s = 0$) and $t_s = 4y$. This is just what we calculated! And if we follow one of the slightly orange lines of S frame simultaneity to the t_G axis, we see that Speedo indeed sees Goslo's elapsed time as just about 3.2 y. This graph shows just what we calculated in a fairly understandable way.

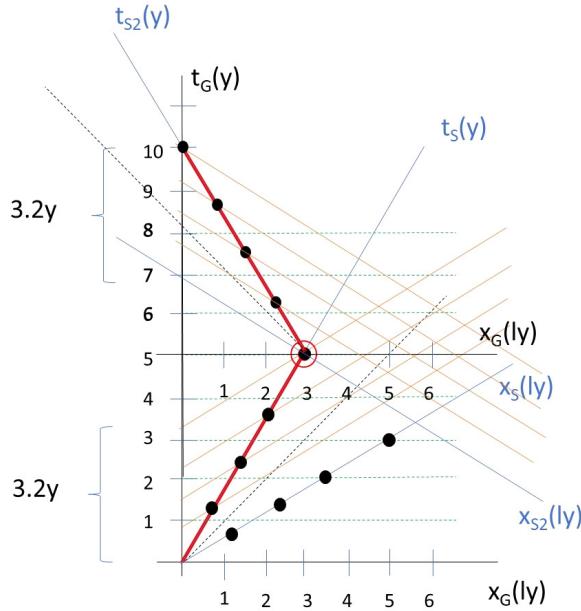
Now let's take on the return trip. And here is the important part, Speedo has to change

reference frames! This might be a quick deceleration followed by an acceleration. But he does change reference frames for the return trip. Let's draw a new diagram from the G frame perspective for the return trip.



The world line in the G frame must go from $3ly$ back to $0ly$. And we already guessed that this will take $5y$. Once again we can use the lorentz transform to find t_{s2} points on this line and to transform them into our diagram. We can use the Lorentz transform, or use the fact that the axes must be symmetric across the c world line to find the x_{s2} axis. I won't calculate points along this axes (though we could) because I just need the lines of simultaneity to finish this particular problem. But those lines of simultaneity are parallel to the transformed x_s axis. So we need at least the axis.

But notice! Something momentous has happened. The lines of simultaneity are very different than for the outbound trip. Let's put the two graphs together.



Now we see something spectacular. Notice that for the return trip we can see that Speedo sees 4 y elapse on his clock on the return trip (count the year dots on the line) but by following the lines of simultaneity we see that Speedo sees Goslo age just 3.2 y. He sees Speedo's clock going slower. Also notice that Goslo sees Speedo's clock tick off only 4 y (again count the dots) while his own ticks off 5 y for the whole return trip. So once again Goslo sees Speedo's clock tick slow. For the legs of the trip the situation is absolutely symmetric. But there is a big difference between Goslo and Speedo's trips. At the turnaround point just Speedo switches reference frame.

Let's think about this switch. Speedo sees Goslo's clock located at the star $t_g\{S\}$ as being ahead of Goslo's clock located at the Earth $t_g\{E\}$ by an amount

$$\Delta t = \frac{\Delta x_{gG}\{E, S\}v_{SG}}{c^2}$$

due to leading clocks lagging trailing clocks. The Earth clock is first as Speedo sees the Earth-Star system move to the left. So the Earth clock will be behind by.

$$\begin{aligned} \Delta t &= \frac{(3ly) 0.6 \left(1 \frac{ly}{y}\right)}{\left(1 \frac{ly}{y}\right)^2} \\ &= 1.8 \text{ y} \end{aligned}$$

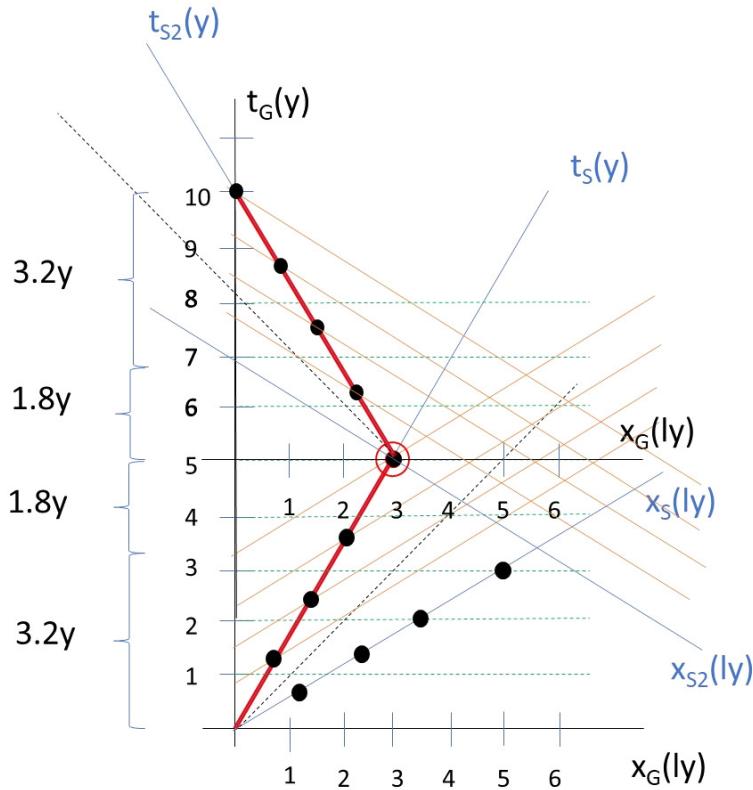
So Speedo does indeed see Goslo's star clock tick off 3.2 y as the star travels to him. But Speedo sees the star's clock start at $t_g\{1, S\} = 1.8 \text{ y}$. So the clock reading as

Speedo and the star are collocated is

$$t_g\{2, S\} = 1.8 \text{ y} + 3.2 \text{ y} = 5.0 \text{ y}$$

and both Speedo and Goslo agree on the clock reading as Speedo reaches the star. (Remember, Goslo sees all his clocks as synchronized in his reference frame, so if the star clock says 5 y, so does the Earth clock in Goslo's frame.)

Now for the return trip we see that the direction has changed, so Speedo now sees the star clock leading and the Earth clock trailing. Once again there will be a 1.8 y difference, but it will be the other way. As soon as Speedo is back up to speed after the turn around he sees a jump in Goslo's clock because the leading clock has switched to the star clock so the "lag" has to switch clocks.



So as Speedo changes direction he sees Goslo's clock go from 3.2 y to $3.2 \text{ y} + 2 \times 1.8 \text{ y} = 6.8 \text{ y}$! And it is Speedo that fires rockets to change direction. So we know it is Speedo that has this reference frame change, not Goslo. The $2 \times 1.8 \text{ y} = 3.6 \text{ y}$ clock shift is a product of changing reference frames, and it makes all the difference.

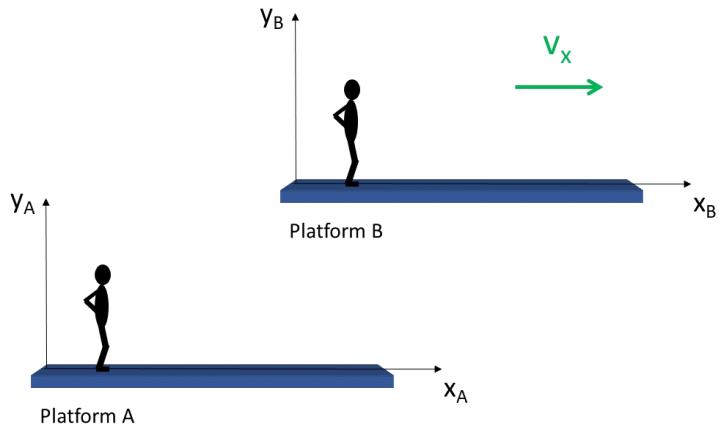
This illustrates one of the important steps in doing relativity problems. You must go

through the process of determining the reference frames, which times are proper times, and which lengths are proper lengths. Mentally setting up the problem is often more work than the math to get an answer. In our twin paradox problem, we need to know which twin is the stationary twin before we can use math to do any calculations. This is going to be a necessary step in our problems.

I'm sure you have noticed, we have a transformation for position, but not yet for velocities. We had a Galilean transformation for velocities. We need to complete our Lorentz transformation. We will take this up next.

Velocity Transforms

Let's review the Galilean Transformation for a minute



For position we have the transform

$$\begin{aligned}x_A &= x_B + v_{AB}t \\y_A &= y_B \\z_A &= z_B\end{aligned}$$

where the subscript a indicates a measured position in the A -frame and the subscript b indicates a measured position in the B -frame. The velocities also transform. Suppose something is moving in frame B as viewed by a person in frame B with speed v_{bB} , but we view the object from frame A . The speed the person in frame A sees for this object is v_{bA} . For the Galilean transformation we just add the velocities in the direction of

motion.

$$\begin{aligned} v_{bAx} &= v_{bBx} + v_{BA} \\ v_{bAxy} &= v_{bBy} \\ v_{bAz} &= v_{bBz} \end{aligned}$$

where in this case the first subscript told us what object was moving and the second subscript told us what reference frame we were using as our view point. We can get the velocity transform from the position transform. First let's identify an object b that is in the B reference frame. Then we can write

$$x_{bA} = x_{bB} + v_{AB}t$$

and now take a derivative with respect to time

$$\begin{aligned} v_{bAx} &= \frac{dx_{bA}}{dt} \\ &= \frac{dx_{bB} - v_{AB}dt}{dt} \\ &= \frac{dx_{bB}}{dt} + v_{AB} \\ &= v_{bBx} + v_{AB} \end{aligned}$$

We want something like this for special relativity. And we have part of it already. We have the coordinate transform part.

$$\begin{aligned} x_a &= \gamma_{BA}(x_b + v_{BA}t_b) & x_b &= \gamma_{BA}(x_a - v_{BA}t_a) \\ y_a &= y_b & y_b &= y_a \\ z_a &= z_b & z_b &= z_a \\ t_a &= \gamma_{BA}\left(t_b + \frac{v_{BA}}{c^2}x_b\right) & t_b &= \gamma_{BA}\left(t_a - \frac{v_{BA}}{c^2}x_a\right) \end{aligned}$$

where here the small letter subscripts tell which reference frame is used to measure the coordinate and the capital subscripts (BA) show the reference frames. But we don't have the velocity transform yet. These equations tell us how to find the position in one reference frame of an object that is in a different reference frame. We called them the Lorentz transformation. Our goal was to find the new equations for special relativity that allow us to find the velocity of an object in one reference frame for an object that is moving in a different reference frame. To do this we will take our Lorentz transform equations and take a derivative. But we know from the Galilean transformation we need two subscripts, one to identify what is moving and one to identify the reference frame. So we might have an object b (maybe Bob walking around in his space ship) moving in reference frame B . We want to know what speed we see when we look from reference frame A . In reference frame A , Bob's speed would still be

$$v_{bAx} = \frac{dx_{bA}}{dt_{bA}}$$

and we know how to find dx_{bA}

$$dx_{bA} = \gamma_{BA} (dx_{bB} + v_{BA} dt_{bB})$$

Note that in taking the derivative we now have a displacement and a duration and we are smart enough at this point in our study of special relativity to know that we need to label the viewing reference frame for these. So now the first subscript is telling us the object (b for Bob) that is moving and the capital letter subscripts are telling us which frame we are using to view the moving object (A). In the Lorentz coordinate transformation small letters told us about which reference frame we were measuring in. That is like where Bob is in the reference frame. The coordinate transformation did not need a viewpoint subscript, but changes in position and time do (think length contraction and time dilation). And velocity is made from changes in position and time (dx and dt are just small Δx and Δt values, so they will need a viewpoint) so we should use the viewpoint subscript.

We have dx_{bA} , we will also need

$$dt_{bA} = \gamma_{BA} \left(t_{bB} + \frac{v_{BA}}{c^2} x_{bB} \right)$$

Putting together our dx_{bA} and our dt_{bA} we get

$$\begin{aligned} v_{bAx} &= \frac{\gamma_{BA} (dx_{bB} + v_{BA} dt_{bB})}{\gamma_{BA} (dt_{bB} + \frac{v_{BA}}{c^2} dx_{bB})} \\ &= \frac{dt_{bB} \left(\frac{dx_{bB}}{dt_{bB}} + v_{BA} \right)}{dt_{bB} \left(1 + \frac{v_{BA}}{c^2} \frac{dx_{bB}}{dt_{bB}} \right)} \\ &= \frac{(v_{bBx} + v_{BA})}{\left(1 + \frac{v_{BA}}{c^2} v_{bBx} \right)} \end{aligned}$$

so

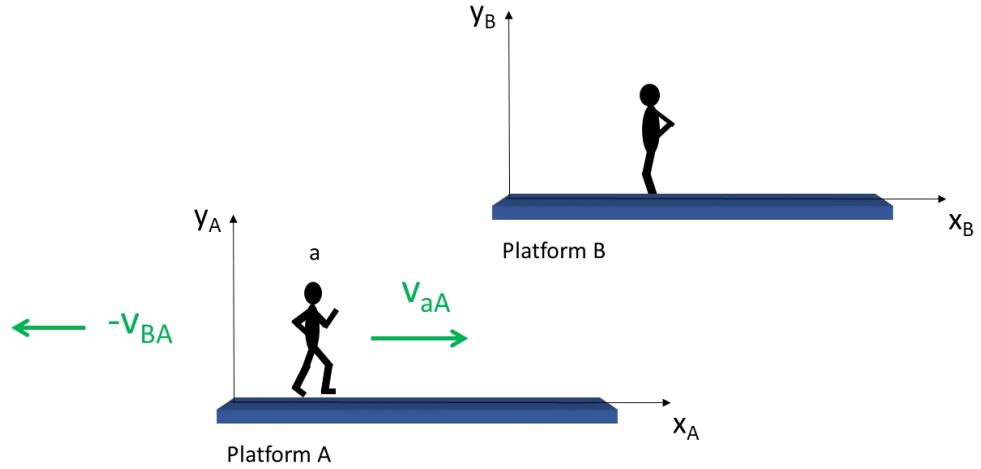
$$v_{bAx} = \frac{(v_{bBx} + v_{BA})}{\left(1 + \frac{v_{BA}}{c^2} v_{bBx} \right)}$$

This notation matches our Galilean subscript system

$$\begin{aligned} v_{bAx} &= v_{bBx} + v_{BA} \quad \text{Galilean} \\ v_{bAx} &= \frac{(v_{bBx} + v_{BA})}{\left(1 + \frac{v_{BA}}{c^2} v_{bBx} \right)} \quad \text{Special Relativity} \end{aligned}$$

Of course we could reverse this. Suppose we let the person in reference frame B stand still and we let the person in reference frame A move. We would find

$$v_{aBx} = \frac{(v_{aAx} - v_{BA})}{\left(1 - \frac{v_{BA}}{c^2} v_{aAx} \right)}$$



For low frame speeds v_{BA} this velocity transform equation has to reduce to the Galilean velocity transformation. At low speeds, $\frac{v_{BAx}}{c^2} v_{aA} \approx 0$ so

$$v_{aBx} \approx \frac{(v_{aAx} - v_{BA})}{(1 - 0)} = (v_{aAx} - v_{BA})$$

which is, indeed, our Galilean velocity transformation.

We found the velocity transformation only for the x -direction so far. We will need the y and z -directions as well. You might think that the velocities might be the same in the y and z -directions. But let's repeat what we have done for the y -direction to check this. We know that $y_b = y_a$ because the y positions don't change under the Lorentz transformation. Ah but the time *does* change! And we know how the time changes because we know dt_{bA} from above so

$$\begin{aligned} v_{bAy} &= \frac{dy_{bA}}{dt_{bA}} \\ v_{bAy} &= \frac{dy_{bB}}{\gamma_{BA} (dt_{bB} + \frac{v_{BA}}{c^2} dx_{bB})} \\ &= \frac{dy_{bB}}{dt_{bB} \gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} \frac{dx_{bB}}{dt_{bB}}\right)} \\ &= \frac{\frac{dy_{bB}}{dt_{bB}}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} \frac{dx_{bB}}{dt_{bB}}\right)} \\ &= \frac{v_{bBy}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} v_{bBx}\right)} \end{aligned}$$

so

$$v_{bAy} = \frac{v_{bBy}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} v_{bBx}\right)}$$

and we could do the same thing for the z component

$$v_{bAz} = \frac{v_{bBz}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} v_{bBx} \right)}$$

To summarize, the *Lorentz velocity transformation* is

$$\begin{aligned} v_{bAx} &= \frac{(v_{bBx} + v_{BA})}{\left(1 + \frac{v_{BA}}{c^2} v_{bBx} \right)} \\ v_{bAy} &= \frac{v_{bBy}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} v_{bBx} \right)} \\ v_{bAz} &= \frac{v_{bBz}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} v_{bBx} \right)} \end{aligned}$$

and the inverse transform is

$$\begin{aligned} v_{aBx} &= \frac{(v_{aAx} - v_{BA})}{\left(1 - \frac{v_{BA}}{c^2} v_{aAx} \right)} \\ v_{aBy} &= \frac{v_{aAy}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} v_{aAx} \right)} \\ v_{aBz} &= \frac{v_{bBz}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2} v_{aAx} \right)} \end{aligned}$$

Notice that the x -component of the an object's velocity does effect the y and z -components of the transformed velocity.

Let's try a problem. Suppose we have a rocket ship going $v_{BA} = 0.8c\hat{i}$ and we have an object, b , in reference frame B that we launch from the space ship (say a "probe") with velocity $v_{bB} = 0.7c\hat{i}$. What is the speed of the probe as viewed from reference frame A ?

The Galilean transformation would give

$$v_{pAx} = (v_{pBx} + v_{BA}) = 0.7c + 0.8c = 1.5c$$

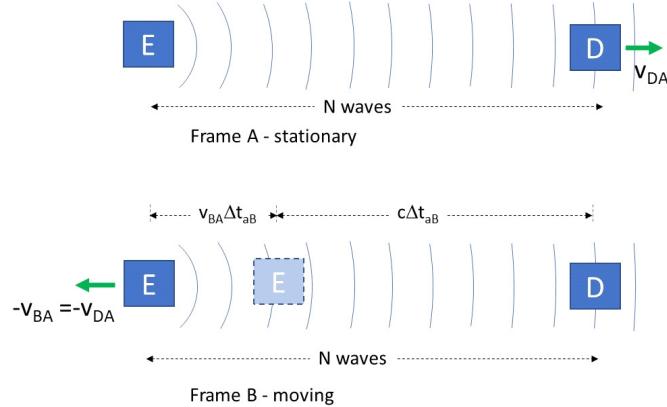
but we are suspicious. We have heard that nothing can go faster than the speed of light in vacuum (something that we will soon show to be true). But it is no surprise that our Galilean transformation failed. Let's try the Lorentz velocity transformation

$$\begin{aligned} v_{pAx} &= \frac{(v_{pBx} + v_{BA})}{\left(1 + \frac{v_{BA}}{c^2} v_{pBx} \right)} \\ &= \frac{0.7c + 0.8c}{1 + \frac{(0.7c)(0.8c)}{c^2}} \\ &= \frac{0.7c + 0.8c}{1 + (0.7)(0.8)} \\ &= 0.96154c \end{aligned}$$

This is in keeping with our statement that nothing can go faster than the speed of light.

Doppler shift

Let's think about wave motion and the Doppler shift. In our relativistic world, we can't have a distinction between reference frames like we did when we derived Doppler shift back in PH123. But we know from experiment that Doppler shift happens with light. Let's try this using our understanding of time dilation to see if we can find how this works.



Our situation is a light source at position E (for emitter) that is stationary in reference to frame A . In frame A we have a moving detector, D , with velocity v_{DA} . The light has made some N waves between E and D of frequency f . In frame A it took (a clock viewed from the A frame) $\Delta t_{aA} = NT = N/f_A$ to create these N waves. This is a proper time because the light emitter (the timing device) is at rest in the A frame. Now let's change reference frames to the B frame where D is stationary and E is moving with velocity $-v_{BA}$ from the perspective of D . In frame B , E has moved to the left. The wavelength of the waves can be found by taking the distance between E and D and dividing by N (regardless from which view point we choose, there are still N waves).

$$\lambda_B = \frac{L_{N \text{ waves}}}{N}$$

The distance L_N will be larger viewed from the B frame by the amount $v_{BA} \Delta t_{aB}$ because E is moving in frame B . So the wavelength will be larger than just the speed of the waves (c) times the time Δt_{aA} . We can say that the B frame is moving with velocity $v_{BA} = -v_{DA}$ so the light travels a distance

$$L_N = c \Delta t_{aB} + v_{BA} \Delta t_{aB}$$

in making N waves. Then the wavelength in the B frame will be

$$\lambda_B = \frac{c\Delta t_{aB} + v_{BA}\Delta t_{aB}}{N}$$

And in frame B the frequency must be

$$f_B = \frac{c}{\lambda_B}$$

where of course c is the same in both frames. We know λ_B so we can just substitute

$$f_B = \frac{c}{\frac{c\Delta t_{aB} + v_{BA}\Delta t_{aB}}{N}}$$

and we know that $\Delta t_{aA} = N/f_A$ so

$$N = \Delta t_{aA}f_A$$

and we can put this into our formula for f_B

$$\begin{aligned} f_B &= \frac{c}{\frac{c\Delta t_{aB} + v_{BA}\Delta t_{aB}}{\Delta t_{aA}f}} \\ &= \frac{c\Delta t_{aA}f_A}{c\Delta t_{aB} + v_{BA}\Delta t_{aB}} \\ &= \frac{cf_A\Delta t_{aA}}{(c + v_{BA})\Delta t_{aB}} \\ &= \frac{f_A}{\left(1 + \frac{v_{BA}}{c}\right)} \frac{\Delta t_{aA}}{\Delta t_{aB}} \end{aligned}$$

But now we have both Δt_{aA} and Δt_{aB} in the same equation. We know how these two relate (time dilation)

$$\Delta t_{aB} = \gamma_{BA}\Delta t_{aA}$$

Then f_B would be

$$\begin{aligned} f_B &= \frac{f_A}{\left(1 + \frac{v_{BA}}{c}\right)} \frac{\Delta t_{aA}}{\gamma_{BA}\Delta t_{aA}} \\ &= \frac{f_A}{\left(1 + \frac{v_{BA}}{c}\right)} \frac{\Delta t_{aA}\sqrt{\left(1 - \frac{v_{BA}^2}{c^2}\right)}}{\Delta t_{aA}} \\ &= \frac{f_A\sqrt{\left(1 - \frac{v_{BA}^2}{c^2}\right)}}{\left(1 + \frac{v_{BA}}{c}\right)} \\ &= \frac{f_A\sqrt{\left(1 - \frac{v_{BA}}{c}\right)\left(1 + \frac{v_{BA}}{c}\right)}}{\left(1 + \frac{v_{BA}}{c}\right)} \\ &= f_A \frac{\sqrt{\left(1 - \frac{v_{BA}}{c}\right)}}{\sqrt{\left(1 + \frac{v_{BA}}{c}\right)}} \end{aligned}$$

So the frequency we would see in frame B is

$$f_B = f_A \frac{\sqrt{(1 - \frac{v_{BA}}{c})}}{\sqrt{(1 + \frac{v_{BA}}{c})}}$$

This is the equation for Doppler shift for light. We will need this for building the universe later.

7 Relativistic Conservation Laws

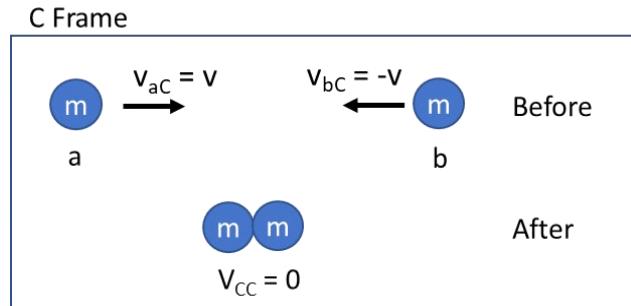
Section 2.8 and 2.9

Fundamental Concepts in the Lecture

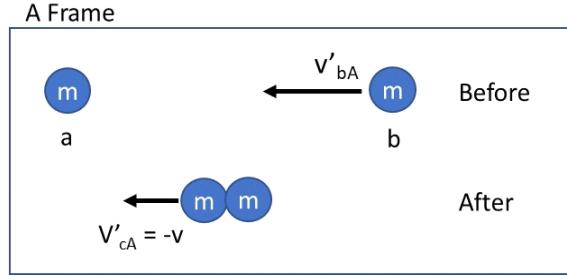
- Relativistic momentum
- Relativistic kinetic energy
- Relativistic total energy

We know from PH121 that momentum is conserved in collisions. We also know that the values of the momenta depend on reference frame. We want the laws of physics, like conservation of momentum, to still work in any reference frame. We could check this with our new Lorentz transformation. Let's take a very specific particle collision.

Two identical particles, each of mass m , move toward each other with speed v . They collide totally inelastically. We can envision doing this experiment in one of our labs, so we could call it the lab frame. But notice that the combined particle cc isn't moving in this frame. In the past we have named frames where an object is stationary in that frame after the object. In this case c is stationary in the frame so let's call the frame C .



Now let's observe this interaction from a frame that moves along with particle a . The a particle will be stationary in this new frame, so let's call it the A frame.



We expect that the values of the momenta would change when we change frames, but that momentum would still be conserved.

In the C frame we know

$$v_{aC\ i} = v$$

$$v_{bC\ i} = -v$$

$$v_{cC\ f} = 0$$

and classically we have momentum

$$\begin{aligned} P_i &= p_{aC\ i} + p_{bC\ i} \\ &= mv_{aC\ i} + mv_{bC\ i} \\ &= mv + m(-v) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} P_f &= 2m(0) \\ &= 0 \end{aligned}$$

and momentum is conserved.

Now let's try the A frame. We know our Lorentz velocity transform,

$$\begin{aligned} v_{bAx} &= \frac{(v_{bBx} + v_{BA})}{\left(1 + \frac{v_{BA}}{c^2}v_{bBx}\right)} \\ v_{bAy} &= \frac{v_{bBy}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2}v_{bBx}\right)} \\ v_{bAz} &= \frac{v_{bBz}}{\gamma_{BA} \left(1 + \frac{v_{BA}}{c^2}v_{bBx}\right)} \end{aligned}$$

so let's use it for our momentum problem. Let's write these for our C and A reference

frames.

$$\begin{aligned} v_{oAx} &= \frac{(v_{oCx} - v_{AC})}{\left(1 - \frac{v_{AC}}{c^2} v_{oAx}\right)} \\ v_{oAy} &= \frac{v_{oCy}}{\gamma_{AC} \left(1 - \frac{v_{AC}}{c^2} v_{oCy}\right)} \\ v_{oAz} &= \frac{v_{oCz}}{\gamma_{AC} \left(1 - \frac{v_{AC}}{c^2} v_{oCz}\right)} \end{aligned}$$

where I have used an nice generic “o” as the object subscript. But we have three particles so we will have to put in their subscripts as we go. Let’s write the first one of these for particle a the x direction with lab and A frame subscripts.

$$v_{aAi} = \frac{v_{aCxi} - v_{AC}}{\sqrt{1 - \frac{v_{AC}}{c^2} v_{aCxi}}}$$

In this particular example, we ride with particle a so $v_{aCxi} = v_{AC} = v$

$$\begin{aligned} v_{aAi} &= \frac{v - v}{\sqrt{1 - \frac{v^2}{c^2}}} \\ &= 0 \end{aligned}$$

and for particle b

$$\begin{aligned} v_{bAi} &= \frac{v_{bCi} - v_{AC}}{\sqrt{1 - \frac{v_{AC}}{c^2} v_{bCxi}}} \\ v_{bAi} &= \frac{-v - v}{\sqrt{1 - \frac{-vv}{c^2}}} \\ &= \frac{-2v}{\sqrt{1 + \frac{v^2}{c^2}}} \end{aligned}$$

and for the combined particle C

$$\begin{aligned} v_{cAx} &= \frac{(v_{cCx} - v_{AC})}{\left(1 - \frac{v_{AC}}{c^2} v_{cCxi}\right)} \\ &= \frac{0 - v}{\sqrt{1 - \frac{(0)v}{c^2}}} \\ &= -v \end{aligned}$$

Now let’s try our classical momentum equations

$$\begin{aligned}
P_{Cxi} &= p_{aCxi} + p_{bCxi} \\
&= mv_{aCxi} + mv_{bCxi} \\
&= m(0) + m \left(\frac{-2v}{\sqrt{1 + \frac{v^2}{c^2}}} \right) \\
&= \frac{-2vm}{\sqrt{1 + \frac{v^2}{c^2}}}
\end{aligned}$$

and

$$\begin{aligned}
P_{cxr} &= 2m(-v) \\
&= -2mv
\end{aligned}$$

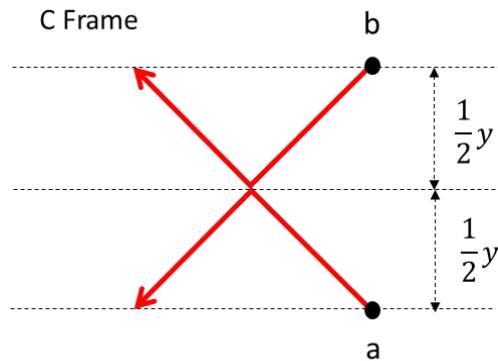
which are not the same. We have a problem. We will need to do more work in developing our theory of special relativity to rescue conservation of momentum.

Relativistic momentum

To remedy our momentum problem, let's hypothesize that the momentum changes in a similar way to our position transform. The relativistic momentum will be larger or smaller by some factor that we will once again call G .

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

Our job would be to find a G that allows the postulates of special relativity to hold, and that reduces to the classical case when speeds are slow. To find G let's take another very specifically chosen particle interaction. Two identical particles with mass m approach with the same velocity, but now the interaction is two dimensional.



Conservation of momentum tells us that the change in momentum for particle a must be the same as the change in momentum for particle b . Let's call our reference frame frame C . Then

$$\begin{aligned}\Delta p_{aCy} &= \Delta p_{bCy} \\ \Delta p_{aCx} &= \Delta p_{bCx}\end{aligned}$$

and for our current problem of finding G , let's look at just the y -part.

$$\begin{aligned}p_{aCyi} &= G_{aC}m_A v_{aCyi} \\ &= G_{aC}m \left(\frac{\frac{1}{2}y_C}{\Delta t_{aci}} \right)\end{aligned}$$

and after the collision in our very special case the speed of particle a is the same only negative

$$\begin{aligned}p_{aCyi} &= G_{aC}m_a v_{aCyi} \\ &= -G_{aC}m \left(\frac{\frac{1}{2}y_C}{\Delta t_{aci}} \right)\end{aligned}$$

because we assume the mass doesn't change, and because the times Δt_{aci} and Δt_{acf} will be the same

$$\begin{aligned}\Delta p_{aCy} &= p_{aCyi} - p_{aCyi} \\ &= G_{aC}m \left(\frac{\frac{1}{2}y_C}{\Delta t_{aci}} \right) - \left(-G_{aC}m \left(\frac{\frac{1}{2}y_C}{\Delta t_{acf}} \right) \right) \\ &= 2G_{aC}m \left(\frac{\frac{1}{2}y_C}{\Delta t_{aci}} \right) \\ &= 2G_{aC}m \left(\frac{y_C}{\Delta t_{ac}\{\text{total}\}} \right)\end{aligned}$$

Likewise,

$$\Delta p_{bCy} = 2G_{bC}m \left(\frac{y_C}{\Delta t_{bc}\{\text{total}\}} \right)$$

Because $\Delta t_{aC}\{\text{total}\} = \Delta t_{bC}\{\text{total}\}$ due to the symmetry of the case we chose to study,

$$\Delta p_{aCy} = \Delta p_{bCy}$$

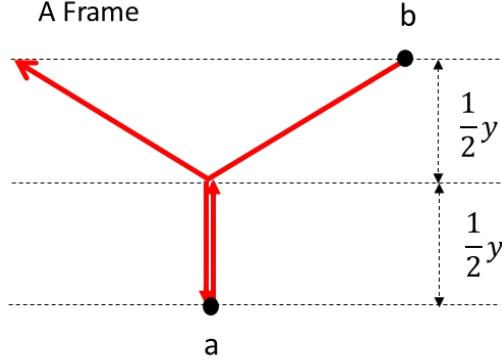
becomes

$$\begin{aligned}2G_{aC}m \left(\frac{y_C}{\Delta t_{aC}\{\text{total}\}} \right) &= 2G_{aC}m \left(\frac{y_C}{\Delta t_{bC}\{\text{total}\}} \right) \\ G_{aC} &= G_{bC}\end{aligned}$$

in the C frame.

But now let's consider a frame, A , where we ride along with $v_{AC} = v_{aCx}$. That is, the A reference frame moves at the speed of the particle a x -component as viewed from

frame C . In this frame particle a is not moving in the x direction (which is why we will name this frame A , it is the frame where a is stationary in x). We would see our collision like this



In this frame we have

$$\Delta p_{aA}y = -2G_{aAm} \left(\frac{y_A}{\Delta t_{aA}\{\text{total}\}} \right)$$

we don't expect the y distance to change because all of our relative motion is in the x -direction, so $y_C = y_A$

$$\Delta p_{aA}y = -2G_{aAm} \left(\frac{y_C}{\Delta t_{aA}\{\text{total}\}} \right)$$

and likewise

$$\Delta p_{bA}y = -2G_{bAm} \left(\frac{y_C}{\Delta t_{bA}\{\text{total}\}} \right)$$

but just looking at the figure reminds us of our boat example that we used to derive time dilation. We expect this case won't be different. We expect that $\Delta t_{aA}\{\text{total}\} \neq \Delta t_{bA}\{\text{total}\}$ so

$$\Delta p_{aA}y = \Delta p_{bA}y$$

becomes

$$\begin{aligned} -2G_{aAm} \left(\frac{y_C}{\Delta t_{aA}\{\text{total}\}} \right) &= -2G_{bAm} \left(\frac{y_C}{\Delta t_{bA}\{\text{total}\}} \right) \\ \frac{G_{aA}}{\Delta t_{aA}\{\text{total}\}} &= \frac{G_{bA}}{\Delta t_{bA}\{\text{total}\}} \end{aligned}$$

and remember that $\Delta t_{aA}\{\text{total}\} = t_{aAf} - t_{aAi}$ and we might try to use the Lorentz transformation to find these times, but we quickly realize that we don't know the actual locations x_{aAi} , x_{aAf} and this will be a problem in using the transformation equations. We can get around this by considering something. The time $\Delta t_{aA}\{\text{total}\}$ has both t_{aAf} and t_{aAi} measured in the same location. Particle A does move in the y -direction, but not at all in the direction of the relative motion. This is like the clock ticks on the

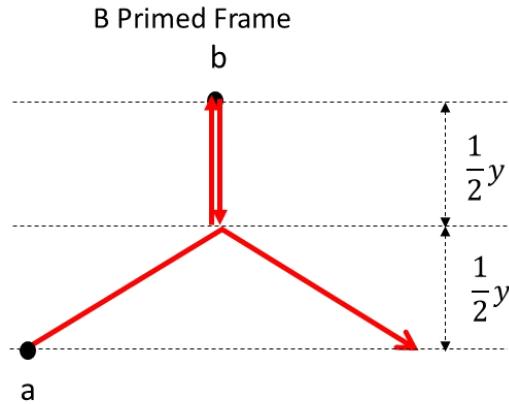
boat in the boat frame. This meets our criteria for begin a proper time

$$\Delta t_{aA}\{\text{total}\} = \Delta t_{ap}$$

By knowing the proper time, we might use the time dilation equation. We would like to do that. But we realize that in our comparison of frame A to frame C , there is the possibility of using a proper time for A , but not for B

$$\Delta t_{aC} = \frac{\Delta t_{aA}}{\sqrt{1 - \left(\frac{v_{AC}}{c}\right)^2}} = \gamma_{AC}\Delta t_{aA}$$

where a is not moving in the x direction in frame A . But it is not true that B is at rest in either frame. We need a different reference frame if we are going to find the proper time for particle b . That's fine, let's consider another reference frame, this time riding along with a frame speed $v_{BC} = v_{bCx}$. We will call this frame B because particle b won't move (in the x direction) in this frame.



We immediately recognize that $\Delta t_{bB}\{\text{total}\}$ is the proper time in this frame because in the B primed frame particle b is not moving in the direction of the relative motion, so it is like the clock situation on the boat frame.

$$\Delta t_{bB}\{\text{total}\} = \Delta t_{bp}$$

and due to symmetry of our carefully designed experiment we can see that for this case $\Delta t_{bp} = \Delta t_{ap}$. We could write time dilation for b as we move from frame B to frame A

$$\Delta t_{bA}\{\text{total}\} = \frac{\Delta t_{bp}}{\sqrt{1 - \left(\frac{v_{BA}^2}{c^2}\right)}}$$

where the speed v_{BA} is still the $v_{AC} = v_{Bx} = v_{Ax}$. This fills in our missing piece from

our prime frame so we can write

$$\frac{G_{aA}}{\Delta t_{aA}\{total\}} = \frac{G_{bA}}{\Delta t_{bA}\{total\}}$$

$$\frac{G_{aA}}{\Delta t_{ap}} = \frac{G_{bA}}{\frac{\Delta t_{bp}}{\sqrt{1 - \left(\frac{v_{BA}^2}{c^2}\right)}}}$$

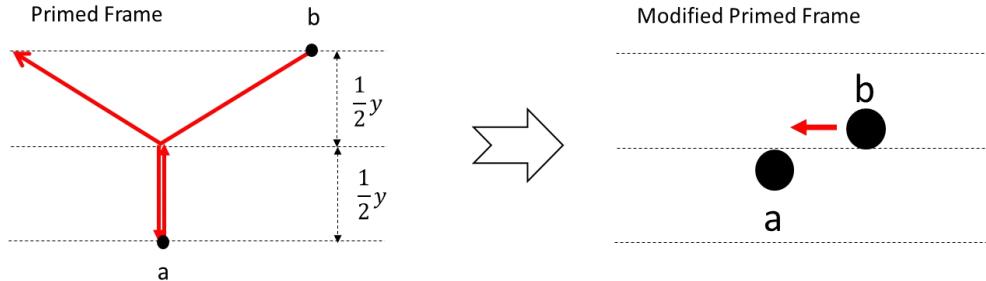
and due to symmetry we can see that it must be true that $\Delta t_{ap} = \Delta t_{bp}$ for our carefully designed situation.

$$\frac{G_{aA}}{\Delta t_{bp}} = \frac{G_{bA}}{\frac{\Delta t_{bp}}{\sqrt{1 - \left(\frac{v_{BA}^2}{c^2}\right)}}}$$

This gives

$$G_{bA} = \frac{G_{aA}}{\sqrt{1 - \left(\frac{v_{BA}^2}{c^2}\right)}}$$

which is a positive step forward. We can relate the G terms from a and b . But we want an actual solution for the G' s. So let's change our situation slightly. Let's say that we reduce the y -distance between the two particles, and let's reduce the y -component of the velocities so that a is stationary and b moves only in the x -direction.



With this modification we can use the requirement that at low speeds our G_{aA} must go to 1. And particle b moves with speed $v_{bA} = v_{BA}$ in the $-x$ direction so

$$G_{bA} = \frac{1}{\sqrt{1 - \left(\frac{v_{BA}^2}{c^2}\right)}}$$

and this is what we wanted! We can write

$$\vec{p}_{bA} = G_{bA} m \vec{v}_{bA}$$

$$= \frac{1}{\sqrt{1 - \left(\frac{v_{BA}^2}{c^2}\right)}} m \vec{v}_{bA}$$

and we can generalize this to any particle in any frame

$$\vec{p} = \frac{1}{\sqrt{1 - (\frac{v^2}{c^2})}} m \vec{v}$$

Often you will see this written as

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

But there is one big caveat. Here γ is a function of v_{bA} the particle speed. In the Lorentz transformation γ is as function of the relative frame speed, v_{BA} . We might do well to keep our subscripts to show the difference when we use γ as a symbol for $1/\sqrt{1 - (\frac{v^2}{c^2})}$.

$$\gamma_{bA} = \frac{1}{\sqrt{1 - \left(\frac{v_{bA}^2}{c^2}\right)}}$$

or

$$\gamma_{BA} = \frac{1}{\sqrt{1 - \left(\frac{v_{BA}^2}{c^2}\right)}}$$

so as not to risk confusion.

Relativistic Kinetic Energy

Our Lorentz transformation broke conservation of momentum. We should be worried about conservation of energy. But rather than test with a particle interaction, lets just re-derive our expression for kinetic energy and see if it is the same using the relativistic quantities that we know so far. We derived the change in kinetic energy back in PH121 as

$$\Delta K = w$$

and we know that work is

$$w = \int \vec{F} \cdot d\vec{x}$$

for a classical universe this would give just

$$\begin{aligned} w &= m \int \frac{d\vec{v}}{dt} \cdot d\vec{x} \\ &= m \int \frac{d\vec{x}}{dt} \cdot d\vec{v} \\ &= m \int \vec{v} \cdot d\vec{v} \\ &= \frac{1}{2} m v^2 + C \end{aligned}$$

but we now know that our classical version of momentum was not universally good,

and classically we can write our force as

$$F = ma = m \frac{dv}{dt} = \frac{dp}{dt}$$

so if our momentum is different for relativity, then forces should be different as well.

We found that the relativistic momentum for any particle in any reference frame is given by

$$\vec{p} = \gamma m \vec{v} = \gamma_{bA} m \vec{v} = \gamma_p m \vec{v}$$

where I didn't include subscripts because it is any particle momentum in any frame. But that is fraught with peril. So let's put a p on as a subscript to let us know this is a momentum type γ . You would have to supply your own subscripts when you do a problem. But we want to do this next problem for any momentum. And we want to take a derivative of the general momentum equation to find the general force equation.

Let's look at dp/dt in some detail

$$\begin{aligned} \frac{dp}{dt} &= \frac{d}{dt} \left(\frac{1}{\sqrt{1 - \left(\frac{v_p^2}{c^2}\right)}} mv \right) \\ \frac{dp}{dt} &= m \left(\frac{1}{\sqrt{1 - \left(\frac{v_p^2}{c^2}\right)}} \frac{dv_p}{dt} + v \left(-\frac{1}{2} \left(1 - \left(\frac{v_p^2}{c^2}\right) \right)^{-\frac{3}{2}} \left(-\frac{2v_p}{c^2} \frac{dv_p}{dt} \right) \right) \right) \end{aligned}$$

We can cancel the twos

$$\frac{dp}{dt} = m \left(\frac{1}{\sqrt{1 - \left(\frac{v_p^2}{c^2}\right)}} \frac{dv_p}{dt} + v_p \left(\left(1 - \left(\frac{v_p^2}{c^2}\right) \right)^{-\frac{3}{2}} \left(\frac{v_p}{c^2} \frac{dv_p}{dt} \right) \right) \right)$$

and we would like a common denominator, so lets work toward that.

$$\begin{aligned} \frac{dp}{dt} &= m \left(\frac{1}{\sqrt{1 - \left(\frac{v_p^2}{c^2}\right)}} \frac{dv_p}{dt} + \left(\frac{\frac{v_p^2}{c^2}}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} \left(\frac{dv_p}{dt} \right) \right) \right) \\ \frac{dp}{dt} &= m \frac{dv_p}{dt} \left(\frac{1}{\sqrt{1 - \left(\frac{v_p^2}{c^2}\right)}} \frac{1 - \left(\frac{v_p^2}{c^2}\right)}{1 - \left(\frac{v_p^2}{c^2}\right)} + \left(\frac{\frac{v_p^2}{c^2}}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} \right) \right) \\ \frac{dp}{dt} &= m \frac{dv_p}{dt} \left(\frac{1 - \left(\frac{v_p^2}{c^2}\right)}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} + \frac{\frac{v_p^2}{c^2}}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} \right) \\ &= \frac{m \frac{dv_p}{dt}}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} \end{aligned}$$

This is just our relativistic force equation. Then our work equation becomes

$$\begin{aligned} w &= - \int_0^x \frac{dp}{dt} dx \\ &= - \int_0^x \frac{m \frac{dv_p}{dt}}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} dx \end{aligned}$$

we can use the fact that $v_p = \frac{dx}{dt}$ so that $dx = v_p dt$ so

$$\begin{aligned} w &= - \int \frac{m \frac{dv_p}{dt}}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} v dt \\ &= - \int_0^v \frac{mv_p dv_p}{\left(1 - \left(\frac{v_p^2}{c^2}\right)\right)^{\frac{3}{2}}} \end{aligned}$$

To solve this, we can use the form

$$\int \frac{x}{(1-x^2)^{\frac{3}{2}}} dx = \frac{1}{\sqrt{1-x^2}}$$

but we need to make our integral look like this form. Let $x = v_p/c$, then $x^2 = v_p^2/c^2$ and $dx = dv_p/c$ then our work integral looks like

$$w = \int_0^x \frac{m(cx)(cdx)}{(1-(x^2))^{\frac{3}{2}}}$$

We can take out the constants, and we have our standard form

$$\begin{aligned} w &= mc^2 \int_0^x \frac{x(dx)}{(1-(x^2))^{\frac{3}{2}}} \\ &= mc^2 \left(\frac{1}{\sqrt{1-x^2}} \right) \Big|_0^x \\ &= mc^2 \left(\frac{1}{\sqrt{1-\frac{v_p^2}{c^2}}} - \frac{1}{\sqrt{1-0}} \right) \\ &= \frac{mc^2}{\sqrt{1-\frac{v_p^2}{c^2}}} - mc^2 \end{aligned}$$

and this must be equal to ΔK . But we started our integral with $v_p = 0$ so this is really just $K_f = K$ where since $K_i = 0$ we can drop the f subscript.

$$\begin{aligned} K &= \frac{mc^2}{\sqrt{1-\frac{v_p^2}{c^2}}} - mc^2 \\ &= \gamma_p mc^2 - mc^2 \end{aligned}$$

It's not clear from looking at this that we will recover $\frac{1}{2}mv_p^2$ for low velocities (like not at all clear). To see that it is true, let's use some of that math that you "learned" in

M215/M316. We can do a series expansion of

$$\left(1 - \frac{v_p^2}{c^2}\right)^{-\frac{1}{2}} \approx 1 + \frac{1}{2} \frac{v_p^2}{c^2} + O(\text{higher}) \quad \text{for } \frac{v_p^2}{c^2} \ll 1$$

and keeping just these first two terms, we get

$$\begin{aligned} K_{low} &\approx mc^2 \left(1 + \frac{1}{2} \frac{v_p^2}{c^2}\right) - mc^2 \\ &\approx mc^2 + \frac{1}{2} mc^2 \frac{v_p^2}{c^2} - mc^2 \\ &\approx \frac{1}{2} mv_p^2 \end{aligned}$$

so for low velocities this works. But the relativistic equation looks very strange. Let's look at the relativistic form further.

Total Energy

It might seem strange that our kinetic energy is the difference between two terms.

$$\begin{aligned} K &= \frac{mc^2}{\sqrt{1 - \frac{v_p^2}{c^2}}} - mc^2 \\ &= \gamma_p mc^2 - mc^2 \end{aligned}$$

what if we take the mc^2 to the other side.

$$\frac{mc^2}{\sqrt{1 - \frac{v_p^2}{c^2}}} = \gamma_p mc^2 = K + mc^2$$

seems to be a total energy of some sort. To see what sort of total it is, let's look at mc^2 . It has the object's mass in it. Our total energy would be equal to this term if we had no kinetic energy, $K = 0$. That would be in a reference frame where the object is not moving. Then our sort-of total energy would be just mc^2 . No motion, but we still have energy. Where is it? Of course you know the answer because this is one of the most famous equations on Earth.

$$E_o = mc^2$$

is the energy tied up in the mass of the object. We call it the *rest energy*. It is this rest energy that gives rise to nuclear energy (and nuclear bombs). So our sort of total energy is a combination of the kinetic energy and the rest energy of the object.

$$E = K + E_o$$

We can now define the relativistic total energy

$$E = \frac{mc^2}{\sqrt{1 - \frac{v_p^2}{c^2}}} = \gamma_p mc^2$$

We should ask, what happened to potential energy? And here again is a strange thing in relativity. The energy tied up in a system configuration will show up in this rest energy term for the system! The idea of potential energy has been swallowed up into our new idea of rest energy of a system. But you may object, rest energy consists of the mass of the system, m , and c^2 . The c^2 can't change, so are you saying what we used to call potential energy of a system actually changes the system mass? And the answer is a surprising yes! We will study this with atoms, where the potential energy due to the Coulomb force between the protons and electrons will end up changing the atom's mass. But it is true for solar systems and galaxies as well. This is another big change in going from Galilean to special relativity.

We started our trip through physics by studying velocities. But you may have noticed that we don't have a conservation of velocity law. Momentum does have a conservation law. This makes momentum more fundamental than velocity! We can write our energy equation in terms of momentum so we have energy in terms of the more fundamental quantity. To do this take our relativistic momentum

$$p = \frac{mv}{\sqrt{1 - \frac{v_p^2}{c^2}}} = \gamma_p mv$$

and square it

$$p^2 = \frac{m^2 v_p^2}{1 - \frac{v_p^2}{c^2}}$$

and multiply by c^2

$$p^2 c^2 = \frac{m^2 v_p^2 c^2}{1 - \frac{v_p^2}{c^2}} = \gamma_p^2 m^2 v_p^2 c^2$$

and we can rearrange

$$\begin{aligned} p^2 c^2 &= \gamma_p^2 m^2 \frac{v_p^2}{c^2} c^4 \\ &= \gamma_p^2 (mc^2)^2 \frac{v_p^2}{c^2} \end{aligned}$$

now

$$\gamma_p = \frac{1}{\sqrt{1 - \frac{v_p^2}{c^2}}}$$

so γ_p^2 would be

$$\gamma_p^2 = \frac{1}{1 - \frac{v_p^2}{c^2}}$$

and we can solve this for v_p^2/c^2

$$\begin{aligned} 1 - \frac{v_p^2}{c^2} &= \frac{1}{\gamma_p^2} \\ \frac{v_p^2}{c^2} &= 1 - \frac{1}{\gamma_p^2} \end{aligned}$$

Using this we can write

$$p^2 c^2 = \gamma_p^2 (mc^2)^2 \frac{v_p^2}{c^2}$$

as

$$\begin{aligned} p^2 c^2 &= \gamma_p^2 (mc^2)^2 \left(1 - \frac{1}{\gamma_p^2}\right) \\ &= \gamma_p^2 (mc^2)^2 - \gamma_p^2 (mc^2)^2 \frac{1}{\gamma_p^2} \\ &= \gamma_p^2 mc^4 - (mc^2)^2 \end{aligned}$$

Now take E^2

$$E^2 = \gamma_p^2 m^2 c^4$$

This is the first term in our expression for $p^2 c^2$!

$$p^2 c^2 = E^2 - (mc^2)^2$$

so that

$$E^2 = p^2 c^2 + (mc^2)^2$$

This is our new relativistic expression of energy in terms of momentum. Note that what we want is conservation of energy and conservation of momentum so we can solve problems. This form for the total energy is nice because it is in terms of momentum. This should make conservation type problems easier!

Notice that if a particle has no mass (i.e. is a photon) we get

$$E_{photon}^2 = p^2 c^2 + 0$$

or the momentum of a photon would be

$$p = \frac{E_{photon}}{c}$$

we used this back in PH220, but it is a result of special relativity.

Universal Speed Limit

Let's look at kinetic energy again and ask the question what kinetic energy it would take to make a particle (or spacecraft, etc.) to the speed of light. We can take

$$K = \frac{mc^2}{\sqrt{1 - \frac{v_p^2}{c^2}}} - mc^2$$

and put in $v_p = c$

$$\begin{aligned} K &= \frac{mc^2}{\sqrt{1 - \frac{c^2}{c^2}}} - mc^2 \\ &= \frac{mc^2}{\sqrt{1 - 1}} - mc^2 \\ &= \frac{mc^2}{0} - mc^2 \\ &= \infty - mc^2 \end{aligned}$$

which undoubtedly is making a mathematician squirm. What we really mean is

$$K = \lim_{v_p \rightarrow c} \frac{mc^2}{\sqrt{1 - \frac{v_p^2}{c^2}}} - mc^2 = \infty$$

and this is a good way to say what we mean. As we accelerate an object toward the speed of light, the energy required to do so becomes infinite. This puts an effective limit on how fast an object can go without doing some general relativity process like warping space (think warp engines from science fiction). Objects with mass cannot have speeds equal to or greater than the speed of light. Of course if $m = 0$ in our equation then we would have a limit that approached zero, and that would be fine. This seems to imply that only massless particles (like photons) can travel at the speed of light.

Sadly this concludes our study of special relativity. And although we will use what we learned for the rest of the semester and into our junior and senior level classes, we won't study special relativity again for a while. You might feel that we have just scratched the surface, and you would be right. If you are interested in relativity, stay tuned for general relativity (well as small taste of it) at the end of this class, and for graduate level courses in relativity.

But there is more fun to come. Next lecture we begin our study of quantum mechanics!

8 Quantization: The Beginnings of Modern Physics

3.1, 3.2, and 3.3

Fundamental Concepts in the Lecture

- Waves and Fields
- Planck's Quantum Hypothesis; Blackbody Radiation
- Photon Theory of Light and the Photoelectric Effect

Waves and Fields Review

So far in our physics adventure we have studied electric and magnetic fields. There are other kinds of fields as we shall see. But let's review what we know from PH220 for a moment. And since we need to recall the ideas of interference, let's review the double slit experiment.

We can start by defining an electric field

$$\vec{E} = \frac{1}{4\pi\epsilon_o} \frac{q}{r^2} \hat{r}$$

where \hat{r} is a unit vector in the radial direction and a magnetic field and of course q is the charge amount. We are familiar with the permittivity of free space ϵ_o from our study of electric fields. We also know

$$\vec{B} = \frac{\mu_o}{2\pi} \frac{I}{r} \hat{\phi}$$

where $\hat{\phi}$ is a unit vector in the azimuthal direction, I is the current, and μ_o is the permeability constant for free space.

We see from the equations that the electric and magnetic fields exist in all of space (they have a value for every r) and we remember from PH220 that they are not really independent. Maxwell's equations tell us that moving electric fields produce magnetic

fields and moving magnetic fields produce electric fields.

$$\begin{aligned}
 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} &= 0 && \text{Gauss's law for electric fields} \\
 \oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} &= 0 && \text{Gauss's law for magnetic fields} \\
 \oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} &= -\frac{d\Phi_B}{dt} && \text{Faraday's law} \\
 \oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} &= \varepsilon_0 \mu_0 \frac{d\Phi_E}{dt} && \text{Ampere-Maxwell Law}
 \end{aligned} \tag{8.1}$$

and of course, relative motion determined which field was moving. So really the electric and magnetic fields were a matter of reference frame. What appeared as a magnetic field in one frame showed up as an electric field in another frame. This is another example of the non-invariance of interpretation that we see in special relativity. Whether we see an electric field or a magnetic field is really a matter of which reference frame we use to observe the experiment. This is why physicists tend to define an electromagnetic field, rather than separate electric and magnetic fields. See, you were doing relativity back in PH220, you just didn't know it!

A great way to make a changing or moving field is to make a wave. We can write a wave in the electric and magnetic fields as

$$\begin{aligned}
 E &= E_{\max} \sin(kr - \omega t + \phi) \\
 B &= B_{\max} \sin(kr - \omega t + \phi)
 \end{aligned}$$

where

$$\begin{aligned}
 k &= \frac{2\pi}{\lambda} \\
 \omega &= 2\pi f
 \end{aligned}$$

are the “wave number” and angular frequency of the wave. For such an electromagnetic wave we found that

$$\frac{E_{\max}}{B_{\max}} = c \tag{8.2}$$

The power in the wave is given by the Poynting vector

$$\vec{\mathbf{S}} = \frac{1}{\mu_0} \vec{\mathbf{E}} \times \vec{\mathbf{B}} \tag{8.3}$$

This gives both the magnitude and the direction of the power. The magnitude can be written in terms of just the electric field because of the relationship between E_{\max} and B_{\max} so that

$$S = \frac{1}{\mu_0} E_{\max} B_{\max} \tag{8.4}$$

can be written as

$$S = \frac{E_{\max}^2}{c \mu_0} \tag{8.5}$$

The intensity of the light is written as

$$I = S_{ave}$$

or

$$I = \frac{1}{2\mu_0 c} E_{\max}^2 \quad (8.6)$$

For a double slit experiment we are mixing two waves. And we know to expect interference fringes. These fringes are not just points, but are patterns that fade from a maximum intensity. We can calculate the intensity pattern. Start with the two fields.

$$\begin{aligned} E_2 &= E_{\max} \sin(k_2 r_2 - \omega_2 t_2 + \phi_2) \\ E_1 &= E_{\max} \sin(k_1 r_1 - \omega_1 t_1 + \phi_1) \end{aligned}$$

and we can calculate the resulting wave

$$\begin{aligned} E_r &= E_2 + E_1 \\ &= E_{\max} \sin(k_2 r_2 - \omega_2 t_2 + \phi_2) + E_{\max} \sin(k_1 r_1 - \omega_1 t_1 + \phi_1) \end{aligned}$$

but we can make this more meaningful if we use a trig identity

$$\sin a + \sin b = 2 \cos\left(\frac{a-b}{2}\right) \sin\left(\frac{a+b}{2}\right)$$

Then we find

$$\begin{aligned} E_r &= E_{\max} \sin(k_2 r_2 - \omega_2 t_2 + \phi_2) + E_{\max} \sin(k_1 r_1 - \omega_1 t_1 + \phi_1) \\ &= 2E_{\max} \cos\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) - (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \\ &\quad \times \sin\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) + (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \\ &= 2E_{\max} \cos\left(\frac{1}{2}[(k_2 r_2 - \omega_2 t_2 + \phi_2) - (k_1 r_1 - \omega_1 t_1 + \phi_1)]\right) \\ &\quad \times \sin\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) + (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \end{aligned}$$

But now we know that we can simplify this because the two sets of light that come through the slits started as one beam, so $\omega_2 = \omega_1 = \omega$, $k_2 = k_1 = k$, $t_2 = t_1 = t$, and $\phi_2 = \phi_1 = \phi_o$.

$$\begin{aligned} E_r &= 2E_{\max} \cos\left(\frac{1}{2}k\delta\right) \sin\left(\frac{kr_2 + kr_1}{2} - \omega t + \phi_o\right) \\ &= 2E_{\max} \cos\left(\frac{1}{2}\frac{2\pi}{\lambda}d \sin\theta\right) \sin\left(k\frac{r_2 + r_1}{2} - \omega t + \phi_o\right) \end{aligned}$$

We have a combined wave that is a traveling wave $\left(\sin\left(k\frac{(r_2+r_1)}{2} - \omega t + \phi_o\right)\right)$ but with amplitude $(2E_o \cos(\frac{1}{2}(\frac{2\pi}{\lambda}d \sin\theta)))$ that depends on our total phase $\Delta\phi = \frac{2\pi}{\lambda}d \sin\theta$.

But the situation is more complicated because of how we detect light. Our eyes, film, and most detectors measure the intensity of the light. We know that

$$\begin{aligned} I &= \frac{E^2}{c\mu_o} \\ &= \frac{4E_o^2}{c\mu_o} \cos^2 \left(\frac{1}{2} \left(\frac{2\pi}{\lambda} d \sin \theta \right) \right) \sin^2 \left(\frac{k(r_2 + r_1)}{2} - \omega t + \phi_o \right) \end{aligned}$$

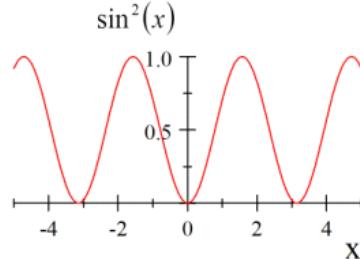
Light detectors collect power for a set amount of time. So most light detection will be a value averaged over a set *integration time*. This means that the detector sums up (or integrates) the amount of power received over the detector time. Usually the integration time is much longer than a period, so we need to time-average our intensity.

$$\begin{aligned} \int_{\text{many periods}} Idt &\propto \int_{\text{many periods}} \frac{4E_o^2}{c\mu_o} \cos^2 \left(\frac{1}{2} \left(\frac{2\pi}{\lambda} d \sin \theta \right) \right) \sin^2 \left(\frac{k(r_2 + r_1)}{2} - \omega t + \phi_o \right) dt \\ &= \frac{4E_o^2}{c\mu_o} \cos^2 \left(\frac{1}{2} \left(\frac{2\pi}{\lambda} d \sin \theta \right) \right) \int_{\text{many periods}} \sin^2 \left(\frac{k(r_2 + r_1)}{2} - \omega t + \phi_o \right) dt \end{aligned}$$

but the term

$$\int_{\text{many periods}} \sin^2 \left(\frac{k(r_2 + r_1)}{2} - \omega t + \phi_o \right) dt = \frac{1}{2} \quad (8.7)$$

To convince yourself of this, think that $\sin^2(x)$ has a maximum value of 1 and a minimum of 0. Looking at the graph



it should be believable that the average value over a period is 1/2. The average over many periods will still be 1/2.

So we have

$$\bar{I} = \int_{\text{many periods}} Idt = \frac{2E_o^2}{c\mu_o} \cos^2 \left(\frac{1}{2} \left(\frac{2\pi}{\lambda} d \sin \theta \right) \right) \quad (8.8)$$

where \bar{I} is the time average intensity. The important thing to notice is that the time varying part has averaged out.

So, usually in optics we ignore the fast fluctuating parts of such calculations because we can't see them and so we write

$$I = I_{\max} \cos^2 \left(\frac{1}{2} \left(\frac{2\pi}{\lambda} d \sin \theta \right) \right)$$

where we have dropped the bar from the I , but it is understood that the intensity we report is a time average over many periods.

We should remind ourselves, our intensity pattern

$$I = I_{\max} \cos^2 \left(\frac{1}{2} \frac{2\pi}{\lambda} d \sin \theta \right)$$

is really

$$I = I_{\max} \cos^2 \left(\frac{\Delta\phi}{2} \right)$$

Which is just our amplitude squared for the mixing of two waves. All we have done to find the intensity pattern is to find an expression for the phase difference $\Delta\phi$.

Our intensity pattern should give the same location for the center of the bright spots as we got back in PH123. Let's check that it works. We used the small angle approximation in PH123. It is still valid, so let's use it again now. For small angles

$$\begin{aligned} I &\approx I_{\max} \cos^2 \left(\frac{\pi d}{\lambda} \theta \right) \\ &\approx I_{\max} \cos^2 \left(\frac{\pi d}{\lambda} \frac{y}{L} \right) \end{aligned}$$

Then we have constructive interference when

$$\frac{\pi d}{\lambda} \frac{y}{L} \approx m\pi$$

where m is an integer. We can write this as

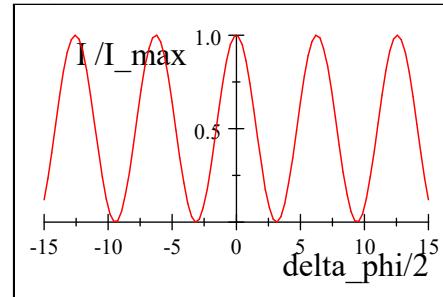
$$y \approx m \frac{L\lambda}{d}$$

which is what we found before.

The plot of normalized intensity

$$\frac{I}{I_{\max}} = \cos^2 \left(\frac{\Delta\phi}{2} \right)$$

verses $\Delta\phi/2$ is given next,

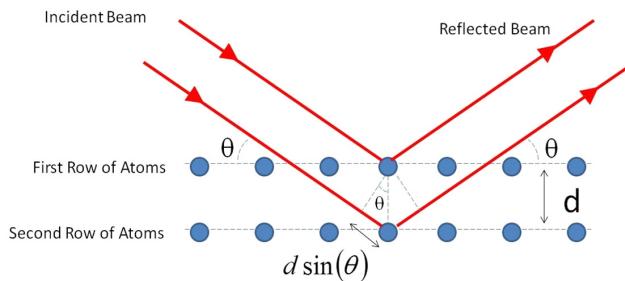


but we will find that we are not quite through with this analysis.

We did this using slits that were small, but on the order of the size of the wavelength

of light. We could apply this to the realm of the very small by using diffraction. Diffraction of X-rays by Crystals. If we make the wavelength of light very small, then we can deal with very small diffraction gratings. X-ray wavelengths are much smaller than visible light wavelengths. If we use X-rays can investigate smaller things. This concept is used to investigate the structure of crystals with X-rays. The crystal lattice of molecules or atoms creates the regular pattern we need for a grating. The pattern is three dimensional, so the patterns are complex.

Let's start with a simple crystal with a square regular lattice. $NaCl$ has such a structure.



If we illuminate the crystal with x-rays, the x-rays can reflect off the top layer of atoms, or off the second layer of atoms (or off any other layer, but for now let's just consider two layers at first). If the spacing between the layers is d , then the path difference will be

$$\delta = 2(d \sin(\theta)) \quad (8.9)$$

then for constructive interference

$$2d \sin(\theta) = m\lambda \quad m = 1, 2, 3, \dots \quad (8.10)$$

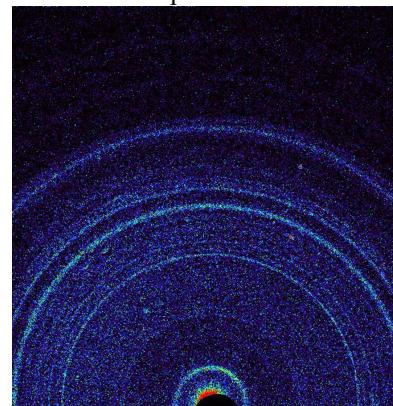
This is known as *Bragg's law*. This relationship can be used to measure the distance between the crystal planes.

A resulting pattern is given in the following figure.



X-ray diffraction pattern of DNA (image courtesy of the National Institute of Health, image in the public domain)

DNA makes an interesting diffraction pattern. If the substance consists of a powder of small crystals, the individual diffraction peaks turn into circles.

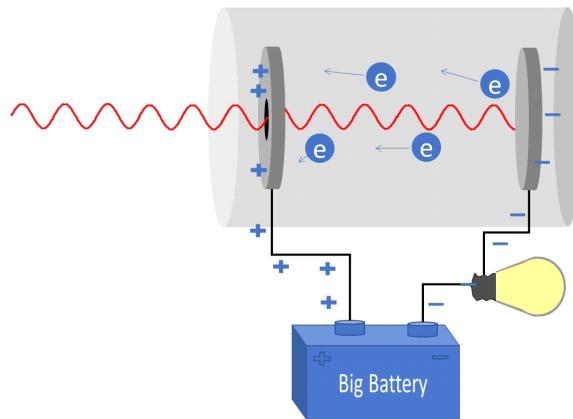


First X-ray View of Martian Soil. <http://photojournal.jpl.nasa.gov/catalog/PIA16217>
We will use this idea of diffraction in our study of quantum mechanics.

Quantum Waves: Photoelectric Effect

ConcepTest 27.2a

But we need to modify our view of light to study quantum mechanics. To do this, I want to talk about a specific experiment (Einstein's explanation of this experiment received the Nobel prize, so it is kind of a big one). To understand, we will need to describe the experiment in detail. Here is a picture of the experimental setup.



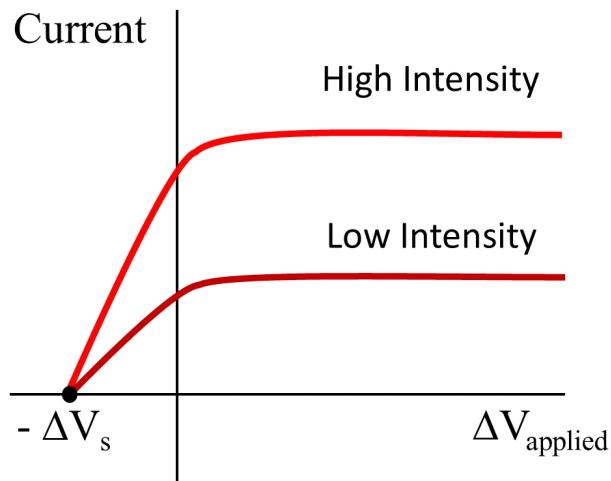
Let's start with the variable power supply. It is just a battery box much like the power supplies we have in our labs, only it can be both positive or negative, and it can achieve very high potentials.

Next let's look at the metal plate on the right. We place it in a glass container, and we remove the air (creating a vacuum). We find that when light strikes the plate we get electrons! They are released when the light energy is absorbed. The electrons get an amount of kinetic energy from the absorbed light. The light must supply enough energy to break the electron away from its nucleus, and then enough to account for the velocity of the electron.

But we know that waves carry energy, and light is a wave, so this should not be too surprising. Now in our setup, we also have a plate on the left with a hole in it for light to pass through. This plate is hooked to the positive side of the power supply. So it has a high potential compared to the right-hand plate. Then the electrons that are freed by light hitting the right-hand plate will be accelerated to the left-hand plate. A current will flow through the light bulb, but only when light hits the plate!

We would expect that if the potential was higher, more of the electrons would be accelerated to the left-hand plate and the current will be greater. We also expect that if we increase the amount of light we would get more electrons, so more electrons would be accelerated to the left-hand plate.

Here is a graph that describes the situation.



It is a graph of current vs. the power supply potential difference. Note that if the intensity of the light is larger, we do get a larger current. If we increase the voltage, we might see a small increase in the current, but the current soon reaches a maximum value. More potential difference does not give more current.

But look what happens when we reverse the polarity of the power supply (switch the wires from plus to minus and from minus to plus). The current starts to fall as we make the potential more negative. What is happening?

Well, each of the electrons is released with some amount of kinetic energy. But now there is a force due to the electric field that tries to prevent the electrons from reaching the left hand plate.

This is like throwing a ball up toward the ceiling. We have the acceleration due to gravity acting downward, the ball's velocity is upward, so the ball slows down. If the ball is thrown with a large initial velocity, it will reach the ceiling, but if it has a smaller initial velocity, it will turn around and head back to us or the floor.

This is just what happens to the electrons. If they are released with a large initial velocity toward the left hand plate, they still make it, but they hit with a slower speed. The slowest electrons stop, turn around, and head back to right hand plate. So the current is reduced.

If we make the potential difference more negative (larger magnitude, but still negative) it is like trying to throw the ball if our class was on Jupiter. The larger gravity would mean that I could never get the ball to hit the ceiling.

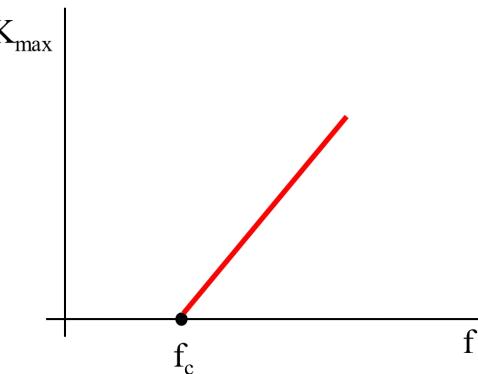
The potential that just stops the current, sending all the electrons back the other way, is called the *stopping potential*.

The maximum kinetic energy of the electrons is related to the stopping potential

$$K_{\max} = e\Delta V_s \quad (8.11)$$

None of this seems strange, but note from the graph that the stopping potential is the same shape for all light intensities! No matter how much light I shine on plate E , if $\Delta V = \Delta V_s$ no current is ever produced. Remember that for waves on strings or sound waves the intensity is proportional to the amplitude of the wave, and it tells us how much energy we are delivering. So if we increase the intensity of light, we expect to have more energy available to release electrons! But it does not happen—ever—for this case.

It gets stranger. If I change the frequency, I still get electrons, that is, until I get below a frequency f_c . Then no matter how much intensity I have, I never get a photoelectron released! I could increase the intensity so much that the absorbed light melts the metal plate, but no electrons come out (well, a few due to thermal effects). If I increase the frequency of light, even if I have almost no intensity, I get photoelectrons, and their kinetic energy increases as frequency increases. But below f_c I never get an electron.



Einstein successfully explained this experiment by theorizing that when the light was emitted an electron may jump from a higher energy state to a lower energy state. This would release energy, giving off a small bit of light. The higher energy state he wrote as

$$E_n = nhf \quad (8.12)$$

He wrote the lower energy state as

$$E_{n-1} = (n - 1) hf \quad (8.13)$$

and the quantized bit of light that was emitted would have energy

$$E = E_n = E_{n-1} = hf \quad (8.14)$$

The quantized bit of light is called a *photon*. Einstein theorized that the energy of the photon was kept locally (in the near vicinity of the center of the photon) so when the photon strikes an electron, it would give up its energy all at once. This explains why increasing the intensity of the light does not increase the kinetic energy of the electrons. Intensity for light waves must be like the superposition of all the photons, and only one photon is absorbed by an individual electron. The Kinetic energy is then

$$K_{\max} = hf - \phi \quad (8.15)$$

where ϕ is something that tells us how much energy is required to free the electrons from the metal atoms. The photon energy $E = hf$ provides enough energy to both free the electron and to give the electron kinetic energy. The ϕ term is the energy that it takes to free the electron. Einstein called ϕ the *work function* of the metal. If the photon provides just this amount of energy, the electrons are freed, but there is no energy left to move them from the metal atoms, so they just recombine (so we will have no current) then

$$K = 0 = hf - \phi \quad (8.16)$$

or

$$f_c = \frac{\phi}{h} \quad (8.17)$$

gives a cutoff frequency just as we saw in the experiment.

ConcepTest 27.2b

ConcepTest 27.2c

ConcepTest 27.2d

The device we described is called a *photocell*. They are used in safety devices for garage door openers or to automatically turn off street lights during the day.

So what do we learn from this experiment? We learn that in some circumstances light acts like a small bundle of energy that travels through space or can be absorbed independently of the total wave. It is this clumpiness of light that will start us thinking in quantum ways.

Blackbody Radiation

ConcepTest 27.1

Long, long, ago we studied radiation as a means of transporting energy (in PH123). We studied Stefan's law

$$\mathcal{P} = \sigma A e T^4 \quad (8.18)$$

where σ is Stefan's constant, A is the area, e is the emissivity, and T is the temperature.

This tells us that when something has a temperature, T it will lose energy with an

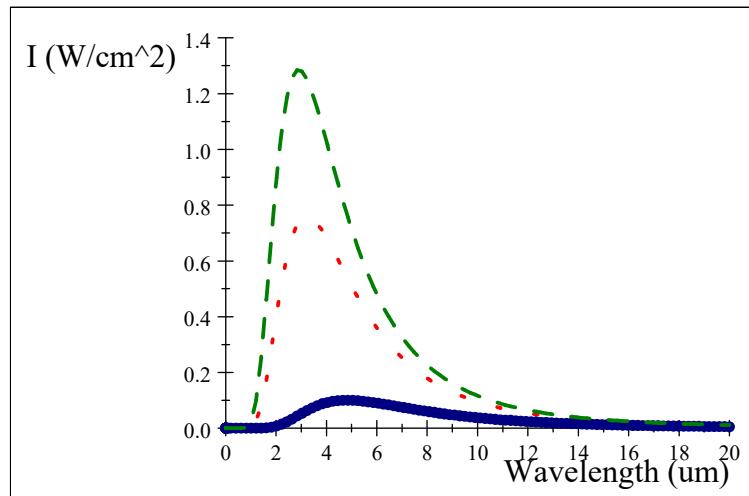
energy loss rate of

$$\frac{\Delta E}{\Delta t} = \mathcal{P}$$

So, when things are hot, we expect them to radiate. But Stefan's law does not talk about the wavelength of the light. In fact, Stefan's law assumes we are talking about *all* wavelengths at once.

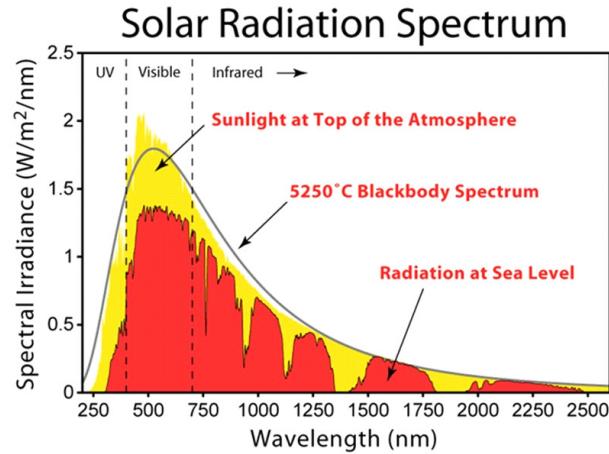
But it is sometimes useful to know how much radiation there is of a specific color. For example. If we have a stove heating element, and it is glowing red, we probably don't want to touch it.

In fact, for an ideal material, the temperature would completely control the distribution of wavelengths that come from it as it glows due to thermal energy. Such an ideal material is called a *black body*. If you tell me what the temperature is, I can tell you what the radiation pattern from a black body will look like. Some example curves at different temperatures are given in the next figure.



Plank Balackboady Curve for $T = 1000$ K (green or dashed), $T = 900$ K (red or dotted), $T = 600$ K (blue or dash dot),

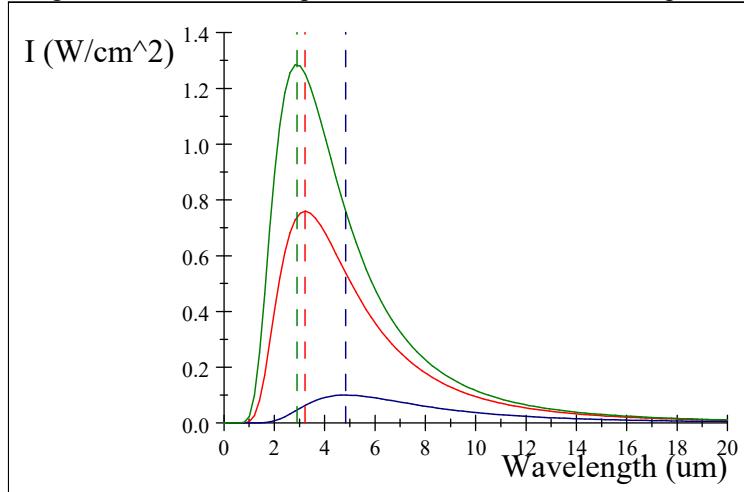
Notice that the curves for different temperatures peak at different wavelengths. You might ask what type of objects display such a radiation curve. and the answer is most objects display a radiation curve that is close to this. The curve is an idealization (for reasons we will study later) but anything warm displays this curve. Here is a graph of the solar radiation



Solar Spectrum (Image courtesy US Government)

Note that the yellow curve (before the Earth's atmosphere absorbs some of the light) matches our ideal curve fairly well. The Sun is nearly an ideal blackbody which shows us that the term "blackbody" doesn't really describe the color of the object, but rather its lack of reflection. The Sun does not reflect light well, but it does emit light well!

In the next figure I have marked the peak of the curve for different temperatures.



Plank Balackbady Curve for $T = 1000\text{ K}$ (green), $T = 900\text{ K}$ (red), $T = 600\text{ K}$ (blue), with verticle lines showing Wein's law predictions for the peak wavelengths.

The equation for the full curve is kind of messy

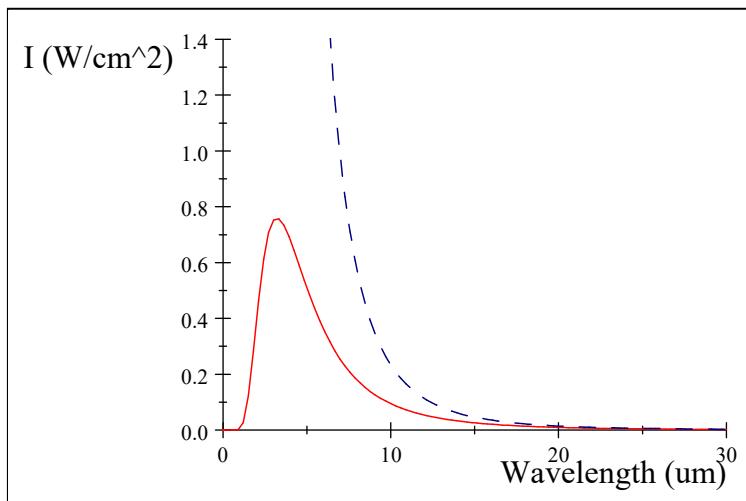
$$I = \frac{2\pi hc^2}{\lambda^5 \left(e^{\frac{hc}{\lambda k_B T}} - 1 \right)} \quad (8.19)$$

The peak of the curve is useful. It gives the wavelength (and therefore the color) of the glow coming from the warm object. To find this we really should take a derivative, but a good approximation would be

$$\lambda_{\max} T = 2.8978 \times 10^{-3} \text{ m K}$$

This is called Wein's law.

Since a warm object that has an intensity that follows this equation is called a *blackbody*, we often call this the blackbody equation. It is shown in the curves above for different T values, and in the next figure as the solid (red) curve. This curve was known from experiment but was not understood. The dashed (blue) curve is the result that was expected from classical electrodynamics. Note that it is quite different. This caused quite a problem. It is even called the *ultraviolet catastrophe*. Classical theory of electromagnetic waves did not predict the right result!



A researcher named Plank solved the problem. He assumed that a black body was made of millions of tiny charged oscillators (we would now talk in terms of standing waves, as we will see later in the course). Then, knowing that standing waves in strings and other things were often quantized (only certain frequencies are possible), he assumed something was quantized, in this case, the energy. This might make some sense if

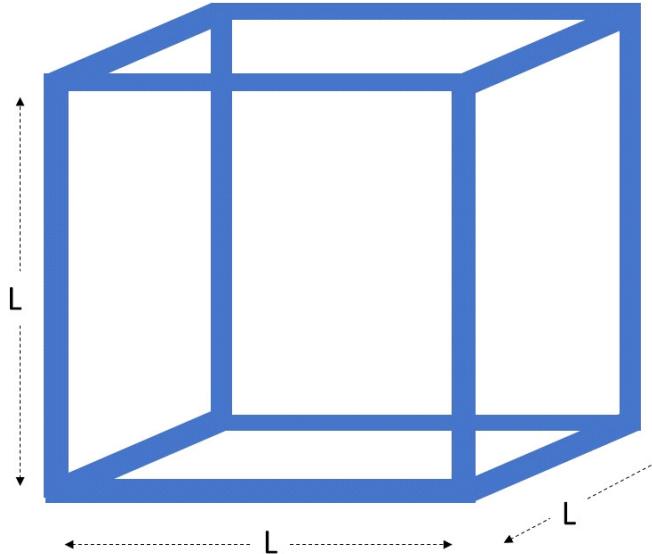
we consider that for a piece of rope that is experiencing a standing wave, the velocity of each piece of rope, and therefore the kinetic energy of that piece, depends on the frequency, but the frequency is quantized for a standing wave! This would mean that if we hold all other variables constant (e.g. don't change the amplitude, or the mass of the rope) only certain values of the kinetic energy are possible. In our case, the equation is quite simple (it is harder for a rope standing wave, were there are bands of allowed energy)

$$E_n = nhf \quad (8.20)$$

here $h = 6.626 \times 10^{-34}$ J s. This is called *Planck's constant*. And n is a positive integer. When Planck made this assumption, he got the messy equation above (Equation 8.19). It looks ugly, but it works! It matched the experimental results. Planck did not know what the resonators were, but he knew the result was useful. It would take many years before others (and Planck, himself) would realize that this picture of quantized light might be real.

Let's see a little of what Planck did (we will derive the black body equation in detail later in our course).

He assumed that light was generated inside a box.



Because light is a wave, he assumed there would be standing waves in the box. We know about standing waves, not all frequencies will make a standing wave. Planck

assumed this would be the case and we would get standing waves with energy values $E_n = nhf$ that would make persistent standing waves. All waves with other energy values would die out. The number of waves with each energy, E , would be given by an equation that we will borrow from statistical mechanics (PH412). It is called the Maxwell-Boltzmann distribution (you should have seen this in PH123 briefly)

$$N(E) = \frac{N}{k_B T} e^{-\frac{E}{k_B T}}$$

But this isn't quite right because the Maxwell-Boltzmann distribution can't describe the energy of the molecules because it a continuous distribution of energy and our energy is quantized. Planck modified the distribution function to be something like this

$$N_n = N \left(1 - e^{-\frac{\varepsilon}{k_B T}}\right) e^{-\frac{n\varepsilon}{k_B T}}$$

where $\varepsilon = hf$. You might wonder where we got this, and the answer is from your future in PH412, Thermodynamics, but in the middle of our course we will take some time to give more details about how Planck got this. For now, let's use this result. The N_n is the number of oscillators with energy E_n . Notice that the energy is in even steps, it is quantized. And we have a definite number N_n of oscillators with the specific energy E_n . The total number of oscillators, N is given by

$$\begin{aligned} N &= \sum_{n=0}^{\infty} N_n \\ &= \sum_{n=0}^{\infty} N \left(1 - e^{-\frac{\varepsilon}{k_B T}}\right) e^{-\frac{n\varepsilon}{k_B T}} \end{aligned}$$

and we can use a trick that you will use in Math 215 or Math 316.

$$\sum_{n=0}^{\infty} e^{nx} = \frac{1}{(1 - e^x)}$$

but this isn't exactly what we need. Let's take a derivative of this with respect to x

$$\frac{d}{dx} \sum_{n=0}^{\infty} e^{nx} = \sum_{n=0}^{\infty} \frac{d}{dx} e^{nx} = \sum_{n=0}^{\infty} n e^{nx}$$

now we need the other side of the equation

$$\begin{aligned} \frac{d}{dx} \frac{1}{(1 - e^x)} &= -1 (1 - e^x)^{-2} (-e^x) \\ &= \frac{e^x}{(1 - e^x)^2} \end{aligned}$$

so that

$$\sum_{n=0}^{\infty} n e^{nx} = \frac{e^x}{(1 - e^x)^2}$$

Maybe your math class did this second identity, mine didn't. But now we have it. We

can find the average energy

$$E_{av} = \frac{1}{N} \sum_{n=0}^{\infty} N_n E_n$$

which says we add up all the energies E_n but weight them with the number of oscillators that have that energy and then divide by the total number of oscillators. So it is a weighted average. Putting in our value for N_n and E_n

$$\begin{aligned} E_{av} &= \frac{1}{N} \sum_{n=0}^{\infty} \left(N \left(1 - e^{-\frac{\varepsilon}{k_B T}} \right) e^{-\frac{n\varepsilon}{k_B T}} \right) (n\varepsilon) \\ E_{av} &= \frac{N \left(1 - e^{-\frac{\varepsilon}{k_B T}} \right)}{N} \sum_{n=0}^{\infty} (n\varepsilon) \left(e^{-\frac{n\varepsilon}{k_B T}} \right) \end{aligned}$$

then if we let

$$x = \frac{\varepsilon}{k_B T}$$

and note that we have

$$E_{av} = \left(1 - e^{-\frac{\varepsilon}{k_B T}} \right) \varepsilon \sum_{n=0}^{\infty} (n) (e^{-nx})$$

we have our form

$$\sum_{n=0}^{\infty} n e^{nx} = \frac{e^x}{(1 - e^x)^2}$$

so that

$$E_{av} = \varepsilon \left(1 - e^{-\frac{\varepsilon}{k_B T}} \right) \frac{e^x}{(1 - e^x)^2}$$

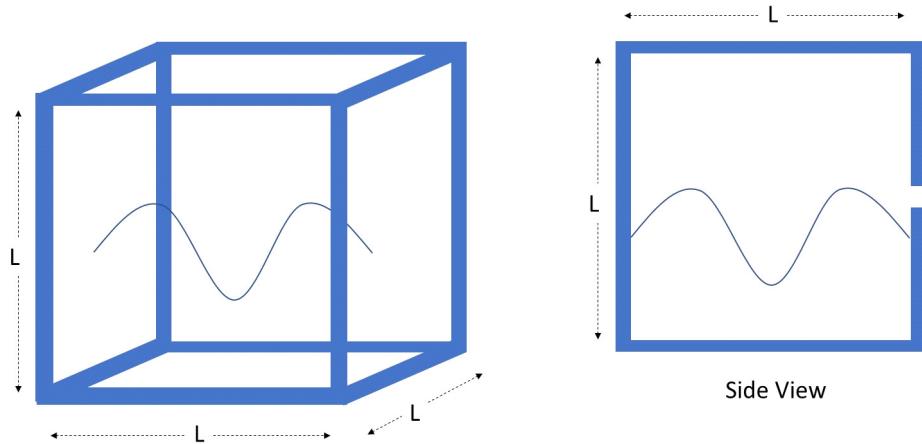
or

$$\begin{aligned} E_{av} &= \varepsilon \left(1 - e^{-\frac{\varepsilon}{k_B T}} \right) \frac{e^{\frac{-\varepsilon}{k_B T}}}{\left(1 - e^{\frac{-\varepsilon}{k_B T}} \right)^2} \\ &= \varepsilon \frac{1}{e^{\frac{\varepsilon}{k_B T}} \left(1 - e^{\frac{-\varepsilon}{k_B T}} \right)} \\ &= \varepsilon \frac{1}{\left(e^{\frac{\varepsilon}{k_B T}} - e^{\frac{\varepsilon}{k_B T}} e^{\frac{-\varepsilon}{k_B T}} \right)} \\ &= \frac{\varepsilon}{\left(e^{\frac{\varepsilon}{k_B T}} - 1 \right)} \end{aligned}$$

and finally

$$E_{av} = \frac{hf}{\left(e^{\frac{hf}{k_B T}} - 1 \right)}$$

Now suppose our box has side length L . Inside the box we could have our standing waves.



Suppose we view these like standing waves on a string. Then the wavelength for a particular mode of oscillation would be

$$\lambda = \frac{2L}{n}$$

Now suppose that there is a small hole in one side of the box. Planck found that the intensity of the light leaving the hole as a function of wavelength was

$$I(\lambda) = \frac{c}{4} u(\lambda)$$

where $u(\lambda)$ is the energy density in the box. This came from averaging over the all the angles the light could leave and taking the amount of light that ended up coming straight out of the hole. I won't repeat this geometric calculation (we will do it later). But let's use it.

The energy density would be

$$u(\lambda) d\lambda = \frac{E_{ave} N(\lambda) d\lambda}{V}$$

and we can find the number of standing waves in the box in one dimension as

$$N_{1D}(\lambda) d\lambda = \frac{dn}{d\lambda} d\lambda$$

and we see that

$$n = \frac{2L}{\lambda}$$

so

$$\left| \frac{dn}{d\lambda} \right| = \frac{2L}{\lambda^2}$$

so

$$N_{1D}(\lambda) d\lambda = \frac{2L}{\lambda^2} d\lambda$$

If we take the box as a cube and allow the same number of standing waves in each

direction we would get

$$N_{3D}(\lambda) d\lambda = \frac{8\pi L^3}{\lambda^4} d\lambda$$

which isn't obvious. Once again we are borrowing from PH412 (or later in our course). But let's assume this is right. It doesn't seem crazy compared to our one dimensional case. Then

$$u(\lambda) = \frac{E_{ave} \frac{8\pi L^3}{\lambda^4}}{V}$$

and $V = L^3$

$$\begin{aligned} u(\lambda) &= \frac{E_{ave} \frac{8\pi L^3}{\lambda^4}}{L^3} \\ &= E_{ave} \frac{8\pi}{\lambda^4} \end{aligned}$$

but we know E_{ave}

$$\begin{aligned} u(\lambda) &= \left(\frac{hf}{\left(e^{\frac{hf}{k_B T}} - 1 \right)} \right) \frac{8\pi}{\lambda^4} \\ &= \frac{8\pi}{\lambda^4} \left(\frac{h \frac{c}{\lambda}}{\left(e^{\frac{hc}{\lambda k_B T}} - 1 \right)} \right) \end{aligned}$$

and the intensity is

$$I(\lambda) = \frac{c}{4} \frac{8\pi}{\lambda^4} \left(\frac{h \frac{c}{\lambda}}{\left(e^{\frac{hc}{\lambda k_B T}} - 1 \right)} \right)$$

from our geometry of the hole so

$$I(\lambda) = \frac{2\pi h c^2}{\lambda^5} \left(\frac{1}{\left(e^{\frac{hc}{\lambda k_B T}} - 1 \right)} \right)$$

We have the Planck function. The math is intriguing, but the point here is that to make this work we had to assume quantized oscillators. These quantized oscillators made standing waves with discrete energies. Oscillations with energies in between the E_n died out and didn't contribute. We saw just this sort of behavior in harmonics in musical instruments in PH123. We have to allow that some energies don't work and a discrete set of energies do work. This is shades of things to come as we study atoms and build their energy shells.

Both the photoelectric effect and the black body radiation curve show that there is more to how light works than we learned in PH123. Light is a wave, but it is not exactly like a water wave or a sound wave. We found that light comes in small packets called photons, and that in the case of thermal equilibrium, not all wavelengths of light would be produced. These are both examples of quantization.

9 Photons and Matter

3.4, 3.5, and 4.1

Fundamental Concepts in the Lecture

- Energy, and Momentum of a Photon
- Compton Effect
- Photon Interactions; Pair Production
- De Broigle Waves

In our last lecture we learned that light comes in small wave packages that superimpose to make large light beams. We called the wave packages “photons.” In this lecture we are going to look at how photons interact with matter. We are going to find that in many ways photons act like particles. At the end of the chapter we will discuss the idea of wave-particle duality.

The discovery of X-rays

In 1895, a physicist, Wilhelm Roentgen, noticed that a fluorescent screen glowed even when it was meters from his gas discharge tube. You have seen a fluorescent screen if you have seen an old fashioned television set. The screen of the TV was a series of fluorescent dots painted on a sheet of glass. Roentgen’s version was a separate thing, no television equipment around it. But Roentgen was surprised, knowing that normal charged particles should not have been able to travel for meters. He placed a black screen in front of the fluorescent screen and was again surprised. The radiation creating the fluorescent glow seemed to go right through the black screen! He tried subjecting his new “rays” to electric and magnetic fields. They did not bend. So they were not charged particles. He had no idea what he had found, so he called the radiation *X*-rays. The *X* standing for an unknown thing.

Another physicist (von Laue) suggested that the *X*-rays were small electromagnetic

waves. He decided that if the wavelengths were short enough, he should be able to use a salt crystal like a diffraction grating to bend the light into patterns. They tried this, and found the predicted pattern (we talked about this in our last lecture).

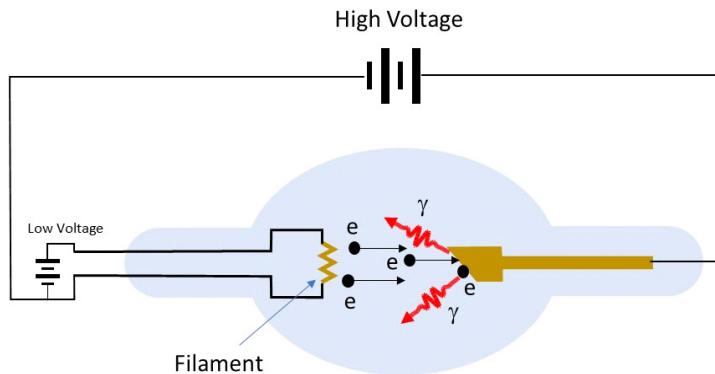
By using the known spacing of the atoms in the salt crystal, they could calculate the wavelength of the electromagnetic waves (about 0.1 nm).

These waves have high frequencies. We now know that this means they have high energies. And because

$$\lambda = \frac{c}{f}$$

they must have small wavelengths. These characteristics explain why they can penetrate most things.

Making X-rays

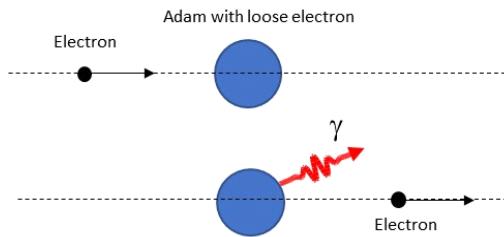


A schematic of an *X*-ray tube is shown above. The idea behind making x-rays is to take fast moving electrons and slow them down in a hurry.

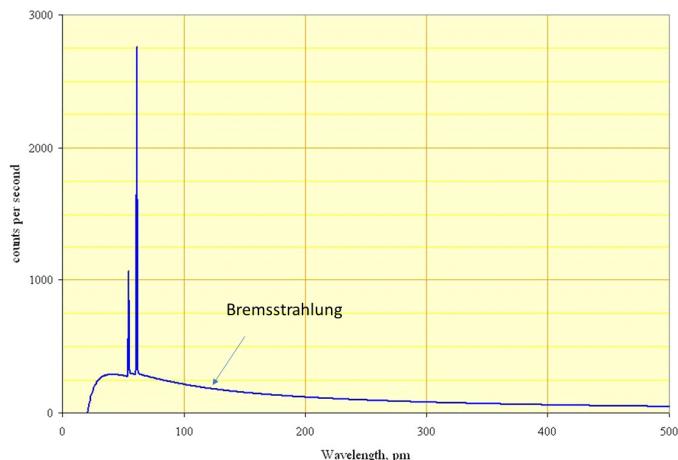
We learned from studying antennas back in PH220 that accelerating electrons make electromagnetic waves. In this apparatus we will accelerate electrons and again we will produce electromagnetic waves. This time we start with very fast electrons, so they have a large kinetic energy. We will recover the lost kinetic energy as electromagnetic waves, *X*-rays.

The way it works is that the wire marked *filament* has a small potential difference. So current flows through the filament and it will heat up due to resistance. A large potential will exist between the filament and the copper rod shown. This potential difference

will cause electrons to leave the filament. This is called an *electron gun* because it “shoots” electrons like bullets. The large potential difference will accelerate the electrons toward the tungsten coating on the copper rod. As the electrons impact the target, they will slow down quickly.



The electrons will nearly impact the target atom’s nucleus. The deceleration may stop the electron, giving the most energy to the created X-ray photon. Or the electron may only lose part of its kinetic energy.



Spectrum of the X-rays emitted by an X-ray tube with a rhodium target, operated at 60 kV. The continuous curve is due to bremsstrahlung, and the spikes are characteristic K lines for rhodium. The curve goes to zero at 21 pm in agreement with the Duane–Hunt law, as described in the text.

If we plot the intensity of the emitted X-rays, we will find a graph much like the one above. The tall peaks we will have to wait to understand, but the rest of the curve comes from this “breaking” of the electrons. The radiation is called *bremsstrahlung* radiation, which means “breaking” or “slowing down” in German.

There are not many calculations that we can do yet with what we know about X-rays, but we can find the minimum wavelength of the X-rays produced. We know that the

kinetic energy of the electrons will be

$$K = e\Delta V \quad (9.1)$$

and we know from Planck that the energy of the photon will be

$$E = hf \quad (9.2)$$

so if the electron is completely stopped all the energy will go into making a photon.

$$e\Delta V = hf \quad (9.3)$$

Then

$$e\Delta V = h\frac{c}{\lambda} \quad (9.4)$$

or

$$\lambda = \frac{hc}{e\Delta V} \quad (9.5)$$

is the minimum wavelength of the X-rays.

Most of us have had an *X*-ray taken, so these once mysterious rays are now quite commonplace. And of course we have already talked about *X*-rays and diffraction through crystals to find the structure of the crystal.

Compton Effect

In 1923, an American physicist named Compton performed an important experiment. He shot X-rays at a carbon target. He measured the wavelength of the X-rays before and after they hit the target and got a surprise. The wavelength changed after the collision (we often say it “shifted” because the peak on a spectrograph would move over on the graph). Compton found that the amount of wavelength change depended on the angle at which the X-rays scattered away from the target.

Compton was familiar with the idea of a photon from Einstein’s work. He theorized that if light did travel in small wave packets like particles, then those small wave packets would carry momentum. And if that were true, the photons could have collisions like billiard balls would. He could use conservation of momentum to explain what he saw!

Think about this for a minute. From what Plank and Einstein taught us the energy of the photon is equal to

$$E = hf$$

and for light in a vacuum

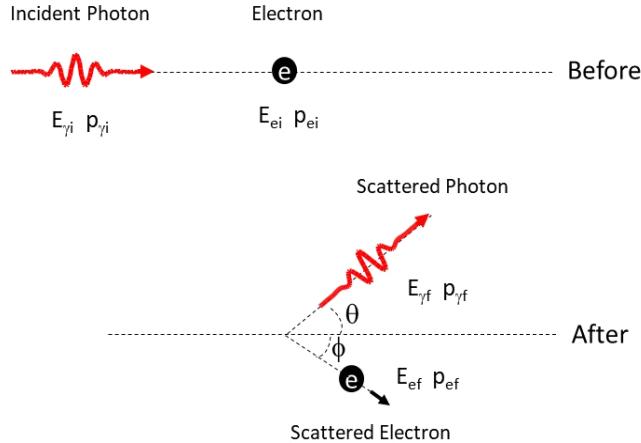
$$c = \lambda f$$

so

$$E = h\frac{c}{\lambda}$$

if the wavelength shifts, the energy changes.

He theorized that the photons hit electrons, as shown in the figure below.



The shift in wavelength, then, is given by

$$\Delta\lambda = \frac{h}{m_e c} (1 - \cos \theta) \quad (9.6)$$

where m_e is the electron mass. The angle θ is as shown in the figure.

We are assuming that the photon not only has energy, but also carries momentum. But from relativity we know that this is true.

$$p = \frac{E}{c} \quad (9.7)$$

which we can now write as

$$\begin{aligned} p &= \frac{h \frac{c}{\lambda}}{c} \\ &= \frac{h}{\lambda} \end{aligned} \quad (9.8)$$

Let's see how this works. For the photon we have

$$E_\gamma = hf = \frac{hc}{\lambda}$$

where we are using γ to indicate a photon. Then

$$p_\gamma = \frac{E_\gamma}{c}$$

For the electron we now know the total energy is the kinetic energy plus the rest energy

$$\begin{aligned} E_e &= K_e + E_{eo} \\ &= K_e + m_e c^2 \end{aligned}$$

and

$$p_e = \frac{m_e v_e}{\sqrt{1 - \frac{v_e^2}{c^2}}}$$

Let's use conservation of energy and conservation of momentum to find the Compton formula. We can write out conservation of energy

$$\begin{aligned} E_{\gamma i} + E_{ei} &= E_{\gamma f} + E_{ef} \\ E_{\gamma i} + m_e c^2 &= E_{\gamma f} + E_{ef} \end{aligned}$$

and conservation of momentum

$$\begin{aligned} p_{\gamma ix} + p_{eix} &= p_{\gamma fx} + p_{efx} \\ p_{\gamma iy} + p_{eiy} &= p_{\gamma fy} + p_{efy} \end{aligned}$$

we can use our angles to make this

$$\begin{aligned} p_{\gamma i} + 0 &= p_{\gamma f} \cos \theta + p_{ef} \cos \phi \\ 0 + 0 &= p_{\gamma f} \sin \theta - p_{ef} \sin \phi \end{aligned}$$

Suppose we measure the energy of the incident photon and the energy and direction of the scattered photon. Then $E_{\gamma i}$, $E_{\gamma f}$, and θ are known. We can eliminate ϕ by solving for the terms in p_{ef} and squaring the equations, and adding the result

$$\begin{aligned} p_{\gamma i} - p_{\gamma f} \cos \theta &= p_{ef} \cos \phi \\ p_{\gamma f} \sin \theta &= p_{ef} \sin \phi \\ (p_{\gamma i} - p_{\gamma f} \cos \theta)^2 &= p_{ef}^2 \cos^2 \phi \\ (p_{\gamma f} \sin \theta)^2 &= p_{ef}^2 \sin^2 \phi \end{aligned}$$

Now we can add these equations together to eliminate ϕ

$$\begin{aligned} p_{ef}^2 \sin^2 \phi + p_{ef}^2 \cos^2 \phi &= (p_{\gamma f} \sin \theta)^2 + (p_{\gamma i} - p_{\gamma f} \cos \theta)^2 \\ p_{ef}^2 (\sin^2 \phi + \cos^2 \phi) &= p_{\gamma f}^2 \sin^2 \theta + p_{\gamma i}^2 - 2p_{\gamma i} p_{\gamma f} \cos \theta + p_{\gamma f}^2 \cos^2 \theta \\ p_{ef}^2 &= p_{\gamma f}^2 \sin^2 \theta + p_{\gamma f}^2 \cos^2 \theta + p_{\gamma i}^2 - 2p_{\gamma i} p_{\gamma f} \cos \theta \\ p_{ef}^2 &= p_{\gamma f}^2 + p_{\gamma i}^2 - 2p_{\gamma i} p_{\gamma f} \cos \theta \end{aligned}$$

Now we know

$$E_{ef}^2 = c^2 p_{ef}^2 + m_e^2 c^4$$

From conservation of energy we know

$$E_{\gamma i} + m_e c^2 - E_{\gamma f} = E_{ef}$$

so our last equation can be written as

$$(E_{\gamma i} + m_e c^2 - E_{\gamma f})^2 = c^2 p_{ef}^2 + m_e^2 c^4 \quad (9.9)$$

and, just to remind ourselves, from conservation of momentum we found

$$p_{ef}^2 = p_{\gamma f}^2 + p_{\gamma i}^2 - 2p_{\gamma i} p_{\gamma f} \cos \theta$$

so our last energy equation (9.9) can be written as

$$(E_{\gamma i} + m_e c^2 - E_{\gamma f})^2 = c^2 (p_{\gamma f}^2 + p_{\gamma i}^2 - 2p_{\gamma i} p_{\gamma f} \cos \theta) + m_e^2 c^4$$

and we can use

$$p_\gamma = \frac{E_\gamma}{c}$$

for the photon momenta. Then

$$(E_{\gamma i} + m_e c^2 - E_{\gamma f})^2 = c^2 \left(\frac{E_{\gamma i}^2}{c^2} + \frac{E_{\gamma f}^2}{c^2} - 2 \frac{E_{\gamma i}}{c} \frac{E_{\gamma f}}{c} \cos \theta \right) + m_e^2 c^4$$

or just

$$(E_{\gamma i} + m_e c^2 - E_{\gamma f})^2 = E_{\gamma i}^2 + E_{\gamma f}^2 - 2E_{\gamma i} E_{\gamma f} \cos \theta + m_e^2 c^4$$

And expanding the left hand side gives

$$m_e^2 c^4 - 2m_e c^2 E_{f\gamma} + 2m_e c^2 E_{i\gamma} + E_{f\gamma}^2 - 2E_{f\gamma} E_{i\gamma} + E_{i\gamma}^2 = E_{\gamma i}^2 + E_{\gamma f}^2 - 2E_{\gamma i} E_{\gamma f} \cos \theta + m_e^2 c^4$$

We can cancel some terms

$$-2c^2 m_e E_{f\gamma} + 2c^2 m_e E_{i\gamma} - 2E_{f\gamma} E_{i\gamma} = -2E_{\gamma i} E_{\gamma f} \cos \theta$$

Now we want to transform this into a nice equation that relates the wavelengths. The path isn't immediately clear. But let's try to isolate the $\cos \theta$ term a bit. Try dividing by $E_{\gamma i} E_{\gamma f}$

$$\frac{-2c^2 m_e E_{f\gamma} + 2c^2 m_e E_{i\gamma} - 2E_{f\gamma} E_{i\gamma}}{E_{\gamma i} E_{\gamma f}} = \frac{-2E_{\gamma i} E_{\gamma f} \cos \theta}{E_{\gamma i} E_{\gamma f}}$$

or just

$$\frac{-2c^2 m_e E_{f\gamma}}{E_{\gamma i} E_{\gamma f}} + \frac{2c^2 m_e E_{i\gamma}}{E_{\gamma i} E_{\gamma f}} - \frac{2E_{f\gamma} E_{i\gamma}}{E_{\gamma i} E_{\gamma f}} = -2 \cos \theta$$

or more simply

$$\frac{-2c^2 m_e}{E_{\gamma i}} + \frac{2c^2 m_e}{E_{\gamma f}} - 2 = -2 \cos \theta$$

Cancelling the two's gives

$$\frac{-c^2 m_e}{E_{\gamma i}} + \frac{c^2 m_e}{E_{\gamma f}} = 1 - \cos \theta$$

More algebra gives.

$$c^2 m_e \left(\frac{1}{E_{\gamma f}} - \frac{1}{E_{\gamma i}} \right) = 1 - \cos \theta$$

or

$$\left(\frac{1}{E_{\gamma f}} - \frac{1}{E_{\gamma i}} \right) = \frac{1 - \cos \theta}{c^2 m_e}$$

and we know that

$$E_\gamma = hf = \frac{hc}{\lambda}$$

so

$$\left(\frac{1}{\frac{hc}{\lambda_f}} - \frac{1}{\frac{hc}{\lambda_i}} \right) = \frac{1 - \cos \theta}{c^2 m_e}$$

$$(\lambda_f - \lambda_i) \frac{1}{hc} = \frac{1 - \cos \theta}{c^2 m_e}$$

$$\lambda_f - \lambda_i = \frac{h}{cm_e} (1 - \cos \theta)$$

Note that the final wavelength, λ_f , is not the same as the initial wavelength. That was

what Compton saw! So then

$$\lambda_f = \lambda_i + \Delta\lambda$$

And finally we have our wavelength shift

$$\Delta\lambda = \frac{h}{cm_e} (1 - \cos \theta)$$

The quantity

$$\frac{h}{cm_e}$$

is sometimes called the Compton wavelength, but this is a bad name. Clearly it is closer to a $\Delta\lambda$.

We can find the direction of the electrons by returning to our momentum equations

$$\begin{aligned} p_{\gamma i} - p_{\gamma f} \cos \theta &= p_{ef} \cos \phi \\ p_{\gamma f} \sin \theta &= p_{ef} \sin \phi \end{aligned}$$

and dividing

$$\frac{p_{ef} \sin \phi}{p_{ef} \cos \phi} = \frac{p_{\gamma f} \sin \theta}{p_{\gamma i} - p_{\gamma f} \cos \theta}$$

and using

$$p_\gamma = \frac{E_\gamma}{c}$$

The \sin / \cos term becomes a tangent.

$$\begin{aligned} \tan \phi &= \frac{\frac{E_{\gamma f}}{c} \sin \theta}{\frac{E_{\gamma i}}{c} - \frac{E_{\gamma f}}{c} \cos \theta} \\ &= \frac{E_{\gamma f} \sin \theta}{E_{\gamma i} - E_{\gamma f} \cos \theta} \end{aligned}$$

But the important point to notice is that we are doing conservation of momentum and energy as though the light were a particle. The photon is participating in what we might think of as a collision. And this is not a normal way to think about something that we described earlier as a wave.

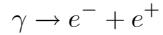
Pair Production

If all this was not strange enough, photons can do stranger things yet. Suppose we have a high energy photon. We know from our study of relativity that rest energy

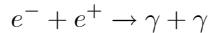
$$E_o = mc^2$$

which strongly tells us that mass is just a form of energy. If our photon has enough energy, it can create matter! This was quite a surprise! But a photon can create a pair of particles (it must be a pair to conserve momentum). One possibility for the pair of

particles is an electron and a positron.



The positron is a particle that has the same mass as an electron and acts like an electron, but it has a positive charge and, significantly, it is antimatter!. These particles form the basis of PET scanners. If the positron and electron don't move apart quickly, they will recombine and out will come two photons again,



where this time we get two photons to conserve momentum. This is called *annihilation*, when matter and antimatter combine to form photons. No mass remains.

Once again, this is a photon acting in a very not-wave-like manner.

Wave-Particle Duality of Light

So now we have studied interference of light which is a wave phenomena, and we have studied Compton scattering, positron-electron annihilation and pair production, and Bremsstrahlung emission that all seem to be particle-like properties of light. Which is it, waves or particles?

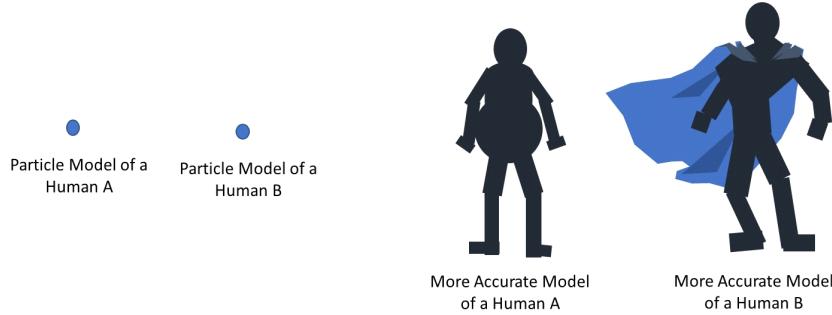
Young's double slit experiment seemed to indicate conclusively that light is a wave. And we successfully did many problems using this wave theory of light back in PH123 (or your second semester physics class). Then the photoelectric effect and the Compton effect were discovered and light seemed to act like a particle again! What is it? Our current theory says it is both! Well, that is a little confusing. It really says that it is a more complicated localized wave. It is not really a wave or a particle, but something else that sometimes looks like a wave and sometimes looks like a particle. Instead of a simple sine wave, it is a short packet of waves.



Before we declare this as too strange, let's realize that a particle model of a human is sometimes a good model

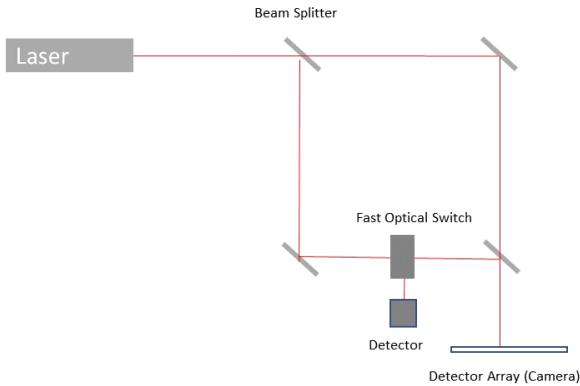


But sometimes viewing a person as a shapeless point is not good enough. For one thing, all particles look alike, and not all humans look alike.



Our particle model from PH121 was designed to make math easier. We could ignore all the internal details of, say, person A if person A is, say, shot out of a circus cannon and is therefore a projectile. Person A can flail his arms or twist, but it won't change his projectile motion. But once we learned about rotational motion, particle model wasn't enough. After all, person A might do artistic somersaults and twists in the air and particle model couldn't describe that. We had to go to extended free-body diagrams and walk away a bit from particle model to describe rotation. Now we will need to do this again. The idea of a particle is not detailed enough to describe a photon. The photon has wave properties (like person A has limbs) and we can't ignore this in all circumstances.

Let's take on a specific experiment to show the range of behavior we should expect from photons.



A laser produces photons. The beam of light must somehow be related to the photons. That beam goes to a beam splitter. Half of the beam goes through the beam splitter and half is reflected off of the beam splitter. A series of mirrors allows the beams to recombine and the recombined waves mix to form an interference pattern on a detector array (something like a camera sensor). But along one path we can place a high speed optical switch. This switch diverts the beam to a detector, something like a silicon photo diode or a photomultiplier. This detector uses photoelectric effect to cause a detection, so basically it is a photon detector. If we turn the switch off, we get two waves mixing at the camera. We see interference. This is a wave phenomena. If we engage the switch, we get detected photons on the photon detector, and this is a particle-like phenomena.

You might say to yourself “wait, maybe the laser knows if the switch is engaged, and if it is, it sends photons and if it isn’t it sends waves. But we could make the beam paths very long, and the switch very fast so that the switch could be engaged long after the waves or photons left the laser. The system still works the same. So it can’t be that we have a decision happening at the laser.

It appears that the wave-like-particle-like nature of light is all part of the way light really is. Neither a particle model nor a wave model is all the way right. But this should not be too surprising (maybe a little weird, but not too surprising). There is no such thing as a “particle.” A particle is an abstraction we used to simplify the math. But no object in the universe is well described in every detail by the ideal of a particle. No one is shocked to find out that divers have limbs and can do twists and turns.² We should also not be too surprised to find that photons have wave properties. Nothing in the universe is really a particle. When our objects get smaller or more fundamental, like photons, the particle model does an even poorer job of describing all the behavior. We will see much more of this failure of particle approximation as we go, starting with the work of

² I can hear in my mind the PBS show about the strange limb-particle duality of divers....

De Broglie.

Wave-particle Duality of Matter

You might be saying to yourselves aren't these dual wave-particle photons quite strange? A physicist named De Broglie also asked himself this question (see, you think like famous physicists!). His conclusion was that maybe light is not so strange. Maybe *everything* really has a wave-particle duality. Let me say this another way. Maybe electrons, protons, and neutrons are really localized wave-like things. Since they have momentum, suppose we use the momentum equation for a photon backwards to find the wavelength of, say, an electron

$$p = \frac{h}{\lambda} \implies \lambda = \frac{h}{p} \quad (9.10)$$

then if our electron is not going too fast,

$$\lambda = \frac{h}{m_e v_e} \quad (9.11)$$

After going this far, he thought he could use the results of the Einstein and Planck to write the energy of these *matter waves* as

$$E = hf \quad (9.12)$$

and then he could use the energy to find the frequency of the waves

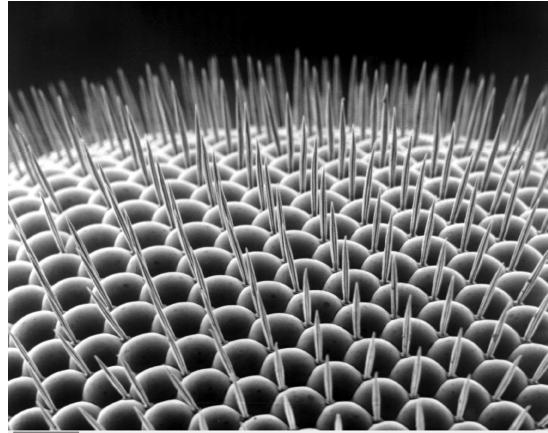
$$f = \frac{E}{h} \quad (9.13)$$

The ideas of De Broglie were not taken well by the scientific community at large until by accident Davison and Germer measured the wavelength of electrons. They were scattering low energy electrons off a nickel target in a vacuum system. The vacuum system sprang a leak, and their nickel target oxidized. They were probably a little annoyed, but this was really the best day of their careers! They heated up the Nickel target to remove the oxidation (blowing hydrogen over it to carry away the oxidation remnants). Once it cooled, they replaced their target, and started up their experiment. But now the electrons scattered into neat patterns! Davison and Germer recognized them as diffraction patterns! The Nickel target had formed a regular crystal structure when it was heated, and now they had diffraction just like X-rays from NaCl crystals! They repeated the experiment, but this time with the intent of measuring the wavelength of the electron waves. Others soon repeated the experiment. And de Broglie's ideas were proven right!

Let's state clearly what this means

All particles have a dual wave-particle nature

We don't see this wave like behavior of matter in everyday experience because the wavelengths are so very small. Electron wavelengths used in electron microscopes are about 100 times smaller than normal light wavelengths use in normal optical systems. But we can make pictures with electron waves! It is also true that in every day experience we don't pass particles through narrow slits. So most matter behavior is well explained by the particle part of the wave-particle duality of matter. It depends on our experiment, which behavior we will see. We will develop our theory of matter waves in our next lecture.



Scanning electron microscope image of an eye on a fruit fly. This image was taken using electron waves. Image courtesy Dartmouth College.

10 De Broglie Waves and Uncertainties

4.2-4.4

Fundamental Concepts in the Lecture

- de Broglie and the Wave nature of Matter
- Electron Microscopes
- Heisenberg uncertainty relations

Question: The first experiments that verified the wave-like nature of electrons through diffraction were performed by directing electrons through

- a. A single slit
- b. A double slit
- c. A crystal lattice
- d. Jello

In the last lecture, we suggested that matter, like electrons and protons, might really be waves in the same way that photons are waves. It is time to develop this theory.

Particle Interference

If, as we said in the last lecture, electrons have a wavelength, then we should be able to make a double slit experiment using electrons. This is hard to do with the wavelength of, say 40keV electrons. Let's see why. We know from de Broglie

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

or, since

$$hc = 1240 \text{ eV nm}$$

we could write the wavelength as

$$\lambda = \frac{hc}{pc}$$

If our electron is not traveling relativistically we can say that

$$\begin{aligned} p &= mv = \sqrt{m^2 v^2} \\ &= \sqrt{2m \frac{1}{2} mv^2} = \sqrt{2mK} \\ &= \frac{1}{c} \sqrt{2(mc^2)K} \end{aligned}$$

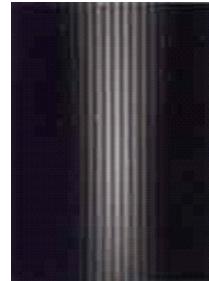
which is convenient because we know $\Delta K = e\Delta V$ for an electron shot out of an electron gun. If we know the accelerating stage voltage for making the electron beam, we can know the kinetic energy and then the momentum. Since we know for electrons $m = 0.511 \text{ MeV}/c^2$

$$\begin{aligned} p &= \frac{1}{c} \sqrt{2(0.511 \times 10^6 \text{ eV}/c^2) c^2 (40000 \text{ eV})} \\ &= 2.0219 \times 10^5 \text{ eV}/c \end{aligned}$$

so then

$$\begin{aligned} \lambda &= \frac{1240 \text{ eV nm}}{2.0219 \times 10^5 \text{ eV}} \\ &= 6.1328 \times 10^{-3} \text{ nm} \\ &= 0.061328 \text{ Å} \end{aligned}$$

It's hard to build slits that will work for such small wavelengths. Claus Jonsson and his group did this back in 1961. They built double, triple, quadruple, and up to five slits. Their double slit had a slit width $a = 0.5 \mu\text{m}$ with a slit spacing of $d = 2 \mu\text{m}$. The distance from the slit to the screen was $L = 0.4 \text{ m}$. Here is one of the results from illuminating the double slit with electrons.



Claus Jonsson, "Electron Diffraction at Multiple Slits," AJP Volume 42, Jan 1974, pp 4-11
(Translated by Deitrich Brandt and Stanley Hirschi)

We can clearly see a diffraction pattern. For a double slit experiment we know that

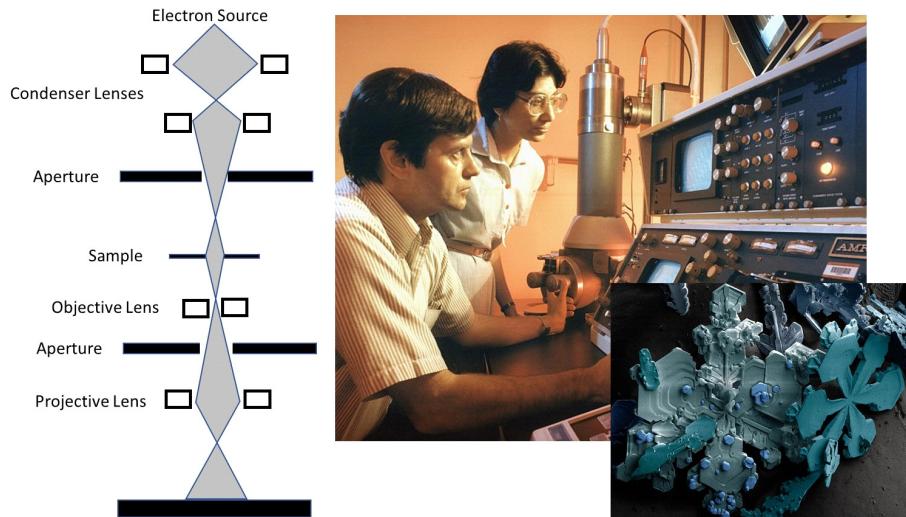
$$y_m = \frac{m\lambda L}{d} \quad m = 0, \pm 1, \pm 2, \dots$$

so that

$$\begin{aligned} y_1 &= \frac{(0.061328 \text{ \AA}) (0.4 \text{ m})}{2 \mu\text{m}} \\ &= 1.2266 \times 10^{-6} \text{ m} \\ &= 1.23 \mu\text{m} \end{aligned}$$

Which is, as Jonsson reported, close to his measured value. Jonsson had to magnify his diffraction pattern in order to record it. A fringe spacing on the order of $1 \mu\text{m}$ is hard to see with your unaided eyes. And so within the uncertainties of the apparatus and the magnification system, the experiment gave consistent values. But just the image itself is a stunning example of the wave nature of electrons!

With this wave nature we can do some fantastic things. A scanning electron microscope (SEM) is one of them. The next figure gives a schematic SEM design and shows what one looks like (complete with very dramatic looking human operators) and shows an image of ice crystals made with electrons waves.



Schematic of a Scanning Electron Microscope (SEM) and a picture of a US Department of Energy SEM. In the lower right hand corner is a SEM image of ice crystals.

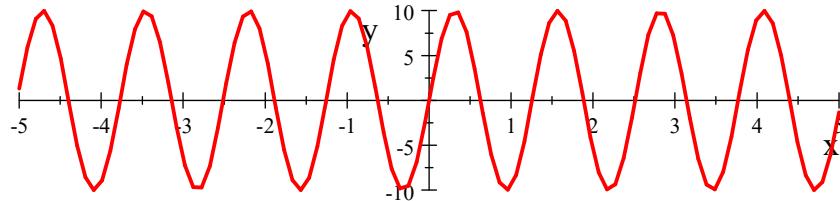
Although it is strange to consider matter as being made of waves, The fact that we can build a device like a SEM is a very good hint that this wave model of matter is not wrong. It is a little like building and using a surfboard to prove that there are ocean waves. Once you are metaphorically hanging 10, it is good evidence that your model works, no matter how unnerving that model might at first appear.

Uncertainty Relationships

When we think of a wave, so far in physics we have mostly thought of something like this

$$y = A \sin(kx - \omega t - \phi)$$

which in practice might look like this



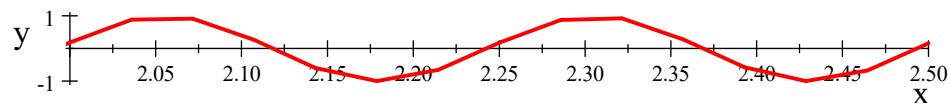
Notice that there is no start or stop to this kind of wave. Our figure starts at $x = -5$ and ends at $x = 5$, but the equation does not! There is a value of y for every x from $-\infty$ to $+\infty$. But our photons are not such waves. They are limited waves at best.



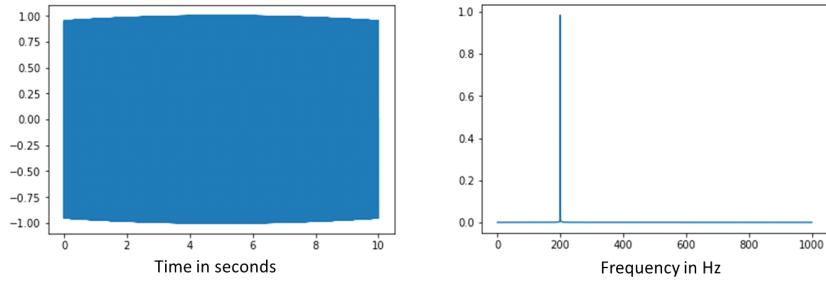
We should investigate what happens when you have a limited wave. I did this investigation using Python. Suppose we have a sine wave with $f = 200$ Hz, but I limit this wave's existence by making it start at $t_i = 0$ and then make it end at $t_f = 10$ s. I could do this in practice by turning on a radio transmission or even an acoustic speaker, and then turning the device off ten seconds later. Our screen resolution is terrible for plotting such a function, but in the figure below you can see that our signal only exists from $t = 0$ to $t = 10$ s.



If I zoom in on a part of the graph, we can see that it is really a sin wave.



Python did equally bad at plotting this. All we see is a blue band.



10s, 200Hz

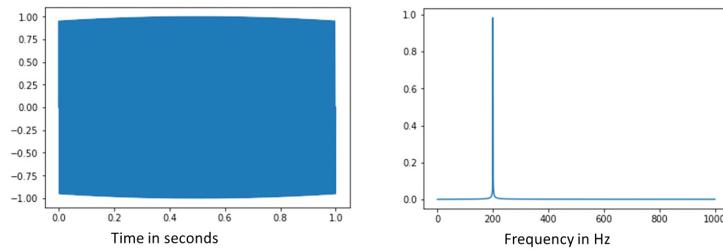
But in the second graph, notice that we have plotted frequency. Python and most scientific programming languages have functions to find which frequencies are in a signal. This function is referred to as a Fourier Transform. In PH123, we would say that these functions perform the job of a spectrometer, so we would call the figure to the right a spectrogram (or just a spectrum).³

We expect only one frequency, 200 Hz, and that is mostly what we get. Since our period for our wave is

$$T = \frac{1}{200 \text{ Hz}} = 0.005 \text{ s}$$

and we have 10 s of data, that is four orders of magnitude more signal than a period.

The whole signal seems very long compared to a period. We expect this to look kind of like an infinite signal. But suppose we take the same wave, but for less time. We limit the wave more.

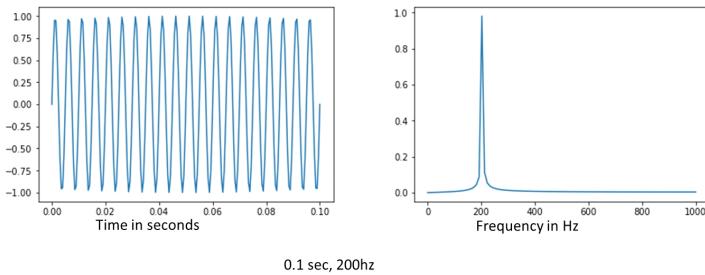


1sec, 200hz

So we still get a blue blur for our wave picture, but now the wave only exists for one instead of ten seconds. If you look closely at the frequency graph, you will notice that the 200 Hz peak representing our wave is a bit wider right at the bottom.

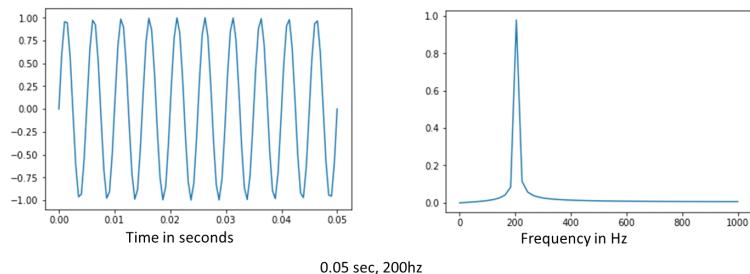
We could limit our wave more, say, so it only lasts $t_f = 0.1 \text{ s}$. We would get a set of graphs that look like this.

³ If the idea of a Fourier transform isn't familiar, see Appendix A.

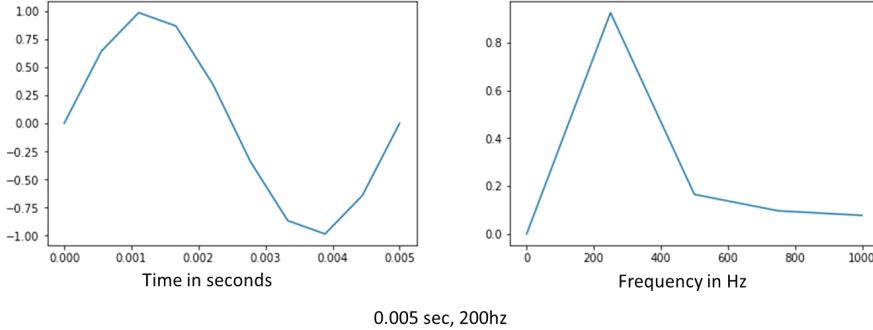


0.1 sec, 200hz

Notice that not only can Python render the wave now, but more importantly the 200 Hz frequency peak is noticeable wider. This is profound! It means that by limiting the wave, we no longer have just one frequency! The graph tells us we have mostly 200 Hz but we also have some 199 Hz and some 201 Hz and some 190 Hz and some 210 Hz, etc. The very fact that the wave does not go on so long requires that we have more than one frequency in the wave. We could say that as Δt gets smaller, our Δf is getting bigger. Here are two more examples with smaller Δt values.



0.05 sec, 200hz



0.005 sec, 200hz

The cost of limiting our waves is that we can't have a single frequency for the wave. For an experimentalist, this means that if you only measure a short segment of the signal,

you have an increased uncertainty in the frequency you will find from that signal. For a photon, it means that we can expect uncertainty in the frequency (or wavelength) because the photon is limited. It is important to know what we mean by uncertainty in this case. In our example above, when we say that Δf increased we really mean that we have more than one frequency. We don't just mean that we don't know the frequency well. We really are mixing more than one frequency.

We could say that we have increased the uncertainty in the frequency by some amount ε such that

$$\Delta f \approx \varepsilon f = \frac{\varepsilon}{T}$$

or the uncertainty in f is some fraction ε of the peak frequency. If we take our last figure and say that $\Delta t \approx T$, we could say

$$\Delta t \Delta f \approx T \frac{\varepsilon}{T} = \varepsilon$$

which gives us an inverse relationship between Δt and Δf

$$\Delta f \approx \frac{\varepsilon}{\Delta t}$$

Or even more interestingly, for a photon $\Delta E = hf$ so

$$\Delta E \approx \frac{h\varepsilon}{\Delta t}$$

and we have an *uncertainty relationship* between the duration, Δt , and the energy in the wave, ΔE . And it tells us just what we have been studying. If the duration of our wave signal is short (small Δt) then our wave signal has to be made of more partial waves with different energies (different frequencies). So we don't have just one frequency. We have a wider *bandwidth* of frequencies centered around a main frequency.

We took the history graph viewpoint in what we have done so far. But we could just as well have taken a snapshot viewpoint and we would have ended up with an uncertainty condition on the length of the whole wave Δx and the wavelength λ . Again modeling the uncertainty as a fraction of the wavelength ϵ we have

$$\Delta x \Delta \lambda \approx \lambda \times \epsilon \lambda = \epsilon \lambda^2$$

where we have used Δx as nearly one wavelength. Once again we don't just mean we don't know λ well. We are actually mixing more than one λ to describe our limited wave.

We could make a measurement of N wavelengths. Then the uncertainty in our wavelength gets better (Think of PH150, measuring millimeter with a meter stick has an uncertainty of about a millimeter - the smallest marking - and measuring a meter with a meter stick has an uncertainty of about a millimeter, but now that millimeter is not as important compared to the whole measurement $1 \text{ mm}/1 \text{ m} = 0.001$ so the

fractional uncertainty is less). We could say that our uncertainty would go down by a factor of $1/N$

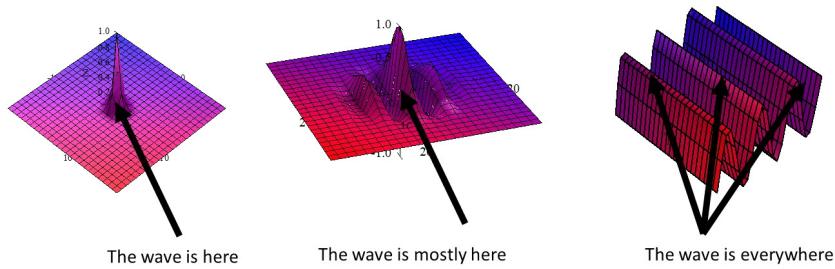
$$\Delta\lambda \approx \frac{\epsilon\lambda}{N}$$

and our distance measurement was $\Delta x = N\lambda$. So then

$$\Delta x \Delta \lambda \approx (N\lambda) \left(\frac{\epsilon\lambda}{N} \right) = \epsilon\lambda^2$$

and our uncertainty relationship is the same. So our wave reasoning can be used as a normal uncertainty. But we really mean that for waves we have more than one partial wave with different wavelengths combined to make an overall wave.

Note that we have done very little in the way of quantum thinking here! These uncertainty relationships apply to classical waves as well. And this is not surprising. Think of a wave on the beach. Knowing exactly where the wave is (Δx) is easy if the wave is very small.

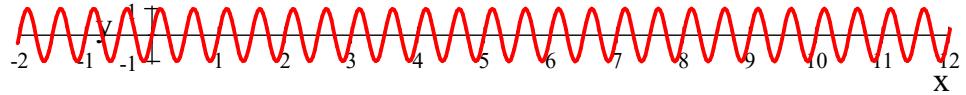


A pulse like wave with small Δx and large $\Delta\lambda$ (left), a more extend wave with larger Δx and smaller $\Delta\lambda$ (middle) and an infinite wave with infinite Δx (because it exists everywhere) and a single λ so $\Delta\lambda$ is zero (right).

For such a limited pulse-like wave $\Delta\lambda$ is larger because you have to have many wavelengths mixed to build a limited wave. You could point to the location of the little wave. But if the wave is an ocean wave that extends across miles of water uniformly, then $\Delta\lambda$ is small, nearly zero and Δx is big. This just says the wave is not all in one location. It is stretched across miles of ocean. And as the extended wave hits the beach, everybody gets wet, not just people in one location. The big wave is not well represented by particle model.

We need a new name for a limited wave. We will call such a wave a *wave packet*.

In quantum mechanics, the wave packet shape tells us something about the position of the particle. A wave packet that looks like this



tells us that the “particle” could be detected anywhere because the wave is everywhere.
A wave packet that looks like this



tells us that the position where the “particle” can be detected must be somewhere close to 4 m.

And of course all real waves have to start and stop at some point, so all real waves are wave packets.

Uncertainty and De Broglie waves

De Broglie waves have a wavelength that is related to the momentum

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

we could write this as

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

Let's find dp by taking a derivative

$$dp = -\frac{h}{\lambda^2} d\lambda$$

The minus sign can be important so we know the different directions of change, but let's look at the magnitude of the change only. If we allow our small changes to be a little bigger, we could, to a good approximation write this as

$$\Delta p = \frac{h}{\lambda^2} \Delta \lambda$$

so

$$\Delta \lambda = \frac{\lambda^2}{h} \Delta p$$

We can use our uncertainty relationship for waves and make it especially for de Broglie waves. Starting with

$$\Delta x \Delta \lambda \approx \epsilon \lambda^2$$

and substituting in our equation for $\Delta\lambda$ we get

$$\begin{aligned}\Delta x \Delta \lambda &\approx \epsilon \lambda^2 \\ \Delta x \left(\frac{\lambda^2}{h} \Delta p \right) &\approx \epsilon \lambda^2 \\ \Delta x \Delta p &\approx \epsilon h\end{aligned}$$

We can use our new wave packet name and say that this last equation tells us that the smaller the wave packet (Δx small) the more uncertain the momentum (Δp large).

We are not quite ready to do the quantum mechanical calculation to find ϵ . That is a problem for PH433. But I will give you the answer now

$$\epsilon = \frac{1}{4\pi}$$

This is the smallest possible uncertainty that we can obtain. This gives our position-momentum uncertainty as

$$\Delta x \Delta p \geq \frac{h}{4\pi}$$

And just to be confusing, let's define a new symbol

$$\hbar = \frac{h}{2\pi} = 6.58 \times 10^{-16} \text{ eV s}$$

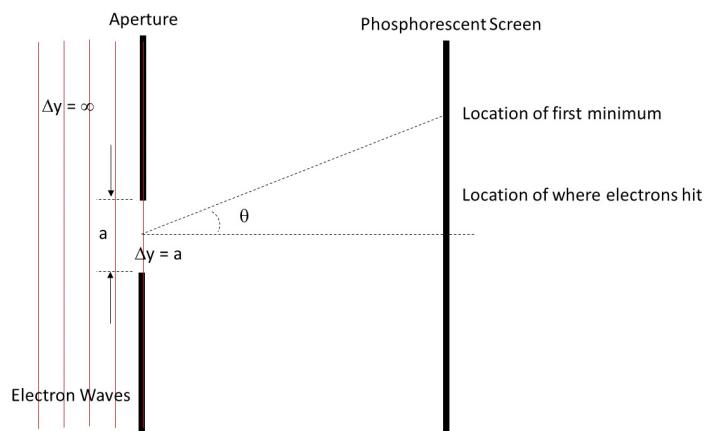
This symbol is pronounced “h bar.” So our position-momentum uncertainty is

$$\Delta x \Delta p \geq \frac{1}{2} \hbar$$

You might have thought that the definition of \hbar should have had a 4π , but that is not our tradition. Often this minimal case is not achieved, so often it is good enough to estimate

$$\Delta x \Delta p \approx \hbar$$

Let's consider a standard experiment to illustrate this uncertainty principle.



Suppose we take a beam of electrons and send them all in the x -direction toward an aperture of size a . The electrons that make it through the aperture will have their

uncertainty in position changed to about a

$$\Delta y = a$$

Then, using our estimate from above

$$\Delta y \Delta p_y \approx \hbar$$

becomes

$$a \Delta p_y \approx \hbar$$

or

$$\Delta p_y \approx \frac{\hbar}{a}$$

We can see that where we would have expected all the momentum to be in the x -direction with $p_y = 0$, we will now have an uncertainty in that p_y . Some electrons will have motion in the y -direction! We can even find an angle where the electrons are likely to go. For small angles

$$\sin \theta \approx \tan \theta = \frac{p_y}{p_x}$$

The wavelength is

$$\begin{aligned} \lambda &= \frac{h}{p_x} \\ p_x &= \frac{h}{\lambda} \end{aligned}$$

so

$$\sin \theta \approx \frac{\frac{h}{a}}{\frac{h}{\lambda}} = \frac{\lambda}{h} \frac{h}{2\pi a} = \frac{\lambda}{2\pi a}$$

Now let's think of a single slit diffraction pattern from our PH123 optics. For the first dark fringe we would have

$$\sin \theta_d = \frac{\lambda}{a}$$

We have most of our electrons going in a direction with an angle that is less than θ_d , which is just what we would expect! We begin to see that diffraction is very connected to our quantum uncertainty relationship!

Let's go back to our energy-time uncertainty relationship.

$$\Delta E \Delta t \approx h \varepsilon$$

We have another ε to find, and once again I am going to quote a future result

$$\varepsilon = \frac{1}{2} \hbar$$

We have the same limiting case for energy-time uncertainty

$$\Delta E \Delta t \geq \frac{1}{2} \hbar$$

and once again we often don't achieve the minimum so a good estimate is

$$\Delta E \Delta t \approx \hbar$$

What we have found is quite profound and bears summarizing

$$\Delta x \Delta p \geq \frac{1}{2} \hbar$$

There is a fundamental limit on how well we can simultaneously determine the position and momentum of a wave packet (object).

$$\Delta E \Delta t \geq \frac{1}{2} \hbar$$

There is a fundamental limit on how well we can simultaneously determine the energy and time coordinate of a wave packet (object).

Notice that we did not use the word “measured” in this summary. It is not that we can’t measure momentum and position with infinite precision (although we can’t!). It is that it is a fundamental impossibility to determine both the momentum and position with infinite precision simultaneously for an object because those quantities don’t have a single value for the object. It is a fundamental limit *because* objects are not really particles. It is *because* they have a wave nature. No effort to improve measurement precision can beat this because it is a part of the wave reality of objects. So this is different than our PH150 measurement uncertainty. For this reason some scientists call this “indeterminacy” instead of “uncertainty.” But I will stick with “uncertainty.” Heisenberg came up with this fundamental limit, so these uncertainty relationships bear his name. These are the Heisenberg uncertainty relations.

Although these uncertainty relationships are just what we should expect from waves, some of the ideas that follow from these ideas of uncertainty can seem strange. Because energy is uncertain for a small amount of time Δt , we can break conservation of energy for a small time. We can have positron-electron pairs simply pop into existence, so long as they pop back out of existence within a time

$$\Delta t \approx \frac{\hbar}{\Delta E_{pair}}$$

and this is totally OK! Waves can do things like this, but particle model would not predict such an outcome.

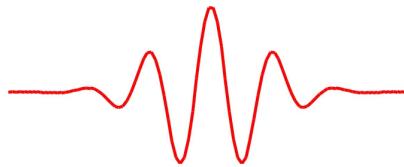
11 Working with Wave Packets

4.5-4.7

Fundamental Concepts in the Lecture

- The mathematical concepts behind wave packets
- Moving wave packets
- Spreading wave packets
- What's in a wave packet?

In the last lecture we introduced the idea of a wave packet. And I used this idea to draw diagrams of photons like this



but really all we did in the last lecture was limit the wave. And we found that limiting our wave mathematically meant that we could not have just one wavelength, but our mathematical description of the wave needed to be a superposition of many waves with different wavelengths. This gave us the starting and stopping of the wave.

But, our picture of a photon wave packet has the amplitude change as well. We need to develop a mathematical way to understand the shape in between the start and stopping points. That is what we will try to do in this lecture.

Building the mathematical concepts behind wave packets

Let's return to PH123 and remember what it is like to mix two waves.

$$\begin{aligned}y_1 &= y_{\max} \sin(k_1 x_1 - \omega_1 t_1 + \phi_{1o}) \\y_2 &= y_{\max} \sin(k_2 x_2 - \omega_2 t_2 + \phi_{2o})\end{aligned}$$

Let's take the case where $\phi_{2o} = \phi_{1o} = \phi_o$, and let's mix the waves in the same location so $x_1 = x_2 = x$. Further let's just let $\phi_o = 0$. And let's say that our waves are not too fast so we can use the non-relativistic approximation that all times are the same.⁴ And let's plot the waves at $t = 0$ so we have

$$\begin{aligned}y_1 &= y_{\max} \sin(k_1 x) \\y_2 &= y_{\max} \sin(k_2 x)\end{aligned}$$

and mixed together we have

$$\begin{aligned}y_r &= y_2 + y_1 \\&= y_{\max} \sin(k_2 x) + y_{\max} \sin(k_1 x)\end{aligned}$$

And let's use our favorite trig identity

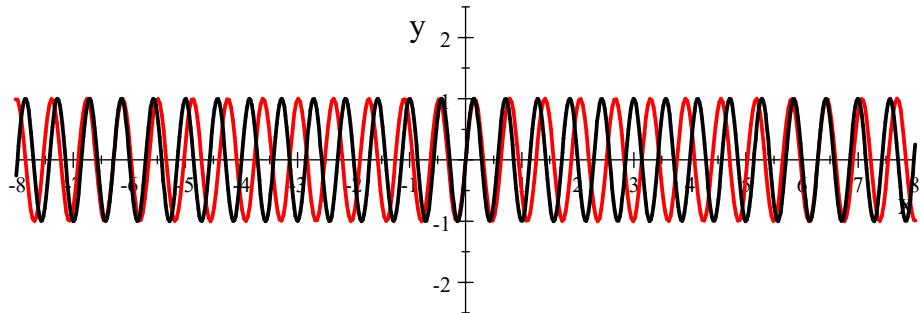
$$\sin a + \sin b = 2 \cos\left(\frac{a-b}{2}\right) \sin\left(\frac{a+b}{2}\right)$$

to write this as

$$\begin{aligned}y_r &= 2y_{\max} \cos\left(\frac{k_2 x - k_1 x}{2}\right) \sin\left(\frac{k_2 x + k_1 x}{2}\right) \\&= 2y_{\max} \cos\left(\frac{x(k_2 - k_1)}{2}\right) \sin\left(\frac{x(k_2 + k_1)}{2}\right) \\&= 2y_{\max} \cos\left(x \frac{\Delta k}{2}\right) \sin(x \bar{k})\end{aligned}$$

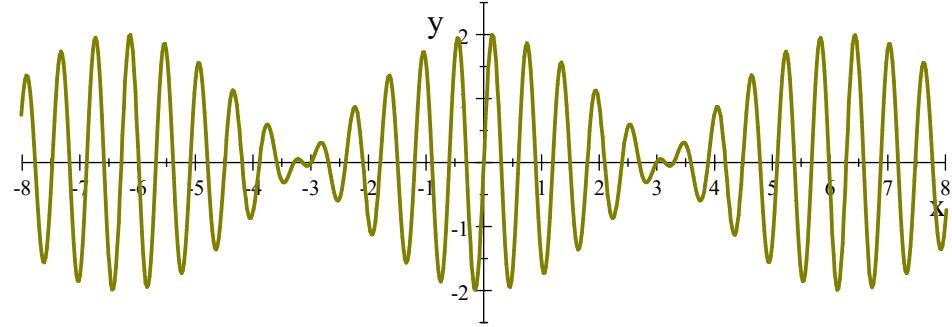
where \bar{k} is the average value of k . Let's plot both of our individual waves and our combined wave. We will need some actual numbers to make the plots. Let's let $k_1 = 10 \frac{\text{rad}}{\text{m}}$ and $k_2 = 11 \frac{\text{rad}}{\text{m}}$ and $A = 1$.

Our individual waves look like



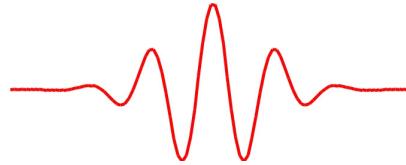
Notice that there are places where the waves are in phase, and places where they are not. The superposition looks like this

⁴ Or do all our experiment in one reference frame so to us all our geoclocks are synchronized.



Where there is constructive interference, the resulting wave amplitude is large, where there is destructive interference, the resulting amplitude is zero. If this were PH123, we would say we have beats. But this time the beating is in the snapshot graph. We have a modulated spatial frequency.

We are part of the way to our goal. We have some of the wave with a different amplitude than the rest of the wave. But we still don't have a wave packet that looks like our photon.



Our uncertainty relation lets us know what to do next. If we want Δx to be smaller, we need more wavelengths. We need to add in more waves. Let's try

$$y_1 = y_{\max} \sin(k_1 x)$$

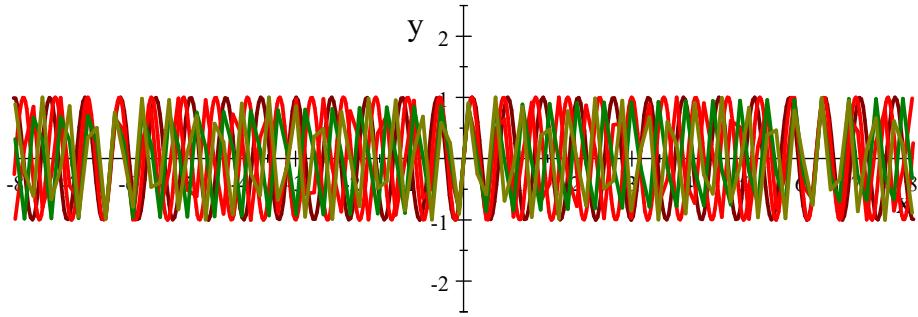
$$y_2 = y_{\max} \sin(k_2 x)$$

$$y_3 = y_{\max} \sin(k_3 x)$$

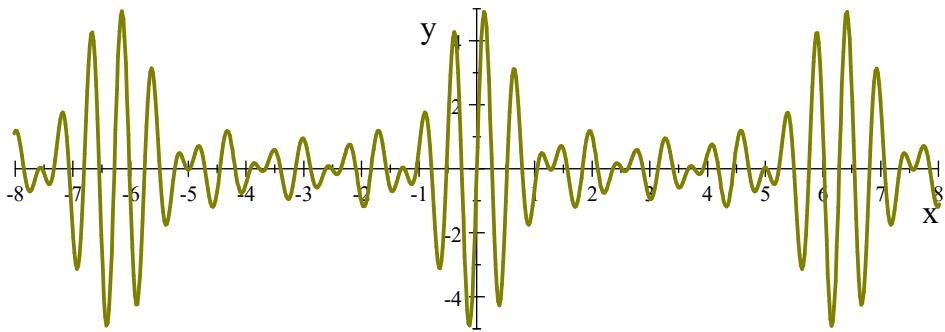
$$y_4 = y_{\max} \sin(k_4 x)$$

$$y_5 = y_{\max} \sin(k_5 x)$$

and let's let $\lambda_1 = 10$, $\lambda_2 = 11$, $\lambda_3 = 12$, $\lambda_4 = 13$, and $\lambda_5 = 14$. We would see the individual waves as



where we can already see regions of constructive interference, and the combined wave looks like



This approach is getting us something that looks more like the shape of our photon, but notice that we always have more than one region that looks right. And these regions that look right go on being reproduced infinitely on either side of the x -axis. So this approach is not enough. We still need to limit the wave. To see how we can fix this let's go back to just two waves and look at our equation for the resultant wave

$$y_r = 2y_{\max} \cos\left(x \frac{\Delta k}{2}\right) \sin(x \bar{k})$$

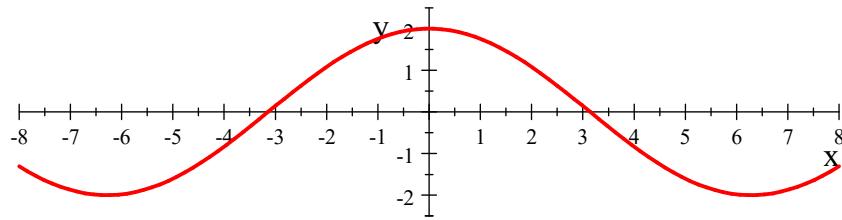
We have an amplitude of

$$A = 2y_{\max} \cos\left(x \frac{\Delta k}{2}\right)$$

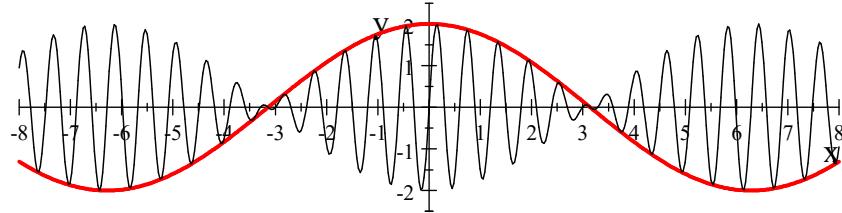
and a wave part

$$\sin(x \bar{k})$$

It is the amplitude that is modifying the wave so we get spots where the wave gets larger. In fact we could plot the amplitude as a function on its own



And now let's plot this amplitude function with the wave part inside it



We can see that the amplitude function seems to contain the wave part of the resultant wave. From the graph it is easy to see the cosine is modifying the sine function, so we have correctly identified which part of our equation is the amplitude and which part is the wave.

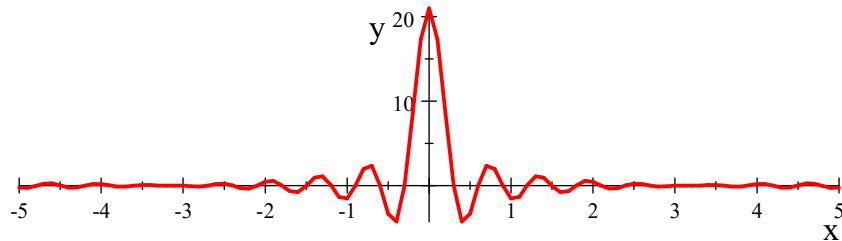
This gives us a clue. The amplitude function

$$A = 2y_{\max} \cos\left(x \frac{\Delta k}{2}\right)$$

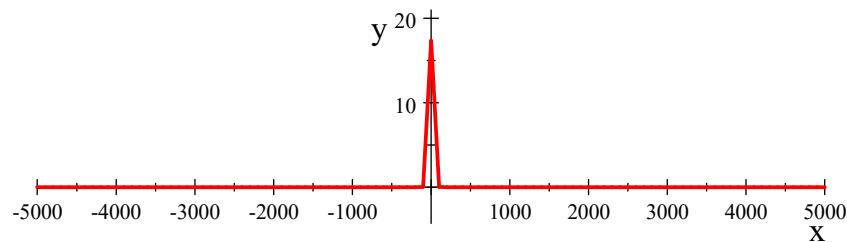
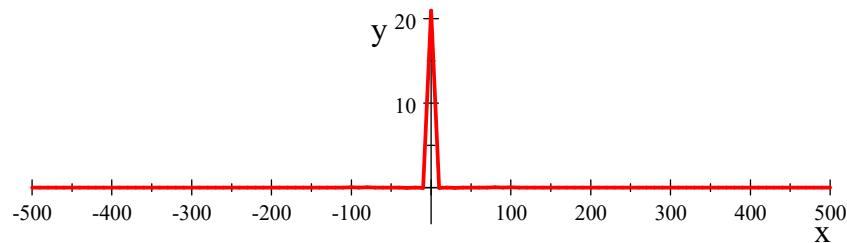
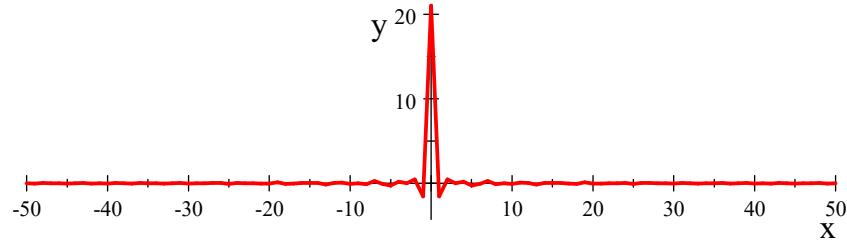
oscillates from large to small for all x , but what if we modified our amplitude function so that it didn't oscillate forever. Suppose we made it be mostly all in one place. Then we would have a limited wave. Maybe we could try adding a function like $1/x$ that we know falls off to zero quickly. Adding this gives an amplitude function of

$$A = \frac{1}{x} 2y_{\max} \cos\left(x \frac{\Delta k}{2}\right)$$

where we added in a $1/x$, to our original amplitude. Let's plot this function



And it does look like our mental picture of a photon a bit. We should check to make sure it doesn't repeat somewhere farther away



This is looking very promising. But this change to our amplitude is not obvious. There must be some implications to such a change.

Let's think, we started with adding waves to make our combined wave look more like a photon wave packet. We did this because if we want Δx to be small, $\Delta\lambda$ must be larger so we added in more waves with different λ values. Mathematically this is

$$y_r(x) = \sum A_i \sin(k_i x)$$

which means for every x value along the wave, the height of the combined wave $y_r(x)$ is made by adding up the y parts of each partial wave at the position x . But up till now we used only a few different wavelengths. What if we used very many wavelengths.

Then we might write this last equation as

$$y_r(x) = \int A(k) \sin(kx) dk$$

Let's make the simple assumption that all the amplitude functions for the partial waves are the same, $A(k) = A_o$, and let's integrate from $k_o - \frac{\Delta k}{2}$ to $k_o + \frac{\Delta k}{2}$. In doing

this we are taking an infinite number of wavelengths with k values that are spread across Δk centered on k_o . An infinite number of k values means an infinite number of wavelengths, so since

$$k = \frac{2\pi}{\lambda}$$

then

$$|\Delta k| = \left| = \frac{2\pi}{\lambda^2} \Delta \lambda \right|$$

so we have a $\Delta \lambda$ that is like

$$\begin{aligned} |\Delta \lambda| &= \left| \frac{\lambda^2}{2\pi} \Delta k \right| \\ &= \left| \frac{2\pi}{2\pi} \frac{\lambda^2}{2\pi} \Delta k \right| \\ &= \left| \frac{2\pi}{k^2} \Delta k \right| \end{aligned}$$

So as our Δk gets larger, so does our $\Delta \lambda$. That should make Δt smaller and we should have a limited wave. Let's try the integral.

$$y_r(x) = \int_{k_o - \frac{\Delta k}{2}}^{k_o + \frac{\Delta k}{2}} A_o \sin(kx) dk$$

We can look up the integral if we don't remember it

$$\int \sin(ax) dx = -\frac{1}{a} \cos ax$$

then we can write our integral as

$$\begin{aligned} y_r(x) &= \int_{k_o - \frac{\Delta k}{2}}^{k_o + \frac{\Delta k}{2}} A_o \sin(kx) dk = \left(-\frac{A_o}{x} \cos kx \right|_{k_o - \frac{\Delta k}{2}}^{k_o + \frac{\Delta k}{2}} \\ &= -\frac{A_o}{x} \cos \left(\left(k_o + \frac{\Delta k}{2} \right) x \right) - \left(-\frac{A_o}{x} \cos \left(\left(k_o - \frac{\Delta k}{2} \right) x \right) \right) \\ &= -\frac{A_o}{x} \left(\cos \left(\left(k_o + \frac{\Delta k}{2} \right) x \right) - \left(\cos \left(\left(k_o - \frac{\Delta k}{2} \right) x \right) \right) \right) \\ &= \frac{A_o}{x} \left(\cos \left(\left(k_o - \frac{\Delta k}{2} \right) x \right) - \cos \left(\left(k_o + \frac{\Delta k}{2} \right) x \right) \right) \end{aligned}$$

and we need another trig identity. I looked it up in the *CRC Standard Math Tables*.

$$\cos a - \cos b = -2 \sin \left(\frac{a+b}{2} \right) \sin \left(\frac{a-b}{2} \right)$$

so

$$y_r(x) = -\frac{2A_o}{x} \sin \left(\frac{(k_o - \frac{\Delta k}{2})x + (k_o + \frac{\Delta k}{2})x}{2} \right) \sin \left(\frac{(k_o - \frac{\Delta k}{2})x - (k_o + \frac{\Delta k}{2})x}{2} \right)$$

and we can write this as

$$y_r(x) = -\frac{2A_o}{x} \sin \left(\frac{(2k_o)x}{2} \right) \sin \left(\frac{(-2\frac{\Delta k}{2})x}{2} \right)$$

or even as

$$y_r(x) = -\frac{2A_o}{x} \sin\left(-\frac{\Delta k}{2}x\right) \sin(k_o x)$$

And knowing

$$\sin(-x) = -\sin x$$

then

$$y_r(x) = \frac{2A_o}{x} \sin\left(\frac{\Delta k}{2}x\right) \sin(k_o x)$$

and we have an amplitude with a $1/x$ in it just like we wanted. And now we can see that this $1/x$ factor did include many more λ values distributed over a $\Delta\lambda$ which makes our Δt limited. We have limited our wave and achieved the our photon like shape. This is a good candidate for a wave packet.

So what did it cost us to get an amplitude function that included a factor of $1/x$? It cost us having an infinite number of spatial frequencies (wave numbers) within a band width Δk . We needed to take every possible wave with a k value between $k_o - \frac{\Delta k}{2}$ to $k_o + \frac{\Delta k}{2}$ and add all these waves together to make our wave packet. This makes the math more challenging, but it represents our new view of a photon fairly well.

But it might not be true that all the amplitudes are the same for the waves we mix. That was a simple first guess. We should allow for $A(k)$ to change. Let's pick,

$$A(k) = A_o e^{-\frac{(k-k_o)^2}{2(\Delta k)^2}}$$

which would give

$$\begin{aligned} y_r &= \int A(k) \cos(kx) dk \\ &= \int A_o e^{-\frac{(k-k_o)^2}{2(\Delta k)^2}} \cos(kx) dk \end{aligned}$$

where for this wave I used a cosine instead of a sine, but $\cos(kx) = \sin(\frac{\pi}{2} - kx)$ is still a wave just with a different ϕ .

This integral is harder to find in a table (but I put it in our table in the appendix).

$$\int_0^\infty e^{-a^2 x^2} \cos(bx) dx = \frac{\sqrt{\pi}}{2|a|} e^{-\frac{b^2}{4a^2}}$$

The result. is

$$y_r = A_o \Delta k \sqrt{2\pi} e^{-\frac{(\Delta k x)^2}{2}} \cos(k_o x)$$

and we see that we do pick up a range of k values (and therefore have a range of λ values) so we again have a limited wave. There are more mathematical forms that can work. The exact form of the wave packet will depend on the place where we put the wave. More on this in the next few lectures.

Motion of the wave packet

So far we have done all our work on forming wave packets at $t = 0$, one snapshot. Let's let these waves move now. When we restore the time dependence we get a traveling wave who's amplitude varies. We can find the amplitude function algebraically. Let's do this for two waves mixing and then try to mentally extrapolate to an infinite number of waves mixed with wave numbers from a band Δk .

We can write out the entire resultant wave in our usual way. Our two waves are

$$\begin{aligned}y_1 &= y_{\max} \sin(k_1 x - \omega_1 t + \phi_o) \\y_2 &= y_{\max} \sin(k_2 x - \omega_2 t + \phi_o)\end{aligned}$$

and the resultant

$$\begin{aligned}y_r &= 2y_{\max} \cos\left(\frac{(k_2 x - \omega_2 t + \phi_o) - (k_1 x - \omega_1 t + \phi_o)}{2}\right) \sin\left(\frac{k_1 x - \omega_2 t + \phi_o + k_2 x - \omega_1 t + \phi_o}{2}\right) \\&= 2y_{\max} \cos\left(\frac{(k_2 - k_1)x}{2} - \frac{(\omega_2 - \omega_1)t}{2}\right) \sin\left(\frac{(k_2 + k_1)x}{2} - \frac{(\omega_2 + \omega_1)t}{2} + \phi_o\right) \\&= 2y_{\max} \cos\left(\frac{\Delta k x}{2} - \frac{\Delta \omega t}{2}\right) \sin(\bar{k}x - \bar{\omega}t + \phi_o)\end{aligned}$$

The first term

$$A = 2y_{\max} \cos\left(\frac{\Delta k x}{2} - \frac{\Delta \omega t}{2}\right)$$

is the amplitude, and the second term is a wave with frequency $\bar{\omega}$ and wave number \bar{k} .

From PH123 we remember that

$$\begin{aligned}k_1 &= \frac{2\pi}{\lambda_1} \\ \omega_1 &= 2\pi f_1\end{aligned}$$

so

$$\frac{\omega_1}{k_1} = \frac{2\pi f_1}{\frac{2\pi}{\lambda_1}} = \lambda_1 f_1 = v_1$$

the wave speed for wave 1 (we could do the same to find the speed of wave 2). Let's give this a special name. Because we took just k_1 and ω_1 , this is the speed of just one of our component waves, not a mix of waves. We call this the *phase speed*. I would prefer to call this a *component speed* because for us there will be more than one component wave making up our mixed wave. But tradition has it that we call this the phase speed.

In our case there is something similar to a phase speed for the combined wave. The combined wave has a wave part with $\omega = \frac{(\omega_2 + \omega_1)}{2}$ and with $k = \frac{(k_2 + k_1)}{2}$. This gives

the speed of the combined wave to be

$$\begin{aligned} v_{\text{wave part}} &= \frac{\frac{(\omega_2 + \omega_1)}{2}}{\frac{(k_2 + k_1)}{2}} = \frac{(\omega_2 + \omega_1)}{(k_2 + k_1)} = \frac{(2\pi f_2 + 2\pi f_1)}{\frac{2\pi}{\lambda_2} + \frac{2\pi}{\lambda_1}} = \frac{2\pi}{2\pi} \frac{f_2 + f_1}{\frac{1}{\lambda_2} + \frac{1}{\lambda_1}} \\ &= \frac{f_2 + f_1}{\frac{\lambda_1}{\lambda_1 \lambda_2} + \frac{\lambda_2}{\lambda_1 \lambda_2}} = \frac{f_2 + f_1}{\frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2}} = \frac{(f_2 + f_1) \lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \end{aligned}$$

and notice this is different than the phase speed of each of the components

$$\begin{aligned} v_1 &= \lambda_1 f_1 \\ v_2 &= \lambda_2 f_2 \end{aligned}$$

But now let's look at the amplitude term. It seems to have a speed as well! It has

$\omega = \Delta\omega$ and $k = \Delta k$ so

$$\begin{aligned} v_{\text{amplitude part}} &= \frac{\Delta\omega}{\Delta k} = \frac{(\omega_2 - \omega_1)}{(k_2 - k_1)} = \frac{(2\pi f_2 - 2\pi f_1)}{\frac{2\pi}{\lambda_2} - \frac{2\pi}{\lambda_1}} = \frac{2\pi}{2\pi} \frac{f_2 - f_1}{\frac{1}{\lambda_2} - \frac{1}{\lambda_1}} \\ &= \frac{f_2 - f_1}{\frac{\lambda_1}{\lambda_1 \lambda_2} - \frac{\lambda_2}{\lambda_1 \lambda_2}} = \frac{f_2 - f_1}{\frac{\lambda_1 - \lambda_2}{\lambda_1 \lambda_2}} = \frac{(f_2 - f_1) \lambda_1 \lambda_2}{\lambda_1 - \lambda_2} \end{aligned}$$

and this speed is different than the speed of the wave part! Let's give this speed a special name. This speed is the speed of the amplitude bumps. Because the amplitude bumps are a result of adding a group of waves, let's call it the *group speed*. For just two waves we can write this as

$$v_{\text{amplitude part}} = v_{\text{group}} = \frac{\Delta\omega}{\Delta k} = \frac{(\omega_2 - \omega_1)}{(k_2 - k_1)}$$

but if we have many many frequencies and wave numbers like in our wave packet case we can write this as

$$v_{\text{group}} = \frac{d\omega}{dk}$$

Let's apply all this to de Broglie waves. The energy in a matter wave looks like the energy for a photon

$$E = hf = \hbar\omega$$

and the momentum is

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

but for a wave

$$\lambda = \frac{2\pi}{k}$$

so we can write momentum for our wave/particle as

$$p = \frac{h}{\frac{2\pi}{k}} = \hbar k$$

to find the group speed, we take

$$v_{\text{group}} = \frac{d\omega}{dk}$$

we can find $d\omega$ and dk

$$\begin{aligned} dE &= \hbar d\omega \\ d\omega &= \frac{dE}{\hbar} \end{aligned}$$

and

$$\begin{aligned} dp &= \hbar dk \\ dk &= \frac{dp}{\hbar} \end{aligned}$$

so

$$v_{group} = \frac{\frac{dE}{\hbar}}{\frac{dp}{\hbar}} = \frac{dE}{dp}$$

For a relativistic particle, this can be more complicated, but let's do the non-relativistic case where

$$E \approx K = \frac{p^2}{2m}$$

then

$$dE = \frac{2pdः}{2m} = \frac{pdः}{m}$$

and

$$v_{group} = \frac{dE}{dp} = \frac{\frac{pdः}{m}}{dp} = \frac{p}{m} = v$$

just as we saw in classical physics (the particle speed is... the particle speed). This last problem gave us a hint on how to interpret these speeds but we should carefully look at what each means.

Let's plot our combined wave as an aid in this.

$$= 2y_{max} \cos \left(\frac{\Delta kx}{2} - \frac{\Delta \omega t}{2} \right) \sin (\bar{k}x - \bar{\omega}t + \phi_o)$$

To plot we need actual numbers so let's say that

$$f_1 = 100 \text{ Hz}$$

and

$$\lambda_1 = 2 \text{ m}$$

. Then

$$v_1 = (2 \text{ m}) (100 \text{ Hz}) = 200.0 \frac{\text{m}}{\text{s}}$$

and let

$$f_2 = 110 \text{ Hz}$$

and

$$\lambda_2 = 2.5 \text{ m}$$

$$v_2 = (2.5 \text{ m}) (110 \text{ Hz}) = 275.0 \frac{\text{m}}{\text{s}}$$

these are the component or phase velocities. The combined wave would have a

wave-part speed of

$$\begin{aligned} v_{\text{wave part}} &= \frac{(f_2 + f_1) \lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \\ &= \frac{(110 \text{ Hz} + 100 \text{ Hz})(2 \text{ m})(2.5 \text{ m})}{(2.5 \text{ m}) + (2 \text{ m})} \\ &= 233.33 \frac{\text{m}}{\text{s}} \end{aligned}$$

which is somewhere in between the two phase velocities. Now for the group velocity

$$\begin{aligned} v_{\text{amplitude part}} &= \frac{(f_2 - f_1) \lambda_1 \lambda_2}{\lambda_1 - \lambda_2} \\ &= \frac{(110 \text{ Hz} - 100 \text{ Hz})(2 \text{ m})(2.5 \text{ m})}{(2 \text{ m}) - (2.5 \text{ m})} \\ &= -100.0 \frac{\text{m}}{\text{s}} \end{aligned}$$

Wow that was a surprise! Let's see what is happening. Let's plot our waves

$$y_1 = 10 \sin \left(\frac{2\pi}{\lambda_1} x - 2\pi f_1 t \right)$$

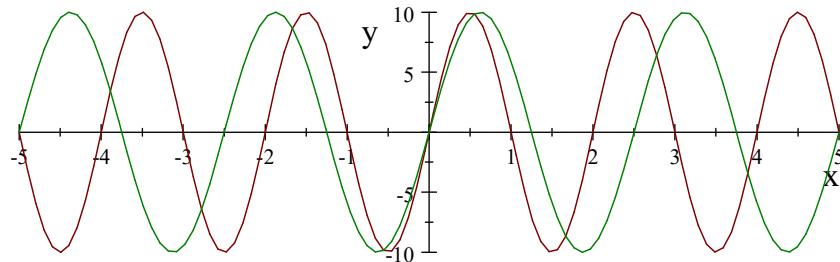
$$y_2 = 10 \sin \left(\frac{2\pi}{\lambda_2} x - 2\pi f_2 t \right)$$

where I let $\phi_o = 0$ and let's put in numbers

$$y_1 = 10 \sin \left(\frac{2\pi}{2 \text{ m}} x - 2\pi (100 \text{ Hz}) t \right)$$

$$y_2 = 10 \sin \left(\frac{2\pi}{2.5 \text{ m}} x - 2\pi (110 \text{ Hz}) t \right)$$

we can plot these at $t = 0$



and we can plot the combined wave, we need

$$\Delta k = k_2 - k_1 = \frac{2\pi}{2.5 \text{ m}} - \frac{2\pi}{2 \text{ m}} = -0.2 \frac{\pi}{\text{m}} = -0.2\pi$$

$$\Delta\omega = \omega_2 - \omega_1 = 2\pi (110 \text{ Hz}) - 2\pi (100 \text{ Hz}) = 20\pi \text{ Hz}$$

and

$$\bar{k} = \frac{\frac{2\pi}{2.5 \text{ m}} + \frac{2\pi}{2 \text{ m}}}{2} = 0.9 \frac{\pi}{\text{m}}$$

$$\bar{\omega} = \frac{2\pi(110 \text{ Hz}) + 2\pi(100 \text{ Hz})}{2} = 210\pi \text{ Hz}$$

so

$$y_r = 2y_{\max} \cos\left(\frac{\Delta kx}{2} - \frac{\Delta \omega t}{2}\right) \sin(\bar{k}x - \bar{\omega}t + \phi_o)$$

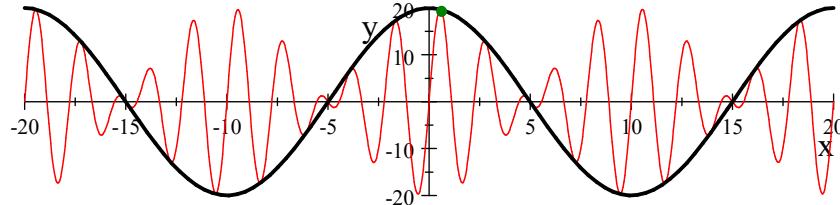
becomes

$$y_r = 20 \cos\left(\frac{-0.2\pi}{2}x \frac{1}{\text{m}} - \frac{20\pi \text{ Hz}t}{2}\right) \sin\left(0.9 \frac{\pi}{\text{m}}x - 210\pi \text{ Hz}t + 0\right)$$

so at $t = 0$

$$y_r = 20 \cos\left(\frac{-0.2\pi}{2}x \frac{1}{\text{m}} - \frac{20\pi \text{ Hz}(0)}{2}\right) \sin\left(0.9 \frac{\pi}{\text{m}}x - 210\pi \text{ Hz}(0) + 0\right)$$

We get the following plot where the combined wave is in red (light line) and the amplitude envelope function is in black (heavy line).



Now let's look at these waves a time $t = 0.05 \text{ s}$ later.

$$y_r = 20 \cos\left(\frac{-0.2\pi}{2}x \frac{1}{\text{m}} - \frac{20\pi \text{ Hz}(0.05 \text{ s})}{2}\right) \sin\left(0.9 \frac{\pi}{\text{m}}x - 210\pi \text{ Hz}(0.05 \text{ s}) + 0\right)$$

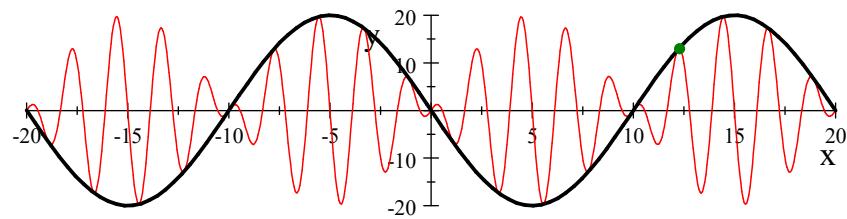
In this amount of time the wave part would move a distance

$$\begin{aligned} \Delta x_{\text{wave}} &= v_{\text{wave part}} \Delta t \\ &= 233.33 \frac{\text{m}}{\text{s}} \times 0.05 \text{ s} \\ &= 11.667 \text{ m} \end{aligned}$$

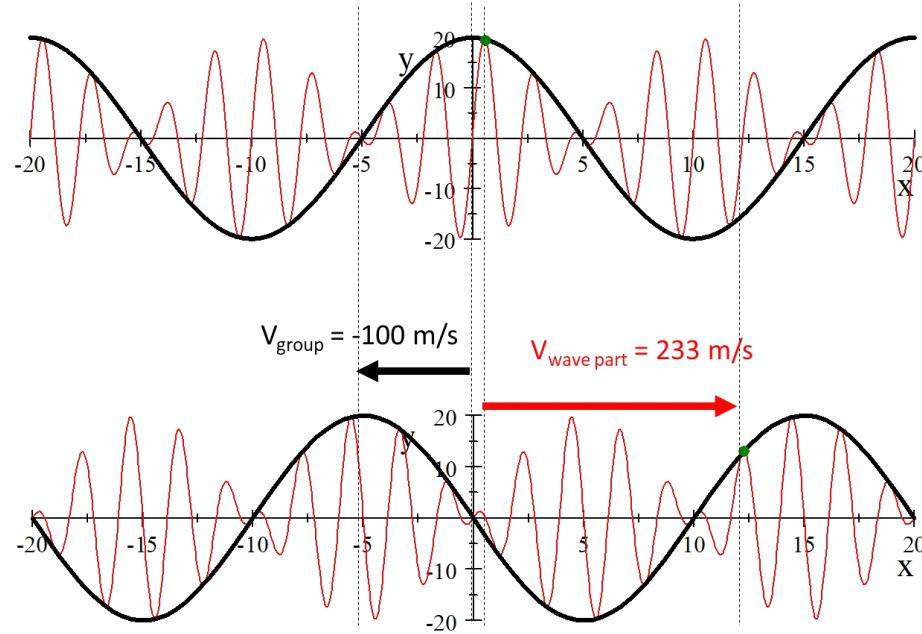
and the amplitude part would move

$$\begin{aligned} \Delta x_{\text{group}} &= v_{\text{wave part}} \Delta t \\ &= -100.0 \frac{\text{m}}{\text{s}} \times 0.05 \text{ s} \\ &= -5.0 \text{ m} \end{aligned}$$

here is what we see

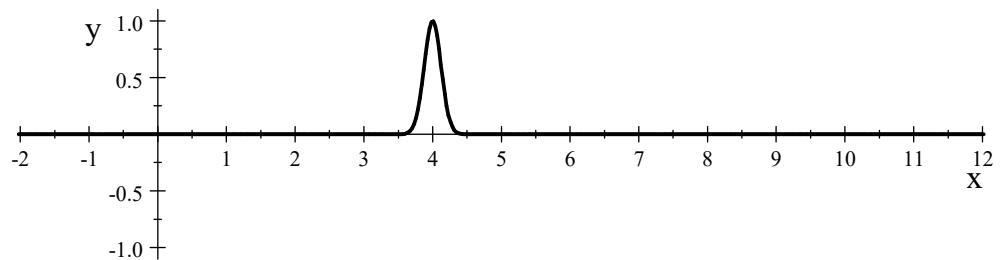


And this sounds crazy, unless we look a lot closer. In the last two figures a particular wave part peak is marked with a little green dot. In the first at $t = 0\text{ s}$ the dot is at $x = 0.62\text{ m}$. That same peak is found later at $t = 0.05\text{ s}$ at $x = 12.287\text{ m}$. It has moved to the right.



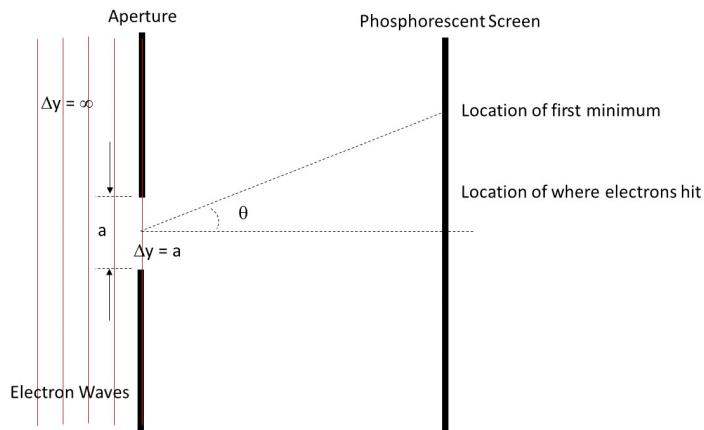
Meanwhile, the envelope function peak was at $x = 0$ at $t = 0$ and it has moved to the *left* to about $x = -0.5$ at $t = 0.05\text{ s}$. The important thing to realize is that the wave peak (red line) that is right under the envelope peak at $x = -5\text{ m}$ is *not the same peak* that was under the envelope curve at $x = 0$ at $t = 0$. The envelope curve is built from interference. The motion of the two original waves is just such that the spot of constructive interference is constructed from different peaks of the two constituent waves at different times. And that spot of constructive interference moves in a different direction than the waves, themselves for this particular case!

Of course this is a very simple situation with just two waves. And having the group velocity turn out to be in the opposite direction of the waves is not normally the case. But the point is, that when we integrate over a bandwidth Δk of infinite waves, we could find a group velocity that is very different than the phase velocities.



Spread of the wave function

Let's consider a single slit experiment again.



In such an experiment we restrict the wave packet so we get a definite Δy_0 with an initial momentum uncertainty Δp_{yo} . This will have an accompanying velocity uncertainty

$$\Delta v_{yo} = \frac{\Delta p_{yo}}{m}$$

So the best we can do to know the velocity is to say

$$v_y = v_{yo} \pm \Delta v_{yo}$$

and the position of the particle later will be

$$\begin{aligned}y &= v_y \Delta t \\&= (v_{yo} \pm \Delta v_{yo}) \Delta t \\&= v_{yo} \Delta t \pm \Delta v_{yo} \Delta t\end{aligned}$$

We have an uncertainty of the y position of the wave packet because of the wave nature of the wave packet. Of course we also have a basic uncertainty in the y position because the particular wave packet could have come from any part of the slit opening so it will have an uncertainty of Δy_o as well. We can estimate the total uncertainty like we did back in PH150

$$\Delta y = \sqrt{(\Delta y_o)^2 + (\Delta v_{yo} \Delta t)^2}$$

and this grows as Δt grows. The growing uncertainty means that the wave packet will spread out. This is just what we see in our double slit experiment. we could write this as

$$\Delta y = \sqrt{(\Delta y_o)^2 + \left(\frac{\Delta p_{yo}}{m} \Delta t \right)^2}$$

and using de Broglie again we can say

$$\Delta p_{yo} = \frac{\hbar}{\Delta y_o}$$

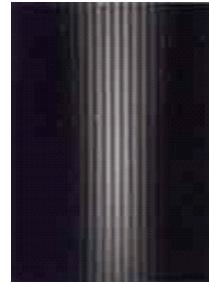
so

$$\Delta y = \sqrt{(\Delta y_o)^2 + \left(\frac{\hbar \Delta t}{m \Delta y_o} \right)^2}$$

and this is really an important discovery. it means that the uncertainty in the second term gets *bigger* when the slit gets *smaller*. And this is an important issue in experimentation. The more we try to confine the wave packet the larger the uncertainty in the position will be as time goes on.

What does the wave packet mean?

If we look at the electron double slit experiment again,



Claus Jonsson, "Electron Diffraction at Multiple Slits," AJP Volume 42, Jan 1974, pp 4-11 (Translated by Deitrich Brandt and Stanley Hirschi)

but this time lower the rate of electrons headed through the slits we find a strange thing. Electrons hitting the screen still make the diffraction pattern, but we can see individual electron hits. The pattern becomes spotty. If we further lower the electron rate. the pattern get's harder to distinguish. There are just more dots in the spots where there were bright fringes were. Fewer, and the dots seem random. But if we are patient, eventually the pattern will become apparent. It just takes more time.

Apparently the diffraction pattern is made of individual electron detections. There are just more detections where the bright fringes would be. The detections are particle nature, the pattern is wave nature.

This is really not so strange. Think of a light diffraction pattern. The intensity is proportional to the electric field amplitude squared. And if we did a light double slit experiment but lowered the intensity to the point where only a few photons at a time pass through the experiment, we would same the same speckle pattern. The detection of the photons is particle nature. The pattern is wave nature. So the probability of a photon begin detected in some part of the pattern is proportional to the amplitude squared.

The same is true for de Broglie wave packets. The wave packet amplitude tells us the probability of an electron detection for any given spot.

12 Wave Packets and the Schrödinger Equation

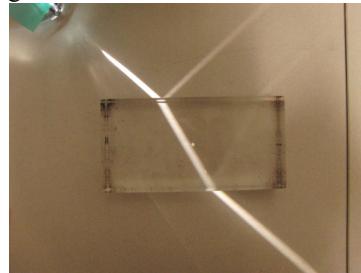
5.1-5.3

Fundamental Concepts in the Lecture

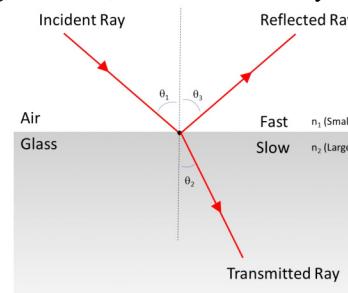
- Boundary conditions
- Schrödinger equation

Wave Packets and Boundaries

Think back to our study of optics in PH123. We learned that light can travel through a material like glass. You might have seen a demonstration of this that looked like this.



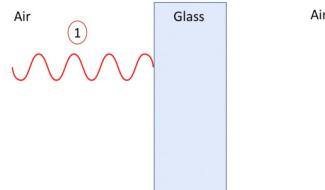
We leaned about incoming, reflected, and transmitted rays of light.



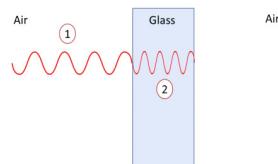
But now with our quantum viewpoint, we should re-examine this situation. Light is a

wave, but a limited wave (a photon) and electrons are waves. How do they interact with their environment?

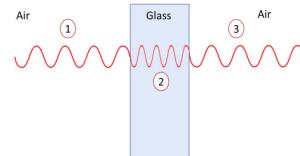
Let's have a light wave be incident on a piece of glass. We will call this incident light wave 1



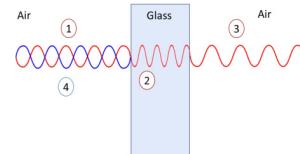
But we know that some of the light energy is transmitted into the glass. We can call this transmitted wave light wave 2



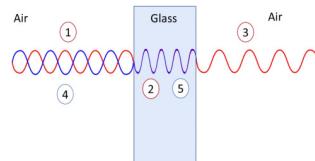
And from optics we know that the wavelength changes at the boundary. In this case, the wavelength is smaller inside the glass than outside the glass. We further know that some of the light will be transmitted across the second boundary. Let's call that wave 3.



But we know there are also reflected waves. At the first boundary we expect a reflection and we can even say that the wave will invert on reflection from the first boundary.

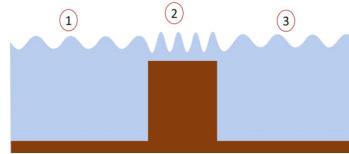


And there will also be a reflection off of the second boundary.

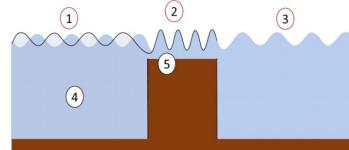


We have what we can identify as five different waves in this situation. And this is

common for all wave phenomena. We could envision a similar situation with a water wave. We would have three transmitted waves

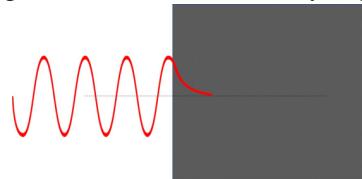


and two reflected waves



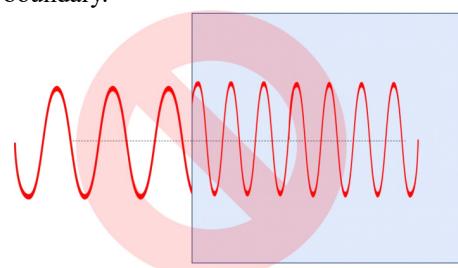
Wherever we have waves and those waves meet boundaries, we expect transmission and reflection.

There is something that happens with light reflection that we didn't cover in PH123. Suppose we have a material that is not transparent. Then the light wave won't transmit—at least not very far. But the light energy does enter the material. The amplitude of the wave decays exponentially until it is zero. The distance from the boundary to where the amplitude is zero is called the penetration depth. But the important idea is that the light wave doesn't immediately stop at the boundary.

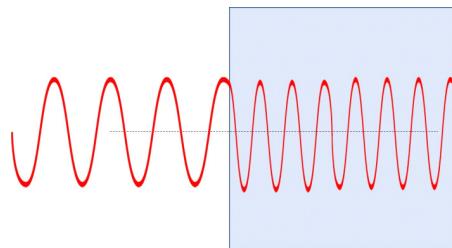


We will find this to be true for matter waves as well.

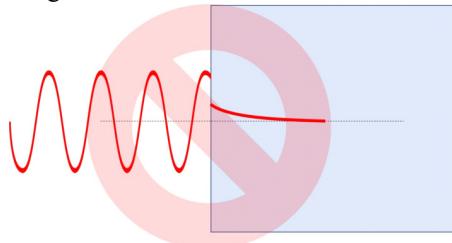
We need to pay attention to our wave at the boundaries. The wave must not experience a discontinuity at the boundary.



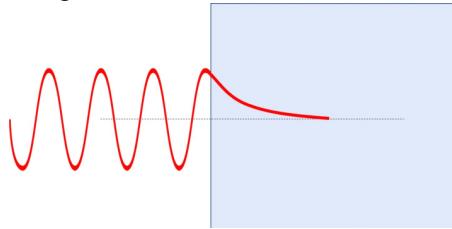
The wave can't have a gap. It must be continuous



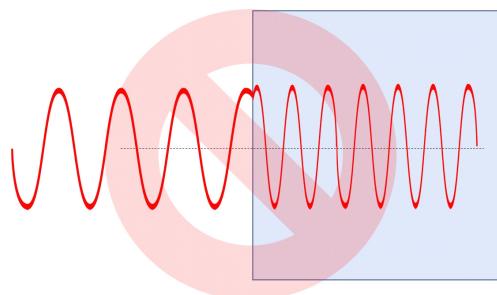
So the waves in the next figure would not work



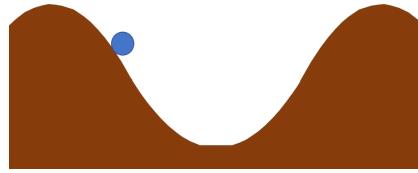
but the waves in this next figure would work



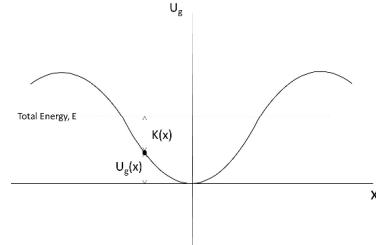
The function must not only be continuous, but must not have any strange discontinuities. It must be differentiable. Here is another example of something that would not work.



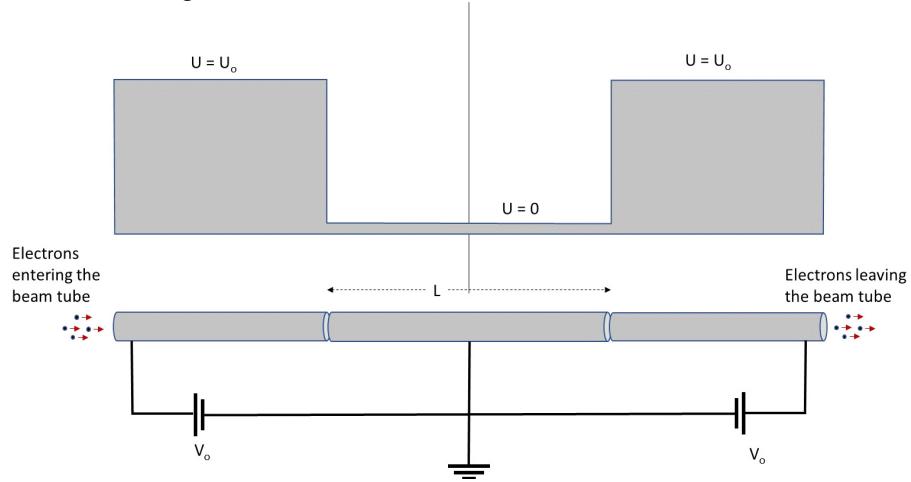
Let's also remember energy diagrams from PH121. If we have a hill



we could describe this situation with an energy graph like this



Now we are ready to take on matter waves and boundaries. Let's consider an accelerator that is accelerating electrons.



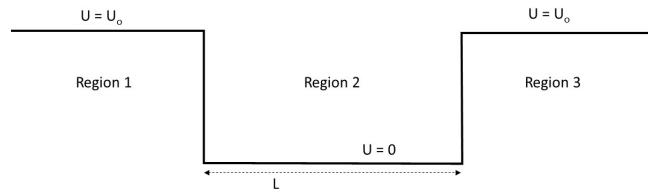
The apparatus is on the bottom of the figure and the energy diagram is on the top. It really is built of a series of pipes, surrounded by electrical systems to adjust the potential. Here is an example from CERN.



LINAC 2 linear accelerator at CERN. This device has multiple drift tubes where applied electric fields create zones of acceleration and deceleration.

The electrons enter the first section of the beam pipe and experience a potential energy $U_o = q\Delta V_o$. Then they reach the middle section where the beam pipe is grounded so $\Delta V = 0$ so $U = 0$. Then in the third segment we again have $U_o = q\Delta V_o$. We would expect some of the electron wave packets to be transmitted across each boundary, and some to be reflected from each of the boundaries.

At the boundaries, the wave packet functions must be continuous. Let's take a specific case



In region 1 where the potential is U_o we might have a wave given by

$$y_1(x) = C_1 \sin\left(\frac{2\pi}{\lambda_1}x - \phi_1\right)$$

where

$$C_1 = 11.5$$

$$\lambda_1 = 4.97 \text{ cm}$$

$$\phi_1 = -65.3^\circ = -1.1397 \text{ rad}$$

But we use y to indicate position along one coordinate axis. Now our potential energy isn't a function of height. So let's change to a new letter, ψ . This is the Greek letter psi, or "pitchfork."

$$\psi_1(x) = C_1 \sin\left(\frac{2\pi}{\lambda_1}x - \phi_1\right)$$

Suppose we measure the wavelength in Region 2 and find

$$\lambda_2 = 10.5 \text{ cm}$$

And suppose that the boundaries between state 1 and 2 of the accelerator is at

$$x_{12} = 0$$

and the boundary between state 2 and state 3 is at

$$x_{23} = L = 20 \text{ cm}$$

We expect that our wave equation will change at the boundaries, and we know we need to match up the different equations for the different regions so they are continuous at the boundaries. For this problem, let's assume that for some reason there are no reflections at the boundaries (not realistic, but it makes the math easier for a first problem in using boundary values to match wave functions). In practical terms, we expect that in Region 2 we will have a similar wave equation

$$\psi_2(x) = C_2 \sin\left(\frac{2\pi}{\lambda_2}x - \phi_2\right)$$

but we will need to adjust C_2 , λ_2 , and ϕ_2 so that $y_2(0) = y_1(0)$ and we will do this again at the second boundary between Regions 2 and 3

$$\psi_3(x) = C_3 \sin\left(\frac{2\pi}{\lambda_3}x - \phi_3\right)$$

and we will need to adjust C_3 , λ_3 , and ϕ_3 so that $\psi_2(L) = \psi_3(L)$. Let's take on the boundary at $x = 0$ first. Set

$$\begin{aligned}\psi_1(0) &= \psi_2(0) \\ C_1 \sin\left(\frac{2\pi}{\lambda_1}(0) - \phi_1\right) &= C_2 \sin\left(\frac{2\pi}{\lambda_2}(0) - \phi_2\right) \\ C_1 \sin(-\phi_1) &= C_2 \sin(-\phi_2) \\ -C_1 \sin(\phi_1) &= -C_2 \sin(\phi_2)\end{aligned}$$

This gives us one equation, but two unknowns, C_2 and ϕ_2 . We can get another equation by demanding that the union be differentiable at the boundary. That is the same as saying that the slope of each wave equation must be the same at the boundary. So we can set the derivatives $d\psi_1/dx = d\psi_2/dx$ at $x = 0$

$$\begin{aligned}\frac{d\psi_1}{dx} &= \frac{d}{dx}\left(C_1 \sin\left(\frac{2\pi}{\lambda_1}x - \phi_1\right)\right) \\ &= C_1 \cos\left(\frac{2\pi}{\lambda_1}x - \phi_1\right) \left(\frac{2\pi}{\lambda_1}\right) \\ &= C_1 \left(\frac{2\pi}{\lambda_1}\right) \cos\left(\frac{2\pi}{\lambda_1}x - \phi_1\right)\end{aligned}$$

and likewise

$$\frac{d\psi_2}{dx} = C_2 \frac{2\pi}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}x - \phi_2\right)$$

setting these equal at $x = 0$ gives

$$\begin{aligned}\left(\frac{d\psi_1}{dx}\right|_{x=0} &= \left(\frac{d\psi_2}{dx}\right|_{x=0} \\ C_1 \left(\frac{2\pi}{\lambda_1}\right) \cos\left(\frac{2\pi}{\lambda_1}(0) - \phi_1\right) &= C_2 \frac{2\pi}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}(0) - \phi_2\right) \\ C_1 \left(\frac{2\pi}{\lambda_1}\right) \cos(-\phi_1) &= C_2 \frac{2\pi}{\lambda_2} \cos(-\phi_2)\end{aligned}$$

We have another equation with the same two unknowns. It is down to algebra. We have two equations

$$\begin{aligned}C_1 \sin(\phi_1) &= C_2 \sin(\phi_2) \\ C_1 \left(\frac{2\pi}{\lambda_1}\right) \cos(-\phi_1) &= C_2 \frac{2\pi}{\lambda_2} \cos(-\phi_2)\end{aligned}$$

and two unknowns ϕ_2 and C_2 . Let's take the two equations and divide them

$$\begin{aligned}\frac{C_1 \sin(\phi_1)}{C_1 \left(\frac{2\pi}{\lambda_1}\right) \cos(-\phi_1)} &= \frac{C_2 \sin(\phi_2)}{C_2 \frac{2\pi}{\lambda_2} \cos(-\phi_2)} \\ \lambda_1 \frac{\sin(\phi_1)}{\cos(-\phi_1)} &= \lambda_2 \frac{\sin(\phi_2)}{\cos(-\phi_2)} \\ \lambda_1 \tan(\phi_1) &= \lambda_2 \tan(\phi_2) \\ \tan(\phi_2) &= \frac{\lambda_1}{\lambda_2} \tan(\phi_1)\end{aligned}$$

so

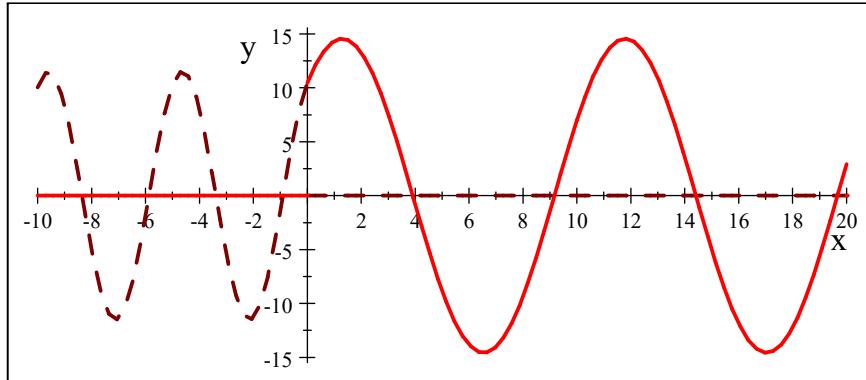
$$\begin{aligned}\phi_2 &= \tan^{-1}\left(\frac{\lambda_1}{\lambda_2} \tan(\phi_1)\right) \\ \phi_2 &= \tan^{-1}\left(\frac{4.97 \text{ cm}}{10.5 \text{ cm}} \tan(-65.3^\circ)\right) \\ &= -0.79974 \text{ rad} \\ &= -45.822^\circ\end{aligned}$$

Now we can solve for C_2 . Using the first of our two equations.

$$\begin{aligned}C_2 &= C_1 \frac{\sin(\phi_1)}{\sin(\phi_2)} \\ C_2 &= (11.5) \frac{\sin(-1.1397 \text{ rad})}{\sin(-0.79974 \text{ rad})} \\ &= 14.568\end{aligned}$$

and we can plot this to see if it worked

$$\begin{aligned}\psi_1(x) &= 11.5 \sin\left(\frac{2\pi \text{ rad}}{4.97 \text{ cm}} x + 1.1397 \text{ rad}\right) \\ \psi_2(x) &= 14.568 \sin\left(\frac{2\pi \text{ rad}}{10.5 \text{ cm}} x + 0.79974 \text{ rad}\right)\end{aligned}$$



The Region one wave is in dark red (and dashed), the Region 2 wave is in light red, and we see right at $x = 0$ the wave is continuous and differentiable. We seem to have done well. Let's take on the next boundary at $x = L$. We will do the same thing. We know C_2 and ϕ_2 , we need to know C_3 and ϕ_3 . But we also need to know λ_3 . We can get λ_3 by symmetry. Since in Region 3 the potential energy is U_o just like it was in Region 1, we can say that $\lambda_3 = \lambda_1$. Think of the light exiting the glass in our previous example. the wavelength in the air region is the same on both sides of the glass region.

Then Set

$$\begin{aligned}\psi_2(L) &= \psi_3(L) \\ C_2 \sin\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right) &= C_3 \sin\left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right)\end{aligned}$$

This gives us one equation, but two unknowns, C_3 and ϕ_3 . We again get another equation by demanding that the union be differentiable at the boundary.

$$\left(\frac{d\psi_2}{dx}\Big|_{x=L}\right) = \left(\frac{d\psi_3}{dx}\Big|_{x=L}\right)$$

We already know $d\psi_2/dx$

$$\frac{d\psi_2}{dx} = C_2 \frac{2\pi}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}x - \phi_2\right)$$

but we need to find $d\psi_3/dx$

$$\begin{aligned}\frac{d\psi_3}{dx} &= \frac{d}{dx} \left(C_3 \sin\left(\frac{2\pi}{\lambda_3}x - \phi_3\right) \right) \\ &= C_3 \cos\left(\frac{2\pi}{\lambda_3}x - \phi_3\right) \left(\frac{2\pi}{\lambda_3}\right) \\ &= C_3 \left(\frac{2\pi}{\lambda_3}\right) \cos\left(\frac{2\pi}{\lambda_3}x - \phi_3\right)\end{aligned}$$

setting these equal at $x = L$ gives

$$\begin{aligned} C_2 \frac{2\pi}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}L - \phi_2\right) &= C_3 \left(\frac{2\pi}{\lambda_3}\right) \cos\left(\frac{2\pi}{\lambda_3}L - \phi_3\right) \\ \frac{C_2}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}L - \phi_2\right) &= \frac{C_3}{\lambda_3} \cos\left(\frac{2\pi}{\lambda_3}L - \phi_3\right) \end{aligned}$$

and our two equations are

$$\begin{aligned} C_2 \sin\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right) &= C_3 \sin\left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right) \\ \frac{C_2}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}L - \phi_2\right) &= \frac{C_3}{\lambda_3} \cos\left(\frac{2\pi}{\lambda_3}L - \phi_3\right) \end{aligned}$$

and two unknowns ϕ_3 and C_3 .

Let's do just what we did before and take the two equations and divide them

$$\frac{C_2 \sin\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right)}{\frac{C_2}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}L - \phi_2\right)} = \frac{C_3 \sin\left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right)}{\frac{C_3}{\lambda_3} \cos\left(\frac{2\pi}{\lambda_3}L - \phi_3\right)}$$

so, canceling the C_2 and C_3 gives

$$\lambda_2 \frac{\sin\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right)}{\cos\left(\frac{2\pi}{\lambda_2}L - \phi_2\right)} = \lambda_3 \frac{\sin\left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right)}{\cos\left(\frac{2\pi}{\lambda_3}L - \phi_3\right)}$$

And using some trig gives

$$\lambda_2 \tan\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right) = \lambda_3 \tan\left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right)$$

Now let's try to isolate ϕ_3

$$\begin{aligned} \frac{\lambda_2}{\lambda_3} \tan\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right) &= \tan\left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right) \\ \left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right) &= \tan^{-1}\left(\frac{\lambda_2}{\lambda_3} \tan\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right)\right) \\ \phi_3 &= \frac{2\pi}{\lambda_3}(L) - \tan^{-1}\left(\frac{\lambda_2}{\lambda_3} \tan\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right)\right) \end{aligned}$$

Then

$$\begin{aligned} \phi_3 &= \frac{2\pi}{4.97 \text{ cm}} (20 \text{ cm}) - \tan^{-1}\left(\frac{10.5 \text{ cm}}{4.97 \text{ cm}} \tan\left(\frac{2\pi}{10.5 \text{ cm}} (20 \text{ cm}) + 0.79974 \text{ rad}\right)\right) \\ &= 24.877 \text{ rad} \end{aligned}$$

This seems unusually high, but we can check it by solving for C_3 first and seeing if we get the same number. Start again with our two equations

$$\begin{aligned} C_2 \sin\left(\frac{2\pi}{\lambda_2}(L) - \phi_2\right) &= C_3 \sin\left(\frac{2\pi}{\lambda_3}(L) - \phi_3\right) \\ \lambda_3 \frac{C_2}{\lambda_2} \cos\left(\frac{2\pi}{\lambda_2}L - \phi_2\right) &= C_3 \cos\left(\frac{2\pi}{\lambda_3}L - \phi_3\right) \end{aligned}$$

This time square both and add them

$$\begin{aligned} C_2^2 \sin^2 \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) + \left(\lambda_3 \frac{C_2}{\lambda_2} \right)^2 \cos^2 \left(\frac{2\pi}{\lambda_2} L - \phi_2 \right) &= C_3^2 \sin^2 \left(\frac{2\pi}{\lambda_3} (L) - \phi_3 \right) \\ &\quad + C_3^2 \cos^2 \left(\frac{2\pi}{\lambda_3} L - \phi_3 \right) \end{aligned}$$

We can take out a C_3^2 on the right hand side

$$\begin{aligned} &C_2^2 \sin^2 \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) + \left(\lambda_3 \frac{C_2}{\lambda_2} \right)^2 \cos^2 \left(\frac{2\pi}{\lambda_2} L - \phi_2 \right) \\ &= C_3^2 \left(\sin^2 \left(\frac{2\pi}{\lambda_3} (L) - \phi_3 \right) + \cos^2 \left(\frac{2\pi}{\lambda_3} L - \phi_3 \right) \right) \end{aligned}$$

and isolate it

$$\begin{aligned} C_2^2 \sin^2 \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) + \left(\lambda_3 \frac{C_2}{\lambda_2} \right)^2 \cos^2 \left(\frac{2\pi}{\lambda_2} L - \phi_2 \right) &= C_3^2 \\ C_3^2 &= \sqrt{C_2^2 \sin^2 \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) + \left(\lambda_3 \frac{C_2}{\lambda_2} \right)^2 \cos^2 \left(\frac{2\pi}{\lambda_2} L - \phi_2 \right)} \end{aligned}$$

putting in numbers gives

$$\begin{aligned} C_3 &= \left((14.568)^2 \sin^2 \left(\frac{2\pi}{10.5 \text{ cm}} (20 \text{ cm}) - (-0.79974 \text{ rad}) \right) \right. \\ &\quad \left. + \left((4.97 \text{ cm}) \frac{14.568}{10.5 \text{ cm}} \right)^2 \cos^2 \left(\frac{2\pi}{10.5 \text{ cm}} (20 \text{ cm}) - (-0.79974 \text{ rad}) \right) \right)^{\frac{1}{2}} \\ &= 7.3576 \end{aligned}$$

Now lets put this back into the first of our two equations and re-solve for ϕ_3

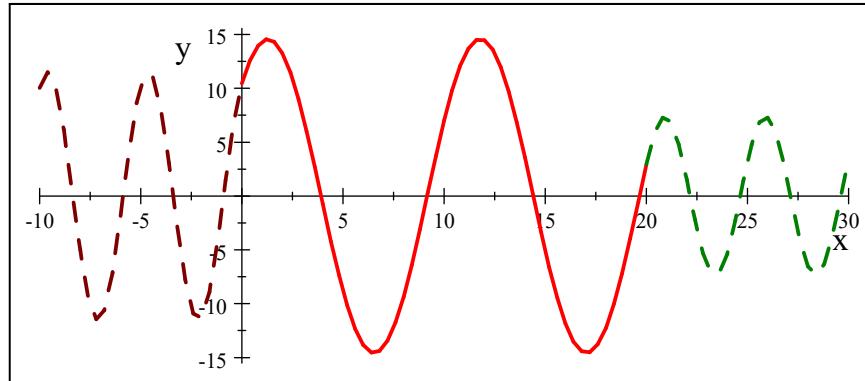
$$\begin{aligned} C_2 \sin \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) &= C_3 \sin \left(\frac{2\pi}{\lambda_3} (L) - \phi_3 \right) \\ \frac{C_2}{C_3} \sin \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) &= \sin \left(\frac{2\pi}{\lambda_3} (L) - \phi_3 \right) \\ \sin^{-1} \left(\frac{C_2}{C_3} \sin \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) \right) &= \left(\frac{2\pi}{\lambda_3} (L) - \phi_3 \right) \\ \phi_3 &= \left(\frac{2\pi}{\lambda_3} (L) - \sin^{-1} \left(\frac{C_2}{C_3} \sin \left(\frac{2\pi}{\lambda_2} (L) - \phi_2 \right) \right) \right) \\ \phi_3 &= \left(\frac{2\pi}{4.97 \text{ cm}} (20 \text{ cm}) - \sin^{-1} \left(\frac{14.568}{7.3576} \sin \left(\frac{2\pi}{10.5 \text{ cm}} (20 \text{ cm}) + 0.79974 \text{ rad} \right) \right) \right) \\ &= 24.877 \text{ rad} \end{aligned}$$

We got the same number. To further check the value, we can plot the wave equations and see if they are continuous and differentiable at L .

$$\psi_1(x) = 11.5 \sin \left(\frac{2\pi \text{ rad}}{4.97 \text{ cm}} x + 1.1397 \text{ rad} \right)$$

$$\psi_2(x) = 14.568 \sin\left(\frac{2\pi \text{ rad}}{10.5 \text{ cm}}x + 0.79974 \text{ rad}\right)$$

$$\psi_3(x) = 7.3576 \sin\left(\frac{2\pi \text{ rad}}{4.97 \text{ cm}}x - 24.877 \text{ rad}\right)$$

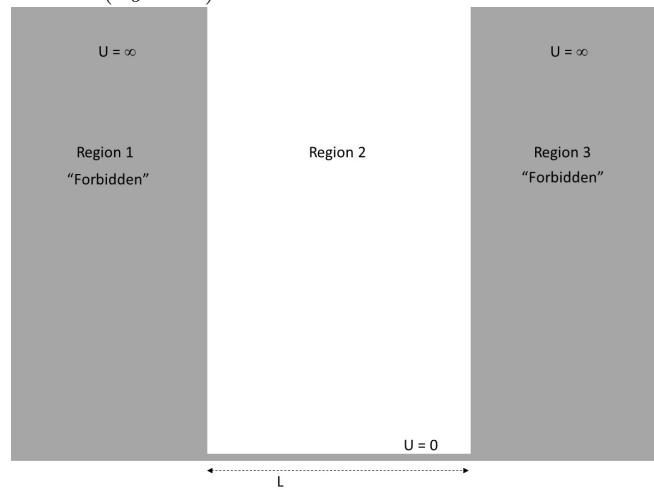


This is just what we expected to see, so we must have good values for ϕ_3 and C_3 .

It's not always true that the wave equation will be the same functional form in all regions, as we will soon see. It is also true that in this example we have ignored any reflected waves. But that is a subject for a later problem.

Particle in a box

Let's repeat our boundary problem, but this time let's make the potential in Region 1 and Region 3 infinite ($U_o = \infty$)



A particle in region 2 could not escape. In our last example there were waves in Region

1 and Region 3, but with infinite potential, a particle/wave can't penetrate into these regions. Think of our poor electrons. If they are in Region 2 of the beam pipe, it would take infinite energy to get them into Region 3 or Region 1.

To deal with this kind of situation, we know that at $x = 0$ the wave

$$\psi_2(0) = 0$$

because the wave

$$\psi_1 = 0$$

in all of Region 1. There is no wave in region 1, and at the boundary the wave functions have to match. It is also true that

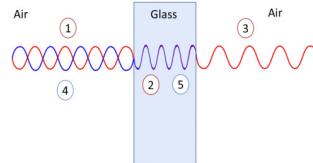
$$\psi_2(L) = 0$$

because

$$\psi_3 = 0$$

in all of Region 3.

With these boundary conditions we should be able to find a wave equation for $\psi_2(x)$ in Region 2. But if we refer back to our optics example, we expect that we could have electrons going to the right, but we could reflect electrons off of the barrier at $x = L$ so we could have electrons going to the left. These correspond to waves 2 and waves 5 in our optics example



and if that is the case we can recognize a PH123 analog of this situation. This is like a guitar string fixed on both ends. We know we will get standing waves.

Let's call the wave going to the right ψ_{2r} and the wave going to the left ψ_{2l} . We could write wave equations for these two waves.

$$\psi_{2r} = C_{2r} \sin(k_2 x - \omega_2 t + \phi_{2r})$$

$$\psi_{2l} = C_{2l} \sin(k_2 x + \omega_2 t + \phi_{2l})$$

And to make things easier, we can do our work for the special case of $t = 0$, knowing that whatever parameter values we find for this special case must be true for all times. We have coefficients C_{20} and ϕ_{20} at the boundary at $x = 0$. At this boundary $\psi_{2r}(0) = 0$

$$0 = C_{2r} \sin(k_2(0) + \phi_{2r})$$

$$0 = \sin(\phi_{2r})$$

so we don't have a constraint on C_{2r} but we do have a constraint on ϕ_{2r} . It should be zero.

$$\phi_{2r} = 0$$

and likewise $\psi_{2r}(L) = 0$ so

$$\begin{aligned} 0 &= C_{2r} \sin(k_2 L) \\ 0 &= \sin(k_2 L) \end{aligned}$$

since $\sin(\theta) = 0$ when $\theta = 0, \pi, 2\pi \dots$ which tells us that

$$k_2 L = n\pi$$

so

$$\frac{2\pi}{\lambda_2} L = n\pi$$

or

$$\lambda_2 = \frac{2\pi}{n\pi} L = \frac{2}{n} L$$

And now we see a very important wave property. Only certain wavelengths will make wave packets that fit in Region 2 with our infinite potential boundary conditions. When only certain values of a physical parameter work in a certain physical situation, we say the system is *quantized*. Of course because frequency and wavelength are related, we expect that only certain values of the frequency are allowed. and

$$f_2 = \frac{v}{\lambda_2} \quad (12.1)$$

$$= \frac{v}{\frac{2}{n} L} \quad (12.2)$$

$$= n \frac{v}{2L} \quad (12.3)$$

But we were looking for the amplitude coefficient C_{2l} which we need to complete our wave equation. We have again

$$0 = C_{2l} \sin(k_2(0) + \phi_{2l})$$

so

$$0 = \sin(\phi_{2l})$$

and once again

$$\phi_{2l} = 0$$

but once again we don't have a way to find C_{2L} . We are going to need some new physics to find the amplitudes to complete the wave equations. Let's assume that for now the symmetry of the situation demands that

$$C_{2l} = C_{2r} = C$$

then

$$\begin{aligned} \psi_{2r} &= C \sin(k_2 x - \omega_2 t) \\ \psi_{2l} &= C \sin(k_2 x + \omega_2 t) \end{aligned}$$

The sum is

$$\psi_r = \psi_{2r} + \psi_{2l} = C \sin(k_2x - \omega_2t) + C \sin(k_2x + \omega_2t)$$

To gain insight into what these two waves produce, we use another of our favorite trig identities

$$\sin(a \pm b) = \sin(a)\cos(b) \pm \cos(a)\sin(b)$$

to get

$$\begin{aligned}\psi_r &= C \sin(k_2x - \omega_2t) + C \sin(k_2x + \omega_2t) \\ &= C \sin(k_2x) \cos(\omega_2t) - C \cos(k_2x) \sin(\omega_2t) + C \sin(k_2x) \cos(\omega_2t) + C \cos(k_2x) \sin(\omega_2t) \\ &= 2C \sin(k_2x) \cos(\omega_2t) \\ &= (2C \sin(k_2x)) \cos(\omega_2t)\end{aligned}$$

This looks like the harmonic oscillator equation

$$\psi = \psi_{\max} \cos(\omega t + \phi_o)$$

with $\phi_o = 0$ and with another complicated amplitude

$$A = 2C \sin(k_2x)$$

Note that we have an oscillatory part and an amplitude part. Let's deal with the amplitude for a while because that will set the shape of the wave packets. We have an amplitude that is dependent on position, and it must be zero at $x = 0$ and $x = L$.

The problem we have been working is given the affectionate name of a *particle in a box* because the particle can't escape. It's a good approximation for many physical systems where one particle is bound to something else (electron bound to an atom, atoms bound to molecules, etc.). So it gives us great physical insight. But we must be able to find the wave equation including its amplitude. We are missing C in what we have done so far. We need to be able to find the C . And notice that since $\psi_2(0) = 0$ and $\psi_2(L) = 0$ and $\psi_1(0) = 0$ and $\psi_3(L) = 0$, we are not going to be able to use matching on the boundaries to find C . We need another strategy. Let's work on the amplitude next.

the Schrödinger equation

Let's continue with our particle in a box problem. We failed to find the amplitude for our wave equation. We could try many different things to find this amplitude, but a physicist named Schrödinger came up with one that has worked well. We will do the non-relativistic case in this class (but if you go on in physics, the relativistic case is really cool). Schrödinger's approach starts with trying derivatives. He gave a symbol for the wave function for a de Broglie wave packet, ψ just like we have done. In our

particle in a box problem then

$$\psi(x, t) = 2C \sin(kx) \cos(\omega t)$$

and we know that we need the amplitude C part so let's work with the envelope for a moment and ignore the $\cos(\omega t)$ wavy part (that is, we will work at $t = 0$).

$$\psi(x, 0) = C \sin(kx) \cos(\omega(0)) = C \sin(kx)$$

Let's write $\psi(x)$ without showing its time dependency (t) as the amplitude of $\psi(x, t)$.

$$\psi(x, t) = \psi(x) \cos(\omega t)$$

And let's find the derivative of the wave function

$$\frac{d\psi(x)}{dx} = 2Ck \cos(kx)$$

and the second derivative

$$\begin{aligned} \frac{d^2\psi(x)}{dx^2} &= -k^2 2C \sin(kx) \\ &= -k^2 \psi(x) \end{aligned}$$

The second derivative gives a constant times the original function. This isn't a surprise if you paid close attention in PH123. We know that this is a property of wave equations. Schrödinger further mused that for a non-relativistic case

$$K = \frac{p^2}{2m} = \frac{(mv)^2}{2m} = \frac{1}{2}mv^2$$

and for a de Broglie wave

$$p = \frac{\hbar}{\lambda} = \frac{\hbar 2\pi}{\lambda 2\pi} = \hbar k$$

so for a non-relativistic de Broglie wave the kinetic energy would be

$$K = \frac{\hbar^2 k^2}{2m}$$

And we can solve for the wave number in terms of K

$$k^2 = \frac{2mK}{\hbar^2}$$

We can put this back into our second derivative.

$$\begin{aligned} \frac{d^2\psi(x)}{dx^2} &= -k^2 \psi(x) \\ &= -\frac{2mK}{\hbar^2} \psi(x) \end{aligned}$$

and now recall that the total non-relativistic energy for a wave packet would be

$$E = K + U$$

where in general we will think of potential energies that arrive from field forces like electric field forces or the strong nuclear force. Gravity will be a special case that we will deal with relativistically (we know the tie to rest mass already). We will save that for later. But the gravitational pull on an electron is small enough that we will ignore it

for now. Then

$$K = E - U$$

and

$$\frac{d^2\psi(x)}{dx^2} = -\frac{2m}{\hbar^2}(E - U)\psi(x)$$

Rearranging gives

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi - U\psi(x)$$

and finally

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U\psi(x) = E\psi(x)$$

This is exciting in and of itself because this last equation is so famous. This is the non-relativistic, time-independent Schrödinger equation. It is time independent because we used just the amplitude of the wave function to get it. But what is more exciting is that we can use the Schrödinger equation to get our missing amplitude.

Let's make an additional notation change before we get the amplitude. We left off the $\cos(\omega t)$. But that is fine, we were after the amplitude so we just wanted the time independent amplitude part. We can retrieve the time independence by writing

$$\psi(x, t) = \psi(x) \cos(\omega t)$$

but here we are going to play a mathematical trick. We can use something from that complex analysis class that you haven't taken

$$e^{-ix} = \cos(x) - i \sin(x)$$

This is a complex number ($i = \sqrt{-1}$). Our equation is real, there is no imaginary part. But notice that the real part of e^{-ix} is $\cos(x)$ so

$$\operatorname{Re}(e^{-i\omega t}) = \cos(\omega t)$$

Then if we write

$$\begin{aligned} \psi(x, t) &= \psi(x) \cos(\omega t) \\ &= \psi(x) \operatorname{Re}(e^{-i\omega t}) \end{aligned}$$

that would work. But the Re is annoying to write. We can just write

$$\psi(x, t) = \psi(x)(e^{-i\omega t})$$

and remember at the end of our calculations we need to take the real part (yes we really do it this way).

To help get our amplitude. Let's delay one more time and talk philosophically. What is this wave function we are building? It seems different than a water wave or a sound wave. The question is hard to answer. We know the wave equation solutions should be like electromagnetic waves from optics. And we know that the square of the amplitude at a particular position of an electromagnetic wave gives us the probability of detecting

a photon at that location.

$$I = \frac{1}{2\mu_0 c} E_{\max}^2$$

Where the electromagnetic wave amplitude is large, there is a high probability of detecting photons. Where it is low, there is a low probability of detecting photons. This suggests that we can treat the amplitude squared of a wave function as a probability

$$P(x) dx = |\psi(x)|^2 dx$$

To do this we need to “remember” some probability theory. The probability of detecting a photon at a mathematical point represented by a single value of x is always zero. Think about this. If the particle is at $x + dx$ no matter how small the dx might be, the photon is still not at x . So finding the probability of a wave-particle at any single x value isn’t very helpful. It is like asking how many students in a class got a score of 91.43530694538475% on the test. The chance of getting that exact score is very low. But we can meaningfully ask how many students got a A- on the exam by binning all the students that got higher than 90% and lower than a 93%. We do the same thing with locating a particle. We find the probability that the wave-particle is within some range of positions Δx or if the range is small, dx . That is what the dx means in our equation. The probability that the particle is in the range of x to $x + dx$.

This should be proportional to our amplitude squared $|\psi(x)|^2$. It is convenient to notice that

$$|\psi(x, t)|^2 = |\psi(x)|^2 |e^{-i\omega t}|^2$$

Here the absolute value symbol means we take the item, say $e^{-i\omega t}$ and multiply it by its complex conjugate, $e^{+i\omega t}$. A complex conjugate is the same quantity but with i replaced by $-i$ so

$$|e^{-i\omega t}|^2 = e^{-i\omega t} e^{+i\omega t} = 1$$

or in general if we have a complex number $z = a + ib$ where a and b are real, then the complex conjugate of z is $z^* = a - ib$. And

$$|z^2| = zz^* = (a + ib)(a - ib) = a^2 + b^2$$

Then, we can write

$$\begin{aligned} |\psi(x, t)|^2 &= |\psi(x)|^2 |e^{-i\omega t}|^2 \\ &= |\psi(x)|^2 (1) \\ &= |\psi(x)|^2 \end{aligned}$$

so it won’t hurt to stick $|\psi(x, t)|^2$ into our probability equation in place of $|\psi(x)|^2$.

We still don’t have our amplitude C value. But we are getting close. We could write the

probability of finding the wave-particle between x_1 and x_2 as

$$P(x_1 : x_2) = \int_{x_1}^{x_2} P(x) dx$$

so in our case

$$P(x_1 : x_2) = \int_{x_1}^{x_2} |\psi(x, t)|^2 dx$$

We know that the probability of the wave-particle being somewhere in the universe has to be 1. If we integrate our probability over all possible states, the integral has to be 1 (the particle/wave has to be somewhere in the universe). So

$$\int_{-\infty}^{\infty} P(x) dx = 1$$

and for us this becomes

$$\int_{-\infty}^{\infty} |\psi(x, t)|^2 dx = 1$$

There is a special name for this condition even though it seems so obvious. A wave function that integrates over all space to 1 is called *normalized*.

Now let's look at the idea of an average. The average value of some quantity x , is given as

$$\bar{x} = \frac{\sum_{i=1}^N n x_i}{\sum_{i=1}^N x_i}$$

we take the number of times a particular value of x is found in the data set. That is $n x_i$. And we sum that up weighting each x_i by the number of times that x_i occurs. Then we have to divide by the total which is $\sum x_i$

We want to find the place where we will detect the wave-particle most often. That is like a classical average location of a particle. That would be each location x_i multiplied by the probability of finding the particle at that location. And we need to divide by the total probability. We write this as

$$\bar{x} = \frac{\int_{-\infty}^{\infty} (x) P(x) dx}{\int_{-\infty}^{\infty} P(x) dx}$$

The (x) in the integral in the numerator is the important part to notice. That is the thing being averaged. It could be that we want the average energy, and that would look like

$$\bar{E} = \frac{\int_{-\infty}^{\infty} (E) P(x) dx}{\int_{-\infty}^{\infty} P(x) dx}$$

In quantum mechanics this average detection location is called an *expectation value* of x . It is often denoted with angle brackets

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} (x) P(x) dx}{\int_{-\infty}^{\infty} P(x) dx}$$

Let's apply this to our particle in a box problem. We found so far that a wave equation like

$$\psi(x, t) = C \sin(kx) \cos(\omega t)$$

should work. We could write this as

$$\psi(x, t) = C \sin(kx) e^{-i\omega t}$$

and we could say that for our wave function the particle must be some where in the universe.

$$\int_{-\infty}^{\infty} |C \sin(kx) e^{-i\omega t}|^2 dx = 1$$

since we found for the particle in a box

$$kL = n\pi$$

then

$$\begin{aligned} \int_{-\infty}^{\infty} \left| C \sin\left(\frac{n\pi}{L}x\right) e^{-i\omega t} \right|^2 dx &= 1 \\ C^2 \int_{-\infty}^{\infty} \sin^2\left(\frac{n\pi}{L}x\right) dx &= 1 \end{aligned}$$

Let's get another result from an integral table (or a symbolic math processor)

$$\int \sin^2(ax) dx = -\frac{1}{4a} (\sin 2ax - 2ax)$$

If we let $a = \frac{n\pi}{L}$ the our integral is just like the table form. Because we know that the wave function is zero for $-\infty < x < 0$ and for $L < x < \infty$ we can integrate from 0 to L . This just says the wave particle must be in the box.

$$\begin{aligned} 1 &= C^2 \int_0^L \sin^2\left(\frac{n\pi}{L}x\right) dx \\ &= \left(-C^2 \frac{1}{4\frac{n\pi}{L}} \left(\sin\left(2\frac{n\pi}{L}x\right) - 2\frac{n\pi}{L}x \right) \right)_0^L \\ &= -C^2 \frac{1}{4\frac{n\pi}{L}} \left(\sin\left(2\frac{n\pi}{L}L\right) - 2\frac{n\pi}{L}L \right) - \left(-C^2 \frac{1}{4\frac{n\pi}{L}} \left(\sin\left(2\frac{n\pi}{L}(0)\right) - 2\frac{n\pi}{L}(0) \right) \right) \\ &= -C^2 \frac{1}{4\frac{n\pi}{L}} \left(\sin\left(2\frac{n\pi}{L}L\right) - 2\frac{n\pi}{L}L \right) - \left(-C^2 \frac{1}{4\frac{n\pi}{L}} \left(\sin\left(2\frac{n\pi}{L}(0)\right) \right) \right) \end{aligned}$$

and we know that at L and at 0 the sine terms must be zero

$$1 = -C^2 \frac{1}{4\frac{n\pi}{L}} \left(-2\frac{n\pi}{L}L \right)$$

so

$$1 = C^2 \frac{1}{2} (L)$$

and

$$C = \sqrt{\frac{2}{L}}$$

and we have our amplitude C value! We can write our wave function as

$$\psi(x, t) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi}{L}x\right) e^{-i\omega t}$$

where $n = 0, 1, 2, \dots$

We don't have just one equation that works, we have a bunch of equations, each with a different n value⁵. We will need to refine our analysis next lecture. But what we have done in this lecture is huge.

1. We have thought about how wave functions for matter waves and even photons must act at boundaries.
2. We have found how to use boundary conditions to find a mathematical form for the wave function in different regions based on the potential energy of the regions.
3. We have found that for a particle in an infinite potential “well” (particle in a box) we could find the form of the equation but not the amplitudes.
4. We solve this by deriving the non-relativistic, time-independent Schrödinger equation
5. We interpreted the amplitude of the wave function as a probability of detection
6. We used the ideas of normalization in probability theory to solve for the amplitude
7. Along the way we used the idea of probabilistic averages to define expectation values

We did something like thirty years of physics in about an hour! We will revisit all this in the next lecture to solidify the procedure. But as of now, we are proper quantum physicists!

⁵ Think of the harmonic series for waves on strings.

13 Using the Schrödinger Equation

5.4-5.6

Fundamental Concepts in the Lecture

- Constant Potential Energy and the Schrödinger Equation
- Free Wavicles (wave-particles)
- Infinite potential well again
- Finite potential well
- Two-dimensional infinite potential well

Constant Potential Energy

Before we return to our particle in an infinite potential well, let's take another look at the Schrödinger equation and do some simple problems. The non-relativistic, time-independent Schrödinger equation we found to be

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U\psi(x) = E\psi(x)$$

If $E > U$ we can write

$$-\frac{d^2\psi(x)}{dx^2} + \frac{2m}{\hbar^2} (U - E)\psi(x) = 0$$

Or, if we take out a minus sign

$$-\frac{d^2\psi(x)}{dx^2} = -\frac{2m}{\hbar^2} (E - U)\psi(x)$$

and then take the constants to the other side

$$\frac{d^2\psi(x)}{dx^2} = -\frac{2m}{\hbar^2} (E - U)\psi(x)$$

And if we define

$$k = \sqrt{\frac{2m(E - U)}{\hbar^2}} = \frac{2\pi}{\lambda}$$

where k is the wave number, we can write the Schrödinger equation as

$$\frac{d^2\psi(x)}{dx^2} = -k^2\psi(x)$$

This really looks like a PH123 problem because this is just the linear wave equation and back in PH123 we showed that sine and cosine functions are solutions to this equation. We can see that energy is going to be important in our study of wave mechanics. The energy in our wavicle determines the wave number (and therefore the wavelength) of our wave.

The Schrödinger equation is linear, which means that a solution to the Schrödinger equation added to another solution to the Schrödinger equation would also be a solution. This is because derivatives of sums of functions are equal to the sum of the derivatives of the individual functions. Since we know sines and cosines are solutions, then the sum of a sine and a cosine function would be a solution

$$\psi(x) = A \sin(kx) + B \cos(kx)$$

Let's see that it works

$$\frac{d\psi(x)}{dx} = Ak \cos(kx) - Bk \sin(kx)$$

and

$$\begin{aligned}\frac{d^2\psi(x)}{dx^2} &= -Ak^2 \sin(kx) - Bk^2 \cos(kx) \\ &= -k^2(A \sin(kx) + B \cos(kx)) \\ &= -k^2\psi(x)\end{aligned}$$

so it does work. Of course if we were to do a problem (we will soon) we would need to find the coefficients A and B just like we did in the last lecture.

But now consider the case when $E < U$. In PH121 we would say that a particle could not travel into a region where U is bigger than the particle energy. This would be like a ball with only a little bit of energy being able to climb tall hills.

But for wave/particles this would be the case of a wave penetrating into a forbidden region. In optics this could be a case where the energy in the wave is absorbed, or even the small penetration into the metal of a mirror before reflection. The expression for k above will be imaginary and that isn't so helpful. If $E < U$ we can instead re-write the Schrödinger equation

$$-\frac{d^2\psi(x)}{dx^2} + \frac{2m}{\hbar^2}(U - E)\psi(x) = 0$$

$$\begin{aligned}-\frac{d^2\psi(x)}{dx^2} &= -\frac{2m}{\hbar^2}(U-E)\psi(x) \\ \frac{d^2\psi(x)}{dx^2} &= +\frac{2m}{\hbar^2}(U-E)\psi(x)\end{aligned}$$

and define

$$\alpha^2 = \frac{2m}{\hbar^2}(U-E)$$

We then have

$$\frac{d^2\psi(x)}{dx^2} = +\alpha^2\psi(x)$$

Sines and cosines don't work for this equation. We need a new function that gives itself when you take the second derivative, but without introducing a minus sign. The function $e^{\alpha x}$ can do this. And so can $e^{-\alpha x}$. So we could have a solution of the form

$$\psi(x) = Ce^{\alpha x} + De^{-\alpha x}$$

where C and D are amplitudes. And just to be complete let's show that it works

$$\begin{aligned}\frac{d\psi(x)}{dx} &= C\alpha e^{\alpha x} - D\alpha e^{-\alpha x} \\ \frac{d^2\psi(x)}{dx^2} &= C\alpha^2 e^{\alpha x} + D\alpha^2 e^{-\alpha x} \\ &= \alpha^2(Ce^{\alpha x} + De^{-\alpha x}) \\ &= \alpha^2\psi(x)\end{aligned}$$

which is what we expected. So in a region with $E < U$, our solution would be

$$\psi(x) = Ce^{\alpha x} + De^{-\alpha x}$$

This function is an exponential and this $e^{-\alpha x}$ is what we call an exponential decay. The amplitude of a wave may decay to zero as it penetrates into a forbidden region. In optics this is like the beam amplitude decaying to zero as it penetrates into the metal of a mirror. The exponential nature of the fall-off seems natural (because nature works that way!).

Free Particle

Let's try a new scenario, lets let $U = 0$. We should get a solution that is a wavicle that is traveling in free open space where there is nothing to pull or push it (remember forces are directly related to potential energies). Then the energy is given by the de Broglie equations

$$E = \frac{\hbar^2 k^2}{2m}$$

and we won't have a quantized k because there is no potential well to cause reflections

to make standing waves form. There are also no boundary conditions because we have assumed no boundary (there is no U , anywhere, so no change in U anywhere). Of course this would mean we have only one wavicle and nothing else in the entire universe, which is never the case. But this might be a good approximation for an electron in a vast open empty region of the universe.

Normally we would use the boundary conditions to find the coefficients A and B . But there is no boundary. But no problem, we have normalization to do this job. So the wavicle must be somewhere in the universe

$$\int_{-\infty}^{\infty} P(x) dx = 1$$

and taking our probability interpretation for the wave function this becomes

$$\int_{-\infty}^{\infty} |\psi(x, t)|^2 dx = 1$$

and we could guess a wave function of the form

$$\psi(x, t) = (A \sin(kx) + B \cos(kx)) e^{-i\omega t}$$

so we would write

$$\int_{-\infty}^{\infty} |(A \sin(kx) + B \cos(kx)) e^{-i\omega t}|^2 dx = 1$$

and separating the complex time dependent term we get $|e^{-i\omega t}|^2 = 1$

$$\int_{-\infty}^{\infty} |A \sin(kx) + B \cos(kx)|^2 dx = 1$$

and taking the square gives

$$\int_{-\infty}^{\infty} (A^2 \sin^2(kx) + AB \sin(kx) \cos(kx) + B^2 \cos^2(kx)) dx = 1$$

and here we see we have another problem. The first term will oscillate all the way from $-\infty$ to $+\infty$. It does not converge at $x = \pm\infty$. The last term is no better, it just shifted along the x -axis. And there is sort of a cross term involved in the square, but that won't converge as well. We will end up adding infinities at the limits, and this is not a comfortable situation. It is hard to see how it will come to be just 1.

Let's play another mathematical trick to make this more tractable. Once again using something from that complex analysis course you haven't taken yet, let's write

$$\begin{aligned} \sin(kx) &= \frac{e^{ikx} - e^{-ikx}}{2i} \\ \cos(kx) &= \frac{e^{ikx} + e^{-ikx}}{2} \end{aligned}$$

so now

$$\begin{aligned}
 \psi(x) &= A \sin(kx) + B \cos(kx) \\
 &= A \left(\frac{e^{ikx} - e^{-ikx}}{2i} \right) + B \left(\frac{e^{ikx} + e^{-ikx}}{2} \right) \\
 &= \frac{Ae^{ikx}}{2i} - \frac{Ae^{-ikx}}{2i} + \frac{Be^{ikx}}{2} + \frac{Be^{-ikx}}{2} \\
 &= \left(\frac{Ae^{ikx}}{2i} + \frac{Be^{ikx}}{2} \right) + \left(-\frac{Ae^{-ikx}}{2i} + \frac{Be^{-ikx}}{2} \right) \\
 &= \left(\frac{A}{2i} + \frac{B}{2} \right) e^{ikx} + \left(\frac{B}{2} - \frac{A}{2i} \right) e^{-ikx}
 \end{aligned}$$

and let's define two new amplitude functions

$$\begin{aligned}
 a &= \left(\frac{A}{2i} + \frac{B}{2} \right) \\
 b &= \left(\frac{B}{2} - \frac{A}{2i} \right)
 \end{aligned}$$

so that

$$\psi(x) = ae^{ikx} + be^{-ikx}$$

and, putting in the time dependence

$$\begin{aligned}
 \psi(x, t) &= ae^{ikx} e^{-i\omega t} + be^{-ikx} e^{-i\omega t} \\
 &= ae^{i(kx-\omega t)} + be^{-i(kx+\omega t)}
 \end{aligned}$$

We can recognize this as two waves, the first one going to the right, and the second one going to the left. Only now we don't just have an agreed upon complex time dependence where we know to take the real part at the end. No, now the a and b coefficients are really complex.

Now let's return to our situation where $U = 0$ everywhere and we have a wavicle, say, moving to the right. Then we know $b = 0$ because the term $ae^{i(kx-\omega t)}$ is a wave moving to the right and $be^{-i(kx+\omega t)}$ is a wave moving to the left.

Given our wave moving to the right, let's try normalization, we have

$$\begin{aligned}
 P(x) &= |\psi(x, t)|^2 \\
 &= |ae^{i(kx-\omega t)}|^2 \\
 &= |a|^2 e^{i(kx-\omega t)} e^{-i(kx-\omega t)} \\
 &= |a|^2
 \end{aligned}$$

which is a constant, so the probability of detecting the particle is the same everywhere! The integral is still awkward because we have a subtraction of infinities at the edges that still must give us just 1. But we can now see why. The probability is constant

everywhere. So there is a small but constant probability that we find the wavicle anywhere.

This situation is called a *free particle*.

This is just what we should have expected from our uncertainty relationship

$$\Delta x \Delta \lambda \geq \frac{\hbar}{2}$$

We have just one wavelength in our $k = 2\pi/\lambda$, so $\Delta \lambda = 0$.

$$\Delta x \geq \frac{\hbar}{2\Delta \lambda}$$

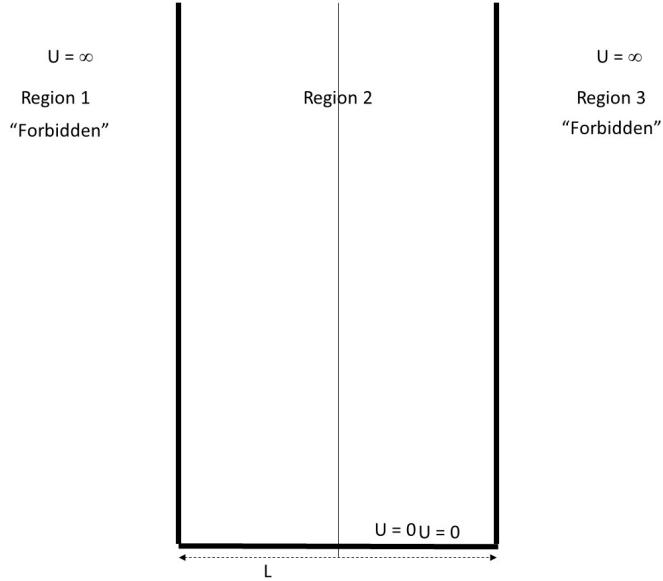
If we know λ precisely then Δx is infinite, meaning the wave is everywhere. Picture a nice sinusoidal wave in the middle of the Pacific ocean. In order to get the (nearly) single wavelength, the wave would have to be spread out everywhere, covering a large part of the ocean.

A wave function of the form

$$\psi(x, t) = ae^{ikx}e^{-i\omega t}$$

is called a plane wave. It is an idealization. Few waves (if any) can actually be infinite in extent. But often this is a good approximation for a wave in a big open volume.

Infinite potential well revisited



Suppose once again we have a wavicle in an infinite potential well. We now know that the Schrödinger equation

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U\psi(x) = E\psi(x)$$

should describe our situation. But we have three regions, and we expect now that we might have different functions in each of the three regions.

In Region 1

$$-\frac{\hbar^2}{2m} \frac{d^2\psi_1(x)}{dx^2} + (\infty)\psi(x) = E\psi(x)$$

and infinity is so big that $U \ggg E$ so we can just ignore the energy E

$$\frac{\hbar^2}{2m} \frac{d^2\psi_1(x)}{dx^2} = (\infty)\psi_1(x)$$

The only function that can satisfy this equation is

$$\psi_1(x) = 0$$

The wave function is zero everywhere in Region 1. We get the same equation with the same result for Region 3.

For Region 2 we have $U = 0$ which is now familiar!

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + 0 = E\psi(x)$$

or

$$\frac{d^2\psi(x)}{dx^2} = -\frac{2mE}{\hbar^2}\psi(x)$$

and we can identify

$$k = \sqrt{\frac{2mE}{\hbar^2}}$$

and this gives just one wavelength, we expect the wave to exist everywhere in Region 2.

But we know that at the boundaries there could be reflections of this wave. So we can have two waves in the box traveling different directions. And if we have two waves in the same region, we could form standing waves! Let's call our two waves ψ_{2r} and ψ_{2l} (one going left, and one going right).

$$\psi_{2r}(x, t) = C_{20} \sin(k_2 x - \omega_2 t + \phi_{20})$$

$$\psi_{2l}(x, t) = C_{2L} \sin(k_2 x + \omega_2 t + \phi_{2L})$$

Before we showed that $\phi_{20} = \phi_{2L} = 0$ so we can simplify to

$$\psi_{2r}(x, t) = C \sin(k_2 x - \omega_2 t)$$

$$\psi_{2l}(x, t) = C \sin(k_2 x + \omega_2 t)$$

We found the sum

$$\psi_r(x, t) = \psi_{2r}(x, t) + \psi_{2l}(x, t) = C \sin(kx - \omega t) + C \sin(kx + \omega t)$$

and we used another of our favorite trig identities

$$\sin(a \pm b) = \sin(a) \cos(b) \pm \cos(a) \sin(b)$$

to get

$$\begin{aligned}\psi_r(x, t) &= C \sin(kx - \omega t) + C \sin(kx + \omega t) \\ &= C \sin(kx) \cos(\omega t) - C \cos(kx) \sin(\omega t) + C \sin(kx) \cos(\omega t) + C \cos(kx) \sin(\omega t) \\ &= 2C \sin(kx) \cos(\omega t) \\ &= (2C \sin(kx)) \cos(\omega t)\end{aligned}$$

But now, realize that we could have done all this with two cosine waves

$$\psi_{2rc}(x, t) = D_{20} \cos(k_2 x - \omega_2 t + \phi_{20})$$

$$\psi_{2lc}(x, t) = D_{2L} \cos(k_2 x + \omega_2 t + \phi_{2L})$$

because a cosine wave is also a solution to the Schrödinger equation, so we should consider a solution that has both of these possibilities

$$\psi(x, t) = A \sin(k_2 x - \omega_2 t + \phi_{20}) + B \cos(k_2 x - \omega_2 t + \phi_{20})$$

or at $t = 0$, with knowing the phase constants would be zero

$$\psi(x, 0) = A \sin(k_2 x) + B \cos(k_2 x)$$

where the amplitudes A and B have to be determined. It is good to start with an addition of all the possible types of solutions to the Schrödinger equation and then use our boundary (and other) conditions to eliminate parts of the solution that can't work by setting their coefficients to zero. We call the combined set of possible solutions a *superposition state*.

For our case let's look at the B coefficient first. At the boundaries $x = 0$ and $x = L$ we know $\psi_2(0, t) = 0$ and $\psi_2(L, t) = 0$ to match $\psi_1 = 0$ and $\psi_3 = 0$ in Regions 1 and 3. So at $x = 0$

$$\begin{aligned}\psi(0, t) &= A \sin(k(0)) + B \cos(k(0)) = 0 \\ 0 &= 0 + B\end{aligned}$$

The only way this can be true is

$$B = 0$$

This seems like a waste of time to have tried this superposition state just to find out the cosine part didn't work. But we won't always know which possible solutions will work

until we try them at the boundaries. Let's look at $x = L$

$$\begin{aligned}\psi(L, t) &= A \sin(k(L)) + B \cos(k(L)) \\ &= A \sin(k(L))\end{aligned}$$

because $B = 0$. Then

$$0 = A \sin(k(L))$$

gives us

$$kL = n\pi$$

as we found before where $n = 1, 2, 3, \dots$. Let's investigate this further. From the Schrödinger equation we found

$$k = \sqrt{\frac{2mE}{\hbar^2}}$$

and

$$k = \frac{n\pi}{L}$$

for Region 2. So then

$$\sqrt{\frac{2mE}{\hbar^2}} = \frac{n\pi}{L}$$

and we can solve this for the energy, E , of the wavefunction

$$\begin{aligned}\sqrt{\frac{2mE}{\hbar^2}}L &= n\pi \\ \frac{2mE}{\hbar^2}L^2 &= n^2\pi^2 \\ E &= \frac{n^2\pi^2\hbar^2}{2mL^2} \\ E &= \frac{n^2\pi^2\frac{\hbar^2}{4\pi^2}}{2mL^2}\end{aligned}$$

And notice that the energy is quantized

$$E = \frac{n^2\hbar^2}{8mL^2} \quad n = 1, 2, 3, \dots$$

with a *quantum number* n . Let's define

$$E_o = \frac{\hbar^2}{8mL^2}$$

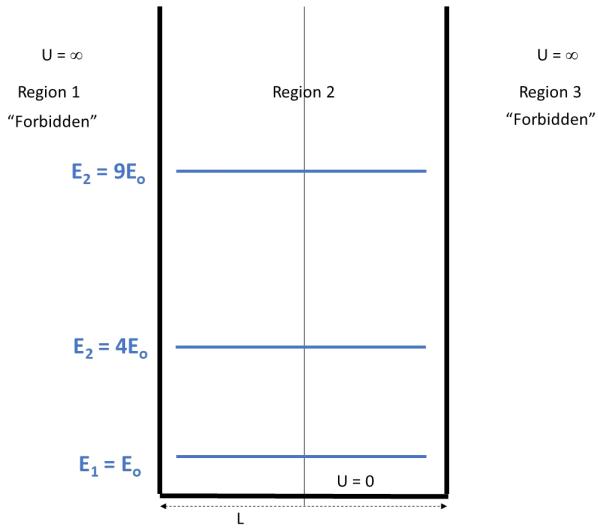
then

$$E = n^2 E_o \quad n = 1, 2, 3, \dots$$

We usually write this as

$$E_n = n^2 E_o \quad n = 1, 2, 3, \dots$$

to remind us that E is quantized. We see that only certain values of energy are possible in the well. Not all energy values give viable de Broglie waves that match the boundary condition. This is very like standing waves on strings. We can draw our energy diagram as



where only the first three energies fit on the diagram at the scale we used. From using the Schrödinger equation we once again have new physics. If we were to insert a detector and detect electrons in an infinite potential well. We would find that the electrons would only have the energies E_n . No in between energies would be found.

We used normalization to find the amplitude A before. Let's repeat it here.

Using our statistical interpretation of the Schrödinger equation we can say where the electrons are most likely to be detected for each allowed energy. We start with the wave function

$$\psi(x, t) = A \sin(kx)$$

and we could say that for our wave function the particle must be some where in the universe.

$$\int_{-\infty}^{\infty} |A \sin(kx) e^{-i\omega t}|^2 dx = 1$$

since we found

$$kL = n\pi$$

then

$$\begin{aligned} \int_{-\infty}^{\infty} & \left| A \sin\left(\frac{n\pi}{L}x\right) e^{-i\omega t} \right|^2 dx = 1 \\ & A^2 \int_{-\infty}^{\infty} \sin^2\left(\frac{n\pi}{L}x\right) dx = 1 \end{aligned}$$

we used the standard form

$$\int \sin^2(ax) dx = -\frac{1}{4a} (\sin 2ax - 2ax)$$

If we let $a = \frac{n\pi}{L}$ the our integral is just like the table form. Because we know that the wave function is zero for $-\infty < x < 0$ and for $L < x < \infty$ we can integrate from 0 to L . This just says the wave particle must be in the box.

$$\begin{aligned} 1 &= A^2 \int_0^L \sin^2 \left(\frac{n\pi}{L} x \right) dx \\ &= \left(-A^2 \frac{1}{4 \frac{n\pi}{L}} \left(\sin \left(2 \frac{n\pi}{L} x \right) - 2 \frac{n\pi}{L} x \right) \right|_0^L \\ &= -A^2 \frac{1}{4 \frac{n\pi}{L}} \left(\sin \left(2 \frac{n\pi}{L} L \right) - 2 \frac{n\pi}{L} L \right) - \left(-A^2 \frac{1}{4 \frac{n\pi}{L}} \left(\sin \left(2 \frac{n\pi}{L} (0) \right) \right) \right) \end{aligned}$$

and we know that at L and at 0 the sine terms must be zero

$$1 = -A^2 \frac{1}{4 \frac{n\pi}{L}} \left(-2 \frac{n\pi}{L} L \right)$$

so

$$1 = A^2 \frac{1}{2} (L)$$

and

$$A = \sqrt{\frac{2}{L}}$$

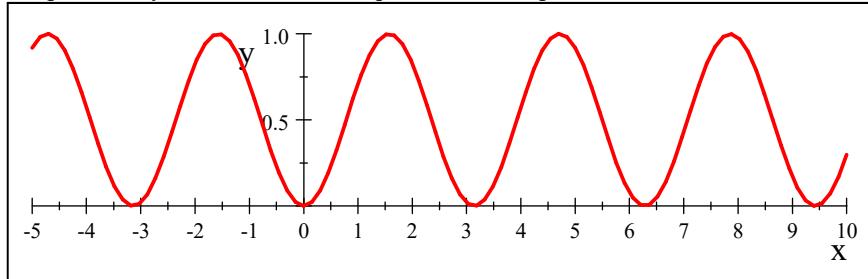
and we have our amplitude ! We can write our wave function as

$$\psi(x, t) = \sqrt{\frac{2}{L}} \sin \left(\frac{n\pi}{L} x \right) e^{-i\omega t}$$

But wait! we can do more. We can interpret this wave function to know where the electrons are most likely to be detected. The probability of detection is proportional to the amplitude squared

$$\begin{aligned} P(x) &= |\psi(x, t)|^2 \\ &= \left| \sqrt{\frac{2}{L}} \sin \left(\frac{n\pi}{L} x \right) e^{-i\omega t} \right|^2 \\ &= \frac{2}{L} \sin^2 \left(\frac{n\pi}{L} x \right) \end{aligned}$$

so our probability function is a sine squared. The shape looks like this $\sin^2 x$

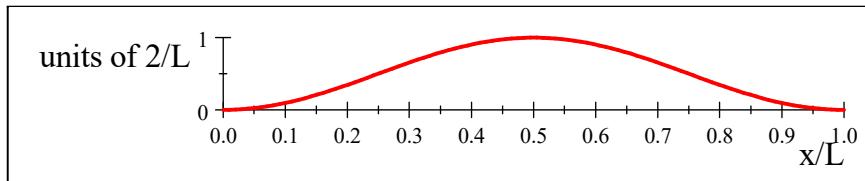


but we don't have an actual value for L . So getting the shape right is harder. But we can

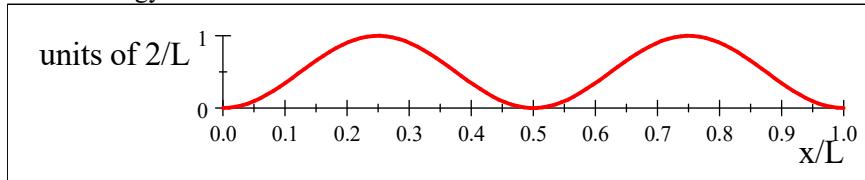
cheat in making a graph. We can define a variable that is $u = x/L$ then

$$P(x) \frac{2}{L} \sin^2(n\pi u)$$

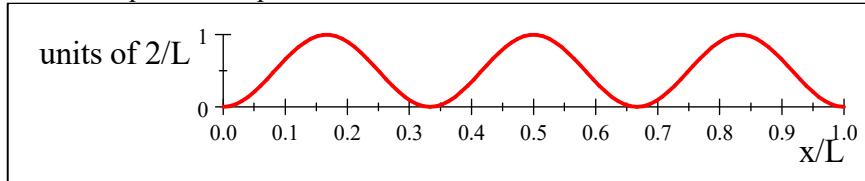
and we can make the y axis in units of $2/L$ and plot from $u(0) = (0)/L = 0$ to $u(L) = (L/L) = 1$. For $n = 1$ we get



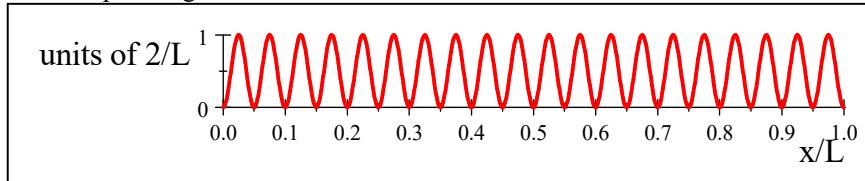
which tells us that we are most likely to detect electrons in the middle of the well for the lowest energy electrons. For $n = 2$ we have



For the second allowed energy the electrons are never detected right in the middle of the well, but rather are most likely detected at the $x = L/4$ and $x = 3L/4$ spots. The electron wave packet has spread out. For $n = 3$

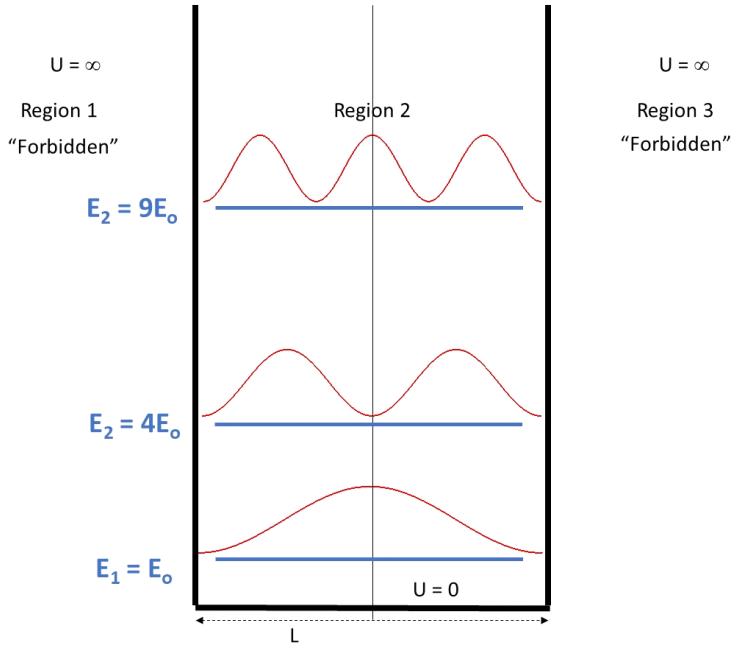


we see the spreading continue. For $n = 20$



we would find that we could detect the electron all over the well, but still the detections would be in clumps.

Let's put this probability of detection on our energy graph.



The spreading of the wave function amplitude reminds us of uncertainty.

$$\Delta x \Delta p \approx \hbar$$

we can use the definition of standard deviation form to find Δx

$$\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

finding $\langle x \rangle$ is easy, it's just $L/2$

$$\langle x \rangle^2 = \frac{L^2}{4}$$

but finding $\langle x^2 \rangle$ is harder. Let's use our average finding equation from last lecture.

$$\begin{aligned} \langle x^2 \rangle &= \frac{\int_0^L (x^2) P(x) dx}{\int_0^L P(x) dx} \\ &= \frac{\int_0^L (x^2) \left| \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi}{L}x\right) e^{-i\omega t} \right|^2 dx}{\int_0^L \left| \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi}{L}x\right) e^{-i\omega t} \right|^2 dx} \\ &= \frac{\frac{2}{L} \int_0^L (x^2) \sin^2\left(\frac{n\pi}{L}x\right) dx}{\frac{2}{L} \int_0^L \sin^2\left(\frac{n\pi}{L}x\right) dx} \\ &= \frac{\int_0^L (x^2) \sin^2\left(\frac{n\pi}{L}x\right) dx}{\int_0^L \sin^2\left(\frac{n\pi}{L}x\right) dx} \end{aligned}$$

we can look up (or ask a symbolic math package) and find

$$\int \sin^2(ax) dx = -\frac{1}{4a} (\sin 2ax - 2ax)$$

$$\int x^2 \sin^2(ax) dx = \frac{\sin(2ax)}{8(a)^3} - \frac{x \cos(2ax)}{4a^2} - \frac{x^2 \sin(2ax)}{4a} + \frac{x^3}{6}$$

in our case $a = \frac{n\pi}{L}$ so

$$\begin{aligned} \int_0^L (x^2) \sin^2\left(\frac{n\pi}{L}x\right) dx &= \left(\frac{\sin(2(\frac{n\pi}{L})x)}{8(\frac{n\pi}{L})^3} - \frac{x \cos(2(\frac{n\pi}{L})x)}{4(\frac{n\pi}{L})^2} - \frac{x^2 \sin(2(\frac{n\pi}{L})x)}{4(\frac{n\pi}{L})} + \frac{x^3}{6} \right|_0^L \\ &= \frac{\sin(2(\frac{n\pi}{L})L)}{8(\frac{n\pi}{L})^3} - \frac{L \cos(2(\frac{n\pi}{L})L)}{4(\frac{n\pi}{L})^2} - \frac{L^2 \sin(2(\frac{n\pi}{L})L)}{4(\frac{n\pi}{L})} + \frac{L^3}{6} \\ &\quad - \left(\frac{\sin(2(\frac{n\pi}{L})(0))}{8(\frac{n\pi}{L})^3} - 0 - 0 + 0 \right) \\ &= \frac{\sin(2n\pi)}{8(\frac{n\pi}{L})^3} - \frac{L \cos(2n\pi)}{4(\frac{n\pi}{L})^2} - \frac{L^2 \sin(2n\pi)}{4(\frac{n\pi}{L})} + \frac{L^3}{6} - (0) \\ &= 0 - \frac{L(1)}{4(\frac{n\pi}{L})^2} - 0 + \frac{L^3}{6} - 0 \\ &= \frac{L^3}{6} - \frac{L^3}{4n^2\pi^2} \end{aligned}$$

and

$$\begin{aligned} \int_0^L \sin^2\left(\frac{n\pi}{L}x\right) dx &= \left(-\frac{1}{4\frac{n\pi}{L}} \left(\sin 2\frac{n\pi}{L}x - 2\frac{n\pi}{L}x \right) \right|_0^L \\ &= -\frac{1}{4\frac{n\pi}{L}} \left(\sin 2\frac{n\pi}{L}L - 2\frac{n\pi}{L}L \right) - \left(-\frac{1}{4\frac{n\pi}{L}} \left(\sin 2\frac{n\pi}{L}0 - 2\frac{n\pi}{L}0 \right) \right) \\ &= -\frac{1}{4\frac{n\pi}{L}} \left(\sin 2\frac{n\pi}{L}L - 2\frac{n\pi}{L}L \right) - 0 \\ &= -\frac{\sin 2\frac{n\pi}{L}L}{4\frac{n\pi}{L}} + \frac{2\frac{n\pi}{L}L}{4\frac{n\pi}{L}} \\ &= -\frac{L}{4n\pi} \sin(2n\pi) + \frac{L}{2} \\ &= \frac{L}{2} \end{aligned}$$

so

$$\begin{aligned} \overline{(x^2)} &= \frac{\frac{L^3}{6} - \frac{L^3}{4n^2\pi^2}}{\frac{L}{2}} \\ &= \frac{\frac{L^3}{6} - \frac{L^3}{4n^2\pi^2}}{\frac{L}{2}} \\ &= \frac{L^2}{3} - \frac{L^2}{2n^2\pi^2} \end{aligned}$$

$$\begin{aligned}\Delta x &= \sqrt{\langle x^2 \rangle - \langle \bar{x} \rangle^2} \\ &= \sqrt{\frac{L^2}{3} - \frac{L^2}{2n^2\pi^2} - \frac{L^2}{4}} \\ &= L\sqrt{\frac{1}{12} - \frac{1}{2\pi^2n^2}}\end{aligned}$$

Now we know that the electrons can go either direction so $\bar{p} = 0$ and then

$$\begin{aligned}K &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}mv(v) \\ &= \frac{1}{2m}(mv)^2 \\ &= \frac{p^2}{2m}\end{aligned}$$

which is the nonrelativistic kinetic energy - momentum relation. Non-relativistic is fine because of our non-relativistic assumptions in the Schrödinger equation so

$$\begin{aligned}p^2 &= 2mK \\ &= 2mE \\ &= 2m\frac{n^2h^2}{8mL^2} \\ &= \frac{n^2h^2}{4L^2}\end{aligned}$$

and none of these change so $\bar{p}^2 = p^2$. Then

$$\begin{aligned}\Delta p &= \sqrt{\langle p^2 \rangle - \langle \bar{p} \rangle^2} \\ &= \sqrt{\frac{n^2h^2}{4L^2} - 0} \\ &= \frac{nh}{2L}\end{aligned}$$

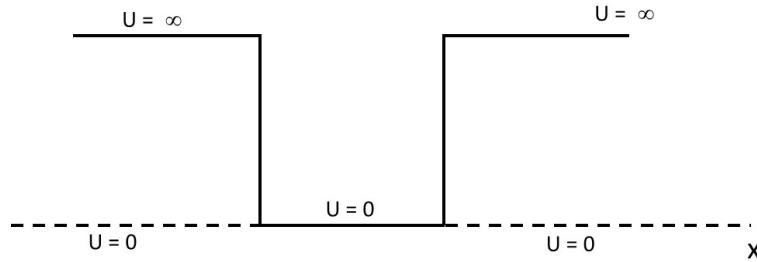
then

$$\begin{aligned}\Delta x \Delta p &= \frac{nh}{2L} L \sqrt{\frac{1}{12} - \frac{1}{2\pi^2n^2}} \\ &= \frac{nh}{2} \sqrt{\frac{1}{12} - \frac{1}{2\pi^2n^2}} \\ &= \frac{h}{2} \sqrt{\frac{n^2}{12} - \frac{1}{2\pi^2}}\end{aligned}$$

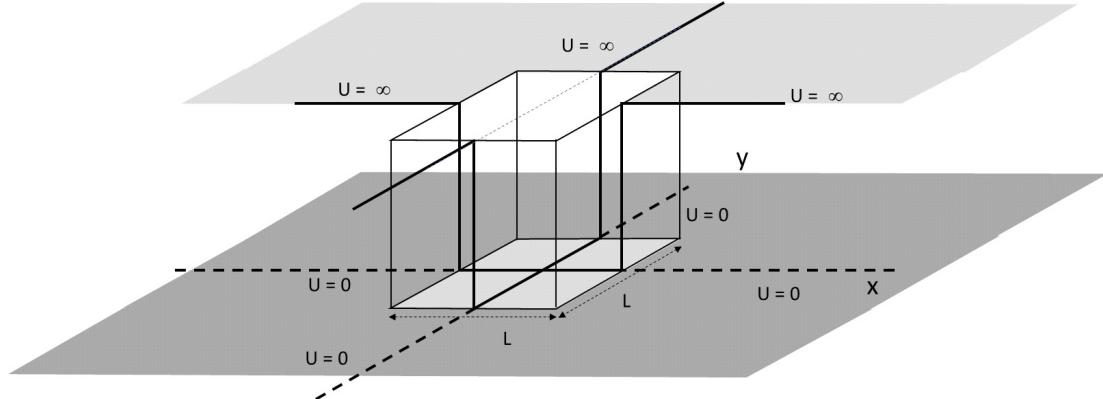
which isn't all that impressive to me. But it does say that the bigger the n the bigger the uncertainty. The spreading of the wave packet that we observed is just what we should expect.

Multi-dimensional infinite potential wells, and finite potential wells

So far we have used only one dimensional examples.



But suppose instead of a group of wavicles traveling in only one dimension, we have particles that could travel in two dimensions. And suppose we have a two dimensional infinite potential well.



where

$$U(x, y) = \begin{cases} 0 & \text{if } 0 \leq x \leq L, 0 \leq y \leq L \\ \infty & \text{if otherwise} \end{cases}$$

We could have waves travel in both the x and the y-directions in the center allowed region. That would modify our Schrödinger equation

$$-\frac{\hbar^2}{2m} \left(\frac{d^2\psi(x, y)}{dx^2} + \left(\frac{d^2\psi(x, y)}{dy^2} \right) \right) + U(x, y) \psi(x, y) = E\psi(x, y)$$

We would have solutions like

$$\psi(x) = A \sin\left(\frac{n_x \pi x}{L}\right)$$

and we could have solutions like

$$\psi(y) = B \sin\left(\frac{n_y \pi y}{L}\right)$$

or we could even have functions that combine x and y dependence. If you have taken M316 you are no doubt saying “ah, this is a separable differential equation.” but some

of us haven't had that math yet, so I will wait a bit before we do a problem like that. For now I will just quote a solution

$$\psi(x, y) = C \sin\left(\frac{n_x \pi x}{L}\right) \sin\left(\frac{n_y \pi y}{L}\right)$$

But this isn't too strange, it is just a combination of the solutions for the two one-dimensional cases. We would need to get the coefficient C from normalization

$$\int \int |\psi(x, y)|^2 dx dy$$

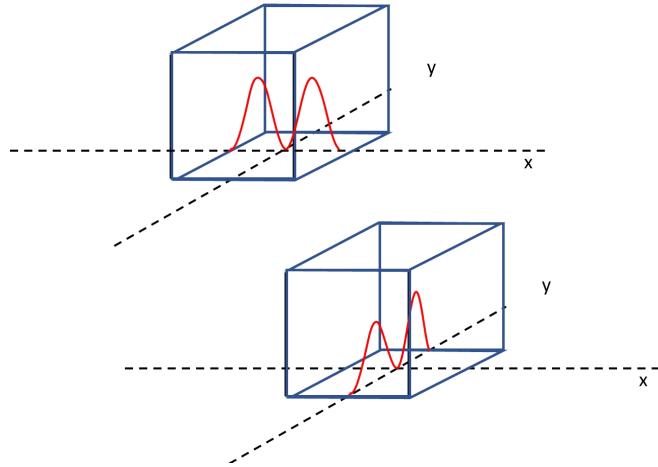
and if we did this problem we would find that

$$C = \frac{2}{L}$$

Note that we are now quantized in two directions, so we have two quantum numbers, n_x and n_y that are in our wave function. The energy for this case would be

$$E = \frac{\hbar^2 \pi^2}{2mL^2} (n_x^2 + n_y^2)$$

(a problem to look forward to in PH433). But think about what we have done for a moment. Since the square well is symmetric, we should expect that we could get the same energy for a wave that is all in the x -direction and for a wave going in the y -direction.



They would look just the same, so they should have the same energy. For example suppose $n_x = 1$ and $n_y = 2$, then

$$E = \frac{\hbar^2 \pi^2}{2mL^2} (1^2 + 2^2) = 5 \frac{\hbar^2 \pi^2}{2mL^2}$$

but now suppose $n_x = 2$ and $n_y = 1$

$$E = \frac{\hbar^2 \pi^2}{2mL^2} (2^2 + 1^2) = 5 \frac{\hbar^2 \pi^2}{2mL^2}$$

we get the same energy. There are two ways to get the same energy. But the two ways

would have different wave functions.

$$\psi_{12}(x, y) = C \sin\left(\frac{\pi x}{L}\right) \sin\left(\frac{2\pi y}{L}\right)$$

$$\psi_{21}(x, y) = C \sin\left(\frac{2\pi x}{L}\right) \sin\left(\frac{\pi y}{L}\right)$$

When this happens for a system we say that states that have the same energy for different wave functions are *degenerate*.

You might wonder if we could have a three-dimensional square well? And of course we could and you might guess that we would pick up an additional quantum number n_z and another term in the wave function that includes n_z . Our energies would change as well. And we could have more degenerate states. But we will take on that problem later in our course.

But what if the potential outside of our well isn't infinite? Well, we still get quantized wave functions. We could do this problem, and we will in three dimensions! well, almost. We will see what our Schrödinger equation would look like and we will use the solutions to those equations that early researchers found. We could do this problem here, but we need an additional insight. That will come from studying the quantum phenomena that we call *tunneling*. And we will do that in our next lecture. stuff.

14 Barriers, Oscillators, and Early Hints about the Structure of the Atom

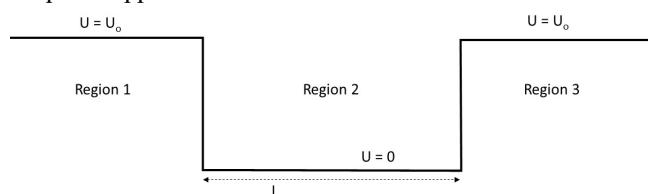
6.1-6.3

Fundamental Concepts in the Lecture

- Quantum Barriers
- Quantum Harmonic Oscillator Ground State
- Atoms are not homogeneous (Thompson's Model)
- Atoms have hard nuclei (Rutherford's Model)

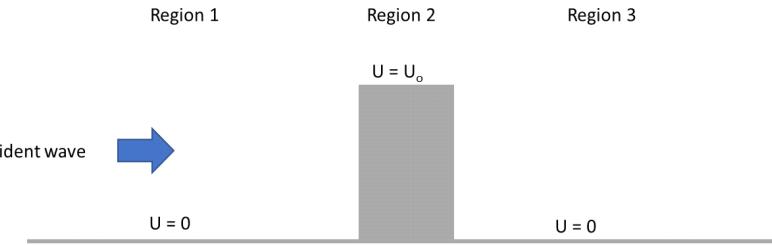
Barriers and Tunneling

In learning about wave functions and boundaries we did problems where the potential energy stepped up or stepped down.

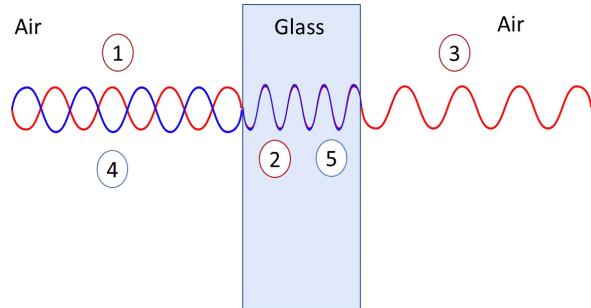


We entered the problems we did with quite a lot of information. That is because doing a finite potential well is quite difficult, enough so that the solution is in the American Journal of Physics. If you are an American Physical Society member you can download the paper. It is likely available from the Library as well. But if you want to wait, you will do this problem in our junior level quantum mechanics class (PH433). With some higher math, you can find the energy states and the wave functions.

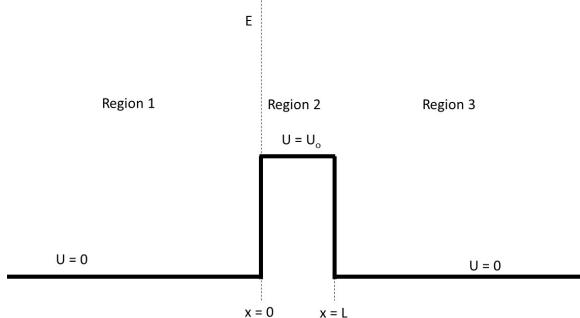
But with the help of knowing λ_1 and λ_2 we did the problem of matching the wave functions at the boundary. Let's take on another one of these, but let's change the situation so that in Regions 1 and 3 $U = 0$ and in Region 2 $E < U_o$. That would mean that the electrons or the wavicles could classically never get into Region 2, much less into Region 3. There is a barrier. Our schematic diagram might look like this.



We can see from our previous optics case that we should expect reflections.



The energy diagram would look like this.



We would expect the Schrödinger equation to apply

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U\psi(x) = E\psi(x)$$

where in Region 1

$$\begin{aligned}-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + (0)\psi(x) &= E_1\psi(x) \\ \frac{d^2\psi(x)}{dx^2} &= -\frac{2mE_1}{\hbar^2}\psi(x) \\ k_1^2 &= \frac{2mE_1}{\hbar^2}\end{aligned}$$

and in Region 2

$$\begin{aligned}-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U_o\psi(x) &= E\psi(x) \\ -\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} &= E\psi(x) - U_o\psi(x) \\ \frac{d^2\psi(x)}{dx^2} &= -\frac{2m}{\hbar^2}(E - U_o)\psi(x) \\ \frac{d^2\psi(x)}{dx^2} &= \frac{2m}{\hbar^2}(U_o - E)\psi(x) \\ \frac{d^2\psi(x)}{dx^2} &= \alpha^2\psi(x)\end{aligned}$$

and we can see that

$$\alpha^2 = \frac{2m}{\hbar^2}(U_o - E)$$

which is like our k values, but for an attenuation.

In Region 3 again we have

$$\begin{aligned}-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + (0)\psi(x) &= E_3\psi(x) \\ k_3^2 &= \frac{2mE_3}{\hbar^2}\end{aligned}$$

And from what we have done before we recognize that in Region 1 we will have a solution of the form

$$\psi_1(x, t) = Ae^{i(k_1 x - \omega t)} + Be^{-i(k_1 x + \omega t)}$$

which is really two waves, one going to the right and one reflected wave going to the left. We are free to choose A , the incident amplitude by choosing the settings on the electron gun that is generating the electrons. The amplitude A we will consider known.

In Region 2 the wave function will look like

$$\psi_2(x) = Ce^{\alpha x}e^{-i\omega t} + De^{-\alpha x}e^{-i\omega t}$$

which again includes a reflected term. We have to match ψ_1 and ψ_2 at the boundaries. In Region 3 we should have a solution like in Region 1, but we have already used the symbol E for energy so let's make the coefficients F and G

$$\psi_3(x, t) = Fe^{i(k_3 x - \omega t)} + Ge^{-i(k_3 x + \omega t)}$$

In Region 3 we don't expect a reflected wave going to the left. That tells us right away

that $G = 0$

$$\psi_3(x, t) = Fe^{i(k_3x - \omega t)}$$

and we know that with $U = 0$

$$k_3 = \sqrt{\frac{2mE}{\hbar^2}} = k_1$$

so let's call $k_3 = k_1 \equiv k$

$$\psi_3(x, t) = Fe^{i(kx - \omega t)}$$

We have to match the wave functions at the boundaries. Let's define some terms to help us think about the physics. The reflection coefficient tells us how much of the wave packet is reflected at $x = 0$.

$$R = \frac{|\psi_1(x, t)|_{reflected}^2}{|\psi_1(x, t)|_{incident}^2} = \frac{|Be^{-i(k_1x + \omega t)}|^2}{|Ae^{i(k_1x - \omega t)}|^2} = \frac{|B|^2}{|A|^2}$$

The transmission coefficient tells us how much of the wave packet gets into Region 2

$$T = \frac{|\psi_2(x, t)|_{transmitted}^2}{|\psi_1(x, t)|_{incident}^2} = \frac{|Ee^{i(k_3x - \omega t)}|^2}{|Ae^{i(k_1x - \omega t)}|^2} = \frac{|F|^2}{|A|^2}$$

and we must have conservation of energy so

$$R + T = 1$$

assuming no dissipation of the energy into thermal energy or other energy forms.

Before we solve for the coefficients, it is interesting (and important!) to look at

$$\psi_2(x) = Ce^{\alpha x}e^{-i\omega t} + De^{-\alpha x}e^{-i\omega t}$$

and

$$\alpha = \sqrt{\frac{2m}{\hbar^2}(U_o - E)}$$

We expect that the wave going to the right should diminish. That is, the particles are less likely to be detected the further we get into Region 2. We can look at

$$\psi_2(x) = De^{-\alpha x}$$

and see where it has diminished to $1/e$ of its original value. That would be when

$$x = \frac{1}{\alpha} = \frac{\hbar}{\sqrt{2m(U_o - E)}}$$

we usually give this particular depth a symbol

$$\delta = \frac{\hbar}{\sqrt{2m(U_o - E)}}$$

We will call this the “penetration depth.” Actually the wave goes a bit farther into the material. But this definition of “penetration depth” is a ball-park estimate for how far the wave goes into the forbidden region of the barrier (Region 2). And now for the astonishing part, if the thickness $L < \delta$, then the wave can get through! Classically the electron wouldn't have enough energy to escape, but a quantum electron is a waveicle

that can slosh over the barrier if the barrier is thin enough. We call this *quantum tunneling*.

But back to the slog-it-out-math. We need at $x = 0$

$$\begin{aligned}\psi_1(0, t) &= \psi_2(0, t) \\ Ae^{i(k(0)-\omega t)} + Be^{-i(k(0)+\omega t)} &= Ce^{\alpha(0)}e^{-i\omega t} + De^{-\alpha(0)}e^{-i\omega t}\end{aligned}$$

which becomes

$$A + B = C + D$$

and we need

$$\begin{aligned}\left(\frac{d\psi_1}{dx}\right|_{x=0} &= \left(\frac{d\psi_2}{dx}\right|_{x=0} \\ \left(k_1 Ae^{i(kx-\omega t)} - k_1 Be^{-i(kx+\omega t)}\right|_{x=0} &= (\alpha Ce^{\alpha x}e^{-i\omega t} + -\alpha De^{-\alpha x}e^{-i\omega t}|_{x=0} \\ ik_1 Ae^{i(k(0)-\omega t)} - ik_1 Be^{-i(k(0)+\omega t)} &= \alpha Ce^{\alpha(0)}e^{-i\omega t} + -\alpha De^{-\alpha(0)}e^{-i\omega t} \\ ikA - ikB &= \alpha C + -\alpha D\end{aligned}$$

now at $x = L$

$$\begin{aligned}\psi_2(L, t) &= \psi_3(L, t) \\ Ce^{\alpha L}e^{-i\omega t} + De^{-\alpha L}e^{-i\omega t} &= Ee^{i(k_1 L - \omega t)} \\ Ce^{\alpha L} + De^{-\alpha L} &= Fe^{ik_1 L}\end{aligned}$$

and finally

$$\begin{aligned}\left(\frac{d\psi_2}{dx}\right|_{x=L} &= \left(\frac{d\psi_3}{dx}\right|_{x=L} \\ (\alpha Ce^{\alpha x}e^{-i\omega t} + -\alpha De^{-\alpha x}e^{-i\omega t}|_{x=L} &= (kF e^{i(kx-\omega t)}|_{x=L} \\ \alpha Ce^{\alpha L}e^{-i\omega t} + -\alpha De^{-\alpha L}e^{-i\omega t} &= ikF e^{ik(kL - \omega t)} \\ \alpha Ce^{\alpha L}e^{-i\omega t} + -\alpha De^{-\alpha L}e^{-i\omega t} &= ikE e^{ikL}e^{-i\omega t} \\ \alpha Ce^{\alpha L} + -\alpha De^{-\alpha L} &= ikE e^{ikL}\end{aligned}$$

So we have four equations

$$\begin{aligned}A + B &= +C + D \\ ikA - ikB &= +\alpha C - \alpha D \\ 0 + Ce^{\alpha L} + De^{-\alpha L} &= +Fe^{ikL} \\ \alpha Ce^{\alpha L} + -\alpha De^{-\alpha L} &= ikF e^{ikL}\end{aligned}$$

with E , U_o , m , and A known and

$$\alpha = \sqrt{\frac{2m}{\hbar^2}(U_o - E)}$$

$$k = \sqrt{\frac{2mE}{\hbar^2}}$$

so really α and k are known. We also know L . If we have numbers, it is likely better at this point to have a computer solve the system of equations. Writing our equations with the A terms on the right of the equal sign

$$\begin{aligned} +B - C - D + 0 &= -A \\ ikA - ikB - \alpha C + \alpha D + 0 &= -ikA \\ 0 + Ce^{\alpha L} + De^{-\alpha L} - Ee^{ikL} &= 0 \\ 0 + \alpha Ce^{\alpha L} + -\alpha De^{-\alpha L} - ikEe^{ikL} &= 0 \end{aligned}$$

and we could put these into a matrix solver

$$\left[\begin{array}{cccc} 1 & -1 & -1 & 0 \\ -ik & -\alpha & \alpha & 0 \\ 0 & e^{\alpha L} & e^{-\alpha L} & -e^{ikL} \\ 0 & \alpha e^{\alpha L} & -\alpha e^{-\alpha L} & -ik e^{ikL} \end{array} \right] \left[\begin{array}{c} B \\ C \\ D \\ F \end{array} \right] = \left[\begin{array}{c} -A \\ -ikA \\ 0 \\ 0 \end{array} \right]$$

Suppose the waves are electrons so

$$m = 9.1093829 \times 10^{-31} \text{ kg}$$

and suppose we adjust our electron flow so that

$$A = 10$$

and suppose we have a copper conducting layer in a piece of microelectronics with an oxide layer that is about

$$L = 5 \text{ nm}$$

thick. Further suppose that the electrons have energy of

$$E = 7 \text{ eV}$$

and that the oxide layer is, to a good approximation, a square barrier with

$$U_o = 10 \text{ eV}$$

We also know

$$\begin{aligned} \hbar &= 1.054571726 \times 10^{-34} \text{ Js} \\ &= 4.13566752 \times 10^{-15} \text{ eV s} \end{aligned}$$

and then

$$\begin{aligned} k &= \sqrt{\frac{2mE}{\hbar^2}} \\ &= \sqrt{\frac{2(9.1093829 \times 10^{-31} \text{ kg})(7 \text{ eV})}{(1.054571726 \times 10^{-34} \text{ Js})^2}} \\ &= 1.3555 \times 10^{10} \frac{1}{\text{m}} \\ &= 13.555 \frac{1}{\text{nm}} \end{aligned}$$

and

$$\begin{aligned} \alpha &= \sqrt{\frac{2m}{\hbar^2}(U_o - E)} \\ &= \sqrt{\frac{2(9.1093829 \times 10^{-31} \text{ kg})}{(1.054571726 \times 10^{-34} \text{ Js})^2}(10 \text{ eV} - 7 \text{ eV})} \\ &= 8.8736 \frac{1}{\text{nm}} \end{aligned}$$

then our matrix

$$\begin{bmatrix} 1 & -1 & -1 & 0 \\ -ik & -\alpha & \alpha & 0 \\ 0 & e^{\alpha L} & e^{-\alpha L} & -e^{ikL} \\ 0 & \alpha e^{\alpha L} & -\alpha e^{-\alpha L} & -ike^{ikL} \end{bmatrix} \begin{bmatrix} B \\ C \\ D \\ F \end{bmatrix} = \begin{bmatrix} -A \\ -ikA \\ 0 \\ 0 \end{bmatrix}$$

becomes

$$\begin{bmatrix} 1 & -1 & -1 & 0 \\ -i\left(\frac{13.6}{\text{nm}}\right) & -8.8736\frac{1}{\text{nm}} & 8.8736\frac{1}{\text{nm}} & 0 \\ 0 & e^{44.368} & e^{-44.368} & -e^{i(67.775)} \\ 0 & \left(\frac{8.87}{\text{nm}}\right)e^{44.368} & -\left(\frac{8.87}{\text{nm}}\right)e^{-44.368} & -i\left(\frac{13.6}{\text{nm}}\right)e^{i(67.775)} \end{bmatrix} \begin{bmatrix} B \\ C \\ D \\ F \end{bmatrix} = \begin{bmatrix} -10 \\ -i\left(\frac{136}{\text{nm}}\right) \\ 0 \\ 0 \end{bmatrix}$$

after putting in the numbers. The answer is

$$\begin{bmatrix} B \\ C \\ D \\ F \end{bmatrix} = \begin{bmatrix} 4.0002 - 9.1651i \\ 8.1192 \times 10^{-39} + 4.7848 \times 10^{-38}i \\ 14.0 - 9.1651i \\ -1.7749 \times 10^{-19} + 9.7107 \times 10^{-19}i \end{bmatrix}$$

It is just fine that the amplitudes are complex because we switched from sines and cosines to complex exponentials. With the coefficients, we can find our wave functions, but we still need $\omega = 2\pi f$ for our time dependence. In regions 1 and 3 we know

$K = E - U = (7 \text{ eV} - 0) = 7 \text{ eV}$ and

$$K = \frac{1}{2}mv^2 = \frac{1}{2}m(\lambda f)^2$$

so

$$\frac{2K}{m} = (\lambda f)^2$$

$$\begin{aligned}
 \frac{1}{\lambda} \sqrt{\frac{2K}{m}} &= f \\
 \frac{2\pi}{\lambda} \sqrt{\frac{2K}{m}} &= 2\pi f \\
 k \sqrt{\frac{2K}{m}} &= \omega \\
 \omega &= \left(13.555 \frac{1}{\text{nm}} \right) \sqrt{\frac{2(7\text{eV})}{9.1093829 \times 10^{-31} \text{kg}}} \\
 &= 2.127 \times 10^{16} \frac{\text{rad}}{\text{s}}
 \end{aligned}$$

then our wave function $\psi_1(x, t)$ would be

$$\begin{aligned}
 \psi_1(x, t) &= 10e^{i(13.555 \frac{1}{\text{nm}} x - (2.127 \times 10^{16} \frac{\text{rad}}{\text{s}})t)} \\
 &\quad + (4.0002 - 9.1651i)e^{-i(13.555 \frac{1}{\text{nm}} x + (2.127 \times 10^{16} \frac{\text{rad}}{\text{s}})t)}
 \end{aligned}$$

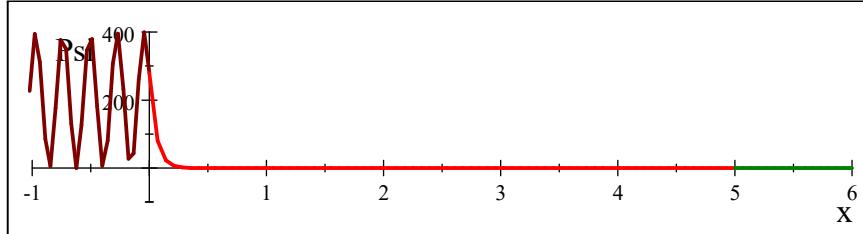
and in region 2

$$\begin{aligned}
 \psi_2(x, t) &= (8.1192 \times 10^{-39} + 4.7848 \times 10^{-38}i)e^{8.8736 \frac{1}{\text{nm}} x} e^{-i(2.127 \times 10^{16} \frac{\text{rad}}{\text{s}})t} \\
 &\quad + (14.0 - 9.1651i)e^{-8.8736 \frac{1}{\text{nm}} x} e^{-i(2.127 \times 10^{16} \frac{\text{rad}}{\text{s}})t}
 \end{aligned}$$

and in region 3

$$\psi_3(x, t) = (-1.7749 \times 10^{-19} + 9.7107 \times 10^{-19}i)e^{i(13.555 \frac{1}{\text{nm}} x - (2.127 \times 10^{16} \frac{\text{rad}}{\text{s}})t)}$$

Notice that each of these is still a wave. But the amplitudes and wavelengths are different. We can get some insight into how the waves move if we plot the wave function amplitudes squared ($|\psi|^2$) to tell us where we are likely to detect the electrons.

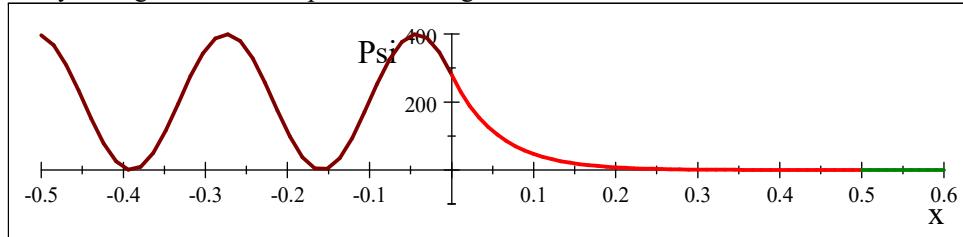


And what we see is instructive. On the left we expect constructive and destructive interference. And that is what we see. In Region 2 we expect to see fewer and fewer detections as we move away from the boundary. And that is what we see. In Region 3 we expect to see whatever particles that make it to be free. They will have a constant amplitude because they could be anywhere beyond the barrier.

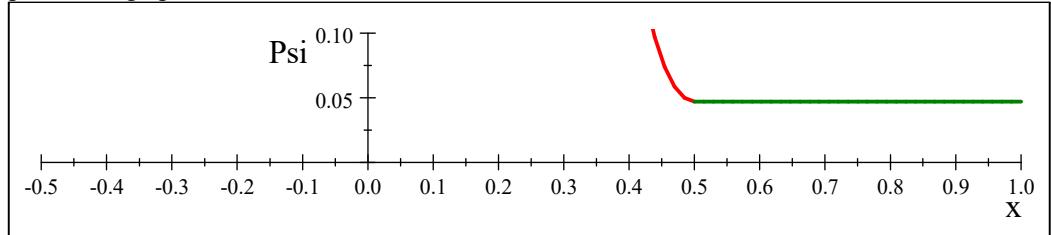
We could make the barrier narrower, say, $L = 0.5 \text{ nm}$. Then the amplitudes are

$$\begin{bmatrix} B \\ C \\ D \\ F \end{bmatrix} = \begin{bmatrix} 3.9983 - 9.1633i \\ 3.9223 \times 10^{-4} + 0.00231i \\ 13.998 - 9.1656i \\ 0.21615 - 1.7955 \times 10^{-2}i \end{bmatrix}$$

and we again see constructive and destructive interference in Region 1, exponential decay in Region 2, and free particles in Region 3.



You might wonder if the curves at $x = L$ really match (I did). So I zoomed in on that part of the graph.



Notice that amplitude after $x = 0.5$ is not zero. There is still a wave there. Some of the wavicles got through. We don't see that wave part because we are just plotting the amplitude. But there are waves leaving the barrier, so the electrons are getting through. Classically this wouldn't happen. We would just get reflection off of the barrier. But in the quantum picture we do get penetration of the barrier and we can even have wavicles escape. This phenomena is part of everyday electronics. Tunneling diodes use this quantum mechanical effect to work. So this is likely happening somewhere in your phone and computer right now.

We have used the Schrödinger equation to solve some interesting problems. But we are really working our way to building atoms. We will start this process in our next lecture, and spend a few lectures to come up with a quantum model for atomic structure.

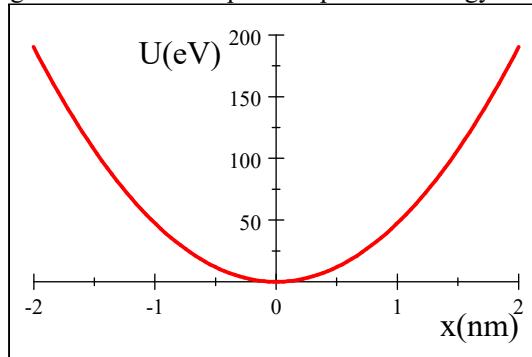
Harmonic Oscillator

Let's try an different kind of potential well. We remember a harmonic oscillator from

PH123. A mass on a spring is an example, but so is a diatomic molecule or even an electron that has been disturbed and oscillates around its equilibrium position. The potential energy for a non-relativistic harmonic oscillator is

$$U = \frac{1}{2}\kappa x^2$$

where κ is a “spring constant.” We can plot this potential energy



We can only calculate the lowest energy or *ground state* for a quantum harmonic oscillator until PH433. But we can do that much, so let's see how it goes. and we can use the Schrödinger equation

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U\psi(x) = E\psi(x)$$

and put in our potential energy

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + \left(\frac{1}{2}\kappa x^2\right)\psi(x) = E\psi(x)$$

We can guess a solution (here an expression that means someone figured it out and I am quoting the right answer from them).

$$\psi(x) = Ae^{-\alpha x^2}$$

This is not an obvious solution, and it is only a partial solution. The complete solution is

$$\psi(x) = N_n H_n(\alpha x) Ae^{-\alpha x^2}$$

where $H(x)$ is one of the Hermite polynomials and where N_n is an amplitude coefficient. You might say to your self that you have never heard of the Hermite polynomials, and for most people that would be true. You will see them again in PH433. But for now we need to know that

$$H_0 = 1$$

then for the ground state wave function we can write

$$\psi_0(x) = N_0 e^{-\alpha x^2}$$

and try it in our Schrödinger equation. First let's take a derivative or two

$$\frac{d\psi_0(x)}{dx} = -N_0 2x\alpha e^{-\alpha x^2}$$

and we have a product of two functions that have x in them

$$\frac{d\psi_0(x)}{dx} = -2N_0\alpha \left(xe^{-\alpha x^2} \right)$$

Then the second derivative is

$$\begin{aligned} \frac{d^2\psi_0(x)}{dx^2} &= -2N_0\alpha \left(-x\alpha 2xe^{-\alpha x^2} + e^{-\alpha x^2} \right) \\ &= 4\alpha^2 x^2 N_0 e^{-\alpha x^2} - 2N_0\alpha e^{-\alpha x^2} \\ &= (4\alpha^2 x^2 - 2\alpha) N_0 e^{-\alpha x^2} \end{aligned}$$

so our Schrödinger equation becomes

$$\begin{aligned} -\frac{\hbar^2}{2m} \left((4\alpha^2 x^2 - 2\alpha) N_0 e^{-\alpha x^2} \right) + \left(\frac{1}{2}\kappa x^2 \right) N_0 e^{-\alpha x^2} &= EN_0 e^{-\alpha x^2} \\ -\frac{\hbar^2}{2m} (4\alpha^2 x^2 - 2\alpha) + \left(\frac{1}{2}\kappa x^2 \right) - E &= 0 \\ -\frac{2\alpha^2 \hbar^2}{m} x^2 + \frac{\alpha \hbar^2}{m} + \frac{1}{2}\kappa x^2 - E &= 0 \\ \left(\frac{\alpha \hbar^2}{m} - E \right) + \left(\frac{1}{2}\kappa - \frac{2\alpha^2 \hbar^2}{m} \right) x^2 &= 0 \end{aligned}$$

We want this to be zero for all x . That means we need the two terms to become zero simultaneously. So

$$\left(\frac{\alpha \hbar^2}{m} - E \right) = 0 \quad (14.1)$$

and

$$\left(\frac{1}{2}\kappa - \frac{2\alpha^2 \hbar^2}{m} \right) x^2 = 0 \quad (14.2)$$

for all x , even if $x \neq 0$. This means that

$$\left(\frac{1}{2}\kappa - \frac{2\alpha^2 \hbar^2}{m} \right) = 0$$

The first equation (14.1) gives the condition

$$\frac{\alpha \hbar^2}{m} = E$$

and the second (14.2) gives

$$\begin{aligned} \frac{1}{2}\kappa k &= \frac{2\alpha^2 \hbar^2}{m} \\ \frac{1}{4\hbar^2} \kappa m &= \alpha^2 \\ \alpha &= \frac{\sqrt{\kappa m}}{2\hbar} \end{aligned}$$

using this we can write

$$\begin{aligned}\frac{\alpha\hbar^2}{m} &= E \\ \frac{\sqrt{\kappa m}\hbar^2}{2\hbar m} &= E \\ \frac{\sqrt{\kappa m}\hbar}{2m} &= E \\ \frac{\hbar}{2}\sqrt{\frac{\kappa}{m}} &= E\end{aligned}$$

so our energy for the ground state is

$$E_0 = \frac{\hbar}{2}\omega_o$$

and we can write our wave function as

$$\psi_0(x) = N_0 e^{-\left(\frac{\sqrt{\kappa m}}{2\hbar}\right)x^2}$$

To finish this off, we need to normalize $\psi_0(x)$

$$\int_{-\infty}^{\infty} N_0 e^{-\left(\frac{\sqrt{\kappa m}}{2\hbar}\right)x^2} dx = 1$$

We can't limit our integration to the classical range of the harmonic oscillator because we now know that our wave equation can penetrate forbidden zones.

$$\begin{aligned}\int_{-\infty}^{\infty} \left| N_0 e^{-\left(\frac{\sqrt{\kappa m}}{2\hbar}\right)x^2} \right|^2 dx &= 1 \\ N_0^2 \int_{-\infty}^{\infty} e^{-2\left(\frac{\sqrt{\kappa m}}{2\hbar}\right)x^2} dx &= 1 \\ N_0^2 \int_{-\infty}^{\infty} e^{-\left(\frac{\sqrt{\kappa m}}{\hbar}\right)x^2} dx &= 1\end{aligned}$$

We need an integral table again.

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}$$

so we can write

$$\begin{aligned}a &= \frac{\sqrt{\kappa m}}{\hbar} \\ &= \frac{\sqrt{m}\sqrt{m}\sqrt{\kappa}}{\sqrt{m}\hbar} \\ &= \frac{m\sqrt{\kappa}}{\hbar\sqrt{m}} \\ &= \frac{m\omega_o}{\hbar}\end{aligned}$$

so

$$\begin{aligned}N_0^2 \int_{-\infty}^{\infty} e^{-\left(\frac{\sqrt{\kappa m}}{\hbar}\right)x^2} dx &= N_0^2 \sqrt{\frac{\pi}{\frac{m\omega_o}{\hbar}}} = 1 \\ N_0^2 \sqrt{\frac{\pi\hbar}{m\omega_o}} &= 1\end{aligned}$$

$$N_0^2 = \sqrt{\frac{m\omega_o}{\pi\hbar}}$$

$$N_0 = \left(\frac{m\omega_o}{\hbar\pi}\right)^{\frac{1}{4}}$$

so our wave function is given by

$$\psi_0(x) = \left(\frac{m\omega_o}{\hbar\pi}\right)^{\frac{1}{4}} e^{-\left(\frac{m\omega_o}{2\hbar}\right)x^2}$$

but this is only the ground state. We would have to do this all again for each excited state. In PH433 we will show you a trick to avoid the nasty math involved in finding all the other states. The Hermite polynomials have a generating function just for doing this. But for now we will stick to the ground state wave function.

We found for the ground state

$$E_0 = \frac{\hbar}{2}\omega_o$$

but this is a special case of the more general

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega_o$$

We can plot our ground state function. Here it is again.

$$\psi_0(x) = \left(\frac{m\omega_o}{\hbar\pi}\right)^{\frac{1}{4}} e^{-\left(\frac{m\omega_o}{2\hbar}\right)x^2}$$

We will need some numbers to make the plot. Suppose we have an electron bound to an atom. It might oscillate around its equilibrium position when an electromagnetic field wave goes by. Suppose it has $\kappa = 95 \text{ eV/nm}^2$ and we know that $m = 0.511 \text{ MeV}/c^2$ then

$$a = \frac{\sqrt{\kappa m}}{\hbar} = \frac{\sqrt{\kappa mc^2}}{\hbar c} = \frac{\sqrt{(95 \text{ eV/nm}^2)(0.511 \text{ MeV})}}{197.3 \text{ eV nm}}$$

$$= \frac{35.314}{\text{nm}^2}$$

so then our wave function would be

$$\psi_0(x) = \left(\frac{35.314}{\pi \text{ nm}^2}\right)^{\frac{1}{4}} e^{-\left(\frac{35.314}{2 \text{ nm}^2}\right)x^2}$$

and the energy would be

$$E_0 = \frac{\hbar}{2} \sqrt{\frac{\kappa}{m}}$$

$$= \frac{\hbar c}{2} \sqrt{\frac{\kappa}{mc^2}}$$

$$= \frac{197.3 \text{ eV nm}}{2} \sqrt{\frac{(95 \text{ eV/nm}^2)}{(0.511 \text{ MeV})}}$$

$$= 1.345 \text{ eV}$$

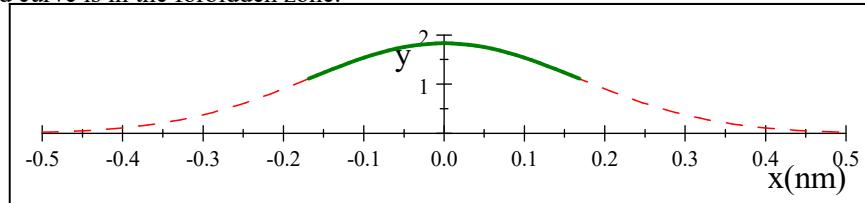
Classically if we had $U = E_o$ then

$$U = \frac{1}{2} \kappa x_{\max}^2$$

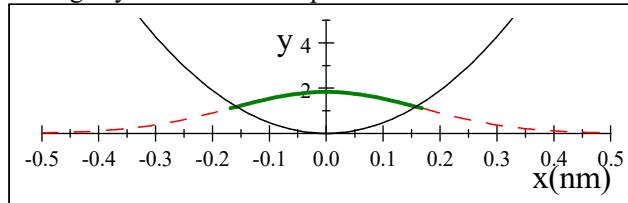
gives

$$\begin{aligned} x_{\max} &= \sqrt{\frac{2E_o}{k}} \\ x_{\max} &= \sqrt{\frac{2(1.3451 \text{ eV})}{(95 \text{ eV/nm}^2)}} \\ &= 0.16828 \text{ nm} \end{aligned}$$

so in our plot the solid green part is in the classical allowed region and the dashed red curve is in the forbidden zone.



If we add in the potential well, we can see that indeed the dashed red part of the wave function is extending beyond the classical potential well.

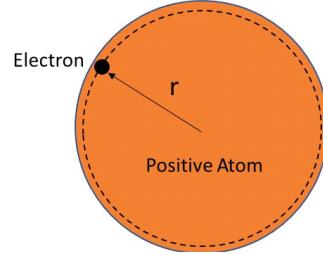


It might not be satisfying to not be able to do the higher states, but we have to leave something for PH433. We are going to leave potential wells for a while, but they will come back. Atoms are essentially electrical potential energy wells, and nuclei are potential energy wells for nuclear forces. So more potential wells to come.

Ancient thoughts on atomic structure

In Newton's day, atoms were thought to be structureless little balls. Particle model worked for the experiments Newton and his generation did. Our ideal gas law uses this assumption. But once it was discovered that there were charged parts of matter, Newton's atomic theory could no longer explain the new experiments. Atoms were

electrically neutral. But if photo-electric effect or Compton scattering knocks out an electron, the atom is left positively charged and a negative electron is lost. These seem like different parts that made up the original atom. Thompson proposed a “raisin pudding” theory of the atom that had electrons and positive charge carriers mixed uniformly together. This could explain how you could knock out an electron. The atom was made of parts, but the parts were all mixed up.



Let's look at one way the Thompson model fails. If we assume a net positive atom with charge Zq_e where q_e is an electron charge and we assume an electron with charge $-q_e$ then, using Gauss's law (just for review) we would have a field due to the atom at the electron's location of

$$\begin{aligned}\int \mathbf{E} \cdot d\mathbf{a} &= \frac{Q_{inside}}{\epsilon_0} \\ E \int d\mathbf{a} &= \frac{Q_{inside}}{\epsilon_0} \\ E4\pi r^2 &= \frac{Q_{inside}}{\epsilon_0} \\ E &= \frac{Zq_e}{4\pi\epsilon_0 r^2}\end{aligned}$$

where we have let the electron be just on the outer rim so very nearly all the Zq_e positive charge is inside the Gaussian surface. Then the electron would feel a force

$$\begin{aligned}\vec{F} &= q_e \vec{E} \\ &= -\frac{Zq_e^2}{4\pi\epsilon_0 r^3} \vec{r}\end{aligned}$$

This force is holding the electron in place in the Thompson model. but notice that it looks like a restoring force. We could model this like a spring force with

$$F = -kr$$

with

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$$

This means we should see an oscillation of the electron about an equilibrium position at

a particular frequency

$$\begin{aligned} f &= \frac{1}{2\pi} \sqrt{\frac{k}{m}} \\ &= \frac{1}{2\pi} \sqrt{\frac{Zq_e^2}{4\pi\epsilon_0 r^3 m}} \end{aligned}$$

and because moving electrons cause electromagnetic waves with the same frequency as their oscillation, such a Thompson model atom should radiate at this frequency. We can test this for hydrogen where $Z = 1$ and $r \approx 0.53 \text{ nm}$.

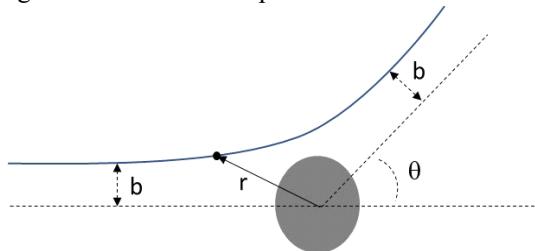
$$\begin{aligned} f &= \frac{1}{2\pi} \sqrt{\frac{(1)(1.602 \times 10^{-19} \text{ C})^2}{4\pi(8.85 \times 10^{-12} \text{ F m}^{-1})(0.53 \text{ nm})^3 (9.11 \times 10^{-31} \text{ kg})}} \\ &= 2.076 \times 10^{14} \text{ Hz} \end{aligned}$$

which gives a wavelength of

$$\begin{aligned} \lambda &= \frac{c}{f} = \frac{3 \times 10^8 \frac{\text{m}}{\text{s}}}{2.076 \times 10^{14} \text{ Hz}} \\ &= 1.4451 \times 10^{-6} \text{ m} \\ &= 1.4451 \mu\text{m} \end{aligned}$$

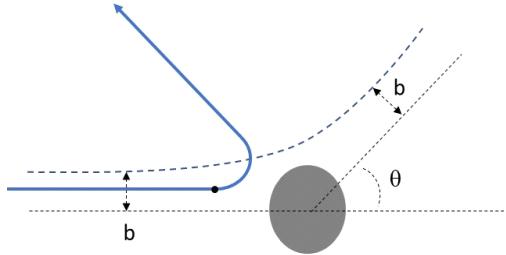
Strangely, Hydrogen has several frequencies that are produced (we will understand that soon) the first of which is about $91 \text{ nm} = 0.091 \mu\text{m}$. There are wavelengths in the micron region at $\lambda = 1.87 \mu\text{m}$, $1.2814 \mu\text{m}$, $1.0935 \mu\text{m}$, $1.005 \mu\text{m}$ and others, but we see that our particular wavelength is not in the list, and our analysis would not predict a list.

Rutherford tested the Thompson model by shooting positively charged α -particles at a thin sheet of gold foil. To his astonishment, some of the α -particles went through the gold as though there were nothing there. Others were reflected back. Rutherford interpreted this to mean that matter must be full of holes. He viewed the atom like a small planetary system. He postulated a nucleus (the “Sun” of the little planetary system) with electrons flying around in orbits like planets.



He could consider each electron hitting the nucleus if it came in a distance from the nucleus about the size of the nuclear radius. The distance, of the approach of the α -particle from the axis of the nucleus is called the *impact parameter* and the apparent circular

area of the nucleus is called the *nuclear cross section*. A particular impact parameter, b , we get a If an α -particle comes in at less than the distance b from the axis, the particle will scatter with an angle that is larger than θ .



We know now that accelerating charges will radiate (remember our antenna). An orbiting electron is an accelerating charge (not speeding up, but changing direction). That would mean that all atoms would constantly emit radiation-a result not backed up by our experience. So from Maxwell's electromagnetic wave theory of light, we have a disaster in our theory of the atom. There must be more to it (which most of you already know from your chemistry courses).

15 The Bohr Model of the Atom

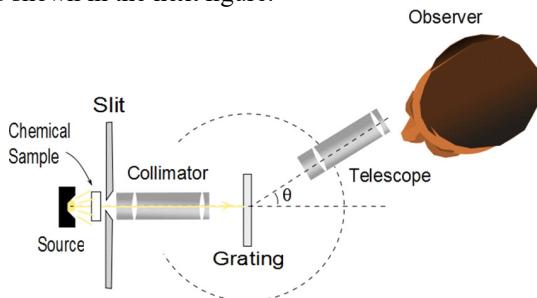
6.4-6.7

Fundamental Concepts in the Lecture

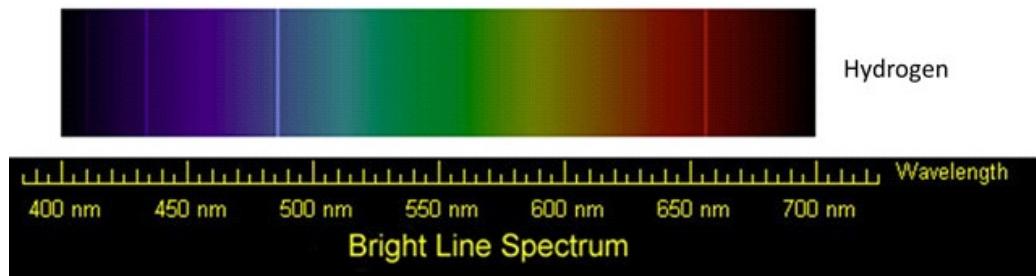
- Atomic spectra
- Bohr Model of the Atom

Atomic Spectra give a clue

We all know that we can make rainbows using a prism or a diffraction grating. If we do this in an organized way, our device is called a spectrometer. A very simple version of a spectrometer is shown in the next figure.



For a long time it has been known that if you take a gas and electrically excite it, you get not only a blackbody radiation spectrum from the gas (and the apparatus, and the lights in the room, etc.) but also some specific colors of light at some specific angles (an therefore at some wavelengths). You get individual bright lines in the spectrum. Let's take Hydrogen as an example. As far back as 1885, Balmer was studying these lines of emission from Hydrogen.

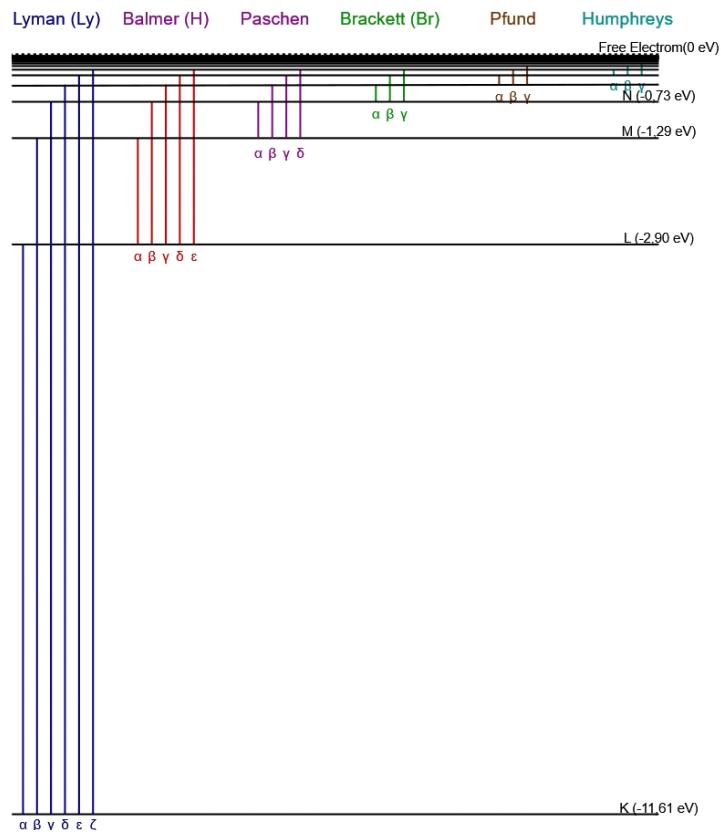


You can see the broad rainbow that comes from the black body emission, but in that emission you can see some bright lines (a red, and blue are prominent in this image, and there is a fainter violet line visible). He even found a formula for calculating what the wavelengths would be for these lines for Hydrogen gas

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right) \quad n = 3, 4, 5, \dots \quad (15.1)$$

The quantity $R_H = 1.0973732 \times 10^7 \frac{1}{\text{m}}$ is called the *Rydberg constant*.

Balmer had limited equipment and he found only some of the emission lines of Hydrogen. Most are not visible to our eyes. Here is a chart with more of the hydrogen spectrum.



It's an oddly designed chart that will make more sense as we go. But from our study of a particle in a box, we can recognize discrete energy levels and guess that they may represent standing waves in the wave functions of something.

For these other lines beyond the ones Balmer found we can modify the Balmer equation

$$\lambda = \lambda_{\text{limit}} \frac{n^2}{n^2 - n_o^2}$$

where for each case the n values start at a particular n_o and increase

$$n = n_o + 1, n_o + 2, n_o + 3, \dots$$

The particular form of this equation is not so important right now as is the fact that only certain values are possible for the wavelength. Think of wave functions in potential wells. We found only certain values for the wavelength created stable wave functions. Could there be something like a standing wave in the atom that is allowing absorption or emission at specific wavelengths?

To be complete, we need to state that the emission spectra do have the lines that we

have seen, but the spectrometer picture above actually is shining light through the gas of atoms. In this case the continuum spectrum from thermal emission has holes in it.



we get dark lines where the emission spectrum has bright lines. This is called an *absorption spectrum* and the bright lines version is called an *emission spectrum*. Some of the dark lines in the absorption spectrum are in the same places as the bright lines in the emission spectrum for a particular gas—but not all. A successful model for atoms would have to explain this.

These spectral lines were known about for a long time and people looked for patterns to try to figure out what was causing them. Two patterns emerged.

Bohr Misses the Clue, but hits on quantization

Niels Bohr went beyond Rutherford in forming a theory of the atom. He didn't get everything right, but he did get some things right and his model is still useful in some ways. This is a little bit like Galilean relativity is still useful, but we know it is not a complete theory. We will study Bohr's model in some detail. To explain what Rutherford had found, and to explain the atomic spectra, Bohr made a few assumptions to make his theory work. For hydrogen, they are as follows:

1. The electron moves in circular orbits around the proton under the influence of the Coulomb force of attraction
 - a. The Coulomb force produces the centripetal acceleration (this is not surprising to us)
2. Only certain electron orbits are stable (this is a little surprising, this is one of his big assumptions)
 - a. These are the orbits in which the atom does not emit energy in the form of electromagnetic radiation
 - b. Therefore, the energy of the atom remains constant and classical mechanics can be used to describe the electron's motion
3. Radiation is emitted by the atom when the electron “jumps” from a more energetic initial state to a lower state (another surprising assumption)
 - a. The “jump” cannot be treated classically
 - b. The frequency emitted in the “jump” is related to the change in the atom’s energy
 - c. It is generally not the same as the frequency of the electron’s orbital motion

- d. The frequency is given by $E_i - E_f = hf$
4. The size of the allowed electron orbits is determined by a condition imposed on the electron's orbital angular momentum

The last assumption takes some explaining. Bohr assumed that the angular momentum of the electron was quantized! That is, it can only have values that are integer multiples of a constant, in this case, \hbar .

Let's review angular momentum. It is

$$L = I\omega \quad (15.2)$$

where I is the moment of inertia, and ω is the angular speed. For a point mass rotating about a point r away we have

$$I = m_e r^2 \quad (15.3)$$

and we need to remember

$$\omega r = v_t \quad (15.4)$$

where v_t is the tangential speed. Then we can find the angular momentum of the electron as it orbits

$$L = I\omega = m_e r^2 \frac{v_t}{r} = m_e v_t r \quad (15.5)$$

so assumption 4 tells us that

$$m_e v_t r = n\hbar \quad n = 1, 2, 3, \dots \quad (15.6)$$

Now Bohr had no idea why the angular momentum was quantized. But he knew that if he made this assumption, he was able to explain the Hydrogen absorption and emission spectra.

Let's use this idea of quantized angular momentum to do a problem. Let's find how much energy is in the hydrogen proton-electron system. We will want to find the potential energy of the electron's orbit (think of a satellite orbiting the Earth) and the electron's kinetic energy.

The potential energy of the electron would be

$$U = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r} = -k \frac{1}{4\pi\epsilon_0} \frac{e^2}{r} \quad (15.7)$$

and the kinetic energy of the electron is

$$K = \frac{1}{2} m_e v^2 \quad (15.8)$$

so the total energy of an atom at rest is

$$E = \frac{1}{2}m_e v^2 - \frac{1}{4\pi\epsilon_0} \frac{e^2}{r} \quad (15.9)$$

We can use Newton's second law to find the kinetic energy . The sum of the forces is

$$\Sigma F = m_e a = m_e \frac{v^2}{r} \quad (15.10)$$

but there is only one force, and this is the Coulomb force so

$$m_e \frac{v^2}{r} = \frac{1}{4\pi\epsilon_0} \frac{e^2}{r^2} \quad (15.11)$$

so

$$(r) m_e \frac{v^2}{r} = k_e \frac{e^2}{r^2} (r) \quad (15.12)$$

$$m_e v^2 = k_e \frac{e^2}{r} \quad (15.13)$$

$$\left(\frac{1}{2}\right) m_e v^2 = k_e \frac{e^2}{r} \left(\frac{1}{2}\right) \quad (15.14)$$

$$K = \frac{1}{4\pi\epsilon_0} \frac{e^2}{2r} \quad (15.15)$$

then the total energy of the atom is

$$E = \frac{1}{4\pi\epsilon_0} \frac{e^2}{2r} - \frac{1}{4\pi\epsilon_0} \frac{e^2}{r} \quad (15.16)$$

$$= -\frac{1}{4\pi\epsilon_0} \frac{e^2}{2r} \quad (15.17)$$

The energy is negative, meaning that the electron would have to gain energy to escape the atom. The electron is bound to the proton.

Bohr wanted to find the stable orbits of the electrons. Let's take on this problem too.

He started with

$$L = m_e v_t r = n\hbar \quad (15.18)$$

and solved for v_t

$$v_t = \frac{n\hbar}{m_e r} \quad (15.19)$$

and squaring gives

$$v_t^2 = \left(\frac{n\hbar}{m_e r}\right)^2 \quad (15.20)$$

We found before that the centripetal force was just the Coulomb force. For a circular orbit $v = v_t$

$$F_c = F_E \quad (15.21)$$

$$m_e \frac{v_t^2}{r} = \frac{1}{4\pi\epsilon_0} \frac{e^2}{r^2} \quad (15.22)$$

let's solve this for v_t^2

$$v_t^2 = \frac{1}{4\pi\epsilon_o} \frac{e^2}{rm_e} \quad (15.23)$$

We have two different expressions for v_t^2 . They must be equal, so

$$\left(\frac{n\hbar}{m_e r} \right)^2 = \frac{1}{4\pi\epsilon_o} \frac{e^2}{rm_e} \quad (15.24)$$

We can solve this for r

$$\frac{n^2\hbar^2}{m_e^2 r^2} = \frac{k_e e^2}{rm_e} \quad (15.25)$$

$$\frac{n^2\hbar^2}{m_e r} = \frac{1}{4\pi\epsilon_o} \frac{e^2}{1} \quad (15.26)$$

$$\frac{4\pi\epsilon_o n^2 \hbar^2}{e^2 m_e} = r \quad (15.27)$$

which gives a quantized orbit radius! This is just what Bohr assumed we would find!

He assumed only certain orbits were stable. We call the smallest radius (with $n = 1$) the *Bohr radius*.

$$r_1 = a_o = \frac{4\pi\epsilon_o \hbar^2}{e^2 m_e} = 0.0529 \times 10^{-9} \text{ m} \quad (15.28)$$

then the others are just multiples

$$r = n^2 a_o \quad n = 1, 2, 3, \dots \quad (15.29)$$

This should really be reminding us of allowed wave functions in potential wells.

Now that we know the radius for each orbit, let's find the energy of the orbits using our energy equation

$$E = -\frac{1}{4\pi\epsilon_o} \frac{e^2}{2r}$$

and put in our result for the radius.

$$E = -\frac{1}{4\pi\epsilon_o} \frac{e^2}{2n^2 a_o} \quad (15.30)$$

$$= -\frac{1}{4\pi\epsilon_o} \frac{e^2}{2n^2 \frac{1}{4\pi\epsilon_o} \frac{\hbar^2}{e^2 m_e}} \quad n = 1, 2, 3, \dots \quad (15.31)$$

$$= -\frac{1}{4^2 \pi^2 \epsilon_o^2} \frac{e^4 m_e}{2n^2 \hbar^2} \quad n = 1, 2, 3, \dots \quad (15.32)$$

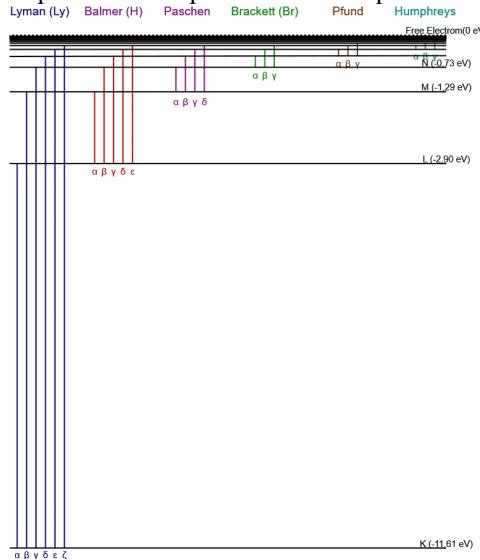
$$= -\frac{1}{32\pi^2 \epsilon_o^2 \hbar^2} \frac{m_e e^4}{n^2} \quad n = 1, 2, 3, \dots \quad (15.33)$$

and we see that the energy is quantized. For hydrogen, we can substitute in values to get

$$E = -\frac{13.6}{n^2} \text{ eV} \quad n = 1, 2, 3, \dots \quad (15.34)$$

When $n = 1$ we have the lowest energy state, called the *ground state*. It has energy $E_1 = -13.6 \text{ eV}$. The next state is $E_2 = -13.6 \text{ eV}/4 = -3.4 \text{ eV}$. We can represent these states graphically by picturing them like a building. We will draw a line

representing the ground floor or ground state. Above it will be the second floor (second energy state) and the third floor (third energy state) etc. Only, for the atom the floors are all different sizes. And this is a picture type we already know from our study of wave functions in the infinite potential well problem and the quantum harmonic oscillator.



The horizontal lines are the energy states, the vertical arrows are the transitions that take an electron from one state to another. They represent absorption or emission of photons. the photons would have the energy represented by the difference of energies of the two states participating in the transition. Bohr's model correctly predicts the emission lines for hydrogen with wavelengths that are very close to the observed wavelengths.

Notice that for large values of n we can get almost no energy between states. They are really continuous, no longer so quantized. The top line of the graph shows the energy of an electron infinitely far away at rest (no energy).

Suppose we want to get rid of the electron all together. This is called *ionization*. The energy required to remove the electron is 13.6 eV.

Lets look at what happens when an electron goes from one stable state to another. We call this a *transition*. Say our electron jumps from the state with the *quantum number* $n = 3$ to $n = 1$, the ground state. There is an energy difference of

$$E_3 - E_1 = -1.512 \text{ eV} - (-13.6 \text{ eV}) = 12.088 \text{ eV} \quad (15.35)$$

Then the energy of the produced photon would be

$$hf = E_3 - E_1 = 12.088 \text{ eV}$$

and the frequency of the photon would be

$$f = \frac{E_3 - E_1}{h} = \frac{12.088 \text{ eV}}{6.63 \times 10^{-34} \text{ J s}} = 2.92113417 \times 10^{15} \text{ Hz} \quad (15.36)$$

In general

$$f = \frac{E_f - E_i}{h} \quad (15.37)$$

You may wonder why this theory had so much success, since most of the assumptions were unfounded when Bohr made them. Here is an example of what the theory could do. Let's find a formula for the emission spectra, based on Bohr's theory. We want to find the inverse of the wavelength

$$\frac{1}{\lambda} \quad (15.38)$$

but we know that

$$\frac{1}{\lambda} = \frac{f}{c} = \frac{E_f - E_i}{ch} \quad (15.39)$$

or, using the Bohr theory expression for energy

$$\frac{1}{\lambda} = \frac{-\frac{1}{32\pi^2\epsilon_o^2\hbar^2}\frac{m_e e^4}{n_f^2} - \left(-\frac{1}{32\pi^2\epsilon_o^2\hbar^2}\frac{m_e e^4}{n_i^2}\right)}{ch} \quad (15.40)$$

Taking out everything that is not an n gives

$$\frac{1}{\lambda} = \frac{m_e e^4}{32\pi^2\epsilon_o^2\hbar^2(ch)} \left(\frac{1}{n_i^2} - \frac{1}{n_f^2} \right) \quad (15.41)$$

$$= \frac{m_e e^4}{32\pi^2\epsilon_o^2\hbar^2(h\frac{2\pi}{2\pi})c} \left(\frac{1}{n_i^2} - \frac{1}{n_f^2} \right) \quad (15.42)$$

$$= \frac{m_e e^4}{(64\pi^3)\epsilon_o^2(\hbar^3)c} \left(\frac{1}{n_i^2} - \frac{1}{n_f^2} \right) \quad (15.43)$$

This looks a little familiar. If $n_i^2 = 2^2$ and if

$$R_H = \frac{m_e e^4}{64\pi^3\epsilon_o^2\hbar^3c} \quad (15.44)$$

$$= 1.0973731534 \times 10^7 \text{ m}^{-1} \quad (15.45)$$

then Bohr has just derived the equation for the quantized wavelengths associated with the Hydrogen spectrum. When you do the calculation, the value Bohr got for R_H matched very well with what Balmer found experimentally. Bohr knew he was on to something. Bohr's model reproduces the strange numeric relationships that were observed between the frequencies of the emission lines!

If we allow for atoms that have more than one electron, then the Bohr model can also explain why some of the emission lines don't show up as absorption lines. An electron may wish to jump to a new state, but if there is already an electron in that state, we can see that the transition can't happen. So far so good for Bohr!

Bohr's Correspondence Principle

Bohr stated something that is obvious in a way, but it needed to be stated. Quantum mechanics is in agreement with classical physics when the energy differences between quantized levels are very small. In other words, if we go outside the atom (n is very large) then the electron should act just like as previous experiments had shown. That is, the electron should still look particle-ish in experiments that had previously succeeded by viewing the electron as a particle. We can see this works in the Bohr model for high n . This will have to be part of our quantum model for the atom.

Extension to “Hydrogen-Like” Atoms

The Bohr theory really only explained Hydrogen. For a scientific theory to be of practical value, it must predict behavior beyond that which it was designed to explain. So the fact that Bohr correctly found the relationship between the Balmer lines was good, but not good enough. But the Bohr model was also able to explain a second group of lines for Hydrogen, the Lyman series. And for Hydrogen, the successes kept piling up.

But more impressively, the Bohr model was extended to atoms that were “like” hydrogen. That is, when the atom only has one electron. To get atoms that only have one electron, we have to remove electrons. Then we would have ionized atoms. To use the Bohr theory for these atoms, we just look for the electron charge e and substitute in Ze^2 wherever there was just e^2 for the nucleus (the proton for Hydrogen). The symbol Z is the atomic number of the element (which is the number of protons). Let’s take an example,

$$\begin{aligned} F &= \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \\ &= \frac{1}{4\pi\epsilon_0} \frac{(Ze)(e)}{r^2} \end{aligned}$$

but this time instead of just e for a proton we have Ze for the nucleus charge. And this force must once again be due to a centripetal acceleration times the electron mass

$$m_e \frac{v_t^2}{r} = \frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r^2}$$

so once again

$$\frac{1}{2} m_e v_t^2 = \frac{1}{8\pi\epsilon_0} \frac{Ze^2}{r^3}$$

so our atomic number shows up in our kinetic energy term

$$K = \frac{1}{8\pi\epsilon_0} \frac{Ze^2}{r}$$

and the total energy with the change to Ze as the nuclear charge becomes

$$E = \frac{1}{2}m_e v^2 - \frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r}$$

And with our expression for K from the force considerations

$$\begin{aligned} E &= \frac{1}{8\pi\epsilon_0} \frac{Ze^2}{r} - \frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r} \\ &= \frac{1}{8\pi\epsilon_0} \frac{Ze^2}{r} - \frac{2}{8\pi\epsilon_0} \frac{Ze^2}{r} \\ &= -\frac{1}{8\pi\epsilon_0} \frac{Ze^2}{r} \end{aligned}$$

and once again use the assumption of angular momentum quantization

$$m_e v_t r = n\hbar \quad (15.46)$$

and solved for v

$$v_t = \frac{n\hbar}{m_e r} \quad (15.47)$$

and squaring

$$v_t^2 = \left(\frac{n\hbar}{m_e r} \right)^2 \quad (15.48)$$

so that

$$K = \frac{1}{2}m_e v_t^2 = \frac{1}{2}m_e \left(\frac{n\hbar}{m_e r} \right)^2$$

and this must be equal to our previous expression for K

$$\frac{1}{2}m_e \left(\frac{n\hbar}{m_e r} \right)^2 = \frac{1}{8\pi\epsilon_0} \frac{Ze^2}{r}$$

and we can solve this for the stable orbit distance, r

$$\begin{aligned} \frac{1}{2}m_e \frac{n^2\hbar^2}{m_e^2 r^2} &= \frac{1}{8\pi\epsilon_0} \frac{Ze^2}{r} \\ \frac{1}{2} \frac{n^2\hbar^2}{m_e r} &= \frac{1}{8\pi\epsilon_0} Ze^2 \\ \frac{8\pi\epsilon_0}{2Ze^2} \frac{n^2\hbar^2}{m_e} &= r \\ r &= \frac{4\pi\epsilon_0\hbar^2}{Ze^2 m_e} n^2 \end{aligned}$$

and we see the only effect will be that the orbits move in because the Coulomb force has become stronger with the additional nuclear charge.

The energy for the orbit will be

$$\begin{aligned} E &= -\frac{1}{8\pi\epsilon_0} \frac{Ze^2}{\frac{4\pi\epsilon_0\hbar^2}{Ze^2m_e}n^2} \\ &= -\frac{Z^2e^4m_e}{32\pi^2\epsilon_0^2\hbar^2} \frac{1}{n^2} \end{aligned}$$

and we can compare this to the hydrogen case

$$E_n = -\frac{1}{4\pi\epsilon_0} \frac{m_e e^4}{2\hbar^2} \left(\frac{1}{n^2} \right) \quad \text{Hydrogen} \quad (15.49)$$

becomes

$$E_n = -\frac{1}{4\pi\epsilon_0} \frac{m_e Z^2 e^4}{2\hbar^2} \left(\frac{1}{n^2} \right) \quad \text{Atom with atomic number } Z \quad (15.50)$$

This all worked very well.

But when we try to do this for elements with more than one electron, it does not go so well. Something was missing

Bohr Model - What doesn't work

There was an initial difficulty that the Bohr model assumed the hydrogen atom was in vacuum, but the emission lines were measured with the apparatus in air. This was a difficulty easily solved once the mistake was understood. But that wasn't the end of troubles for the Bohr model.

We assumed in deriving the Bohr model that the proton never moved. But real objects don't act this way. The Sun, for example, moves because the planets orbit around them. The Earth as well wiggles in its orbit of the Sun because the Moon orbits around it but really the Moon and Earth both orbit the center of mass of the Moon-Earth system. We can fix this problem in the Bohr model somewhat artificially by using the “reduced mass” of

$$m = \frac{m_e m_p}{m_e + m_p}$$

in the calculations instead of m_e . (We will explain why this works in detail later in our course.)

But a more serious challenge came when the angular momentum of the ground state electron was experimentally measured to be $L = 0$. A founding assumption for Bohr was that

$$L = m_e v r = n\hbar \quad (15.51)$$

and neither n nor \hbar could be zero. Something was clearly wrong.

We can now add on a problem from the uncertainty principle, that a planar orbit, say, in the xy plane would leave $v_x = 0$ and $z = 0$ so $\Delta p = 0$ and $\Delta z = 0$ and we know that can't be for an electron wavicle. Bohr did his work before Heisenberg, so this wasn't discovered as a problem until later. But we, knowing how to use uncertainty relations for wavicles, can see right away that this can't work.

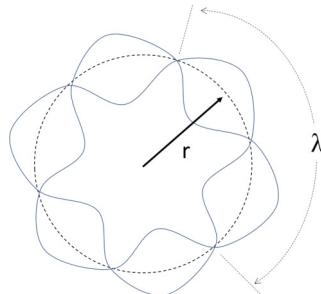
De Broigle Discovers the Clue

Remember that Bohr told us that the angular momentum of the electron was quantized,

$$m_e v_t r = n\hbar$$

but why?

From our work with De Broigle waves, we might have an idea. We now know that electron waves act in some ways like waves on strings. If they have a set of boundary conditions in a potential well we get standing waves that we can call stable states. So maybe the stable states of the electron would be standing waves! Since the atom is round, they would look a little more like the waves around the lip of a vibrating cup or the edge of a bell.



To make the wave work, there would have to be an integer number of electron wavelengths. That is

$$2\pi r = n\lambda \quad n = 1, 2, 3, \dots$$

Now remember that the de Broigle wavelength for an electron is

$$\lambda = \frac{h}{p_e} = \frac{h}{m_e v_e}$$

so our condition is

$$2\pi r = \frac{nh}{m_e v_e} \quad n = 1, 2, 3, \dots$$

Now let's do a little algebra

$$\begin{aligned}\frac{1}{2\pi} 2\pi r &= \frac{1}{2\pi} \frac{n\hbar}{m_e v_e} \\ r &= \frac{n\hbar}{m_e v_e} \\ m_e v_e r &= n\hbar\end{aligned}$$

which is just what Bohr had proposed! Quantum mechanics gives a reason for Bohr's quantized angular momentum.

It looks like the atom is a potential well and that the electrons are standing waves at allowed states in that potential well! We will build a quantum model of the electron states in an atom in our next lecture.

16 Quantum Model for the Atom

7.1-7.3

Fundamental Concepts in the Lecture

- We can form a one dimensional model of hydrogen that let's us solve the Schrödinger equation
- Quantizing the angular momentum
- Three dimensional solutions of the Schrödinger equation for hydrogen

One Dimensional first look at Quantum Hydrogen

Atoms are three dimensional objects (at least!). You can bet that going from one dimensional potential wells to three dimensional potential wells would mean more math. So before we take on a three dimensional atom, let's do a useful approximation. Let's assume a one dimensional atom. Because our atom will be hydrogen, we will imagine a proton at $x = 0$ and an electron that is allowed to exist at some location $x > 0$. We know the form of the potential energy for such a situation from PH220

$$\begin{aligned} U &= \frac{1}{4\pi\epsilon_0} \frac{q_1 q_1}{r} \\ &= \frac{1}{4\pi\epsilon_0} \frac{e^2}{x} \end{aligned}$$

then the Schrödinger equation

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U\psi(x) = E\psi(x)$$

becomes

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} - \frac{1}{4\pi\epsilon_0} \frac{e^2}{x} \psi(x) = E\psi(x)$$

and we are back to looking up solutions to differential equations in books. We do have boundary conditions. At $x = 0$ the term $\frac{1}{4\pi\epsilon_0} \frac{e^2}{x} \psi(x)$ will be infinite unless the wave function goes to zero at $x = 0$. And as $x \rightarrow \infty$ it seems that the electron should be closer to the nucleus if it is a bound state so we expect $\psi(\infty) \rightarrow 0$. This leads us (after

looking it up in a math book) to suggesting a wave function of the form

$$\psi(x) = Axe^{-bx}$$

where we will have to find b and A based on our boundaries, or normalization. We know what to do by now to see if this works. We take derivatives

$$\begin{aligned}\frac{d\psi(x)}{dx} &= A(x(-b)e^{-bx} + e^{-bx}) \\ &= Ae^{-bx}(1 - bx)\end{aligned}$$

and

$$\begin{aligned}\frac{d^2\psi(x)}{dx^2} &= A(e^{-bx}(-b) + (1 - bx)(-b)e^{-bx}) \\ &= A(-be^{-bx} + (1 - bx)(-b)e^{-bx}) \\ &= -Abe^{-bx}(1 + (1 - bx)) \\ &= -Abe^{-bx}(2 - bx)\end{aligned}$$

We can substitute our second derivative that we found into the Schrödinger equation and also put in our form for $\psi(x)$ wherever it shows up

$$\begin{aligned}-\frac{\hbar^2}{2m} \left(\frac{d^2\psi(x)}{dx^2} \right) - \frac{1}{4\pi\epsilon_0} \frac{e^2}{x} [\psi(x)] &= E[\psi(x)] \\ -\frac{\hbar^2}{2m} (-Abe^{-bx}(2 - bx)) - \frac{1}{4\pi\epsilon_0} \frac{e^2}{x} [Axe^{-bx}] &= E[Axe^{-bx}]\end{aligned}$$

or

$$-\frac{\hbar^2}{2m} (-Abe^{-bx}(2 - bx)) - \frac{e^2}{4\pi\epsilon_0} (Ae^{-bx}) = E(Axe^{-bx})$$

The A and e^{-bx} terms cancel leaving

$$-\frac{\hbar^2}{2m} (-b(2 - bx)) - \frac{e^2}{4\pi\epsilon_0} = Ex$$

Rearranging gives

$$\frac{\hbar^2 b}{2m} (2 - bx) - \frac{e^2}{4\pi\epsilon_0} = Ex$$

and now let's collect terms in x

$$\begin{aligned}2\frac{\hbar^2 b}{2m} - bx\frac{\hbar^2 b}{2m} - \frac{e^2}{4\pi\epsilon_0} &= Ex \\ 2\frac{\hbar^2 b}{2m} - \frac{e^2}{4\pi\epsilon_0} &= x \left(E + b\frac{\hbar^2 b}{2m} \right) \\ 2\frac{\hbar^2 b}{2m} - \frac{e^2}{4\pi\epsilon_0} - x \left(E + b\frac{\hbar^2 b}{2m} \right) &= 0\end{aligned}$$

Like for the harmonic oscillator, we want this equation to be zero for all x , so we need

$$2\frac{\hbar^2 b}{2m} - \frac{e^2}{4\pi\epsilon_0} = 0$$

and

$$\left(E + b \frac{\hbar^2 b}{2m} \right) = 0$$

Let's solve for b using the first of these

$$2 \frac{\hbar^2 b}{2m} = \frac{e^2}{4\pi\epsilon_o}$$

$$b = \frac{me^2}{4\pi\epsilon_o\hbar^2} = \frac{1}{a_o}$$

where a_o is the first Bohr radius. That is encouraging. Let's try the second equation and solve for energy

$$E + b^2 \frac{\hbar^2}{2m} = 0$$

$$E = -b^2 \frac{\hbar^2}{2m}$$

and now we know b

$$E = - \left(\frac{me^2}{4\pi\epsilon_o\hbar^2} \right)^2 \frac{\hbar^2}{2m}$$

$$= - \frac{m^2 e^4}{16\pi^2 \epsilon_o^2 \hbar^2} \frac{1}{2m}$$

$$= - \frac{me^4}{32\pi^2 \epsilon_o^2 \hbar^2}$$

which is the energy of the first Bohr state. This is really cool, but you might guess that we are not done. We really do need three dimensions. But let's finish this 1D Bohr state wave function. We still need A so we normalize

$$\int_0^\infty |Axe^{-bx}|^2 dx = 1$$

$$A^2 \int_0^\infty x^2 e^{-2bx} dx = 1$$

and we are back to looking things up in integral tables. I found one

$$\int_0^\infty x^n e^{-cx} dx = \frac{n!}{c^{n+1}}$$

and though I don't like the idea of factorials in my equations, we can use this. Set $n = 2$ and $c = 2b$ then

$$A^2 \int_0^\infty x^2 e^{-2bx} dx = A^2 \frac{2!}{(2b)^3} = 1$$

so then

$$A^2 = \frac{(2b)^3}{2} = 4b^3$$

$$A = 2b^{\frac{3}{2}}$$

so

$$\psi(x) = 2b^{\frac{3}{2}} x e^{-bx}$$

and we found b earlier, so

$$\psi(x) = 2a_o^{-\frac{3}{2}} x e^{-x/a_o}$$

which is fine but we could put in a_o as well

$$\psi(x) = 2 \left(\frac{me^2}{4\pi\epsilon_o\hbar^2} \right)^{-\frac{3}{2}} xe^{-\frac{4\pi\epsilon_o\hbar^2}{me^2}x}$$

And we can plot the shape of this. Remember that we have our nucleus (a proton) at $x = 0$. Then the positive x direction is away from the nucleus. And the factor

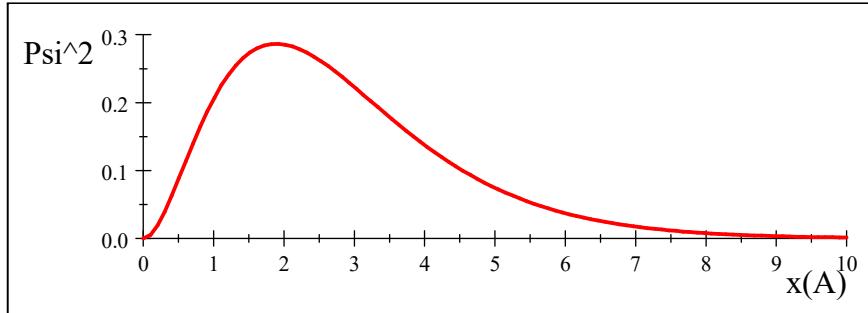
$$\begin{aligned} \frac{4\pi\epsilon_o\hbar^2}{me^2} &= \frac{4\pi (8.854187817 \times 10^{-12} \text{ F m}^{-1}) (1.05457266 \times 10^{-34} \text{ J s})^2}{(9.1093897 \times 10^{-31} \text{ kg}) (1.60217733 \times 10^{-19} \text{ C})^2} \\ &= 5.2918 \times 10^{-11} \text{ m} \\ &= 0.52918 \text{ \AA} \end{aligned}$$

and

$$\begin{aligned} 2 \left(\frac{me^2}{4\pi\epsilon_o\hbar^2} \right)^{-\frac{3}{2}} &= 2 \left(\frac{(9.1093897 \times 10^{-31} \text{ kg}) (1.60217733 \times 10^{-19} \text{ C})^2}{4\pi (8.854187817 \times 10^{-12} \text{ F m}^{-1}) (1.05457266 \times 10^{-34} \text{ J s})^2} \right)^{-\frac{3}{2}} \\ &= \frac{2.0}{\left(1.8897 \frac{1}{\text{\AA}} \right)^{\frac{3}{2}}} \end{aligned}$$

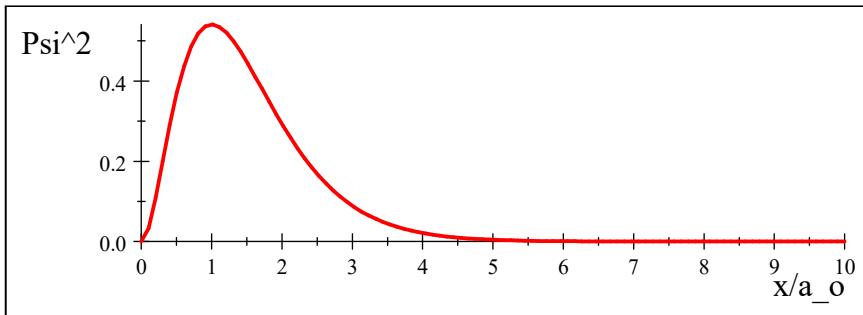
then the amplitude squared would be

$$\psi^2(x) = \frac{4.0}{\left(1.8897 \frac{1}{\text{\AA}} \right)^3} x^2 e^{-2(0.52918 \text{ \AA})x}$$



It is instructive to plot this with our units on the x axis spaced a_o apart.

$$\psi(x) = 2a_o^{-\frac{3}{2}} xe^{-x/a_o}$$



and this shows easily that the peak probability of detection is right at the Bohr radius. But this solution is very different than what Bohr's model would give us! Notice that the electron can be detected right up to the nucleus and has a significant probability of detection at four times the Bohr radius. This is a very different solution even though it preserves the good characteristics of the Bohr model. We could try for a first excited state, but really we know atoms are not one dimensional. So let's move to the three dimensional case.

Fixing the angular momentum

Bohr had to assume quantized angular momentum. We expect that we will need this to be true in our case. It is strange to think of waves having angular momentum. With linear momentum we found that the momentum was really more fundamental than mass or velocity. After all, there is no velocity conservation law in the universe, but momentum is conserved. So as strange as it seems, photons that don't have mass do have momentum and can push things. The formula $p = mv$ was really a special case of a more general relationship. We are going to find that angular momentum is also fundamental. So our intuition from PH121 is going to get stretched. Our classical relationship $L = r \times p$ is once again a special case and there is a more general relationship. We need to investigate a little bit of this extension of the concept of angular momentum now.

This is because we need to fit this new angular momentum quantization (on top of wavelength and energy quantization) into our Schrödinger equation. Let's develop a way to do this. Classically

$$L = r \times p$$

which puts L perpendicular to the distance vector from the object we are orbiting and to the momentum vector. So for a planetary orbit, if r and p are in the xy plane L is in the z direction. It is worth "remembering" that the energy of an orbiting object sets

the average orbital distance, but the shape of the orbit comes from angular momentum. And angular momentum is conserved in an orbit just like energy. The three components of the angular momentum can be any number classically.

Bohr guessed that angular momentum

$$L = n\hbar \quad n = 1, 2, 3, \dots$$

but we have seen that this doesn't match experiment. We will see (in PH433) that a better guess is

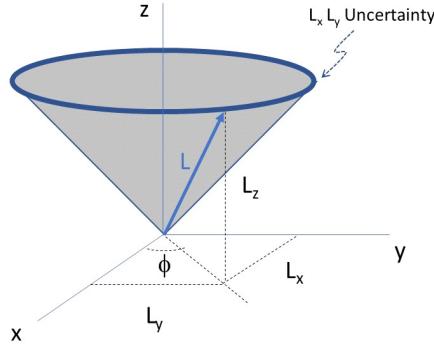
$$L = \sqrt{l(l+1)}\hbar \quad l = 0, 1, 2, 3, \dots$$

This works because when $l = 0$ then $L = 0$ and we are in agreement with experiment. It still gives quantization (only certain values work) but it is more complicated than just an n . That is not too strange, we have found for different potential wells we got energies that went like n , but some went like n^2 . This is just a little more complicated.

Now let's look at components of the angular momentum. Let's define a z -component of the angular momentum

$$L_z = m_l \hbar \quad m_l = 0, \pm 1, \pm 2, \pm 3, \dots \pm l$$

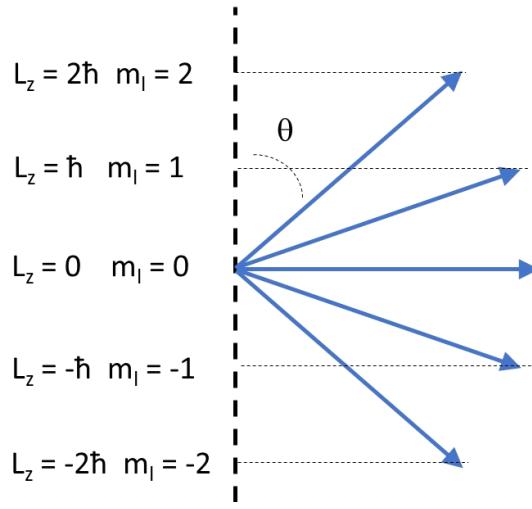
This number, m_l is traditionally called the *magnetic quantum number* and l is traditionally called the *angular momentum quantum number*. These numbers will pop out of our three dimensional solution to the Schrödinger equation just like the n in the harmonic series popped out of our solution for wave equations for waves on strings or and n^2 popped out in our solution for the energy states of the infinite potential well. The numbers l and m_l are, in a way, like the components of the angular momentum. The m_l is related to L_z . But l is a little different. In classical mechanics the angular momentum has three components, L_x , L_y , and L_z . It turns out in quantum mechanics we can't specify all three angular momentum components at the same time. If we know L_z precisely, then L_x and L_y are indeterminate (think uncertainty). In the next figure, if we know L_z then we see that the values for L_x and L_y are spread around a circle and can take any value on the circle. They are, in our current use, "uncertain." And once again it is not that we don't know them well, it is that with the restriction that they must together form the circle, L_x and L_y , will take on all the values that would form that circle.



If we consider the angle ϕ where $\tan \phi = \frac{L_y}{L_x}$ the we can say that ϕ will take on any value from 0 to 2π . It is uncertain. And then we could declare

$$\Delta L_z \Delta \phi \geq \hbar$$

and we have yet another uncertainty relationship. But the strangeness doesn't end there. Remember that the L_z is quantized with quantum number m_l . So it can only take on certain values.



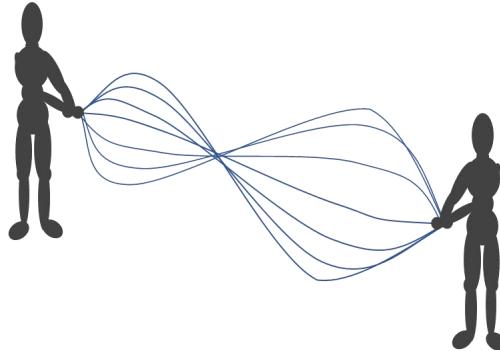
We could define an angle θ measured from the positive z axis and then

$$\cos \theta = \frac{L_z}{L} = \frac{m_l}{\sqrt{l(l+1)}}$$

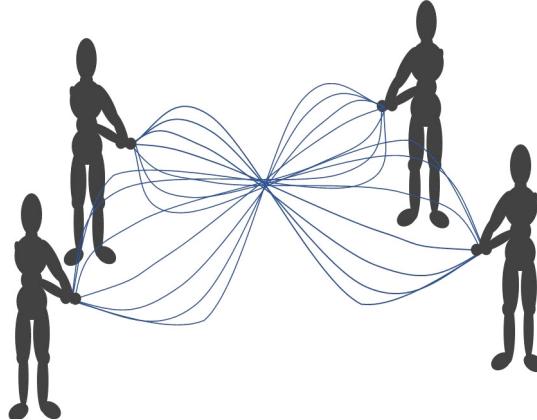
and we see that only some values of θ are allowed. This seems strange, but remember that we are dealing with waves and when we have standing waves only some values of the quantum numbers will make those standing waves.

As a very rough analogy, you could envision standing waves on strings. Let's think of

using jump ropes and make standing waves. But we could have our rope turners make a three dimensional standing wave by rotating the rope while making it oscillate.



But notice that with this structure, we could add in another set of people with another rope if we match the nodes up just right.



But we would have to line up the oscillating ropes just right or the two ropes would start to hit each other and destroy the standing waves. We probably couldn't add in a third rope because there just isn't room to fit in another wave without the waves colliding. We could envision another rope in the z -direction if we could find rope holders that could float. Our hydrogen electron waves will form something like this. Some values of angular momenta would make the waves that would conflict with themselves, other waves, or the boundaries and would therefore not make standing wave patterns. I hope this helps to see why only some values of angular momentum are possible in our atom. We have to remember we are dealing with waves, and the standing waves that we make from our specific potential well will only have certain forms that don't conflict

and ruin the standing wave pattern. With de Broglie waves, ruining the stable wave function for the electron would require the electron to go somewhere else (like leave the atom). So if we continue to think of electrons using the wave nature, we can see that angular momentum quantization is natural, maybe even for our angular momentum. This analogy isn't perfect, but maybe it will help.

The idea that angular momentum is quantized is given a strange name *space quantization*. Perhaps this is because it means only certain directions for the angular momentum are possible. This will give the shape of the electron orbitals as we will see.

Hydrogen wave functions

Now it is time to get the actual hydrogen wave functions. Of course we start with the Schrödinger equation

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} + U(x) = E\psi(x)$$

but we already have a problem. This equation is one dimensional. Let's fix that now

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi(x, y, z)}{\partial x^2} + \frac{\partial^2 \psi(x, y, z)}{\partial y^2} + \frac{\partial^2 \psi(x, y, z)}{\partial z^2} \right)$$

$$+U(x, y, z)\psi(x, y, z) = E\psi(x, y, z)$$

and this works because derivatives are linear, they just add, and we now have a wave function that is a function of x , y , and z . But this simple fix turns out to be mathematically challenging. That is because

$$\begin{aligned} U(x, y, z) &= -\frac{e^2}{4\pi\epsilon_0 r} \\ &= -\frac{e^2}{4\pi\epsilon_0 \sqrt{x^2 + y^2 + z^2}} \end{aligned}$$

is a function of all three variables x , y , and z . Our best hope for a solution is to use separation of variables so that we have three Schrödinger equation parts, one that is only a function of x and one that is only a function of y and one that is only a function of z . But the $U(x, y, z)$ is messing that up. But U is only a function of r , and we do have spherical symmetry, so maybe spherical coordinates will let us separate the equations. However, the derivatives in spherical coordinates are more complicated. In spherical coordinates our Schrödinger equation looks like this .

$$-\frac{\hbar^2}{2m} \left[\left(\frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} \right) + \left(\frac{1}{r^2} \frac{\partial^2 \psi}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right] + U(r)\psi = E\psi$$

You might be wondering where on Earth this came from. And the answer is from your math courses. Either you already have seen this transformation from Cartesian to spherical coordinates, or you soon will. I will let the math course prove that this is the

right equation. We will just use it.

For those who have taken Math316 let's pause and notice that if we define a new operator for second derivatives ∇^2 such that

$$\nabla^2\psi = \frac{1}{r^2 \sin \theta} \left[\sin \theta \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{\sin \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right]$$

we can rewrite our Schrödinger equation. Let's try this by writing out the derivative of the products in this last equation

$$\nabla^2\psi = \frac{1}{r^2 \sin \theta} \left[\sin \theta \left(r^2 \frac{\partial^2 \psi}{\partial r^2} + 2r \frac{\partial \psi}{\partial r} \right) + \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial^2 \psi}{\partial \theta^2} + \frac{\partial \psi}{\partial \theta} \cos \theta \right) + \frac{1}{\sin \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right]$$

and taking the $1/(r^2 \sin \theta)$ inside we have

$$\nabla^2\psi = \left[\left(\frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} \right) + \left(\frac{1}{r^2} \frac{\partial^2 \psi}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right]$$

which looks very like the left hand side of our Schrödinger equation. So

$$-\frac{\hbar^2}{2m} \left[\left(\frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} \right) + \left(\frac{1}{r^2} \frac{\partial^2 \psi}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right] + U(r) \psi = E\psi$$

could be written as

$$-\frac{\hbar^2}{2m} [\nabla^2\psi] + U(r) \psi = E\psi$$

and you will sometimes see it this way. If you are just now taking M316, then you now know you are in for a treat. But the question for us now is to find a solution to this three dimensional non-relativistic Schrödinger equation in spherical coordinates.

The trick is to now assume a wave function that is the product of three functions, but each of only one variable

$$\psi(r, \theta, \phi) = R(r) \Theta(\theta) \Phi(\phi)$$

Let's put this into our new Schrödinger equation to try out our guess. We will have to rearrange a bit to make it fit on the page. We start with

$$-\frac{\hbar^2}{2m} \left[\left(\frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} \right) + \left(\frac{1}{r^2} \frac{\partial^2 \psi}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right] + U(r) \psi = E\psi$$

but everywhere there is a ψ we put in $\psi = R(r) \Theta(\theta) \Phi(\phi)$. We get

$$\begin{aligned} & -\frac{\hbar^2}{2m} \left(\frac{\partial^2 (R(r) \Theta(\theta) \Phi(\phi))}{\partial r^2} + \frac{2}{r} \frac{\partial (R(r) \Theta(\theta) \Phi(\phi))}{\partial r} \right) \\ & -\frac{\hbar^2}{2m} \left(\frac{1}{r^2} \frac{\partial^2 (R(r) \Theta(\theta) \Phi(\phi))}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} \frac{\partial (R(r) \Theta(\theta) \Phi(\phi))}{\partial \theta} \right) \\ & -\frac{\hbar^2}{2m} \left(\frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 (R(r) \Theta(\theta) \Phi(\phi))}{\partial \phi^2} \right) \\ & + U(r) R(r) \Theta(\theta) \Phi(\phi) = ER(r) \Theta(\theta) \Phi(\phi) \end{aligned}$$

Where I had to distribute the $-\hbar^2/(2m)$ to make it all fit on the page.

We can take out some of the functions that act like constants in the partial derivatives

(another M316 thing, but we know partial derivatives form PH150).

$$\begin{aligned}
& -\frac{\hbar^2}{2m} \left(\Theta(\theta) \Phi(\phi) \frac{\partial^2(R(r))}{\partial r^2} + \frac{2}{r} \Theta(\theta) \Phi(\phi) \frac{\partial(R(r))}{\partial r} \right) \\
& -\frac{\hbar^2}{2m} \left(R(r) \Phi(\phi) \frac{1}{r^2} \frac{\partial^2(\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} R(r) \Phi(\phi) \frac{\partial(\Theta(\theta))}{\partial \theta} \right) \\
& -\frac{\hbar^2}{2m} \frac{1}{r^2 \sin^2 \theta} R(r) \Theta(\theta) \frac{\partial^2(\Phi(\phi))}{\partial \phi^2} \\
& + U(r) (R(r) \Theta(\theta) \Phi(\phi)) \\
& = E(R(r) \Theta(\theta) \Phi(\phi))
\end{aligned}$$

Now let's divide each term by $(R(r) \Theta(\theta) \Phi(\phi))$

$$\begin{aligned}
& -\frac{\hbar^2}{2m} \left(\frac{\Theta(\theta) \Phi(\phi)}{R(r) \Theta(\theta) \Phi(\phi)} \frac{\partial^2(R(r))}{\partial r^2} + \frac{2}{r} \frac{\Theta(\theta) \Phi(\phi)}{R(r) \Theta(\theta) \Phi(\phi)} \frac{\partial(R(r))}{\partial r} \right) \\
& -\frac{\hbar^2}{2m} \left(\frac{R(r) \Phi(\phi)}{R(r) \Theta(\theta) \Phi(\phi)} \frac{1}{r^2} \frac{\partial^2(\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} \frac{R(r) \Phi(\phi)}{R(r) \Theta(\theta) \Phi(\phi)} \frac{\partial(\Theta(\theta))}{\partial \theta} \right) \\
& -\frac{\hbar^2}{2m} \frac{1}{r^2 \sin^2 \theta} \frac{R(r) \Theta(\theta)}{R(r) \Theta(\theta) \Phi(\phi)} \frac{\partial^2(\Phi(\phi))}{\partial \phi^2} \\
& + U(r) \frac{(R(r) \Theta(\theta) \Phi(\phi))}{R(r) \Theta(\theta) \Phi(\phi)} \\
& = E \frac{(R(r) \Theta(\theta) \Phi(\phi))}{R(r) \Theta(\theta) \Phi(\phi)}
\end{aligned}$$

to get

$$\begin{aligned}
& -\frac{\hbar^2}{2m} \left(\frac{1}{R(r)} \frac{\partial^2(R(r))}{\partial r^2} + \frac{2}{r} \frac{1}{R(r)} \frac{\partial(R(r))}{\partial r} \right) \\
& -\frac{\hbar^2}{2m} \frac{1}{r^2} \left(\frac{1}{\Theta(\theta)} \frac{\partial^2(\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta} \frac{1}{\Theta(\theta)} \frac{\partial(\Theta(\theta))}{\partial \theta} \frac{1}{\sin^2 \theta} \frac{1}{\Phi(\phi)} \frac{\partial^2(\Phi(\phi))}{\partial \phi^2} \right) \\
& + U(r) = E
\end{aligned}$$

which we can regroup so all the $R(r)$, $\Theta(\theta)$ and $\Phi(\phi)$ functions are grouped together.

$$\begin{aligned}
& -\frac{\hbar^2}{2m} \left(\frac{1}{R(r)} \frac{\partial^2(R(r))}{\partial r^2} + \frac{2}{r} \frac{1}{R(r)} \frac{\partial(R(r))}{\partial r} \right) \\
& -\frac{\hbar^2}{2m} \left(\frac{1}{\Theta(\theta)} \frac{1}{r^2} \frac{\partial^2(\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{r^2 \sin \theta} \frac{1}{\Theta(\theta)} \frac{\partial(\Theta(\theta))}{\partial \theta} \right) \\
& -\frac{\hbar^2}{2m} \frac{1}{r^2 \sin^2 \theta} \frac{1}{\Phi(\phi)} \frac{\partial^2(\Phi(\phi))}{\partial \phi^2} \\
& + U(r) = E
\end{aligned}$$

And let's do more rearranging, let's multiple through by $r^2 \sin^2 \theta$ and put all the terms

in $\Phi(\phi)$ on the right hand side

$$\begin{aligned}
 & -r^2 \sin^2 \theta \frac{\hbar^2}{2m} \left(\frac{1}{R(r)} \frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r R(r)} \frac{\partial (R(r))}{\partial r} \right) \\
 & - \frac{r^2 \sin^2 \theta}{r^2} \frac{\hbar^2}{2m} \left(\frac{1}{\Theta(\theta)} \frac{\partial^2 (\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta} \frac{1}{\Theta(\theta)} \frac{\partial (\Theta(\theta))}{\partial \theta} \right) \\
 & + U(r) r^2 \sin^2 \theta - Er^2 \sin^2 \theta \\
 & = \frac{\hbar^2}{2m \Phi(\phi)} \frac{1}{\partial \phi^2} \frac{\partial^2 (\Phi(\phi))}{\partial \phi^2}
 \end{aligned}$$

Notice that the right hand side of the equal sign is just a function of ϕ and the left hand side is a function of r and θ . The only way for this to be true independent r, θ, ϕ is for each side to be a constant. And further notice that ϕ is an angular direction. We expect that this direction of motion will produce angular momentum. And we know from our discussion of space quantization that angular momentum is quantized. So although each side must be equal to a constant, we suspect that there will be more than one value for that constant that will work (more than one standing wave that can form). Those values will correspond to the allowed angular momenta. As a terrible spoiler, we give this constant the symbol $-m_l^2$. So

$$\begin{aligned}
 \frac{1}{\Phi(\phi)} \frac{\partial^2 \Phi(\phi)}{\partial \phi^2} &= -m_l^2 \\
 \frac{\partial^2 \Phi(\phi)}{\partial \phi^2} &= -m_l^2 \Phi(\phi)
 \end{aligned} \tag{16.1}$$

and we could solve this for the $\Phi(\phi)$ part of the wave function. We would expect a wave in the ϕ direction. And then m_l would be a quantum number just like n was for waves on strings. The quantum number m_l is going to tell us which wave solutions make standing wave patters that work for our hydrogen potential well like n told us which frequencies made standing waves on violin strings.

But for now let's just replace $\frac{1}{\Phi(\phi)} \frac{\partial^2 \Phi(\phi)}{\partial \phi^2}$ with $-m_l^2$ in our Schrödinger equation and rearrange again

$$\begin{aligned}
 & -r^2 \sin^2 \theta \left(\frac{1}{R(r)} \frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r R(r)} \frac{\partial (R(r))}{\partial r} \right) \\
 & - \frac{r^2 \sin^2 \theta}{r^2} \left(\frac{1}{\Theta(\theta)} \frac{\partial^2 (\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta} \frac{1}{\Theta(\theta)} \frac{\partial (\Theta(\theta))}{\partial \theta} \right) \\
 & + \frac{2m}{\hbar^2} U(r) r^2 \sin^2 \theta - \frac{2m}{\hbar^2} Er^2 \sin^2 \theta \\
 & = -m_l^2
 \end{aligned}$$

Now let's put all the terms with $\Theta(\theta)$ on the right hand side

$$\begin{aligned} & -r^2 \left(\frac{1}{R(r)} \frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r} \frac{1}{R(r)} \frac{\partial (R(r))}{\partial r} \right) + \frac{2m}{\hbar^2} U(r) r^2 \sin^2 \theta - \frac{2m}{\hbar^2} E r^2 \sin^2 \theta \\ &= \left(\frac{1}{\Theta(\theta)} \frac{\partial^2 (\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta} \frac{1}{\Theta(\theta)} \frac{\partial (\Theta(\theta))}{\partial \theta} \right) - \frac{m_l^2}{\sin^2 \theta} \end{aligned}$$

and notice that on the left hand side of the equal sign we have only a function of r and on the right we have only a function of θ . For these two sides to be equal for all r and θ we need each side to be equal to another constant. Once again this constant will be associated with angular motion and therefore angular momentum. And we know that the angular momentum is quantized. So once again we expect that more than one value for this constant will work. And once again as a large spoiler let's write this constant as $l(l+1)$ following what we found from space quantization, then

$$\left(\frac{1}{\Theta(\theta)} \frac{\partial^2 (\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta} \frac{1}{\Theta(\theta)} \frac{\partial (\Theta(\theta))}{\partial \theta} \right) - \frac{m_l^2}{\sin^2 \theta} = -l(l+1)$$

and

$$\left(\frac{\partial^2 (\Theta(\theta))}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta} \frac{\partial (\Theta(\theta))}{\partial \theta} \right) - \frac{m_l^2}{\sin^2 \theta} = -l(l+1) \Theta(\theta) \quad (16.2)$$

And we could solve this for the $\Theta(\theta)$ part of our wave function. But for now, let's replace the left hand term with the right hand term in our Schrödinger equation and rearrange again

$$\begin{aligned} & -r^2 \left(\frac{1}{R(r)} \frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r} \frac{1}{R(r)} \frac{\partial (R(r))}{\partial r} \right) - \frac{2m}{\hbar^2} (U(r) - E) r^2 \\ &= l(l+1) \end{aligned}$$

We can multiply both sides by $R(r)$

$$\begin{aligned} & -R(r) r^2 \left(\frac{1}{R(r)} \frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r} \frac{1}{R(r)} \frac{\partial (R(r))}{\partial r} \right) + \frac{2m}{\hbar^2} (U(r) - E) r^2 R(r) \\ &= -l(l+1) R(r) \end{aligned}$$

Then dividing by r^2 gives

$$\begin{aligned} & - \left(\frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r} \frac{\partial (R(r))}{\partial r} \right) + \frac{2m}{\hbar^2} (U(r) - E) R(r) \\ &= -\frac{l(l+1)}{r^2} R(r) \end{aligned}$$

and moving the $\hbar^2/2m$ back in place

$$\begin{aligned} & -\frac{\hbar^2}{2m} \left(\frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r} \frac{\partial (R(r))}{\partial r} \right) - E R(r) + U(r) R(r) \\ &= \frac{l(l+1)}{r^2} \frac{\hbar^2}{2m} R(r) \end{aligned}$$

We can almost fit our radial part of the Schrödinger equation on one line of the page

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2 (R(r))}{\partial r^2} + \frac{2}{r} \frac{\partial (R(r))}{\partial r} \right) + [U(r)] R(r) + \frac{l(l+1)}{r^2} \frac{\hbar^2}{2m} R(r) = E R(r)$$

and putting in our Coulomb potential we have

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial r^2} (R(r)) + \frac{2}{r} \frac{\partial}{\partial r} (R(r)) \right) + \left(\left[-\frac{e^2}{4\pi\epsilon_0 r} \right] + \frac{l(l+1)}{2mr^2} \right) R(r) = E(r) R(r) \quad (16.3)$$

The mass is still the reduced mass. We can find equations that solve this differential equation for the $R(r)$ part of the . We will find clever ways to do that in PH433. But for now, we will just quote the results. The first thing to note is that m_l isn't in the radial equation. But E and U are in the radial equation. So all wave functions with the same l but different m_l values will have the same energy. These states are called *degenerate* (they have different quantum numbers, but have the same energy). We will get an additional quantum number, n , from solving the radial part of the Schrödinger equation that labels the energies.

$$E_n = -\frac{me^4}{32\pi^2\epsilon_0^2\hbar^2 n^2}$$

which looks much like or energy quantization for an infinite potential well.

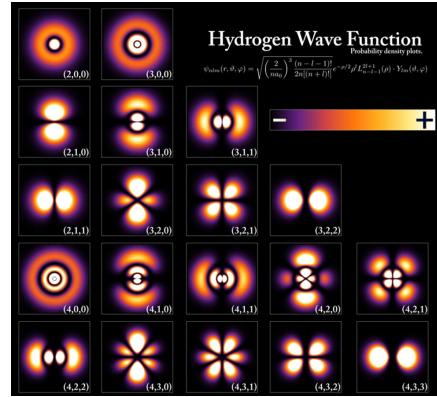
Sadly, we won't actually launch off into a differential equations class at this point. But we know

$$\psi(r, \theta, \phi) = R(r) \Theta(\theta) \Phi(\phi)$$

and the pieces of this wave function have been found using partial differential equation techniques to solve equations (16.1), (16.2) and (16.3). The parts are given in the next table for a few of the possible quantum numbers n , l , and m_l .

n	l	m_l	$R(r)$	$\Theta(\theta)$	$\Phi(\phi)$
1	0	0	$\frac{2}{(a_o)^{\frac{3}{2}}} e^{-\frac{r}{a_o}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2\pi}}$
2	0	0	$\frac{1}{(2a_o)^{\frac{3}{2}}} \left(2 - \frac{r}{a_o} \right) e^{-\frac{r}{2a_o}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2\pi}}$
2	1	0	$\frac{1}{\sqrt{3}(2a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o} \right) e^{-\frac{r}{2a_o}}$	$\sqrt{\frac{3}{2}} \cos \theta$	$\frac{1}{\sqrt{2\pi}}$
2	1	± 1	$\frac{1}{\sqrt{3}(2a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o} \right) e^{-\frac{r}{2a_o}}$	$\pm \frac{\sqrt{3}}{2} \sin \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$
3	0	0	$\frac{2}{(3a_o)^{\frac{3}{2}}} \left(1 - \frac{2r}{3a_o} + \frac{2r^2}{27a_o^2} \right) e^{-\frac{r}{3a_o}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2\pi}}$
3	1	0	$\frac{2}{9\sqrt{2}(3a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o} - \frac{r^2}{6a_o^2} \right) e^{-\frac{r}{3a_o}}$	$\sqrt{\frac{3}{2}} \cos \theta$	$\frac{1}{\sqrt{2\pi}}$
3	1	± 1	$\frac{2}{9\sqrt{2}(3a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o} - \frac{r^2}{6a_o^2} \right) e^{-\frac{r}{3a_o}}$	$\pm \frac{\sqrt{3}}{2} \sin \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$
3	2	0	$\frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2} \right) e^{-\frac{r}{3a_o}}$	$\sqrt{\frac{5}{8}} (3 \cos^2 \theta - 1)$	$\frac{1}{\sqrt{2\pi}}$
3	2	± 1	$\frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2} \right) e^{-\frac{r}{3a_o}}$	$\pm \sqrt{\frac{15}{4}} \sin \theta \cos \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$
3	2	21	$\frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2} \right) e^{-\frac{r}{3a_o}}$	$\sqrt{\frac{15}{4}} \sin^2 \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$

A plot of these functions is hard, but doable. Usually we plot $|\psi|^2$ so the plot gives the probability of where the electron will be detected. Here is one attempt.



If you are thinking that these look a lot like pictures of electron orbitals, you are right! The orbitals you studied in High School are the wave functions (squared) of the electrons.

Not every value of the quantum numbers for electrons in hydrogen atoms make real wave functions that work. This is a little like when we did standing waves back in PH123 and found that the even quantum numbers for pipes closed on one end don't make standing waves. We won't do the math to show this here (more for PH433 to do!) but the pattern is

n	$1, 2, 3, \dots$
l	$0, 1, 2, \dots, n-1$
m_l	$\pm 1, \pm 2, \pm 3, \dots, \pm l$

Finding the probability density is still a little tricky. We need to remember that in spherical coordinates the volume element is

$$dV = r^2 \sin \theta dr d\theta d\phi$$

and we would need to integrate over all space. We can try that for the radial functions in the problems.

What we have done is monumental, but its kind of like building only part of the monument ourselves. We borrowed solutions from our future selves (and past colleagues). But we now can explain that atoms are potential wells, and we can explain Bohr's assumption of quantization because we should just expect quantization from a wavicle in a potential well.

We could check one of the radial parts, say the ground state radial part, of the wave function to make sure it really works

$$R(r) = \frac{2}{a^{\frac{3}{2}}} e^{-\frac{r}{a_O}}$$

if we take derivatives

$$\frac{\partial}{\partial r} (R(r)) = \frac{\partial}{\partial r} \left(\frac{2}{a^{\frac{3}{2}}} e^{-\frac{r}{a_o}} \right) = -\frac{2}{a^{\frac{3}{2}} a_o} e^{-\frac{r}{a_o}}$$

$$\frac{\partial^2}{\partial r^2} (R(r)) = \frac{\partial^2}{\partial r^2} \left(\frac{2}{a^{\frac{3}{2}}} e^{-\frac{r}{a_o}} \right) = \frac{2}{a^{\frac{3}{2}} a_o^2} e^{-\frac{r}{a_o}}$$

and take the ground state energy

$$\begin{aligned} E_1 &= -\frac{me^4}{32\pi^2\epsilon_o^2\hbar^2} = -\frac{me^2me^2\hbar^2}{2m(4\pi\epsilon_o\hbar^2)(4\pi\epsilon_o\hbar^2)} \\ &= -\frac{\hbar^2}{2ma_o^2} \end{aligned}$$

and put all this into equation (16.3)

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial r^2} (R(r)) + \frac{2}{r} \frac{\partial}{\partial r} (R(r)) \right) + \left(\left[-\frac{e^2}{4\pi\epsilon_o r} \right] + \frac{l(l+1)}{2mr^2} \right) R(r) = E(r) R(r)$$

then we get

$$\begin{aligned} &-\frac{\hbar^2}{2m} \left(\frac{2}{a^{\frac{3}{2}} a_o^2} e^{-\frac{r}{a_o}} + \frac{2}{r} \left(-\frac{2}{a^{\frac{3}{2}} a_o} e^{-\frac{r}{a_o}} \right) \right) \\ &+ \left(-\frac{e^2}{4\pi\epsilon_o r} + \frac{\sqrt{(0)((0)+1)}}{2mr^2} \right) \left(\frac{2}{a^{\frac{3}{2}}} e^{-\frac{r}{a_o}} \right) \\ &= -\frac{\hbar^2}{2ma_o^2} \left(\frac{2}{a^{\frac{3}{2}}} e^{-\frac{r}{a_o}} \right) \end{aligned}$$

or

$$-\frac{\hbar^2}{2m} \left(\frac{2}{a^{\frac{3}{2}} a_o^2} e^{-\frac{r}{a_o}} + \frac{2}{r} \left(-\frac{2}{a^{\frac{3}{2}} a_o} e^{-\frac{r}{a_o}} \right) \right) + \left(-\frac{e^2}{4\pi\epsilon_o r} \right) \left(\frac{2}{a^{\frac{3}{2}}} e^{-\frac{r}{a_o}} \right) = -\frac{\hbar^2}{2ma_o^2} \left(\frac{2}{a^{\frac{3}{2}}} e^{-\frac{r}{a_o}} \right)$$

and simplifying

$$\begin{aligned} &-\frac{\hbar^2}{2m} \left(\frac{2}{a_o^2} + \frac{2}{r} \left(-\frac{2}{a_o} \right) \right) + \left(-\frac{e^2}{4\pi\epsilon_o r} \right) (2) = -\frac{\hbar^2}{2ma_o^2} (2) \\ &-\frac{\hbar^2}{2m} \left(\frac{2r}{a_o^2 r} - \frac{4a_o}{ra_o^2} \right) + \left(-\frac{e^2}{4\pi\epsilon_o r} \right) (2) = -\frac{\hbar^2}{2ma_o^2} (2) \\ &-\frac{\hbar^2}{2m} \left(\frac{2r - 4a_o}{a_o^2 r} \right) + \left(-\frac{e^2}{4\pi\epsilon_o r} \right) (2) = -\frac{\hbar^2}{2ma_o^2} (2) \\ &-\frac{\hbar^2}{2m} \left(\frac{r - 2a_o}{a_o^2 r} \right) + \left(-\frac{e^2 m \hbar^2}{4\pi\epsilon_o \hbar^2 m r} \right) + \frac{\hbar^2}{2ma_o^2} = 0 \\ &-\frac{r\hbar^2 - 2a_o\hbar^2}{2ma_o^2 r} - \frac{2\hbar^2 a_o}{2a_o^2 mr} + \frac{\hbar^2 r}{2a_o^2 mr} = 0 \\ &-r\hbar^2 + 2a_o\hbar^2 - 2\hbar^2 a_o + \hbar^2 r = 0 \\ &-r\hbar^2 + \hbar^2 r + 2a_o\hbar^2 - 2\hbar^2 a_o = 0 \\ &0 = 0 \end{aligned}$$

so indeed it works.

There is a little more to this story. We still need the angular parts. And there is the

surprise of electron angular momentum (spin) still to come. But we have developed (with help) a new model for the atom (well hydrogen) that matches all that we know about atoms (well, hydrogen atoms).

17 Details of the Quantum Model for the Atom

7.4-7.7

Fundamental Concepts in the Lecture

- Radial probability densities
- Angular probability densities
- Electron intrinsic spin
- Spectroscopic notation (cause we need another notation...)

Radial probability densities

If you have taken chemistry you know they make a big deal of whether an electron is an “outer” or “inner” electron. We should be able to find the highest probability of detection for an electron wave, and sort of mentally equate that to the location of the electron (it is where it would be detected most often, so this isn’t really so bad). We would need to integrate $|\psi^2| = |R(r)\Theta(\theta)\Phi(\phi)|^2$ and we now need to integrate over a volume interval in spherical coordinates

$$dV = r^2 \sin \theta dr d\theta d\phi$$

We would have

$$P = \int \int \int |R(r)\Theta(\theta)\Phi(\phi)|^2 r^2 \sin \theta dr d\theta d\phi$$

and this could be daunting. But if we just want to know the distance from the nucleus where there is the most probability of detection, we could write the probability of being at a particular position r away from the nucleus as just

$$P(r) dr = |R(r)|^2 r^2 dr \int_0^\pi |\Theta(\theta)|^2 \sin \theta d\theta \int_0^{2\pi} |\Phi(\phi)|^2 d\phi$$

where we are, in effect, averaging out all the angular dependence, and if we demand that each of the parts of the wave function $R(r)$, $\Theta(\theta)$, $\Phi(\phi)$ are individually normalized, then the two integrals are both equal to 1 so we are left with just

$$P(r) dr = |R(r)|^2 r^2 dr$$

And we could take a particular form of $R(r)$ from our chart and plot it. For the ground state

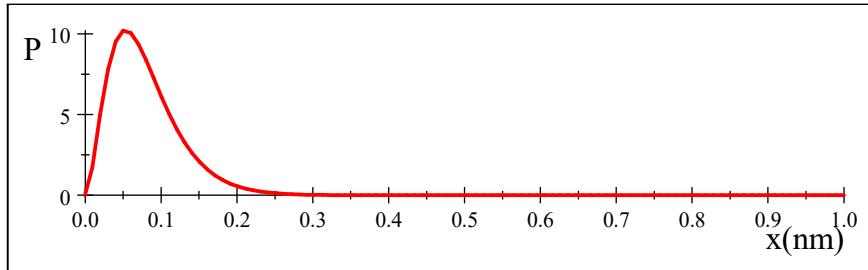
$$R(r) = \frac{2}{a_o^{\frac{3}{2}}} e^{-\frac{r}{a_o}}$$

and

$$a_o = \frac{4\pi\epsilon_0\hbar^2}{e^2 m_e} = 0.0529 \text{ nm} \quad (17.1)$$

so then

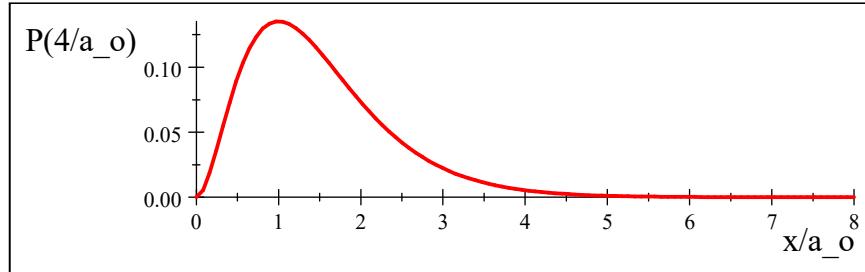
$$P(r) = \left| \frac{2}{(0.0529 \text{ nm})^{\frac{3}{2}}} e^{-\frac{r}{0.0529 \text{ nm}}} \right|^2 r^2$$



and we see that we are most likely to detect an electron at a little over a tenth of an angstrom away from the nucleus. We now know the trick of plotting with $u = r/a_o$. Then our function looks like

$$\begin{aligned} P(r) &= \left| \frac{2}{a_o^{\frac{3}{2}}} e^{-\frac{r}{a_o}} \right|^2 r^2 \\ &= \left| \frac{2}{a_o^{\frac{3}{2}}} e^{-\frac{ua_o}{a_o}} \right|^2 (ua_o)^2 \\ &= \left| \frac{2}{(a_o)^{\frac{3}{2}}} e^{-u} \right|^2 u^2 (a_o)^2 \\ &= \frac{4}{(a_o)^3} e^{-2u} u^2 (a_o)^2 \\ &= \frac{4u^2}{(a_o)} e^{-2u} \end{aligned}$$

and we can make our vertical access in units of $4/a_o$ so the plot looks like this



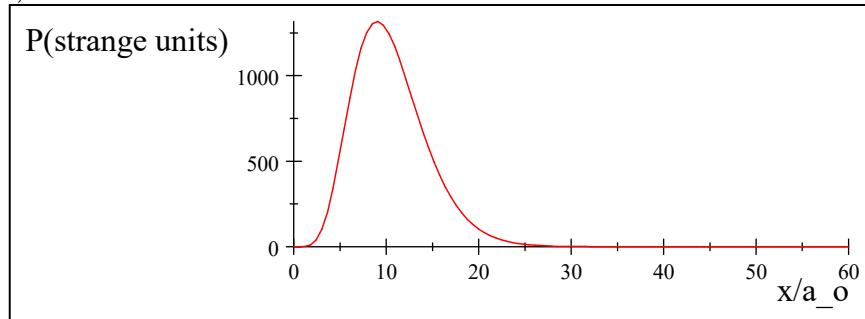
and this is just the ground state. We will designate a state by its quantum numbers. The ground state has $n = 1$, $l = 0$, $m_l = 0$ so we call this the $(1, 0, 0)$ state. We could try another, say the $(3, 2, 1)$ state (look at the table in section 16.3). Then

$$R(r) = \frac{4}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2}\right) e^{-\frac{r}{3a_o}}$$

and we will again let $u = r/a_o$ so $r = ua_o$

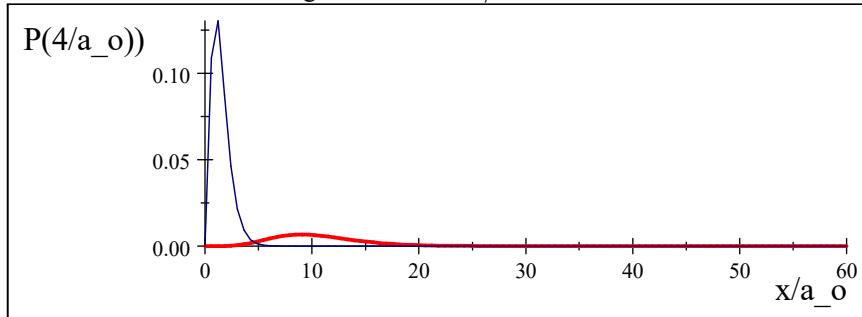
$$\begin{aligned} P(r) &= \left| \frac{4}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2}\right) e^{-\frac{r}{3a_o}} \right|^2 r^2 \\ P(r) &= \left| \frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{(ua_o)^2}{a_o^2}\right) e^{-\frac{u}{3}} \right|^2 (ua_o)^2 \\ &= \left| \frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{u^2 a_o^2}{a_o^2}\right) e^{-\frac{u}{3}} \right|^2 (ua_o)^2 \\ &= \frac{4}{7290(3a_o)^3} (u^4) e^{-2\frac{u}{3}} (ua_o)^2 \\ &= \frac{4}{7290(3a_o)^3} (u^6 a_o^2) e^{-2\frac{u}{3}} \\ &= \frac{4}{196830a_o} (u^6) e^{-2\frac{u}{3}} \end{aligned}$$

so if we make our vertical axis again in units of $\frac{4}{196830a_o}$ then our plot looks like $(u^6) e^{-2\frac{u}{3}}$



and we can see that when $x = a_o$ $u = 1$ so our ground state had a peak near $x = a_o$ and

now we have a peak near $10a_o$. The higher electron states are indeed more “outer” just like we learned in Chemistry. Here is a plot of both the ground state and the $(3, 2, 1)$ state and we can see that the electron in the ground state is more often found nearer the nucleus with the vertical axis again in units of $4/a_o$.

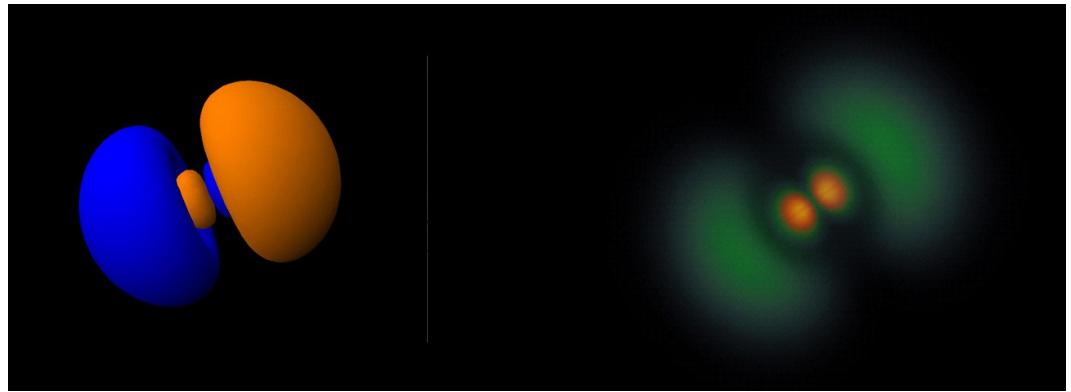


The $(1, 0, 0)$ (thin blue line) and $(3, 2, 1)$ (thick red line) radial probability functions plotted together in distance units of x/a_o and with vertical axis units of $4/a_o$. The $n = 3$ electron is more likely to be detected farther away from the nucleus than the $n = 1$ case.

Angular probability densities

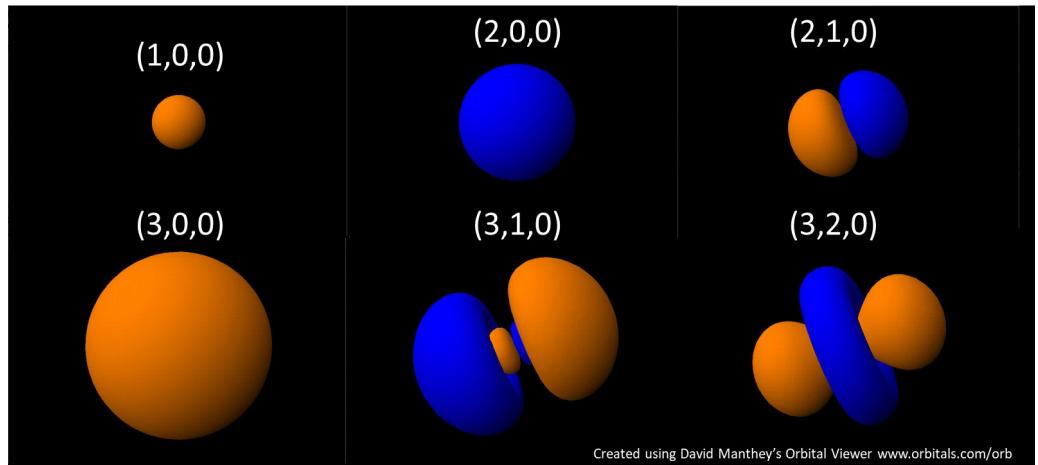
Looking back at our table of wave function pieces we note that the wave function $\Phi(\phi)$ does depend on ϕ , but $|\Phi(\phi)|^2$ will always lose the ϕ dependence. The probability density does not depend on ϕ . We call this being *cylindrically symmetric*.

For the ground state $(1, 0, 0)$ and first excited state $(2, 0, 0)$ the function $\Theta(\theta)$ is a constant. That means all angles, θ , are equally likely (think uncertainty relation) so the $(1, 0, 0)$ and $(2, 0, 0)$ look like fuzzy balls with the most probable detection at different distances, r , from the nuclear center. But with larger n values we expect the wave functions to not be symmetric. Here is an example for the $(3, 1, 0)$ state.



Hydrogen (3,1,0) orbitals plotted using David Manthey's Orbital Viewer www.orbitals.com/orb, and Mathematica.

The plot on the left makes a surface of the peak of the probability function (amplitude squared) it looks like a smooth surface. This is how we usually see electron probability functions graphed. But the plot on the right might be a better way to think about the probability function. It shows the same structures, but you can see that the probability fades into the maximum and back out of the maximum surfaces. In either case, we just plot the amplitude squared, so you don't see that there is an underlying wave. In the next figure (figure ??) on the left are the the (1, 0, 0) and (2, 0, 0) states plotted like the left hand case in the last figure. Once again, the software used picked the peak value of $|\psi|^2$ to plot as a surface. So the surface of the ball shaped (1, 0, 0) function is the surface over which the probability of detection is constant. That gives us our orbital visualization shape



First six hydrogen orbitals plotted using David Manthey's Orbital Viewer www.orbitals.com/orb

Now let's look at the $(2, 1, 0)$ wave function. The $\Phi(\phi)$ function is still constant, but now

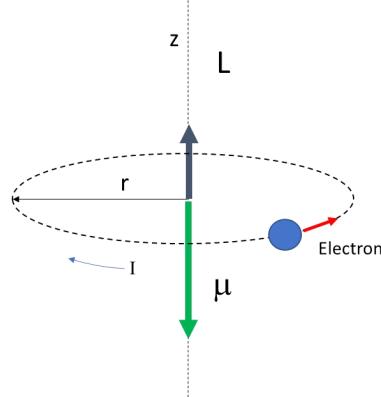
$$\Theta(\theta) = \sqrt{\frac{3}{2}} \cos \theta$$

which is not spherically symmetric. We can see the $(2, 1, 0)$ state also plotted in the figure (??). Its shape is very different.

The figure also has the $(3, 0, 0)$, $(3, 1, 0)$ and $(3, 2, 0)$ states plotted. Notice that as n gets bigger the $(n, 0, 0)$ states get bigger, and notice that the standing wave patterns get more complicated as n increases and l increases. Like we did with our analogy with waves on ropes, we are building standing wave shapes that can fit in the geometry of our spherical potential well. Those shapes are more complicated as l and m_l get bigger.

Angular momentum and space quantization

It would be good to pause a moment and look once again at angular momentum quantization. We can get some benefit in this case from using the Bohr model because it gives the same result as our quantum model but without the lovely difficult math. Let's model an electron orbiting the atom. That moving electron is a current, and current loops create magnetic fields. So we would expect the electron (and therefore the atom) to have a magnetic moment μ .



The magnetic moment is defined as the current multiplied by the area of the current loop

$$\mu = IA$$

and the current in the Bohr model is very simple. It is one electron charge, e , circling the atom in a period

$$T = \frac{2\pi r}{v_t}$$

so our current is

$$I = \frac{e}{\frac{2\pi r}{v_t}} = \frac{ev_t}{2\pi r}$$

and we could put this in terms of momentum by writing $p = mv_t$

$$I = \frac{emv_t}{2\pi rm} = \frac{ep}{2\pi rm}$$

then the magnetic moment is

$$\begin{aligned}\mu &= IA \\ &= \left(\frac{-ep}{2\pi rm}\right)(\pi r^2) \\ &= \left(\frac{-e}{2m}\right)(rp) \\ &= -\left(\frac{e}{2m}\right)L\end{aligned}$$

and of course both μ and L are vectors and because of the minus sign, they point in opposite directions. We could take the z component of these vectors

$$\mu_z = -\left(\frac{e}{2m}\right)L_z$$

but we know at some point we have to go beyond the Bohr model, so let's use our quantum version of L_z and see if this is enough to get a semi-reasonable result.

$$L_z = m_l\hbar$$

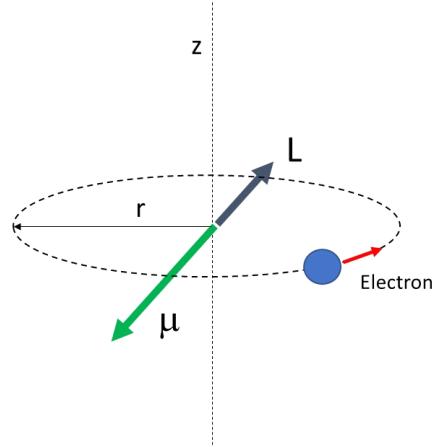
Then the magnetic moment z -component would be

$$\mu_z = -\left(\frac{e}{2m}\right)m_l\hbar$$

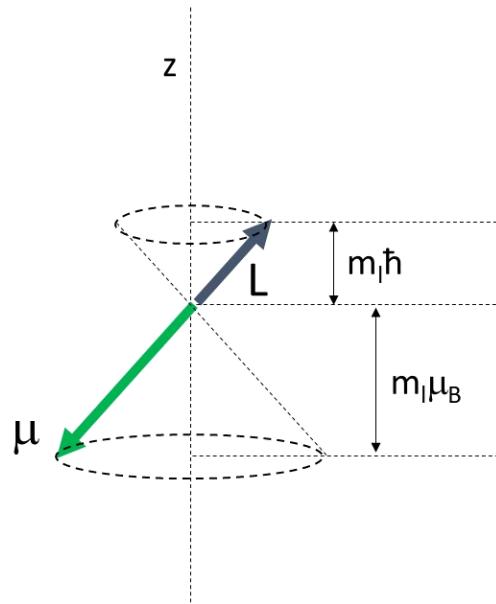
and now we know that the magnetic moment z component is quantized. This isn't so much a surprise because we have seen space quantization before. We know that this means the atom can only align the orbit of the electron at certain angles. And it means that if we know μ_z precisely, we don't know μ_x and μ_y precisely. They are uncertain, meaning they might take on many different values with the restriction that

$$\mu = \sqrt{\mu_z^2 + \mu_x^2 + \mu_y^2}$$

so we might get something like this



and over time we would get something like what is shown in the next diagram where we would see the μ_x and μ_y values (and therefore the L_x and L_y) form a circle.



It is tradition to define

$$\mu_B = \frac{e\hbar}{2m}$$

and we call this the *Bohr magneton*. Then

$$\mu_z = -\mu_B m_l$$

If we place the atom in an external magnetic field, we would expect that the magnetic

moment would want to line up with the field and the atom will turn so that this is the case. That would be the lowest energy. Think of taking a small magnet and placing next to a large magnet. It would take work to turn the small magnet over. It prefers to be oriented with the external magnetic field. But now we know that we can't perfectly align our atom. We can make the z component align, but we get uncertainty in the μ_x and μ_y components. They won't be exactly zero.

This is a semi-classical way to get μ_z so it isn't really the right reasoning, but the value we got is good and the insight into what is happening is not terrible. We can do better in the upper division quantum mechanics courses. But we need this result to understand how atoms behave in magnetic fields. A clever experiment made use of this atomic magnetic moment. And to understand how we need to know that if we put the magnetic dipole in a non uniform external magnetic field. the non-uniformity can push the dipole. We will look at this experiment in our next lecture.

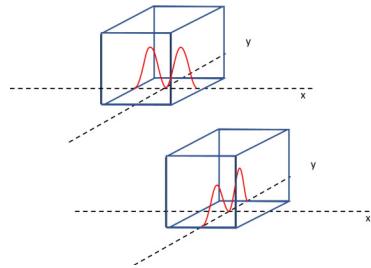
18 Hydrogen Atoms in Magnetic Fields

7.8-7.9, 8.1

Fundamental Concepts in the Lecture

- Intrinsic Spin.
- Spectroscopic Notation
- Zeeman Effect
- Fine Structure

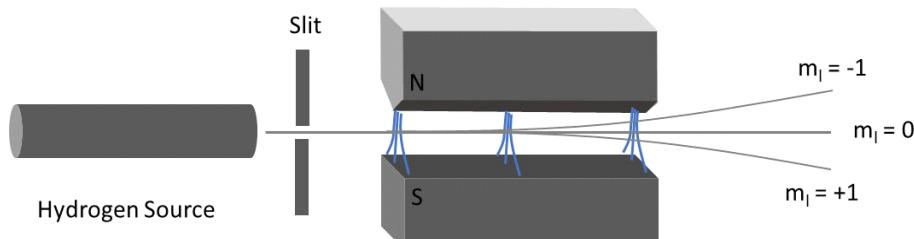
Back when we studied the infinite square well we briefly considered a two-dimensional square well and found that we could make standing waves with the same energy in different ways.



And we now know that we call such waves *degenerate*, meaning that they have the same energy. We use energy to do problems in physics, and these waves all have the same energy, so in problems they would all look alike. We need some way to make the degenerate wave states visible. And in our last lecture, we found that electrons in orbitals have magnetic moments. And this gives us an idea of how to see the degenerate states. If we push on the electron orbitals with external magnetic fields, we could slightly change the energy of the standing wave, and then it would be distinguishable from the other standing wave states in physics experiments! And that is just what we are going to do to investigate the inner workings of our atoms.

Stern-Gerlach Experiment

We now know quite a lot about angular momentum in Hydrogen atoms. We could design an experiment to see the different angular momentum states. Suppose we could produce Hydrogen atoms in an excited state, say, the $(2, 1)$ state so that we could have quantum numbers of $(2, 1, -1)$, $(2, 1, 0)$, and $(2, 1, +1)$. We could do this by heating the hydrogen atoms in an oven. The thermal energy would be used to excite the electrons into higher orbitals. Then we could allow the hot hydrogen to shoot out through a region with a strong non-uniform magnetic field.



The hydrogen will eventually decay back to the ground state, so if we want to see the excited states we can't make the hydrogen atoms move far. But if we keep the distances small, we can see the effect of the atom's angular momentum. Because some of the atoms will have $m_l = 0$, we would predict that these atoms will go straight through the region with the B -field. These atoms have zero angular momentum, and therefore will have zero magnetic moment. They will be unaffected by the external magnetic field.

But for the atoms with $m_l = \pm 1$ the atoms will feel a push from the field because they do have a magnetic moment. The $m_l = +1$ atoms will go one direction and the $m_l = -1$ atoms will go the other direction. We expect to detect the hydrogen atoms in three clumps. One clump for each state

$$(2, 1, -1), (2, 1, 0), (2, 1, +1)$$

Two researchers, Stern and Gerlach, came up with this idea and performed an experiment. And what they found was that our predication was wrong. They got six groups of electrons! The angular momentum did split the atoms into groups, but apparently we missed some angular momentum somewhere.

The answer to this dilemma is that we dealt with the angular momentum of the electron orbits. Now it is not true that the electrons orbit, but the wave function of the electron about the nucleus has orbital angular momentum. But a little like an orbiting planet can not only move around the star it orbits, but can also spin as it goes, we could have an amount of intrinsic angular momentum attributed to the electron. Because the solution

came from the analogy with planetary orbits, this intrinsic angular momentum is called *spin angular momentum*. But the electrons are waves, so it's not correct to say that the electrons turn on an axis like a planet. Instead we think of this as another contribution to the total angular momentum of the atom, but one that is part of the intrinsic nature of the electron. Electrons have a specific spin angular momentum in the same way they have an amount of charge. It is independent of them being in an atom. We won't go through the modification of the Schrödinger equation to show how the spin angular momentum quantum number arrives. But we will define a quantum number

$$m_s = \pm \frac{1}{2}$$

so that we gain an additional spin angular momentum of

$$S_z = \pm \frac{1}{2} \hbar = m_s \hbar$$

and we get an additional magnetic moment of

$$\mu_s = -\frac{e}{m} \mathbf{S}$$

that accounts for the additional splitting of our hydrogen atoms. The atoms in our experiment would have quantum numbers of the form

$$\left(2, 1, -1, +\frac{1}{2}\right), \left(2, 1, -1, -\frac{1}{2}\right), \left(2, 1, 0, +\frac{1}{2}\right), \\ \left(2, 1, 0, -\frac{1}{2}\right), \left(2, 1, +1, +\frac{1}{2}\right), \left(2, 1, +1, -\frac{1}{2}\right)$$

Six total just as the experiment demonstrated.

You might wonder what you would see without the magnetic field. And that would be just one group of hydrogen atoms. You might also notice that we don't walk around all day in strong non-uniform magnetic fields. So most of the time we wouldn't see this effect. But such experiments show us how atoms are made. And so they are valuable. In this case, we can see that the $(2, 1)$ state has a six-fold degeneracy that is broken (so we see all six states) in a non-uniform magnetic field. For most of the time we might not see an obvious result of all these quantum numbers. But internally, these quantum numbers are the clue to how atoms (and chemistry) work.

Spectroscopic notation and Selection rules

Because you only see the effects of some of the quantum numbers for hydrogen atoms when they are in magnetic fields, you might guess that it took a while before they were discovered. There was a previous way to denote the quantum state of the electrons that was used before we even knew there were quantum states. It comes from naming the emission lines that we studied before. The spectral emissions also depend on the

hydrogen atom's electron orbital structure. So there should be a tie between our wave functions and the spectral lines. But because of the degeneracy of our electron states when they are not in a magnetic field, the spectral notation only deals with n , the principle quantum number, and l . For each n we have

l	0	1	2	3	4	5	6
Symbol	s	p	d	f	g	h	i

The letters don't make much sense. Originally they stood for sharp, principle, diffuse, and fundamental. These were words that described the emission lines. But after the first four you can see they just went down the alphabet. You have likely heard of these from high school chemistry. We put the value of n with the symbol for l to describe the state of electrons. You know a $1s$ electron is in the lowest "shell." So our principal and angular momentum quantum numbers l are telling us something about where the electrons will be detected in the atom!

Selection rules

If we did quite a lot more math (we won't) we could use our solutions to the Schrödinger equation to find probability of an electron transitioning from one quantum state to another. The results of these calculations give a series of rules that tell us what transitions are most likely. They are called "selection rules." Our first one is that

$$\Delta l = \pm 1$$

So if we have a hydrogen atom in the $(3, 2, 1, \frac{1}{2})$ state it could transition to the $(2, 1, 1, \frac{1}{2})$ state because

$$\Delta l = l_{final} - l_{initial} = 2 - 1 = 1$$

and this is "allowed" by our probability calculations. That is, it is a transition that is highly probable. We could also transition to the $(1, 1, 0, \frac{1}{2})$ state. But we are very unlikely to see a transition to the $(2, 0, 1, \frac{1}{2})$ because this would be a Δl of 2 and that is less likely.

This selection rule is a great success of our quantum model of the atom. We see just these same transition probabilities in experiment. So something seems to be right in our model! But without the math it is less satisfying to simply state the rule. Let's motivate it a little by recalling that when an electron transitions from a higher state to a lower state it emits a photon. But in the transition we might lose angular momentum, say, going from an $l = 2$ to and $l = 1$ state. We still need conservation of angular momentum. So the angular momentum must go somewhere. And of course it goes into the photon. Photons have spin angular momentum of $m_s = 1$. This is why a Δl of ± 1

is so much more likely. The loss of electron angular momentum must match the gain of photon angular momentum.

Zeeman Effect

We can get another selection rule by considering another experiment. Once again we will place our hydrogen atoms in an external magnetic field. The atoms may have a net magnetic moment, and if they do they will tend to align with the external field. We know from PH220 that there is a potential energy of the form

$$U = -\mu_L \cdot \mathbf{B}$$

and the potential energy is higher when the magnetic moment is anti-aligned with the field. So the atoms will tend to align their magnetic moment with the field to reach the lowest energy state (like a ball rolling down a hill). We could put this potential energy in the Schrödinger equation and solve it again with this new potential, but let's not. We can use what we know and our quantum numbers to get a good idea of what will happen. We can write U as

$$U = -\mu_z B = m_l \mu_B B$$

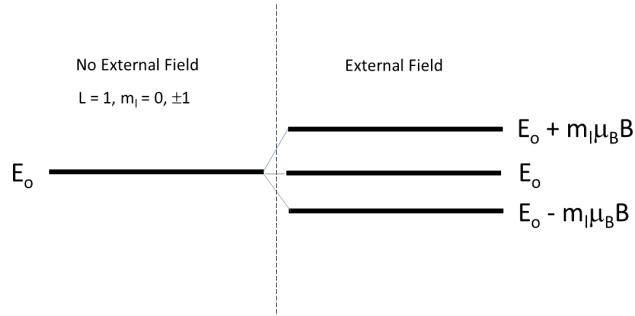
With no external magnetic field the electron has some energy that we can call E_0 . But when the external magnetic field is turned on we would the electron would have more energy due to the external magnetic field

$$E = E_0 \pm U$$

This just modifies our Schrödinger equation so that where there was one solution with $E = E_0$ now we have three solutions with energies

$$\begin{aligned} E &= E_0 - U \\ E &= E_0 \\ E &= E_0 + U \end{aligned}$$

and we could represent this with our energy diagram with the E_0 line broken into three lines.

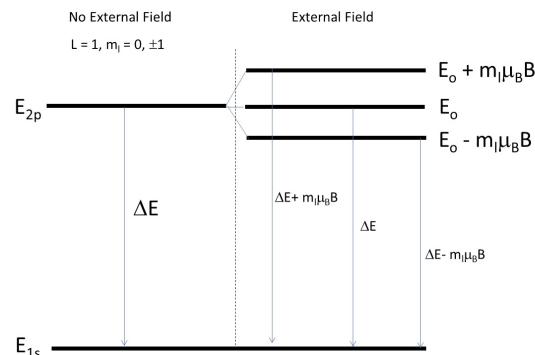


This will show up in the spectral lines of the atom. In the absence of the external field, we would get transitions from, say, the $2p$ state to the ground state where the energy of the photons would be

$$\Delta E = E_{2,1} - E_{1,0} = E_o - E_{1,0}$$

But if our atom is in an external magnetic field we could get

$$\begin{aligned}\Delta E &= E_{2,1} - E_{1,0} \\ &= E_o \pm U - E_{1,0} \\ &= \Delta E \pm U\end{aligned}$$



And because the energy is different, we expect the frequency to be different

$$E = hf = \frac{hc}{\lambda}$$

and

$$|dE| = \left| \frac{hc}{\lambda^2} d\lambda \right|$$

so we could write

$$\Delta\lambda = \frac{\lambda^2}{hc} \Delta E$$

and see that the change in the wavelength is proportional to the change in the energy.

These changes are small, but can be easily seen in spectrographs.

We have done it again. We have used magnetic fields to look carefully at the electron orbitals. Degenerate states were forced into different energies by the external magnetic field and that made them visible. Once again we would need to find the wave functions and calculate the transition probabilities to derive this (we won't in this class), but we would get a new selection rule

$$\Delta m_l = 0, \pm 1$$

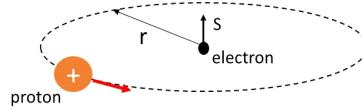
and transitions where Δm_l is 2 or more are much less likely to happen.

Fine Structure

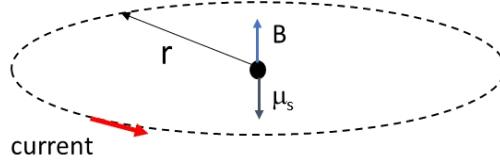
You would think we would be done with magnetic fields and hydrogen atoms, but there is one more consideration we need to make. And this one is harder to think about. Once more, let's use the Bohr model (Because it is easier, like using Newton's laws is easier than special relativity). But it is not too bad under some circumstances.

Lets say that in our Bohr model the electron and protons move like particles in the atom, so they are making magnetic fields. The proton magnetic field will affect the electron because the electron has spin angular momentum so it has a spin magnetic moment. Notice that this is different than our Zeeman case. We don't have an external magnetic field. We are talking about magnetic fields made by the charged wavicles that are part of the atom.

The calculation is easier to do if we consider a strange view point. Think of sitting on a back porch watching the sun go down. We know that this isn't what really happens. The Earth is going around the Sun, so it is the Earth's movement that makes it look like the Sun is going down. But from a back porch where all we care about is how much light we are getting from the Sun, it is perfectly fine to model this as though the Earth is stationary and the Sun moves. The coordinates are strange, but we can totally do the math to make it work. We could do the same to find the magnetic field due to the relative motion between the electron wavicle and the proton in the hydrogen atom. All we care about is the view point from the electron, because we want the potential energy due to the magnetic field as seen from the electron (as though we were viewing from a tiny back porch on the electron). We can do the calculation by taking the coordinate system where the electron is stationary and the proton is moving. We are using Bohr model so we can treat the wavicle electron and proton like a particle. From this view point the situation is as shown in the next figure.



This amounts to a current loop



So we will have a magnetic field upward (think right hand rule) and our electron magnetic moment will be affected by this proton magnetic field. We could have for the electron

$$S_z = \pm \frac{1}{2} \hbar$$

and this would give a magnetic dipole potential energy of

$$U_s = \pm \mu_B B_p$$

where we can estimate the proton magnetic field as

$$B_p = \frac{\mu_0 I}{2r} = \frac{\mu_0 e}{2r T} = \frac{\mu_0}{2r} \frac{ev_t}{2\pi r}$$

this would give an energy separation for the split states of

$$\begin{aligned} \Delta E &= 2\mu_B B \\ &= 2\mu_B \frac{\mu_0}{2r} \frac{ev_t}{2\pi r} \\ &= \mu_B \frac{\mu_0 ev_t}{2\pi r^2} \end{aligned}$$

Now let's use the Bohr model quantization of orbital angular momentum

$$mr v_t = n\hbar$$

to write

$$v_t = \frac{n\hbar}{mr}$$

so

$$\begin{aligned} \Delta E &= \mu_B \frac{\mu_0 e}{2\pi r^2} \left(\frac{n\hbar}{mr} \right) \\ &= \mu_B \frac{\mu_0 e n \hbar}{2\pi m r^3} \end{aligned}$$

and recalling that

$$\mu_B = \frac{e\hbar}{2m}$$

we have

$$\begin{aligned}\Delta E &= \mu_B \frac{\mu_o e}{2\pi r^2} \frac{n\hbar}{mr} \\ &= \frac{e\hbar}{2m} \frac{\mu_o e n \hbar}{2\pi m r^3} \\ &= \frac{\mu_o e^2 n \hbar^2}{4\pi m^2 r^3}\end{aligned}$$

and we know the Bohr radius for the ground state orbit

$$r_1 = a_o = \frac{4\pi\epsilon_o\hbar^2}{e^2 m} n^2 \quad (18.1)$$

so

$$\begin{aligned}\Delta E &= \frac{\mu_o e^2 n \hbar^2}{4\pi m^2 \left(\frac{4\pi\epsilon_o\hbar^2}{e^2 m} n^2 \right)^3} \\ &= \frac{\mu_o e^2 n \hbar^2 e^6 m^3}{4\pi m^2 (64\pi^3 \epsilon_o^3 \hbar^6 n^6)} \\ &= \frac{\mu_o e^8 m}{256\pi^4 \epsilon_o^3 \hbar^4 n^5}\end{aligned}$$

It is traditional to define a constant

$$\alpha = \frac{e^2}{4\pi\epsilon_o\hbar c}$$

and now remember from PH220 that

$$c = \frac{1}{\sqrt{\epsilon_o \mu_o}}$$

so that

$$\alpha = \frac{e^2 \sqrt{\epsilon_o \mu_o}}{4\pi\epsilon_o\hbar}$$

This is called the *fine structure constant*. And our energy split causes line splitting in the spectrum that is known as the *fine structure*. We can write ΔE in terms of the fine structure constant. First let's compute α^4

$$\begin{aligned}\alpha^4 &= \frac{e^8 \epsilon_o^2 \mu_o^2}{256\pi^4 \epsilon_o^4 \hbar^4} \\ &= \frac{e^8 \mu_o^2}{256\pi^4 \epsilon_o^2 \hbar^4}\end{aligned}$$

This looks a lot like our energy

$$\begin{aligned}\Delta E &= \frac{\mu_o^2 e^8}{256\pi^4 \epsilon_o^2 \hbar^4} \frac{m}{\mu_o \epsilon_o n^5} \\ &= \alpha^4 m c^2 \frac{1}{n^5}\end{aligned}$$

The value of α is approximately

$$\alpha = \frac{1}{137}$$

and for a hydrogen atom in the $n = 2$ state we would have

$$\begin{aligned}\Delta E &= \left(\frac{1}{137}\right)^4 (0.511 \text{ MeV}) \frac{1}{2^5} \\ &= 4.533 \times 10^{-11} \text{ MeV} \\ &= 4.533 \times 10^{-5} \text{ eV}\end{aligned}$$

We used the Bohr model so our estimate can't be exact, but just by chance it is really good. The experimental value is 4.54×10^{-5} eV. Don't take this match too seriously, It just happened to work out. Our Bohr approximation can only be taken to give order of magnitude type values, and this one happened to be close but it is really not an indication of something right about the Bohr model. It is just luck. We really should (but won't) do the math using waves to show this value is predicted by the relativistic Schrödinger equation.

You might wonder what all this fuss about hydrogen atoms is all about. Though hydrogen is prevalent in the universe, there are other elements like carbon and iron. But what we have been doing with hydrogen is to build intuition on how larger atoms will work. We expect that they will have energy levels with a series of quantum numbers (n, l, m_l, m_s) and that the electrons will follow a pattern of "filling" the atom with electrons that have different quantum numbers. We won't prove this in our class, but Wolfgang Pauli did long ago. It turns out that in an atom, no two electrons can have the same set of quantum numbers (n, l, m_l, m_s) . This is called the *Pauli exclusion principle* and it is where we will start our next lecture.

19 Many Electron Atoms

8.1 - 8.4

Fundamental Concepts in the Lecture

- Pauli Exclusion Principle
- Electron states in atoms
- The periodic table
- Electron Screening
- Optical transitions
- Properties of the elements

Pauli exclusion principle

In our last lecture we found that the atom has a detailed structure of possible standing waves that we call orbitals around the nucleus. We found the details of the orbital structure using magnetic fields. But now it is time to figure out what those orbitals do.

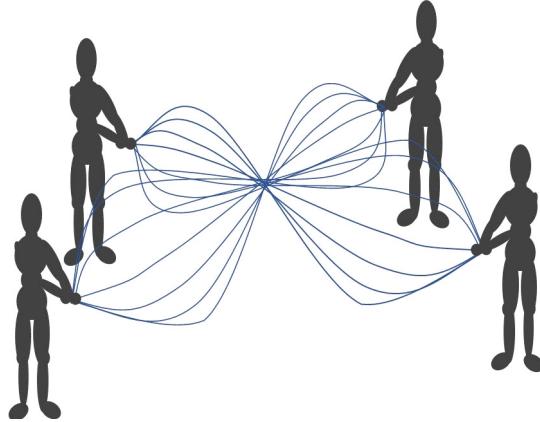
We also found that we could borrow the results of some lengthy numerical calculations and use quantum numbers and selection rules taken from the results of those calculations and gain understanding of how the hydrogen atom works. We are going to continue this process, avoiding the numerical work, but using the results to gain insight.

The first of these results comes from Wolfgang Pauli who had to do his work experimentally because digital computers had not been invented yet. Pauli noticed experimentally that not all the emission lines show up as absorption lines. This led him to postulate that no two electrons can have the same set of quantum numbers (n, l, m_l, m_s) . This has become known as the *Pauli exclusion principle*.

This principle explains that we can't have all the electrons in the ground state. Since in the Hydrogen ground state there are two possible quantum number sets $(1, 0, 0, \frac{1}{2})$

and $(1, 0, 0, -\frac{1}{2})$, we can only have two electrons in the ground state. This is really an expression of the idea that it takes space for a standing wave, and too many standing waves in a volume of space will interfere with each other, destroying themselves.

Think again of trying to make standing waves on jump ropes. We found before that if we added jump ropes, we could make more than one standing wave in the same general place. But after a while, adding more waves would make the standing waves interfere with each other.



Only certain combinations of waves could exist in the same location. Our electron standing waves are a lot like this. Pauli's exclusion principle just tells us that the electron waves interfere with each other if we put two electrons in exactly the same standing wave form in the same location. Only one particular standing wave type can fit in a particular space.

Pauli's principle even tells us that existing standing waves will resist the introduction of more standing waves beyond what will fit. In physics, we have a name for something that pushes back on an object. It is a force. So the electron standing waves can exert a force on other electrons, forcing them into higher energy states where they can form new standing waves without conflict.

And this becomes the basis of filling the electron states in the atom. Each electron will have its own unique set of quantum numbers representing a unique standing wave pattern. So for every unique set of quantum numbers we could have an electron. If we try lithium, we have three total electrons. Then two will likely be in a ground state, and one will be forced into the next state up. The electrons, like balls in a gravitational potential well, will tend to the lowest energy state. So we would guess we would have electrons in the $(1, 0, 0, \frac{1}{2})$, $(1, 0, 0, -\frac{1}{2})$, and $(2, 0, 0, +\frac{1}{2})$ or $(2, 0, 0, -\frac{1}{2})$ states. Three in all because lithium has three electrons.

We would expect that electrons would move to as low and energy as possible. So the lithium atom is very unlikely to have electrons with quantum numbers $(1, 0, 0, \frac{1}{2})$, $(1, 0, 0, -\frac{1}{2})$, and $(5, 1, 0, +\frac{1}{2})$, for example under normal conditions (like on a planet surface, or in free space). That last electron won't want to stay in a $5p$ state if there is a $2s$ state that isn't occupied.

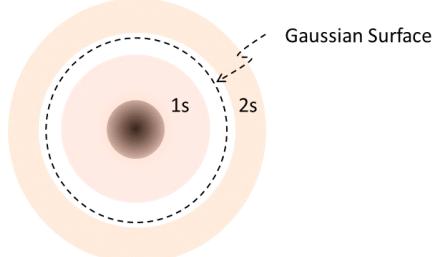
Then atoms with more than one electron would seem easy to understand. You look for the possible quantum numbers, and fill the states from the lowest energy to the highest. But there are some complications.

Screening in Many-Electron Atoms

Let's consider an atom like Lithium again. It would have two electrons in a $1s$ state and one in a $2s$ state. But let's think about the potential energy for that last electron. The potential energy would look like

$$U = \frac{1}{4\pi\epsilon_0} \frac{e(3e)}{r}$$

if the other two electrons weren't there. But they are there and we know from our study of Hydrogen that they are likely to be detected closer to the nucleus than the $2s$ electron. We can use Gauss' law to think about this situation.



The electric field felt by the $1s$ electrons is due to the positive nucleus and has a charge of $3e$. But for the $2s$ electrons, we could take a Gaussian surface in between the most probable radii for the $1s$ and $2s$ electrons and find that the $1s$ electrons would diminish the total charge inside the Gaussian surface. If this was a Bohr model atom, they would reduce the charge inside to just e . We would get Bohr energy states like

$$\begin{aligned} E_n &= (-13.5 \text{ eV}) \frac{Z_{eff}^2}{n^2} \\ &\approx (-13.5 \text{ eV}) \frac{(3-2)^2}{2^2} \\ &= -3.375 \text{ eV} \end{aligned}$$

But because we have a quantum atom things are less exact. The electrons don't just

exist right at the most probable radius for the $1s$ state. The wave function extends into the $2s$ state. But just the same, the field would be less, and we know that the field is what generates the force that causes the potential energy

$$F = -\frac{dU}{dr}$$

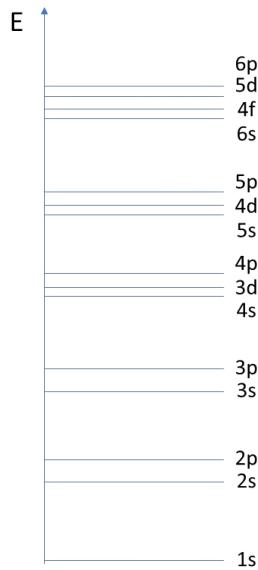
$$eE = -\frac{dU}{dr}$$

$$E = -\frac{dV}{dr}$$

So the potential is less because the field is less. This effect gets more pronounced as we examine larger atoms with more electrons. And it gets a little harder to calculate because the p and d orbitals are stranger shapes. A $2s$ orbital actually extends all the way to the nucleus, so it will feel more than the effect of just one proton in the Lithium case. But if we have a $2p$ electron in an atom, its orbital goes to zero at the nucleus location, so we would expect to feel more of the charge from the $1s$ electrons.

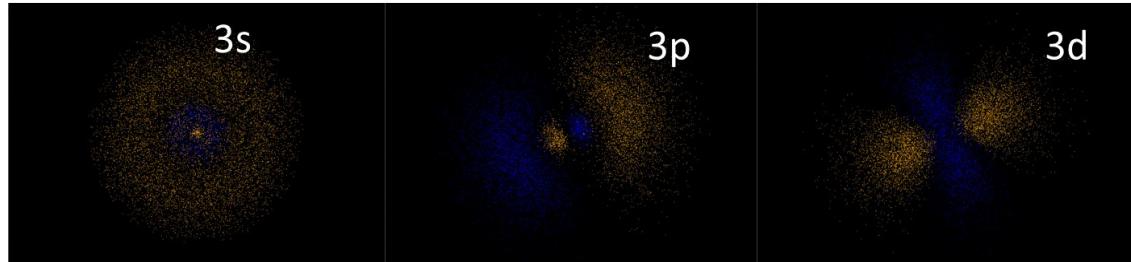
This canceling of part of the nuclear charge due to more inner electron orbitals is called *screening*. And it makes the calculations of energy states more exciting. But we can see that because of screening, inner electrons will be more tightly bound than outer electrons because the outer electrons see a lower effective nuclear charge. The field at their location is less, so the binding force is less.

Extensive numerical calculations give the following general relative energy levels for many-electron atoms.



Notice the screening effects. The s and p orbitals have higher probabilities closer to the

nucleus so their energy levels are drug down the energy diagram. The d orbitals have less probability of being near the nucleus. So we see that the $4s$ and $4p$ levels are low enough to be right around the $3d$ level. This pattern persists as we go up the energy levels. This is a somewhat generic treatment. For actual atoms, a numerical calculation needs to be done to see exactly what the energy of each state will be. From such a calculation we can see that the s state goes deep into the atom and even into the nucleus, but the p and d states appear to have a node at the nucleus location. Here are the $3s$, $3p$, and $3d$ orbitals plotted as point clouds so you can see into the orbital structure.



The $3s$, $3p$, and $3d$ orbitals for hydrogen plotted as point clouds using David Manthey's Orbital Viewer www.orbitals.com/orb

Keeping with our somewhat generic treatment, we know from Hydrogen that all the electrons with the same n quantum number will have about the same average distance for their most probable detection location. The s states are more uniform, the p have lobes that extend further, but the average distance is about the same. The set of orbitals that have about the same average distance from the nucleus is called a *shell*. And because these groups tend to share n quantum numbers, the shells can be related directly to those n values. The shell symbols are as follows.

n	1	2	3	4	5
Shell	K	L	M	N	O

The various l quantum number values for a given n quantum number are called *subshells*. So in the L shell we could have two subshells, $2s$ and $2p$. And using the Pauli exclusion principle and the tight relationship between subshells and the l quantum number we can see that each subshell can have

$$2(2l + 1)$$

electrons where the $(2l + 1)$ comes from the number of m_l possibilities for each l and the extra 2 out front comes from the two m_s possible values.

Combining all this, we can see that our generic order for filling atoms with electrons

goes as follows.

n	l	Subshell	Capacity $2(2l + 1)$
1	0	$1s$	2
2	0	$2s$	2
2	1	$2p$	6
3	0	$3s$	2
3	1	$3p$	6
4	0	$4s$	2
3	2	$3d$	10
4	1	$4p$	6
5	0	$5s$	2
4	2	$4d$	10
5	1	$5p$	6
6	0	$6s$	2
4	3	$4f$	14
5	2	$5d$	10
6	1	$6p$	6
7	0	$7s$	2
5	3	$5f$	14
6	2	$6d$	10

Optical Transitions

@@@@@@@ Not Done @@@@@@@@

The Periodic Table of the Elements

The Periodic Table of the Elements was developed based on chemical properties. But those chemical properties of the elements must come from the atomic structure. And the atomic structure comes from quantum waves forming orbitals. We can map what we know about orbitals to our periodic table to show how the orbital structure makes the chemical properties happen.

Alkalies																			Inert Gases		
1s	1 H	Alkaline Earths																		2 He	
2s	3 Li	4 Be																			
3s	11 Na	12 Mg																			
4s	19 K	20 Ca	Transition Metals																		
5s	37 Rb	38 Sr	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn									
6s	55 Cs	56 Ba	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr	Halogens		
7s	87 Fr	88 Ra	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe			
		5d	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Uub	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn			
		6d																			
		4f	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb					
		5f	89 Ac	90 Th	91 Pa	92 Np	93 Pu	94 Cm	95 Am	96 Cf	97 Bk	98 Es	99 Fm	100 Md	101 No						
		Lanthanides (Rare Earths)																		Actinides	

Notice that the Alkalies all have s states for their uppermost electrons. The lower states are all filled, and they end the filling on an s state. But in each case the final s subshell is not full. This leaves an electron without a balancing electron with the opposite spin state. Helium has a full s subshell, but Hydrogen does not. And this creates the chemical difference between hydrogen and helium. Hydrogen is very reactive. It has a single electron in a subshell that isn't full. This electron can participate in bonds. But Helium doesn't easily participate in bonds at all. Its highest subshell is full. And these properties follow as we go down the columns in the periodic table. All the Alkali elements have an unmatched electron in a s subshell and are therefore reactive. All the Inert gasses have full subshells and are not reactive.

Note the structure of the table. The first row has just hydrogen and helium. That takes care of all the $1s$ states where there are no more electrons. But in the next row we have more electrons. Lithium and Beryllium take the two $2s$ electrons and we are out of $2s$ states. We then start on the $2p$ states and those atoms are across the chart in the section leading up to the halogens. As we go down the chart we see that all the rows on the far right hand side are p states. Going down the chart we get larger atoms with more electrons. Eventually we will start filling d states. And this gives us the transition metals.

Remember in our generic energy diagram that the $4f$ state energy got mixed up with the $6s$ and $6p$ states because of screening. We see this in our periodic table. Really the table should look like this.

The table is a standard periodic table where each element is represented by a small square containing its atomic number. Key features include:
- **Alkalies (1s):** Hydrogen (H) and Lithium (Li) through Francium (Fr).
- **Alkaline Earths (2s):** Beryllium (Be) and Magnesium (Mg) through Barium (Ba).
- **Lanthanides (Rare Earths) (4f):** A group of 15 elements from Lanthanum (La) to Lutetium (Lu).
- **Transition Metals (3d, 4d, 5d):** Groups of elements with partially filled d-subshells.
- **Halogens (2p, 3p, 4p, 5p):** Groups of elements with one electron in their outer p-subshell.
- **Inert Gases (He, Ne, Ar, Kr, Xe, Rn, Uuo):** Elements with full outer shells, shown in yellow boxes.
- **Actinides (5f):** A group of 15 elements from Actinium (Ac) to Ununquadium (Uuo).

so that the $4f$ states show up in the right order for filling the electron energy states. But this makes the table too wide to easily go in books. So we tend to put the Lanthanide or Rare Earth metals below the rest of the chart. But this is only for convenience in chart making. The $4f$ states belong with the $6s$ states energy wise. The same is true for the $5f$ Actinides.

If we were chemists, we would spend quite a long time on this showing that the energy states for the highest populated subshell produce the chemical bonding properties of the elements. As we are physicists (at least in this class), let's just summarize some results.

Intent gasses

The column that has Helium on top is the Inert gasses They have filled subshells and are very stable. They are generally nonreactive.

Halogens

The column marked with p subshells next to the inert gasses all are one electron short of a filled subshell. They all have their highest populated subshell of the form np^5 . The n is their shell number, and the 5 says they have five electrons in this highest energy subshell. The p subshell can have 6 electrons, so we are missing one. This is the source of their reactivity. If they can borrow an electron from another element their p subshell will be more stable. We won't prove this, but a filled subshell is able to fit in the atomic potential well at a slightly lower energy. So like balls like to roll down hills to get to a lower energy state, atoms like to have filled shells to get to a lower energy state. This may seem strange, but remember that our potential well comes from both Coulomb force and from angular momentum. If we can get the angular momentum of the electrons to, say, cancel out, we will affect the energy state of the whole shell (think of all that angular momentum based splitting that we found in magnetic fields in our last lecture). There is more to it than just this, but it does make sense that a combination of electron waves could have a different energy than those waves would

have independently. And it turns out that the combined wave functions in a filled shell cancels out enough to lower the energy of all the electrons in the shell.

So halogens readily form bonds with atoms that can share an electron. Chlorine readily bonds with Sodium. The Sodium atom has one electron in its outer most *s* state. That electron is less tightly bound because the nuclear charge is screened. It will share this electron with the chlorine atom so the chlorine atom has a full *p* state and the sodium atom will effectively have a full *2s* state so it will be more stable (lower energy). As we look at the other *p* state elements, we change the reactivity and the bonding. Oxygen, for example is missing two electrons in it's uppermost *p* subshell. And we can see why it might bond with two hydrogens to fill this subshell.

Transition metals

This is the first three rows of the section where we are filling *d* states. Some exciting things happen here. Let's look at Copper. We would expect the configuration $4s^23d^9$ for its final subshells. But a *d* state would like $2(2(2) + 1) = 10$ electrons. A filled subshell has a lower energy state (and therefore is more stable) so the *3d* state borrows an electron from the *4s* state producing $4s^13d^{10}$. Thus copper has an upper single *s* state electron and like the Alkali elements, this electron is loosely bound and can readily be freed. This is one reason copper is a good conductor.

Rare Earth metals

These elements have the same thing going on as the transition metals. They tend to have an inner subshell, the *4f*, that is being filled. But a higher *n* subshell is already filled (*6s*). We should expect some of the same properties as the Transition metals.

The higher number of electrons in the *f* states has effects on paramagnetism. You may have heard of rare earth magnets.

Actinides

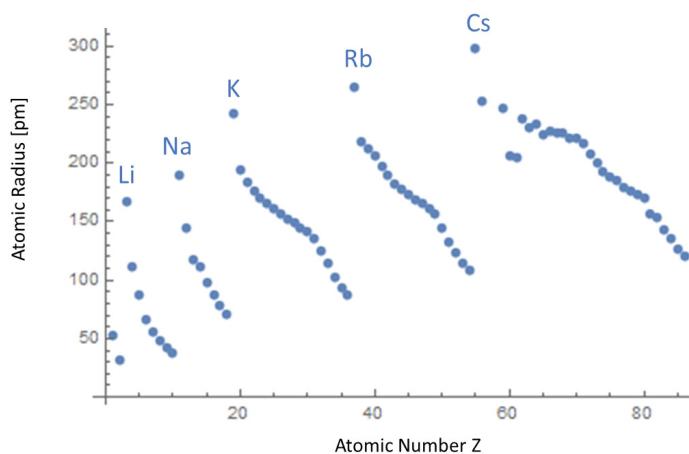
We would expect these atoms to act much like the rare earth group. But here we are getting to the end of what we can do with our quantum atom structure so far. We find that many of the actinides are radioactive. That is not predicted by our electron

structure. We will need to study more atomic structure, specifically the nuclear structure, to understand this.

Graphs of atomic properties.

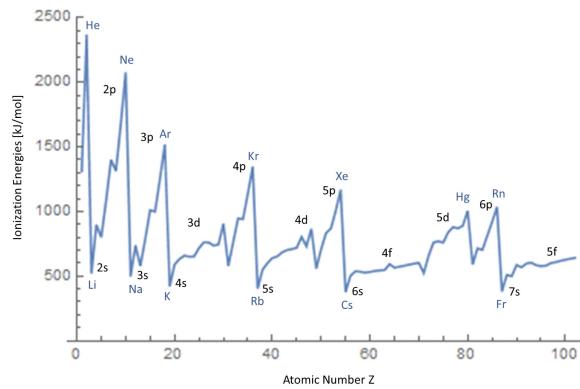
We would expect our quantum structure to be radially visible in the properties of the elements. And we can see that is true by graphing properties by element Z number.

For example, we expect the atomic radius to decrease as the nuclear charge increases. The nuclei with more protons have a stronger electrical force (or deeper potential well) and therefore electrons in these atoms are held more tightly. So as we fill p and d and f shells, the atomic radii should decrease. But s states are large, their orbital probabilities have a broad fall off as we saw for hydrogen atoms. So the atomic size should increase whenever we hit an s state. This is just what we see.



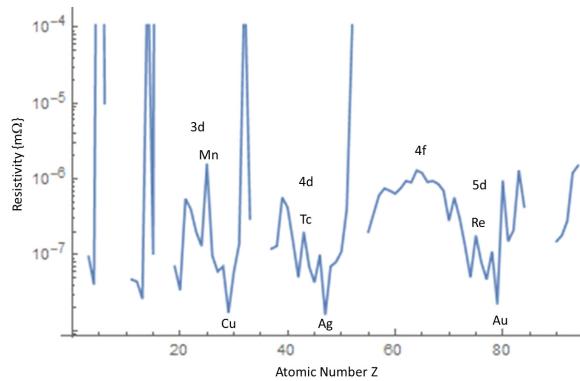
Atomic Radii Plotted by Atomic Number. Atomic properties plotted using the Mathematica ElementData function.

How tightly the atom holds on to its outer electrons should also show us signs of our quantum atomic structure. Filled shells should require more energy to remove the electrons. So as shells fill, we should see the ionization energy increase. Here is a graph of the ionization energy.



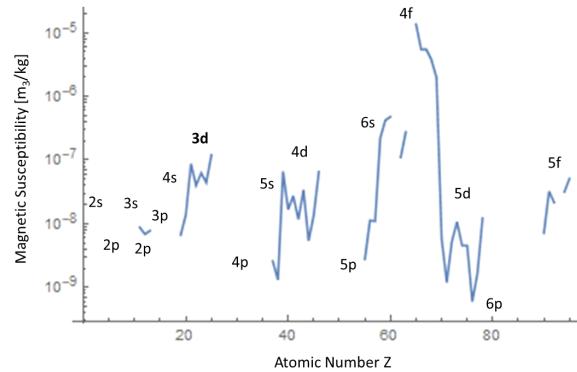
We see the inert gasses require the most energy to remove an electron. We can see the Alkalies require the least. Those *s* subshell electrons are easy to remove.

More complicated properties that involve larger quantities of atoms are harder to analyze, but we can still see some structure in our curves that is due to our quantum model of the atom. For copper, we found the subshell structure allowed a relatively free electron to participate in current flow. Then resistivity should show effects of subshell structure. Here is the electric resistivity plotted as a function of atomic number.



From this plot we could predict the sort of subshell structure we see in *Cu* would also be present for *Ag* and *Au*.

And one more bulk property plot, this time for magnetic susceptibility, shows some patterns that are due to our quantum atomic structure.



Atomic Magnetic Susceptibility Plotted by Atomic Number. Atomic properties plotted using the Mathematica ElementData function.

We could make many more such plots. But for now we can say that our new quantum model of the atom seems to be very successfully explaining many experimental results.

20 Electron Inner Structure

8.5-8.7

Fundamental Concepts in the Lecture

- Absorption edges
- X-ray transitions of inner electrons
- Moseley's law
- Addition of Angular Momenta
- Stimulated emission
- Lasers

Addition of angular momentum

You might be getting the feeling that with all our quantum numbers and selection rules that we are obscuring the physics of what is going on in the atom. The truth is that the math is doable, but difficult enough that it is hard to keep the physics in mind even if we did the details of calculating the orbitals. But we do need to remind ourselves of the physics of what we are doing. Let's take a moment and think about one of the implications of what we have done.

We started collecting quantum numbers by finding the allowed energy values and the allowed angular momentum values for an electron in an atom (hydrogen). We found that the n, l, m_l , and m_s quantum numbers were the clue to how to build the electron wave structure of the atom. But the l, m_l and m_s quantum numbers do stand for angular momentum. So we could reasonably ask if the atom, itself, has an angular momentum. And of course the answer is yes!

The angular momentum of the atom is largely due to the electron angular momenta. From our study of mechanics (PH121) we know that angular momentum is a vector and angular momenta add like vectors. But now we have quantized vectors. How do we

modify our vector addition so we ensure we end up with allowed angular momentum states for the atom?

We know that orbital angular momentum should follow rules like

$$\begin{aligned} |\mathbf{L}| &= \sqrt{l(l+1)}\hbar \\ L_z &= m_l\hbar \quad m_l = 0, \pm 1, \pm 1 \dots \pm l \\ S_z &= m_s\hbar \quad m_s = \pm \frac{1}{2} \end{aligned}$$

and this must be true for the overall atom as well. We expect the same sort of quantization. But now we have more than one angular momentum to add. That is, when we add angular momenta the result still has to follow our quantization rules for angular momentum.

The first thing to realize is that a filled shell has zero net angular momentum. Let's look at this. A filled ns state will have two electrons with quantum numbers

$$\left(n, 0, 0, \frac{1}{2}\right) \left(n, 0, 0, -\frac{1}{2}\right)$$

Due to what we have called space quantization, the L_z values will be specified, but the L_x and L_y values will be uncertain meaning all L_x and L_y values will be experienced with the restriction that

$$|\mathbf{L}| = \sqrt{L_x^2 + L_y^2 + L_z^2}$$

so it is not surprising that the net angular momentum in the x and y directions sums to zero. But we have to look at the z component of the orbital angular momentum and we need to consider the spin angular momentum. We have

$$\begin{aligned} l_1 &= 0 \\ l_2 &= 0 \\ m_{l1} &= 0 \\ m_{l2} &= 0 \\ m_{s1} &= \frac{1}{2} \\ m_{s2} &= -\frac{1}{2} \end{aligned}$$

We can find the angular momentum for each angular momentum part. The z -components give

$$\begin{aligned} L_{1z} &= 0\hbar = 0 \\ S_{1z} &= \frac{1}{2}\hbar \end{aligned}$$

and

$$\begin{aligned} L_{2z} &= 0\hbar = 0 \\ S_{2z} &= -\frac{1}{2}\hbar \end{aligned}$$

and since L_{1z} and L_{2z} are zero and we expect the x and y -parts to sum to zero then

$$|\mathbf{L}_1| + |\mathbf{L}_2| = 0$$

is really the only possibility. Same with m_l

$$m_{1l} + m_{2l} = 0$$

and

$$m_{1s} + m_{2s} = \frac{1}{2} + \left(-\frac{1}{2}\right) = 0$$

there is no net angular momentum for a filled ns state.

A np state is harder. We would have quantum numbers

l	m_l	m_s
1	-1	$\frac{1}{2}$
1	-1	$-\frac{1}{2}$
1	0	$\frac{1}{2}$
1	0	$-\frac{1}{2}$
1	1	$\frac{1}{2}$
1	1	$-\frac{1}{2}$
Total	0	0

And, adding up the m_l and m_s values for a filled np state we still have no net angular momentum. Remembering our last few lectures we know that there was some energy tied up in the angular momentum of the electrons, but if we cancel out all of that angular momentum then that energy would not be there, and the entire filled shell would have a lower energy. This is just what we expected in our last lecture.

But if we don't have a filled shell we should have a net angular momentum. And the only shells that are not filled are the outer most. So let's look at those.

The detailed math tells us that the orbital angular momentum needs to be added in a specific order to get the quantization right. Let's do an atom with two electrons in its outer shell first so it is easy, then we will deal with three or more outer shell electrons. Let's pick carbon with electron structure $1s^2 2s^2 2p^2$. It is those $2p^2$ electrons that will give carbon a net angular momentum. For a single electron we found that we had a rule that

$$|\mathbf{L}| = \sqrt{l(l+1)} \quad l = 0, 1, 2, \dots n-1$$

We can only have certain l values. We expect that we should have a maximum l for the combined angular momentum of the two $2p^2$ electrons when their angular momentum

vectors point the same direction. We take

$$L_{\max} = l_1 + l_2$$

Because both electrons are p electrons both have $l = 1$ so

$$L_{\max} = 1 + 1 = 2$$

We also expect to have a minimum allowed angular momentum that represents both electron's angular momentum canceling. We find this as follows

$$L_{\min} = |l_1 - l_2|$$

so for carbon

$$L_{\min} = |1 - 1| = 0$$

This gives us a new atomic angular momentum quantum number L with allowed states that range from L_{\min} to L_{\max} but L is still quantized so it goes from L_{\min} to L_{\max} in integer steps.

$$L = L_{\min}, L_{\min} + 1, L_{\min} + 2 \cdots L_{\max}$$

For carbon this would be

$$L = 0, 1, 2$$

This gives the total orbital angular momentum for the atom, but we know we need the z component of this as well. We call the L_z atomic quantum number M_L and its summation rule is easy, we just add up the individual m_l values

$$M_l = m_{1l} + m_{2l}$$

and the permitted values are

$$M_l = 0, \pm 1, \pm 2, \cdots \pm L$$

This is just our normal rules for m_l . For the carbon atom we could have the possibilities

L	M_l
0	0
1	-1, 0, 1
2	-2, -1, 0, 1, 2

The atomic spin angular momentum similarly is

$$M_{S \max} = m_{1s} + m_{2s}$$

$$M_{S \min} = |m_{1s} - m_{2s}|$$

For carbon this is simple, it is just

$$M_{S \max} = \frac{1}{2} + \frac{1}{2} = 1$$

$$M_{S \min} = \left| \frac{1}{2} - \frac{1}{2} \right| = 0$$

and we need integer steps, so these are our only possibilities.

But we are not done. We need to check that the Pauli exclusion principle is still met. No two electrons can have the same quantum numbers. Our quantum numbers for the

$2p^2$ electrons can be

n	l	m_l	m_s
2	1	-1	$\frac{1}{2}$
2	1	-1	$-\frac{1}{2}$
2	1	0	$\frac{1}{2}$
2	1	0	$-\frac{1}{2}$
2	1	1	$\frac{1}{2}$
2	1	1	$-\frac{1}{2}$

Our system tells us that we could have $L = 2$ and $M_l = 2$ and $M_s = 1$, but let's see if

that particular state follows the Pauli exclusion principle. To get $L = 2$ we could take any two of the states. To get $M_l = 2$ we would need the last two states with $m_{1l} = 1$ and $m_{2l} = 1$ so then $m_{1l} + m_{2l} = 1 + 1 = 2$. But in this case M_S can't equal 1 because that would require $m_{1s} = \frac{1}{2}$ and $m_{2s} = \frac{1}{2}$ and that would give two states with exactly the same quantum numbers. That is not allowed by the Pauli exclusion principle. All we have in the last two states is $m_{1s} = \frac{1}{2}$ and $m_{2s} = -\frac{1}{2}$ which would give $M_s = 0$. So the atomic angular momentum state $[L, M_l, M_S]$ of

$$[2, 2, 1]$$

is not allowed. But we could have

$$[2, 2, 0]$$

For carbon the allowed atomic angular momentum states would be

L	M_l	M_S
0	0	0
1	-1	0, 1
1	0	0
1	1	0, 1
2	-2	0
2	-1	0, 1
2	0	0
2	1	0, 1
2	2	0

So far, so good! But remember we need to deal with atoms with more than two electrons in the outer shell. Nitrogen, $1s^2 2s^2 2p^3$, seems like a good atom to take on. We start by defining a partial total orbital angular momentum for just two electrons

$$L_{12 \text{ max}} = l_1 + l_2 = 2$$

just like before. We would have

$$L_{12} = 0, 1, 2$$

But now we have three electrons so we couple in another to make sure the end result

follows our angular momentum quantization rules

$$L_{123\max} = L_{12\max} + l_3 = 2 + 1 = 3$$

We need to think a bit to find $L_{123\min}$. We want the smallest minimum that we can get by coupling the 1, 2 angular momentum with the third electron.

$$L_{12\min} = |l_1 - l_2| = 0$$

We know from before that L_{12} can be 0, 1, or 2. Then calculating $L_{123\min}$ could give any of the values

$$L_{123\min} = |L_{12} - l_3|$$

which are

$$\begin{aligned} L_{123\min} &= |0 - 1| = 1 \\ &= |1 - 1| = 0 \\ &= |2 - 1| = 1 \end{aligned}$$

so the smallest of these is

$$L_{123\min} = 0$$

So our atomic L quantum number can have the values

$$L_{123} = 0, 1, 2, 3$$

and our atomic M_L allowed states can be

$$M_l = 0, \pm 1, \pm 2, \pm 3$$

And the atomic spin states would follow the pattern for the L quantum numbers. The partial spin angular momentum would be

$$M_{S12\max} = S_1 + S_2 = \frac{1}{2} + \frac{1}{2} = 1$$

Then coupling in the third electron gives

$$M_{S123\max} = S_{12\max} + S_3 = 1 + \frac{1}{2} = \frac{3}{2}$$

And for the minimum

$$M_{S12\min} = \left| \frac{1}{2} - \frac{1}{2} \right| = 0$$

and coupling in the third electron gives

$$M_{S123\min} = \left| 0 - \frac{1}{2} \right| = \frac{1}{2}$$

so the allowed atomic spin states are

$$M_{S123} = \frac{1}{2}, \frac{3}{2}$$

because we only get integer steps. We still need to do the Pauli test to show which of these are actually allowed. Some will not be allowed by the Pauli exclusion principle.

This always feels to me like a lot of busy work to find the allowed atomic angular

momentum. But remember that we are avoiding a tremendous amount of numerical calculations by using these quantum number rules. So it is busy work with a wide smile.

We know the possible atomic angular momentum states, but we should ask, which of the possibilities is the atomic ground state? There, of course, would be a large numerical computation to find this. But the result would be the same as we would get using the following rules:

1. Find

$$M_{S \max} \equiv S$$

being careful to make sure it is one of the states possible by considering the Pauli exclusion principle. If it's not we check the next lower state until we find one that works.

2. Next, for that M_S find the maximum value of

$$M_{l \max} \equiv L$$

We still have to make sure that our state is possible according to the Pauli exclusion principle. If it's not we check the next lower state until we find one that works.

3. The resulting S, L values give the lowest energy or ground state.

These rules are called *Hund's Rules*. Let's try this for carbon. Only the outermost electrons will matter, because the lower states will all be full and will therefore have no angular momentum. The upper most electrons are $2p^2$ states. We found that $M_{S \max} = +1$ so our M_S value is $M_S = 1$. Because a p state can have three electrons with $m_s = +\frac{1}{2}$ and three with $m_s = -\frac{1}{2}$, and we only have two electrons in our p state, we can totally have two electrons with $m_s = +\frac{1}{2}$ so the $M_S = 1$ state is allowed by the exclusion principle. So far so good.

Now we want a maximum $M_l = m_{1l} + m_{2l}$ and it would be nice to pick both m_l values as 1. But we can't do that by the exclusion principle or we would have both electrons with $(2, 1, 1, \frac{1}{2})$ so we need one of them in a different state. We choose $(2, 1, 0, \frac{1}{2})$ or $m_{2l} = 0$. This gives $M_l = 1 + 0 = 1$ so our atomic ground state is

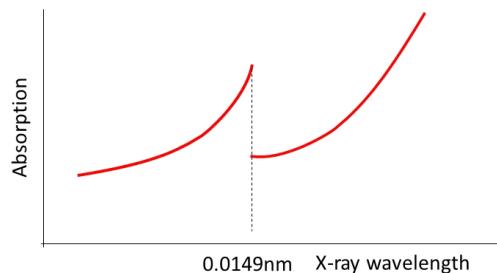
$$[L, S] = [1, 1]$$

This M_l and M_s will have the lowest energy of the many possibilities for two $2p$ electrons, so without additional energy being added to the atom, the two $2p$ electrons will end up in this state.

Transitions involving inner electrons

In what we have done so far we have concentrated on the outer most electron orbitals because they govern chemical reactions. Chemical reactions are important to us living things. So our concentration on these outer shells is justified. But you may have wondered if it is possible to remove an electron from one of the inner shells. These electrons are held much more tightly because the Coulomb force is larger and in addition these shells are full, providing another energy disincentive for an electron to leave. But it is possible if we provide, say, a photon with just the right energy. The energy must match the binding energy of that particular electron.

We could design an experiment where we scatter *x*-rays off of an atom. It is relatively easy to make an *x*-ray source where you can tune the wavelength (and therefore change the energy) of the *x*-rays. So we can irradiate the atom but change the energy until we get just the right one to knock out a $1s$ electron. It's hard to irradiate just one atom, so we could take a thin film of metal atoms (say, mercury held between two microscope slides). Then the *x*-rays that don't have the matching energy would be absorbed by the bulk metal material, but the matching wavelength would knock out an electron. The absorption would stop at that matching frequency. The experimental results look something like this.



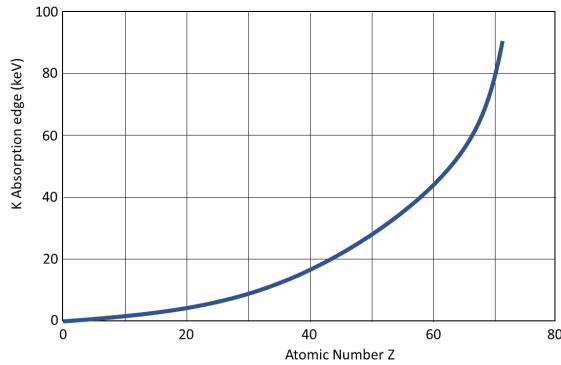
Schematic depiction of an absorption edge for mercury.

The absorption jumps down right at $\lambda = 0.0149 \text{ nm}$. This wavelength has an energy of

$$\begin{aligned} E &= \frac{hc}{\lambda} = \frac{1239.8419 \text{ eV nm}}{0.0149 \text{ nm}} \\ &= 83211. \text{ eV} \end{aligned}$$

which is just what our quantum atomic model gives for a $1s$ electron binding energy for mercury. Because we dealt with a $n = 1$ case, this change in the absorption spectrum is called a *K* absorption edge, but we could target an $n = 2$ electron to get an *L* edge, or in large atoms we could get an *M* edge, etc.

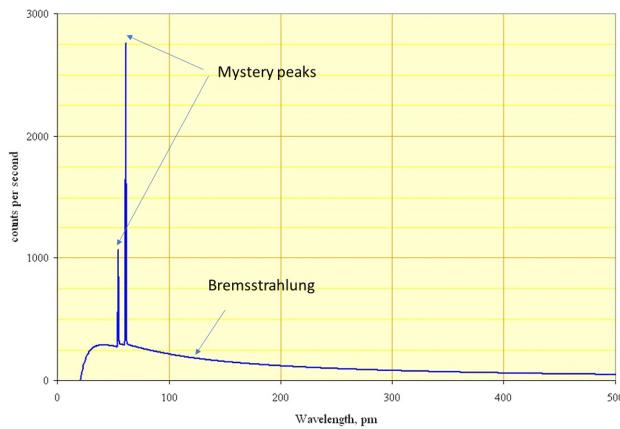
Our quantum model tells us that the $1s$ ionization energies should be large, and should grow with atomic number Z . But we wouldn't expect the sort of structure as we would see with outer electron ionization energy. The outer electrons are lightly bound and come in different shells. The $1s$ shell, is, well, always a $1s$ shell, so if we plot this K shell ionization vs. atomic number we get a smooth curve.



Moseley's Law

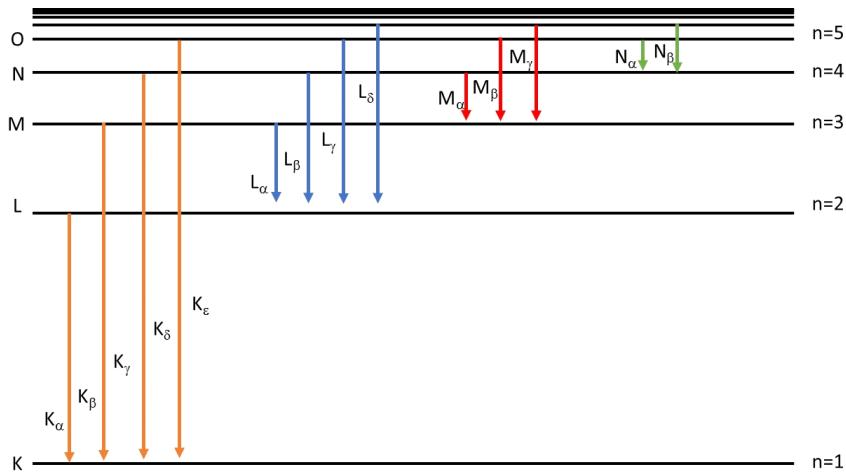
But, you might say, if we remove a $1s$ electron, what happens to that vacancy. Won't one of the other electrons fill that vacancy? and that would leave another vacancy, so wouldn't we get a cascade of electrons shifting downward? That is exactly what happens. And with each transition downward we get photons emitted. So we can see this cascade happen. There is a quirk in the literature in discussing these transitions. They are named for the original vacancy. So if an electron moves down to the $1s$ state from anywhere above on the energy diagram, the transition is called a K -shell transition even though the electron came from somewhere else.

Long ago we studied x -rays. We found that if we stopped electrons they gave off their energy and we called this Bremsstrahlung radiation. But in the plot we used we had not only the broad Bremsstrahlung radiation, but we also saw some large peaks.



X-ray tube with a rhodium target, operated at 60 kV. The continuous curve is due to bremsstrahlung, and the spikes are characteristic K lines for rhodium. The curve goes to zero at 21 pm in agreement with the Duane–Hunt law, as described in the text.

We can now identify these peaks by looking at their energies. They correspond to the ionization energies of the inner electrons. The two mystery peaks that we saw in our graph are from removing $1s$ electrons and then letting higher electrons transition down to the $1s$ shell by emitting photons. There are just two prominent peaks in our experimental graph. But there are often three (or more) such peaks, and they are designated as the K_α , K_β , and K_γ (etc.) x -rays. If we remove an electron from the $n = 2$ shell we would call the transitions the L x -rays. The following figure shows some of the possible transitions



If we think of the inner electrons and Gauss' law again, we can see that the outer electrons mostly don't participate in the Coulomb force for the inner electrons (in the Bohr model they wouldn't matter at all). So we can treat these inner electrons almost as though the other electrons weren't there. Henry Moseley used this idea to predict the transition energies.

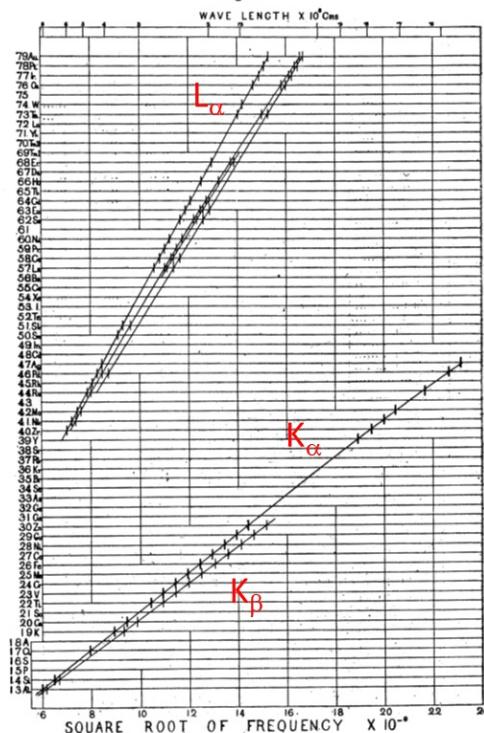
$$\Delta E = E_{n_2} - E_{n_1} = (-13.6 \text{ eV}) (Z - 1)^2 \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right)$$

where n_2 and n_1 are the n quantum numbers for the final and initial energy levels of the transitioning electron. For the K_α transition he got

$$\Delta E_{K_\alpha} = E_2 - E_1 = (-13.6 \text{ eV}) (Z - 1)^2 \left(\frac{1}{2^2} - \frac{1}{1^2} \right) = 10.2 \text{ eV} (Z - 1)^2$$

The frequency of the photon emitted is proportional to ΔE_{K_α} . So our equation says that the frequency of the photon emitted should be proportional to Z^2 . Moseley plotted Z vs. \sqrt{f} and according to our equation this should be a straight line. His data is in the plot below.

High-Frequency Spectra of the Elements. 709
Fig. 8.



From H. G. Moseley, Philos. Mag. (6), 27:703, 1914

The bottom curves are the K series and you see the K_α and K_β lines. The higher

group of lines are the L series. Moseley found he could identify elements this way. If you gave him a sample of the material, he could tell you what element it was. Even better he found several positions on the slope of his line where there should be an atom, but none was known. Technetium was one of these. Technetium is radioactive with a short enough half life (something we will discuss later) that it is not found in nature. But Moseley's chart said there should be an element with $Z = 43$ that would have a particular K_{α} frequency. And sure enough it is possible to make Technetium and Moseley's law correctly predicted the K_{α} frequency.

We have the equipment to do this x-ray work in our laboratories at BYU-Idaho. If you are interested, you can join the research group that does this work.

Stimulated Emission

Here is a transition diagram for helium (below, figure ??). This diagram is somewhat schematic (don't take the energy values to seriously) but it shows what we have been saying. Let's analyze the helium case.

We start with the ground state $1s^2$. We have just two electrons and they are in the same shell so we have quantum numbers

$$\begin{pmatrix} 1, 0, 0, \frac{1}{2} \\ 1, 0, 0, -\frac{1}{2} \end{pmatrix}$$

The m_s values must be different to satisfy the Pauli exclusion principle. We can calculate the atomic angular momentum. We start with

$$L_{\max} = 0 + 0 = 0$$

and

$$L_{\min} = 0 - 0 = 0$$

which gives the only possibility for

$$M_l \max = L = 0$$

and we have just $m_{s1} = \frac{1}{2}$, $m_{s2} = -\frac{1}{2}$ so

$$M_S = S = 0$$

And we started with the Pauli check, so we have found the atomic angular momentum state for helium in the ground state. But what if we excite one of the electrons? Then

we would have $1s^1 2s^1$. Our quantum numbers would be like

$$\left(1, 0, 0, \pm\frac{1}{2}\right)$$

$$\left(2, 0, 0, \pm\frac{1}{2}\right)$$

Our atomic angular momentum still would have

$$L = 0$$

so

$$M_l = 0$$

but now we can have different values for m_{s1} and m_{s2} because the $n = 1$ and $n = 2$ quantum numbers guarantee that the Pauli test will be satisfied for any m_s combination. We could have

$$M_{S \max} = S = m_{s1} + m_{s2}$$

be

$$S = \frac{1}{2} - \frac{1}{2} = 0$$

with two ways that could happen, or

$$S = \frac{1}{2} + \frac{1}{2} = 1$$

We have two cases. The first one is called a singlet state and the second is called a triplet state (because we could have $M_S = -1, 0, +1$ for the $S = M_{s \max} = 1$ case).

We can see in the diagram that we can go beyond the first excited state.

Note that the diagram is split in half. If we have an atomic $S = 0$ starting state we get different energy levels than if we have a $S = 1$ starting state! And of course there would be different emission spectra. Note that in both cases we have many allowed transitions. But some are not allowed. We could do a lot of math (but we won't) to show that this leads to some selection rules on atomic angular momentum

$$\Delta L = 0, \pm 1$$

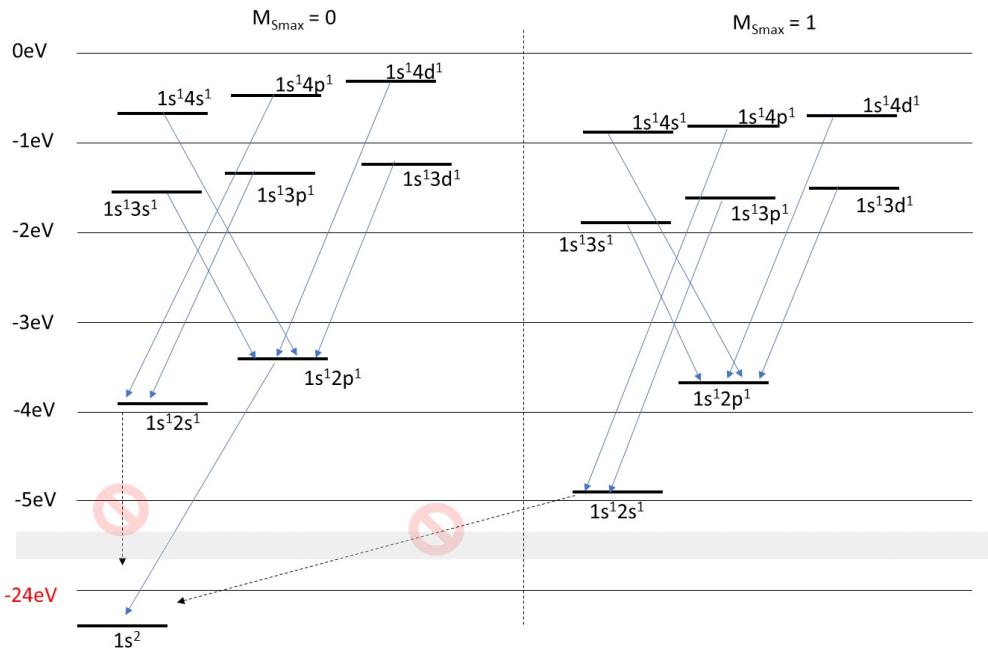
$$\Delta S = 0$$

that must be added to our selection rules we already know

$$\Delta l = \pm 1$$

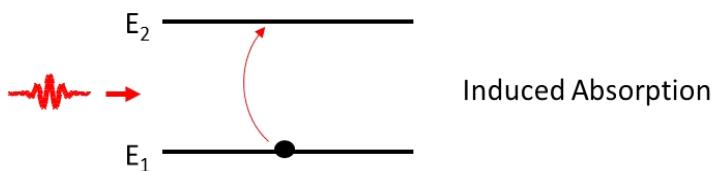
$$\Delta m_l = 0, \pm 1$$

If we have $M_S = 1$ can the electrons never transition to the electron ground state?



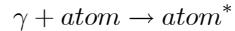
You can guess that the helium always ends up in the ground state eventually. But it takes more time for these transitions to happen because they are less likely. Most transitions are usually almost instantaneous so a long time might be microseconds. Such states are called “metastable” because they really do transition but it takes much more time for them to decay.

So far we have been talking about photons removing electrons or at least pushing them to higher energy states. The photon comes in and the electron absorbs the energy from the photon and jumps to a higher energy subshell. Schematically we could draw this as



Note that the diagram is an energy diagram, so it is not showing position of the photon or the electron. It is just symbolic. The photon doesn't come from the left, for example. Coming from the left is just a symbolic way to show it enters the system. Note that we are using the word absorption slightly differently than we did before. In the previous section it meant that the *x*-rays would become thermal energy in the bulk material. But this absorption is the destruction of the photon with the energy going to excite electron

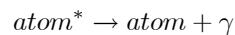
in the atom. We symbolically write this as



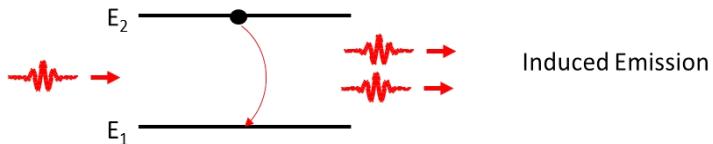
But we also know from what we have done that the electron will eventually fall back down into the energy well and when that happens we get a photon back out. This is called *spontaneous emission*.



We write this as

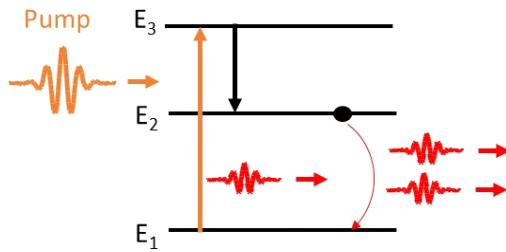


There is another possibility. If the atom is excited, and if another photon of the transition energy difference comes, that new photon can force the electron to jump down and emit a second photon.



This is called *induced emission*. And it has some strange properties. It is not surprising that the two photons have exactly the same energy (and therefore the same wavelength and frequency). But they are also in phase.

If we are clever, we can make this more probable than induced absorption. The trick is to ensure the electron is already excited into the higher energy state. We can use one of those “forbidden” states to accomplish this. You know that although their probability is lower for transition, eventually the transition will happen. But that means the atom is in the excited state for a long time. If you can get many atoms in a material to be in a metastable forbidden state and quickly insert just the right energy photon, you would get a cascade of single frequency in-phase photons out. Ruby crystals were one of the first materials used to do this. A different wavelength light flashed on to excite the atoms in the ruby crystal. These atoms decayed, but were stuck briefly in a metastable state. Then when a matching photon came through the new photon induced the electron to transition through the forbidden transition. and out came another photon. Symbolically this would be.

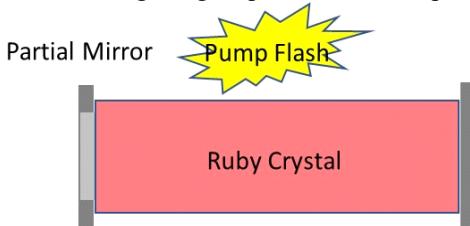


Here is a ruby crystal ready for such an experiment

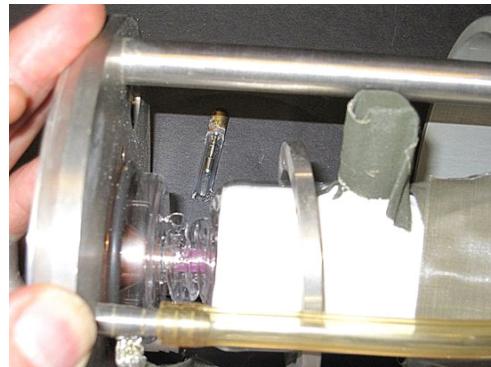


Laser Ruby Crystal (Photo courtesy Guy Immega)

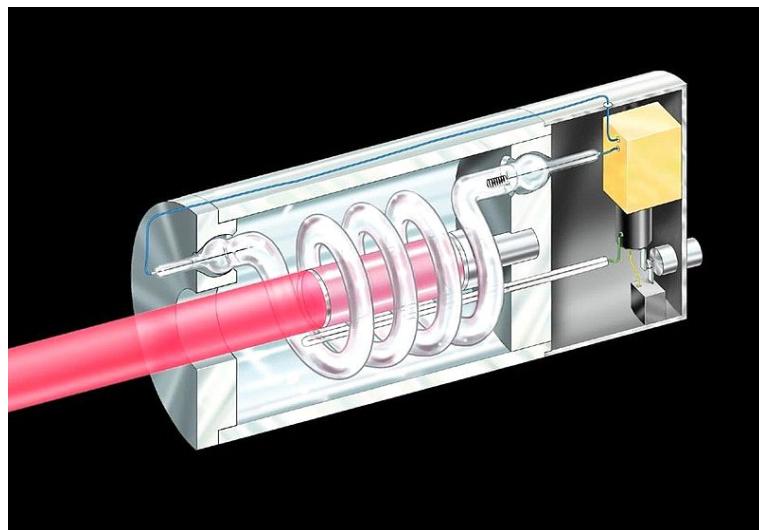
But of course if you put mirrors on both sides of the ruby crystal, the light would pass back and forth through the crystal getting amplified with each pass.



This device is called a light amplification by stimulated emission device, or a LASER. The initial photon that pushes the electron into the original excited state is called the pump photon and in the next figure you can see the ruby crystal surrounded by a curly flash bulb that provides the pump photons.



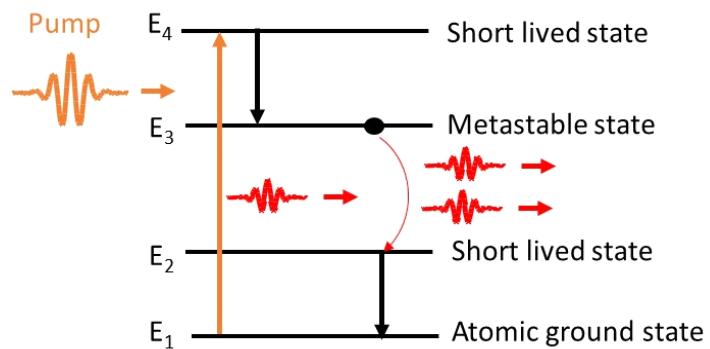
Actual Ruby Laser 1963 (Photo courtesy US National Institute of Standards)



Ruby Laser Design (Photo courtesy US Department of Energy)

Pumping the atoms into the higher energy state is essential, and you need the majority of the atoms to be in the higher energy state or the laser beam will die out due to induced absorption and then spontaneous emission. When more atoms are excited than are in the ground state we call the situation a *population inversion*.

There are much better laser materials than ruby crystals. Good materials use more states. One improvement is to use four states with the metastable lasing state in the middle.



With what we know about atoms so far we can explain so many phenomena. But we have been alluding to chemical bonding for some time. It's time to take that on in our next lecture.

21 Molecular Bonds and their Structure

9.1-9.3

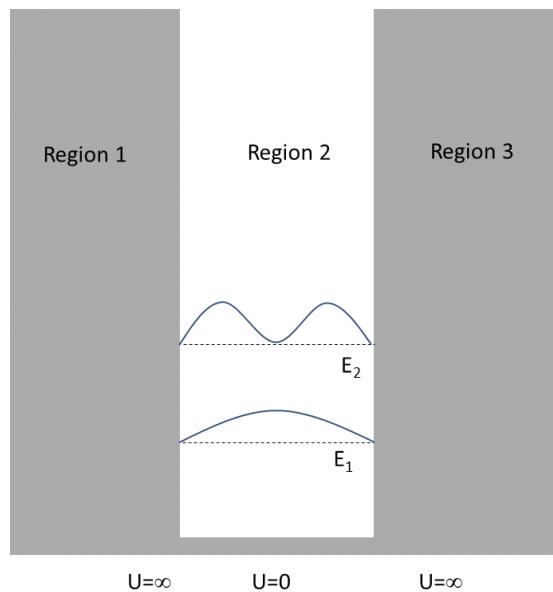
Fundamental Concepts in the Lecture

- Molecular Wave Functions
- Covalent Bonds
- Ionic Bonds

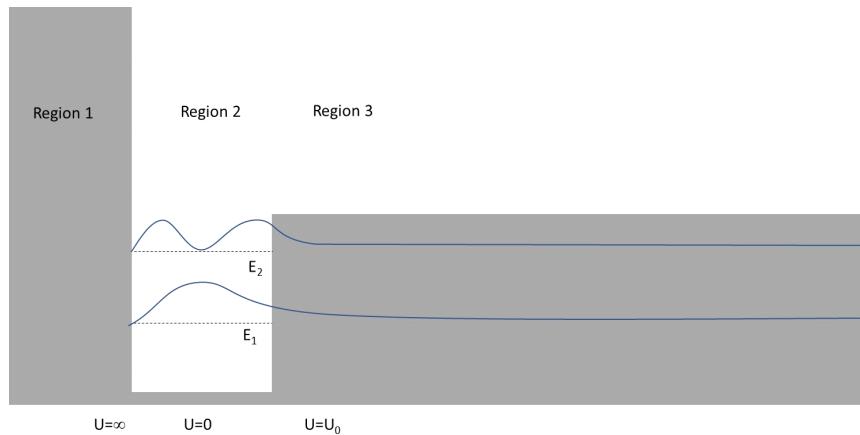
In working with molecular bonds, we have the same problem we had with multi-electron atoms. We can't solve many-bodied problems exactly. We have to do this numerically. For atoms we used the simplest atom, hydrogen, because we could solve the mathematical equations for the simple case and we extended our intuition we gained from solving this problem to guide us in the many-electron atoms. We will do the same with molecules. We will start with the molecule H_2^+ which is two hydrogen atoms, but one atom has lost its electron. We call this a singly ionized hydrogen molecule. Then we will extend our intuition from this simple case to more complicated molecules.

Molecular Wave Functions and the H_2^+ Molecule

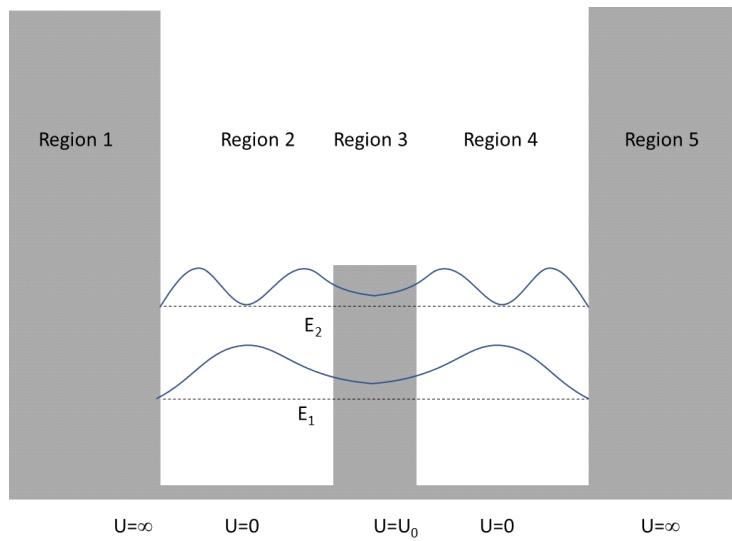
Let's step back a minute in our study of quantum systems and review one dimensional potential wells. Here is a schematic diagram of an infinite potential well with, say, an electron trapped in the well.



A few energy states are sketched in with their probability density functions. But what would happen if we lowered the potential on one side of the well? We expect something like the next sketch



Now the electrons in the well can escape into region 3. We know all about matching boundary conditions for the wave functions and we could do such a problem. But now suppose we have an additional potential well beyond region 3 as shown in the next figure.



In this sketch, we see energy levels and probability density functions, but we recognize that the electrons could pass through region 3 and find themselves in Region 4. We could find the wave functions by matching coefficients at all the boundaries (but we won't in our class). We can guess what sort of result we would get. We expect to see a wave function that spans from the boundary of Region 1 and Region 2 all the way to the boundary of Region 4 and Region 5. The wave function spans the small well in Region 2 and the small well in Region 4. The wave function expanded to include all of the region where the electrons could be detected. We don't have two wave functions for the small wells in the E_1 state. We have one bigger wave function. This is just what we would expect from wave behavior (well, quantum wave behavior). A wave traveling from the right would travel all the way to region 5 and reflect. And, if the wavelength is just right, it would form a standing wave.

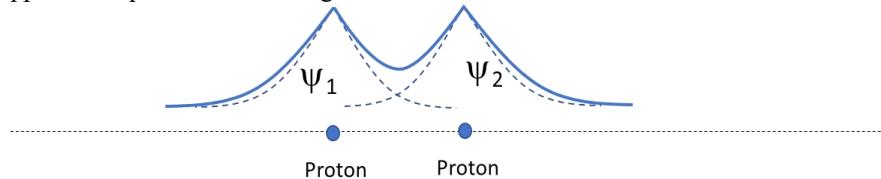
We can use the intuition we have from potential wells to study a H_2^+ molecule. This molecule is two hydrogen atoms that are bound together, but with just one electron. The other electron is missing. The two protons will be like deep Coulomb potential wells for the electrons. If the two protons are far apart, we just have two hydrogen atoms and all we have done with individual hydrogen atoms up till now would apply. Schematically⁶ this might look sort of like this

⁶ "Schematically" means I will draw it like an artist trying to illustrate the point, and not plot it using equations to show exact behavior.

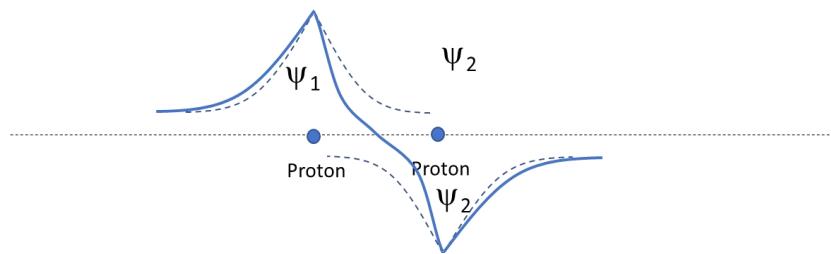


Note that these wave functions are not exactly accurate. We know they are three dimensional and even the radial part peaks away from the nucleus a bit, but our radial wave functions would be something like this. We will ignore the details for now.

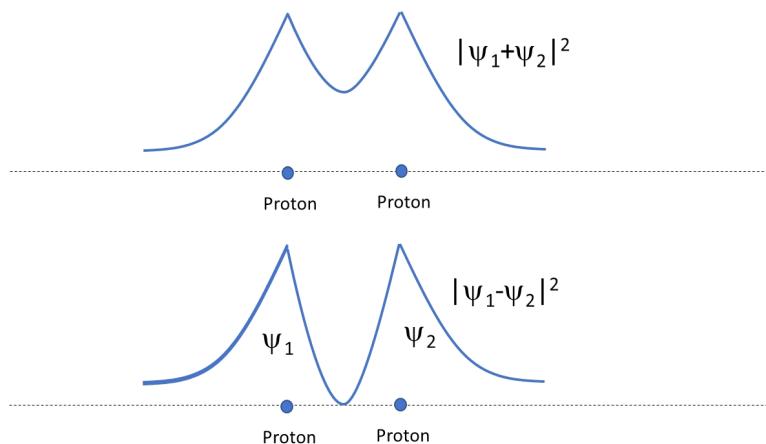
If we allow these somewhat inaccurate wave functions, then we can draw what would happen if we put them closer together.



From our one-dimensional potential well example, we would expect the two protons to create two dips in an overall deep potential well. The potential between the two protons would be higher. And, most importantly, the wave function of our one and only electron in H_2^+ would form around *both* protons! Of course our molecular wave function might have several possible shapes. For example, This pattern might also work.



We would find the probability of detection for our electron and the two wave functions above would make different probability graphs.

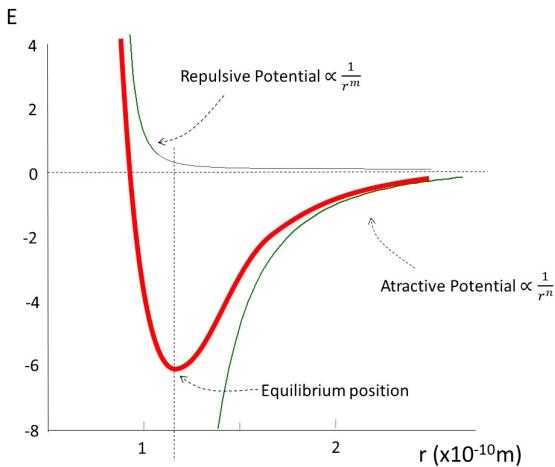


And we know that the orbitals came from these probability functions! *This means our atomic orbitals will change into molecular orbitals that span both nuclei.* Back in high school chemistry you were probably told that in molecules the atoms “share” electrons. Now we can see what this means. The electron states have a complex potential well structure, and the electron wave functions exist in all parts of that potential well structure. The wave function is what is “shared” by the nuclei of all the atoms in the molecule.

The details of forming the potential well may be complicated. For H_2^+ they are bad enough, but for a long carbon chain molecule they can be very complicated. But the basic principles come from the forces involved. We have Coulomb forces from the nuclei, and from the other electrons. Thus in a molecular bond we expect two opposing forces to make up the potential well. And we know that

$$F_{x_i} = -\frac{dU}{dx_i}$$

where x_i is a generic coordinate. So the potential energy is related to our force. We expect something like the schematic in the next figure.



The repulsive potential comes from the electrons being repulsed by the electrons of the other atom and eventually can come from the nuclei repulsing each other. The attractive potential comes from the electrons of one atom begin attracted to the nucleus of the other atom. The potential energy of the system is the sum of these two competing potential energies. We expect the repulsive potential to dominate at very small separation distances. We expect the attraction potential to dominate at larger separation distances (remember we are missing an electron in H_2^+). And we expect that in between there should be an equilibrium where the potential is negative (because the atoms are bound) and where the forces are balanced.

We hinted above that there were two possible ways we could have the two atom wave functions combine $|\psi_1 + \psi_2|^2$ or $|\psi_1 - \psi_2|^2$. In the first case we would have a significant probability in between the two nuclei. This is just what we want for a bond. We want the electron to be detected in between the two positive nuclei so there is an attractive Coulomb force bonding the atoms together. The energy of the molecule would be the potential energy due to the proton U_p plus the potential energy due to the electrons U_e . The sign of U_e depends on the form of the molecular wave function. It turns out that the $\psi_1 - \psi_2$ case gives a positive U_e (repulsive Coulomb force) and the $\psi_1 + \psi_2$ case gives a negative U_e (attractive Coulomb force). We need the negative U_e case to combine with U_p to give a region where

$$U_{mol} = U_p + U_e$$

can be negative and provide a bond. So the $\psi_1 + \psi_2$ forms a bond and the $\psi_1 - \psi_2$ does not.

We didn't prove any of this mathematically. But the calculations can be done

numerically. And these results give us an idea of how molecular bonds are formed. For H_2^+ we find that the equilibrium position is at $r_{eq} = 0.106 \text{ nm}$ with an energy of $E = -16.3 \text{ eV}$.

We can find out how much of this energy is involved in the bond, itself. This is the amount of energy required to separate the two atoms. We call this the bonding energy and it is found by taking the molecular energy and subtracting it from the energy of the two separate atoms. For H_2^+ we have one atom with no electron which has no orbital energy state. So we have

$$\begin{aligned} B &= (E\{H\} + E\{H^+\}) - E\{H_2^+\} \\ &= (-13.6 \text{ eV} + 0) - (-16.3 \text{ eV}) \\ &= 2.7 \text{ eV} \end{aligned}$$

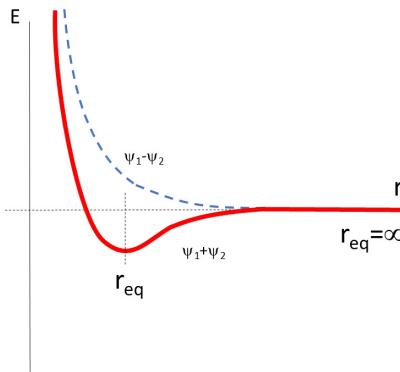
Our numerical calculations also gave $r = 0.106 \text{ nm}$ and we should note that this is

$$\begin{aligned} r_{eq} &= 0.106 \text{ nm} \\ &= 2(0.529177249 \times 10^{-10} \text{ m}) \\ &= 2a_o \end{aligned}$$

And we recall that for hydrogen the $1s$ state maximum value occurs right at $r = a_o$. This means that the maximum probability for the detection location of the electron is right in the middle of the molecule. Once again the electron seems to be “shared” in the molecule.

The H_2 Molecule

If you think we did little math for the H_2^+ molecule, just wait for the H_2 version. We can envision again having two hydrogen atoms and bring them from an infinite distance apart. We again have the possibilities of combining the wave functions as a sum $\psi_1 + \psi_2$ or a difference $\psi_1 - \psi_2$.



The difference, $\psi_1 - \psi_2$, still gives a positive total energy for all values of r . It won't form a bond. We will call this the *antibonding* state. The sum, $\psi_1 + \psi_2$, does have a region where the total energy is negative and has a minimum. It can form a bond. We will call this the *bonding* state. Now that we have two electrons in our molecule we need to consider the Pauli exclusion principle. We now have a molecular orbital which will give us our quantum numbers, and with two electrons we now we need to have different spin quantum numbers. So we can have two electrons in our H_2 molecule so long as one has spin $+\frac{1}{2}$ and the other has spin $-\frac{1}{2}$.

Calculations give the equilibrium separation for H_2 of $r_{eq} = 0.074$ nm, closer than for ionized H_2 . The extra electron Coulomb energy seems to have modified the potential well changing the equilibrium position. The molecular energy at equilibrium is $E = -31.7$ eV.

The bonding energy for H_2 is

$$\begin{aligned} B &= (E(H) + E(H)) - E(H_2^+) \\ &= 2(-13.6 \text{ eV} + 0) - (-31.7 \text{ eV}) \\ &= 4.5 \text{ eV} \end{aligned}$$

The extra electron increased the binding energy. It will be harder to split H_2 than it was to split H_2^+ .

Covalent bonds

The bonding we have just described is a special kind of covalent bond. It is a *homopolar* bond because the two atoms are the same. A covalent bond is formed under the following conditions:

1. When the two atoms are brought together from infinity, their electrons interact and their separate atomic states become molecular states.
2. The molecular wave function has a region with a lower energy than the separate atomic states
3. The Pauli exclusion principle applies, so only electron states that obey the exclusion principle will form.

Of course there is another molecular wave function that forms the antibonding state, but that won't form a bond. Since hydrogen can form a covalent bond, we expect the hydrogen like atoms to also form covalent bonds. And indeed Li_2 (dilithium) will form such a bond with $r_{eq} = 0.267\text{ nm}$ and binding energy of $B = 1.10\text{ eV}$. The equilibrium position is larger than for H_2 and Li_2 is not as tightly bound.

We could also match up lithium and hydrogen with a covalent bond to form LiH . Instead of a binding energy let's give a *dissociation energy*. This is the energy it takes to split the molecule which should be the same as the binding energy. But dissociation energies are given at a particular temperature, meaning that they can often be less than the binding energy because at a particular temperature thermal energy is helping dissociate the two atoms. LiH has a dissociation energy of 2.43 eV and an equilibrium separation of $r_{eq} = 0.160\text{ nm}$.

From our quantum atomic model we would expect s states to act similarly, but that higher s states would have larger most probable detection distances for the electrons, and like in our H_2^+ case we would expect the shared electrons to be in between the two atoms. So we would expect the r_{eq} values to increase with increased n and we would expect the binding energies to go down. Here are some covalent bonds for some hydrogen like atoms⁷

Molecule	Dissociation Energy E_D (eV) at 298K	r_{eq} (nm)
H_2	4.52	0.074
Li_2	1.10	0.267
Na_2	0.80	0.308
K_2	0.59	0.392
Rb_2	0.47	0.422
Cs_2	0.43	0.450

where we have used only homopolar cases. The trend for larger r_{eq} is clear and the accompanying trend in lower dissociation energy is also obvious.

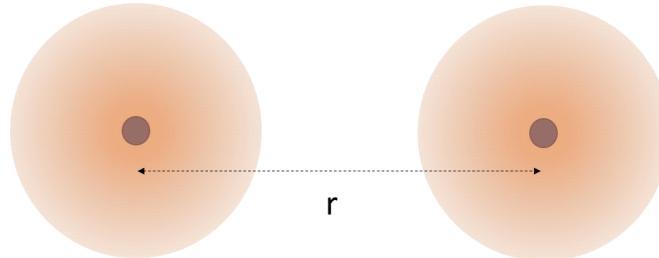
⁷ See <https://cccbdb.nist.gov/> or B. deB. Darwent, "Bond Dissociation Energies in Simple Molecules, NSRDS-NBS 31, US National Bureau of Standards, January 1970.

For non-homopolar molecules we expect the r_{eq} and E_D values to be in between those of the homopolar values.

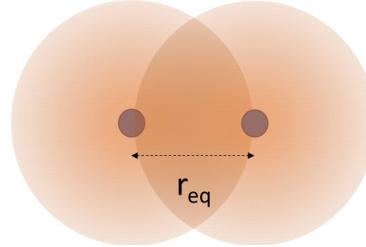
Molecule	Dissociation Energy E_D (eV) at 298K	r_{eq} (nm)
LiH	2.43	0.160
$LiNa$	0.91	0.281
KNa	0.66	0.347

and this is what we see.

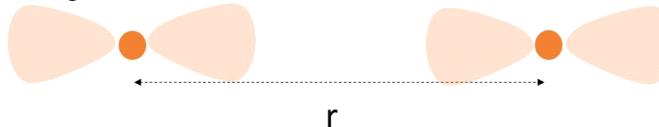
One way to mentally picture these bonds might be to draw an s orbital around a nucleus for each atom like this



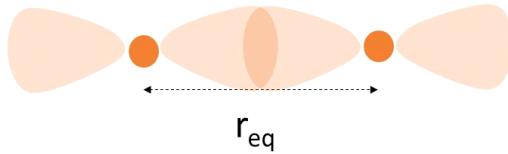
as the atoms become closer together, the orbitals overlap.



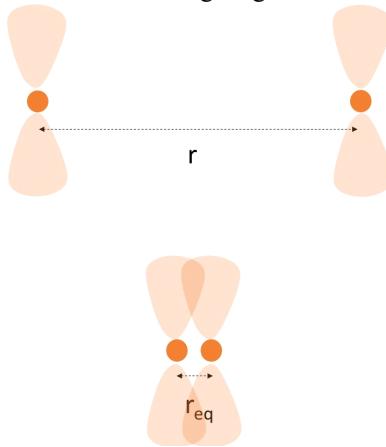
and the overlapping section shows where the probability density will become larger. But be careful. This is *not* what the molecular orbital looks like. But it does give us a way to think about the situation that would cause the molecular orbital to form. Let's think about an atom that is not hydrogen-like, but has, say, electrons in an outer most p shell. Then we might think of our two atoms like this



where the figure eight p shell structure is drawn very schematically. And if we move them together we could think of them like this

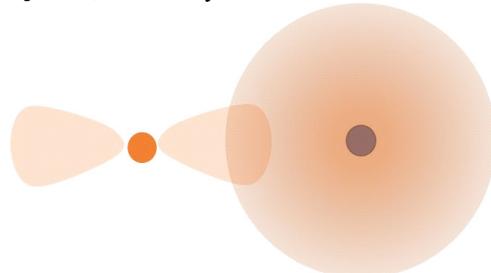


where once again we see that we will increase the electron probability in the middle between the two atoms. This will increase the potential and allow a bond to form. That bond will involve these p states to change into a molecular state, and that molecular state won't look like our picture. But the picture can help us to see that such a molecular state is likely to form. Let's try another p state possibility. Our p states could be oriented differently, say, like in the following diagram.



And once again we would predict a new covalent bond because we have increased the electron probabilities in between the two nuclei (although not directly in the middle). Once again a molecular bond is likely to form, but the orbitals won't look like our picture. From our simple approach we would guess from our pictures that for p states we could have different r_{eq} values for different orientations of the atoms.

How about an s and a p state, could they form a covalent bond?

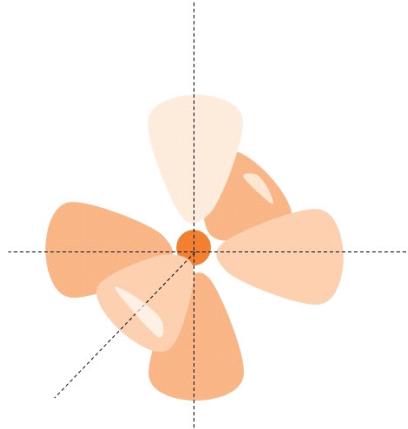


It sure looks like this should work. And it does.

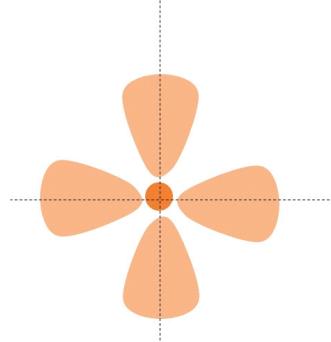
A good example of this is HF . The Fluorine atom has an electron state of $1s^22s^22p^5$. In the $2p$ state four of the electrons are paired in stable orbitals. They are less likely to form bonds. But one electron in the $2p$ state is not paired and it readily forms a bond. Some sp covalent bonds are described in the next table.

Molecule	Dissociation Energy E_D (eV) at 298 K	r_{eq} (nm)
HF	5.90	0.092
HCl	4.48	0.128
HI	3.10	0.160

But we should consider our uncertainty principle. We got the shape of the p states by assuming a particular orientation of our atoms. We could think of allowing some uncertainty in which way we define our z axis and therefore which direction our p orbital will take. We have three possibilities.



And if we look at this along one axis we get

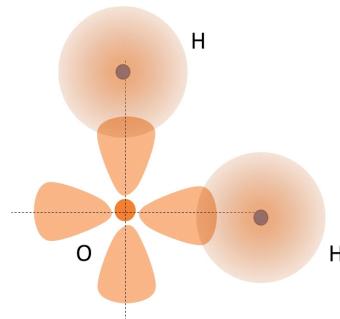


This increases the number of covalent bonding points for an atom. For example we could have oxygen with electron states $1s^22s^22p^4$ and we could have these $2p^4$

electrons in any of the six possible states.

l	m_l	m_s
1	-1	$\frac{1}{2}$
1	-1	$-\frac{1}{2}$
1	0	$\frac{1}{2}$
1	0	$-\frac{1}{2}$
1	1	$\frac{1}{2}$
1	1	$-\frac{1}{2}$

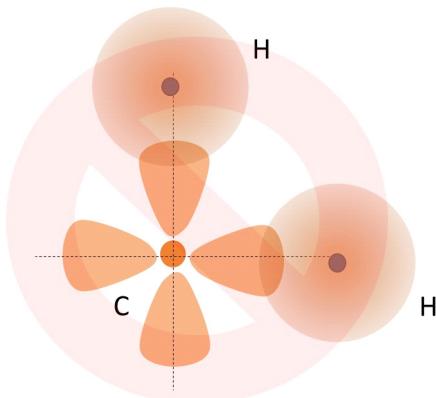
If we allow two to be paired to make a stable p state, but allow two to be unpaired, we could get two bonds!



Once again our picture is a nice way to think of creating the molecular wave functions, but not what they actually look like. This last picture is of an water molecule. It implies that the angle between the two hydrogen atoms would be 90° . The actual angle is 104.5° . So there is something not too bad about this simple graphical model, but there is also more to the story. The molecular orbitals are not just a combination of the atomic orbitals.

But we have discovered the origin of what high school chemistry calls “valence electrons” and why sometimes we have one bond and sometimes we have more than one bond for the same element. It is all in how the electrons are spread out in the allowed states for the upper most shell.

Our simple graphical model has some limitations. Consider carbon, $1s^22s^22p^2$. Our graphical model would expect a molecule like CH_2 to form very like water molecules



but this molecule doesn't usually happen. Instead CH_4 is much more likely. How can this be? The answer is that when we form molecular states we destroy the atomic orbitals and form new molecular orbitals. These don't necessarily look like the atomic orbitals. Chemists say that we have hybridized the states. I'm not sure this is a good term because it implies the new molecular states are made from the old atomic states. But the new molecular states are solutions to the Schrödinger with more than one potential minimum in a complex potential well. They really are their own things. But we can think of them as somewhere between s and p states so we will call them sp hybrid states.

The hybrid states follow some basic rules of formation:

1. As the atoms come together to form a molecule an atom with a configuration $2s^22p^n$ will have one of the $2s$ electrons move to the $2p$ shell giving $2s^12p^{n+1}$
2. The hybrid states can be estimated by taking equal mixtures of the wave functions representing the $2s$ state and each of the used $2p$ states.

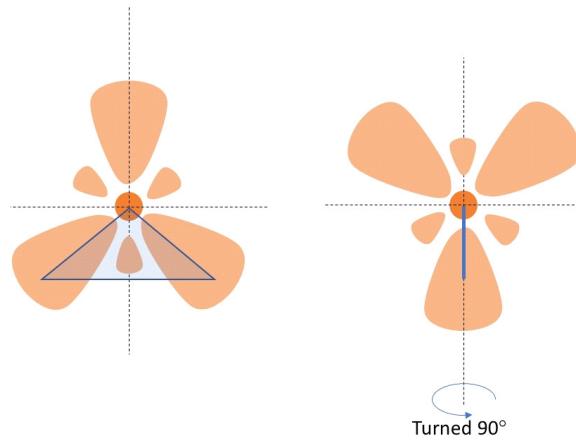
For carbon, we get

$$1s^22s^22p^2 \rightarrow 1s^22s^12p^3$$

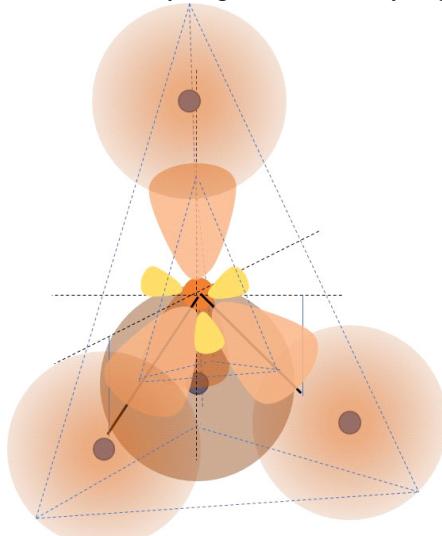
and the wave function could be any of the combinations of

$$\psi = \psi_{2s} \pm \psi_{2px} \pm \psi_{2py}$$

where ψ_{2px} and ψ_{2py} are the p wave functions along the x and y directions. For CH_4 we get a wave function that is a strange tetrahedral shape which is hard to draw.

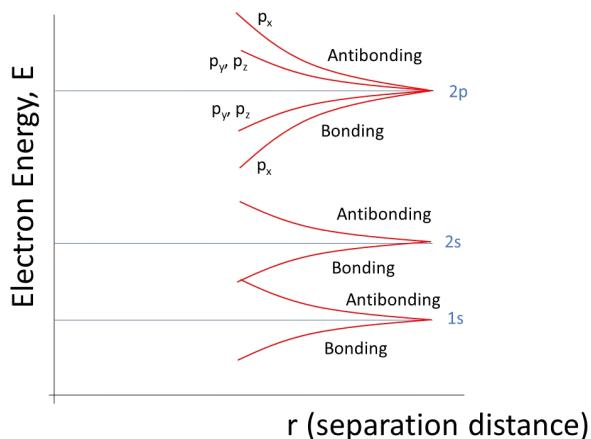


but if you can picture this, add in the hydrogen atoms. And you get something like this.



But once again the actual molecular states don't look like this. We can see that where the s states from hydrogen overlap we will get higher probabilities, but we really need the numerical calculation to get the shape of the molecular orbitals.

We can guess the strength of the bonds by thinking about our bonding and antibonding states. The electrons of both atoms will participate in building molecular orbitals.



If we have two electrons in, say, the 1s state of both atoms then these electrons will form bonding or antibonding states when the atoms are brought together. Since each atom has two 1s electrons, we will form both a bonding and an antibonding pair. Each pair form a molecular orbital. This will happen for each of the electrons in the atom. And the strength of the bond will be determined by how many of the electron pairs in the new orbitals are bonding or antibonding.

Let's take Nitrogen as an example. It's electron state is $1s^2 2s^2 2p^3$. So if we bring two nitrogen atoms together we will fill a 1s bonding state and a 1s antibonding state (see figure above). So far we have used four electrons. Likewise for the four 2s electrons that we have between the two atoms. We will get a 2s bonding and a 2s antibonding state. This takes up another four electrons, with a total of 8 used so far. We have a total of 14 electrons in the molecule, so there are six left. Note that for the 2p states there are p_x , p_y , and p_z bonding and antibonding states. And note from the figure that the bonding states are all lower energy than the antibonding states. So these will fill first. We have six electrons to place, with two per bonding state. That leaves us with a total of five bonding and two antibonding states. Thus overall the N_2 molecule will be bound.

Ionic Bonds

But high school chemistry taught us that there is another kind of bond, one in which an electron is taken from one atom and given to another. Then the Coulomb force keeps the two atoms together. $NaCl$ forms such a bond. We have Na with an electron structure of $1s^2 2s^2 2p^6 3s^1$ so it has an outer s electron that is unpaired. That electron is

easy to remove. It has a binding energy of 5.14 eV. And we have *Cl* with an electron structure of $1s^2 2s^2 2p^6 3s^2 3p^5$ and that leaves its highest shell missing one electron. If it had an extra electron, the whole $3p$ shell would have a 3.61 eV lower energy. So if we had an additional $5.14 \text{ eV} - 3.61 \text{ eV} = 1.53 \text{ eV}$ of energy we could remove the electron from the *Na* atom and attach it to the *Cl* atom. Then the two atoms would be attracted through Coulomb force

$$U = -\frac{1}{4\pi\epsilon_o} \frac{q_{Na^+} q_{Cl^-}}{r}$$

But the Coulomb potential is a potential energy. An atom in a Coulomb potential will move and gain energy by “falling” in the potential like a ball falling down a hill. We could retrieve energy from that “fall.” We need $U = 1.53 \text{ eV}$. Then the atoms will have to “fall” to within

$$\begin{aligned} -\frac{4\pi\epsilon_o U}{q_{Na^+} q_{Cl^-}} &= \frac{1}{r} \\ r &= -\frac{q_{Na^+} q_{Cl^-}}{4\pi\epsilon_o U} \\ &= \frac{e^2}{4\pi\epsilon_o U} \\ &= \frac{(1.60217733 \times 10^{-19} \text{ C})^2}{4\pi (8.854187817 \times 10^{-12} \text{ F m}^{-1})(1.53 \text{ eV})} \\ &= 9.4115 \times 10^{-10} \text{ m} \end{aligned}$$

to gain the energy needed to make the electron switch happen. Many molecules are formed this way. Here are a few.

Molecule	Dissociation Energy E_D (eV)	r_{eq} (nm)
<i>NaCl</i>	4.26	0.236
<i>LiCl</i>	4.86	0.202
<i>KBr</i>	3.97	0.282

The equilibrium separation has an additional term in its calculation for ionic bonds. The atoms must stay far enough apart that their outer orbitals don't overlap much. If they do, then we may violate the Pauli exclusion principle. That would force the electrons to jump to higher energy states to make room for the electrons of the other atom that would have the same quantum numbers. And such jumps require energy. That energy to make room for the electrons of the other atom must be part of the potential energy of the system. And we would have to supply that energy to push the atoms closer together.

This is a nice story, but it is really a story from high school chemistry. We should worry

when we say we lose an electron from one atom and place it in another atom. Wouldn't the wave function of the electron still reach the first atom if they are only a fraction of a nanometer apart?

And the answer is yes, the two atoms really share the electron, they just don't share equally. For NaCl the "sharing" is very unequal. For other molecules it is more equal and for homopolar molecules it is exactly equal. We can envision a NaCl wave function that is very lopsided. This means that being a covalent or ionic bond is really a spectrum. How ionic a bond is could be determined by seeing how polarized the molecule would be. If the electron wave function has a small amplitude around one atom and a larger amplitude around the other, then the first atom will be more positive and the second more negative. If this charge difference is large, the bond is very ionic.

A measure of the charge displacement for a molecule is the electric dipole moment. If our NaCl were purely ionic then we would have

$$p = qr_{eq} = (1.60217733 \times 10^{-19} \text{ C}) (0.236 \text{ nm}) = 3.78 \times 10^{-20} \text{ C nm}$$

The measured value for NaCl is

$$p_{measured} = 3.00 \times 10^{-20} \text{ C nm}$$

so we could say that the fractional ionic nature of the bond would be

$$\frac{3.00 \times 10^{-20} \text{ C nm}}{3.78 \times 10^{-20} \text{ C nm}} = 0.79365$$

or this is a 79% ionic bond.

The bonding structure is an exciting success of our quantum model of atoms. But another great success is the prediction of molecular spectra, and we will take on this topic in our next lecture.

@@@ define electron affinity

22 Molecular Spectra, Vibrations and Rotations

9.4,9.5,9.6

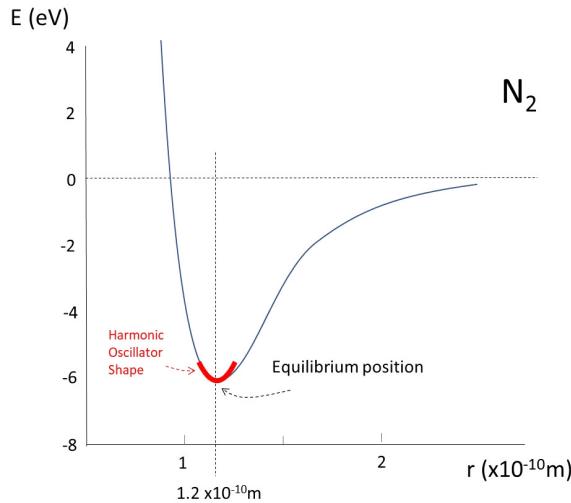
Fundamental Concepts in the Lecture

- Molecular vibration and vibrational states
- Molecular rotation and rotational states
- Combined vibrational and rotational states
- Molecular spectra

Molecular oscillators

We studied quantum oscillators back in section (3). We hinted at the solution to the various allowed energy states and even solved for the ground state wave function. What we found was that a harmonic oscillator has a potential energy and therefore can be treated like a potential well. From this treatment we find that there are allowed energy states associated with wave functions.

We have built diatomic molecules in our last lecture and we have considered how the bonds are formed. But think about those bonds. We have two atoms in equilibrium with attractive and repulsive forces. Only the forces keep the atoms near each other. So there isn't a physical structure keeping them in place. We could easily make the system oscillate by pulling on one of the atoms (or both) with, say, an electromagnetic field (like a light beam). Here is a plot of the potential energy as a function of separation distance for N₂.



Notice that the bottom of the well looks a lot like a harmonic oscillator potential. The match is not exact, but if we use the harmonic oscillator solutions and don't let r be too far from r_{eqlib} , we should have solutions that match the situation fairly well. Let's see what we can do with this idea.

Harmonic oscillator review

When we studied the harmonic oscillator before we remembered from PH123 that classically

$$\omega = \sqrt{\frac{\kappa}{m}}$$

where κ is the spring constant. And we recall that

$$\omega = 2\pi f$$

and the classical potential energy is

$$U_s = \frac{1}{2}\kappa(x - x_{eqlib})^2$$

and the total energy is

$$E = \frac{1}{2}\kappa(x_{max} - x_{eqlib})^2$$

The solutions of the Schrödinger equation for a harmonic oscillator gave energy states of the form

$$E_N = \left(N + \frac{1}{2}\right)\hbar\omega \quad N = 0, 1, 2, 3, \dots$$

The ground state energy would then be

$$E_N = \left(0 + \frac{1}{2}\right)\hbar\omega = \frac{1}{2}\hbar\omega$$

or we could write this as

$$E_N = \frac{1}{2}hf$$

This harmonic oscillator ground state energy is called the *zero point energy* and it is important that it is not zero. We get a potential well and allowed energy states that are in even jumps of $\frac{1}{2}hf$.

$E_5 = \frac{11}{2}hf$	_____
$E_4 = \frac{9}{2}hf$	_____
$E_3 = \frac{7}{2}hf$	_____
$E_2 = \frac{5}{2}hf$	_____
$E_1 = \frac{3}{2}hf$	_____
$E_0 = \frac{1}{2}hf$	Ground State _____
$E = 0$

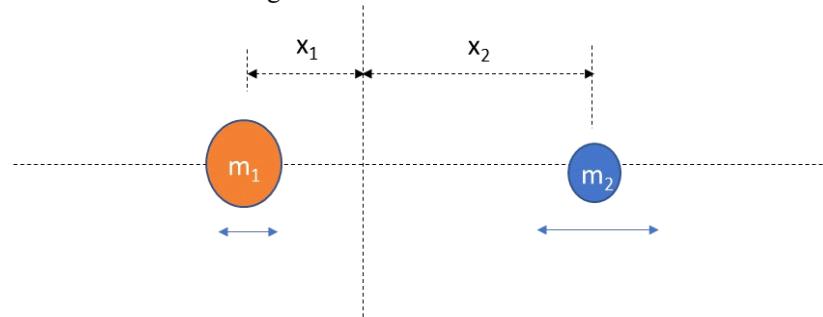
and we have a selection rule.

$$\Delta N = \pm 1$$

That is, jumps of more than one level are very unlikely.

Applying the quantum oscillator to molecular vibrations

Consider a diatomic molecule. We have a problem right from the start because we have two masses that are oscillating.



We didn't do a problem like this back in PH123. We need to find a way to do it now. We have learned that we can take special situations to help us solve such problems and one way to do this is to take special viewpoints, like a center of mass reference frame or a reference frame that is riding along with one of the atoms. We will use both of these techniques in this problem. In the end, we want a reference frame that has one of the atoms stationary and the other one oscillating. We want this because it turns out this makes the math easier. Let's start with a special case, Let's take the case when both atoms are right at their equilibrium position.

For atom 1 that will be $-x_1$ and for atom 2 that will be x_2 . Then at that special moment the energy is all kinetic

$$\begin{aligned} E &= \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 \\ &= \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} \end{aligned}$$

Now let's switch reference frames to the center of mass frame. In this reference frame the total momentum is zero, so it must be true that

$$p_2 = p_1 = p$$

and in that frame

$$\begin{aligned} E &= \frac{p^2}{2m_1} + \frac{p^2}{2m_2} \\ &= \frac{p^2}{2} \left(\frac{1}{m_1} + \frac{1}{m_2} \right) \\ &= \frac{p^2}{2} \left(\frac{m_2}{m_1m_2} + \frac{m_1}{m_1m_2} \right) \\ &= \frac{p^2}{2} \left(\frac{m_2 + m_1}{m_1m_2} \right) \\ &= \frac{p^2}{2 \left(\frac{m_1m_2}{m_2 + m_1} \right)} \\ &= \frac{p^2}{2(M)} \end{aligned}$$

where

$$M = \frac{m_1m_2}{m_2 + m_1}$$

which we have seen before. This is the reduced mass. If we use this mass then our energy is

$$E = \frac{p^2}{2M}$$

and we can treat our two oscillators as one oscillator with mass M .

We should think about this reduced mass. When $m_1 = m_2$

$$M = \frac{m_1m_1}{m_1 + m_1} = \frac{1}{2}m_1$$

This makes sense. Our double oscillator is like a single oscillator with half the mass of the double situation. But from this we can see that if the masses are not equal, we should expect different behavior for each. Larger masses would move less, smaller masses would move more.

We have almost turned our double oscillator into a single oscillator problem. But we need to worry about κ , the spring constant.

We could measure an amount of energy E for a particular separation distance r . We know the potential energy in a spring, and we can model this like a spring force

$$E = \frac{1}{2}\kappa(x - x_{eq})^2$$

And we could find the minimum energy E_{\min} for the equilibrium separation distance R_{eq} . Then

$$E - E_{\min} = \frac{1}{2}\kappa(R - R_{eq})^2$$

From this we can solve for our spring constant, κ

$$\kappa = 2 \frac{(E - E_{\min})}{(R - R_{eq})^2}$$

Let's take a specific case, for H_2 suppose our energy is measured to be

$$E - E_{\min} = 0.50 \text{ eV}$$

and then our difference from equilibrium separation is measured to be

$$R - R_{eq} = 0.017 \text{ nm}$$

then our spring constant is

$$\kappa = 2 \frac{(0.50 \text{ eV})}{(0.017 \text{ nm})^2} = 3460.2 \frac{\text{eV}}{\text{nm}^2}$$

Knowing κ and how to find a reduced mass should allow us to find the frequency of oscillation for our H_2 molecule.

$$\begin{aligned} f &= \frac{1}{2\pi} \sqrt{\frac{\kappa}{m}} \\ &= \frac{1}{2\pi} \sqrt{\frac{2 \frac{(E - E_{\min})}{(R - R_{eq})^2}}{\frac{m_1 m_1}{m_1 + m_1}}} \\ &= \frac{1}{2\pi} \sqrt{\frac{3460.2 \frac{\text{eV}}{\text{nm}^2}}{\frac{1}{2}m}} \\ &= \frac{1}{2\pi} \sqrt{\frac{3460.2 \frac{\text{eV}}{\text{nm}^2}}{\frac{1}{2}(1.008u)}} \end{aligned}$$

but we need to convert from u to some other mass unit. If we recall that

$$1u = 931.49406 \frac{\text{MeV}}{c^2}$$

we write our frequency as

$$f = \frac{1}{2\pi} \sqrt{\frac{3460.2 \frac{\text{eV}}{\text{nm}^2} c^2}{\frac{1}{2}(1.008u) \left(\frac{931.49406 \frac{\text{MeV}}{c^2}}{u} \right) c^2}}$$

Or, canceling some of the c terms,

$$\begin{aligned} f &= \frac{1}{2\pi} \sqrt{\left(3460.2 \frac{\text{eV}}{\text{nm}^2}\right) \left(2.99792458 \times 10^8 \text{ m s}^{-1}\right)^2} \\ &= 1.2953 \times 10^{14} \text{ Hz} \end{aligned}$$

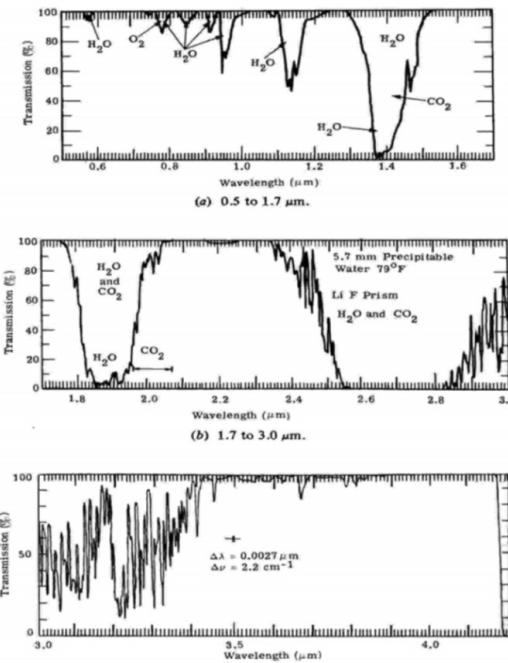
The energy of the transition would be

$$\begin{aligned} \Delta E &= hf = (4.1356692 \times 10^{-15} \text{ eV s}) (1.2953 \times 10^{14} \text{ Hz}) \\ &= 0.53569 \text{ eV} \end{aligned}$$

and the wavelength would be

$$\begin{aligned} \lambda &= \frac{c}{f} = \frac{2.99792458 \times 10^8 \text{ m s}^{-1}}{1.2953 \times 10^{14} \text{ Hz}} \\ &= 2.3145 \times 10^{-6} \text{ m} \\ &= 2.3145 \mu\text{m} \end{aligned}$$

This is in the infrared. And this is great. We find that molecules have spectral lines in the infrared that correspond to vibrations.



Chan, Mun Kit, 5Atmospheric Transmission Windows for High Energy Short Pulse Lasers, Naval Postgraduate School, Monterey, CA. 2003, (<https://calhoun.nps.edu/handle/10945/6224>)
 Different atoms have different masses, so we expect different frequencies for different molecules. And we can use this vibrational spectrum of allowed transitions to identify molecules. Astrophysicists and atmospheric physicists do this all the time.

Having reduced our double molecular oscillator to a simple oscillator we expect energy levels of the form

$$E_N = \left(N + \frac{1}{2} \right) \hbar\omega \quad N = 0, 1, 2, 3, \dots$$

where we have used our reduced mass and combined κ to get ω .

But of course our harmonic model isn't perfect. It only works at the bottom of the actual potential well. So if the molecule is excited, the predicted frequencies won't be right. In particular the nice even spacing of the harmonic oscillator energy levels doesn't actually work for real molecules. but we get approximations for the first few energy states that are not to terribly bad. For much of the matter in the universe, we expect the lower states to be the most frequently occupied. But not for stars or other hot glowing objects.

Obviously there is way more to finding molecular transition than our simple harmonic oscillator model. If you go on in physical chemistry you might do that work. But this is enough for us now. For the physicists you will do more in PH473.

Rotational molecular states

Of course if we can make our molecule vibrate, we could also rotate the whole thing. And that would take rotational energy. That is a new energy possibility. But rotation means angular momentum, and we already know that angular momentum is quantized. Let's see how this works for molecules.

Classically we would have

$$K_{rot} = \frac{1}{2} I \omega^2$$

where I is the moment of inertia and ω is now the angular speed. And

$$K_{rot} = \frac{1}{2} I \omega^2 = \frac{L_{mol}^2}{2I}$$

where for the molecule we have an angular momentum of

$$L_{mol} = I\omega$$

But we know that we can't specify all the components of L_{mol} due to uncertainty. Molecules are not really little rigid barbells that can be spun. They are complex wave structures. You might despair of finding a way to express these molecular rotational states. But there are books on this topic. And for simple systems we can almost do the problems exactly. Let's borrow from those treatments.⁸ The state energies are of the

⁸ For example, *Physics of Atoms and Molecules*, B. H. Bransden and C. J. Joachain, Longman Scientific &

form

$$E_L = \frac{L_{mol} (L_{mol} + 1) \hbar^2}{2I} \quad L = 0, 1, 2, 3, \dots$$

Note that this is very like the classical rotational kinetic energy but with L_{mol}^2 quantized.

We get a selection rule of

$$\Delta L_{mol} = \pm 1$$

where once again we mean that unit jumps are more probable but the other jumps happen just with less frequency. We get an energy graph structure that is something like this

$$E_3 = 20 \frac{\hbar^2}{2I} \quad \text{---}$$

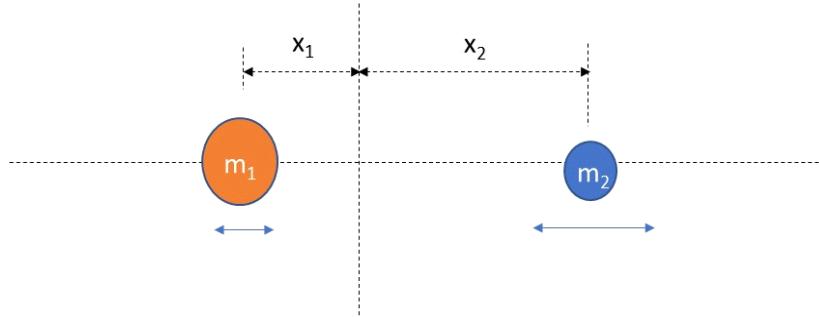
$$E_3 = 12 \frac{\hbar^2}{2I} \quad \text{---}$$

$$E_2 = 6 \frac{\hbar^2}{2I} \quad \text{---}$$

$$E_1 = 2 \frac{\hbar^2}{2I} \quad \text{---}$$

$E_0 = 0$ Ground State

with single jumps most probable.



For our diatomic molecule we could write

$$I = Mr_{eq}^2$$

where M is the reduced mass

$$M = \frac{m_1 m_2}{m_1 + m_2}$$

and where

$$r_{eq} = x_1 + x_2$$

so that

$$E_L = \frac{L_{mol} (L_{mol} + 1) \hbar^2}{2M r_{eq}^2}$$

and we could define

$$B = \frac{\hbar^2}{2M r_{eq}^2}$$

(just because it is tiresome to always calculate it) so that

$$E_L = B L_{mol} (L_{mol} + 1)$$

Then if we excited an molecule rotationally, so it goes from $L_i = L$ to $L_f = L + 1$ we would expect the excitation energy to be like

$$\begin{aligned} \Delta E &= E_f - E_i = B L_f (L_f + 1) - B L_i (L_i + 1) \\ &= B (L + 1) ((L + 1) + 1) - B L (L + 1) \\ &= B (L + 1) (L + 2) - B L (L + 1) \\ &= B (L + 1) L + B (L + 1) 2 - B L (L + 1) \\ &= 2B (L + 1) \end{aligned}$$

so our transitions depend on L , they are not evenly spaced and the emitted or absorbed photons would have energies like $2B$, $4B$, $6B$ and so forth.

Let's try a transition. Say we take our old friend H_2 and find the lowest radiation energy emitted due to rotation. We need a numerical B value so let's do that first.

$$B = \frac{\hbar^2}{2M r_{eq}^2}$$

we need $M = \frac{1}{2}m$ and we have $r_{eq} = 0.074\text{ nm}$ from our previous study of H_2 so

$$\begin{aligned} B &= \frac{\hbar^2}{2\frac{1}{2}mr_{eq}^2} \\ &= \frac{\hbar^2}{mr_{eq}^2} \\ &= \frac{(\hbar c)^2}{mc^2 r_{eq}^2} \end{aligned}$$

Numerically this is

$$\begin{aligned} B &= \frac{(197.326972\text{ eV nm})^2}{((1.008)(931.49406\text{ MeV})) (0.074\text{ nm})^2} \\ &= 7.573 \times 10^{-3}\text{ eV} \end{aligned}$$

So now we take

$$\Delta E = 2B (L + 1)$$

and for the $L = 0$ case

$$\begin{aligned}\Delta E &= 2B \\ &= 2 \times 7.573 \times 10^{-3} \text{ eV} \\ &= 1.5146 \times 10^{-2} \text{ eV}\end{aligned}$$

so the frequency of the emitted photon would be

$$f = \frac{\Delta E}{h} = \frac{1.5146 \times 10^{-2} \text{ eV}}{4.1356692 \times 10^{-15} \text{ eV s}} = 3.6623 \times 10^{12} \text{ Hz}$$

and let's find the wavelength for the emitted photon.

$$\begin{aligned}\lambda &= \frac{c}{f} = \frac{2.99792458 \times 10^8 \text{ m s}^{-1}}{3.6623 \times 10^{12} \text{ Hz}} \\ &= 8.1859 \times 10^{-5} \text{ m} \\ &= 81.859 \mu\text{m}\end{aligned}$$

which is infrared, but much larger than the vibrational states. This is the region of the electromagnetic spectrum known as the far infrared. It is the stuff of thermal cameras.

Vibration and Rotation

Of course in real life a molecule that is vibrating could be rotating as well. So we are unlikely to see just a rotation or just a vibration spectrum. Real spectra are usually mixed. A first order approximation to finding an equation for such a mixed spectrum would be to add together our vibrational and rotational energy equations.

$$E_L = BL(L+1) \quad L = 0, 1, 2, 3, \dots$$

$$E_N = \left(N + \frac{1}{2}\right)hf \quad N = 0, 1, 2, 3, \dots$$

so our combined spectrum would be

$$E_{L,N} = BL(L+1) + \left(N + \frac{1}{2}\right)hf$$

An energy level would be indexed by two quantum numbers, L and N . This combined spectrum must obey *both* selection rules

$$\Delta L = \pm 1$$

$$\Delta N = \pm 1$$

Let's try one. For an initial state N, L let's find a final state $N+1, L \pm 1$ the possible

energies for the absorbed photon are

$$\begin{aligned}\Delta E &= E_{N+1,L+1} - E_{NL} \\ \Delta E &= E_{N+1,L-1} - E_{NL}\end{aligned}$$

Let's start with the first one

$$\begin{aligned}\Delta E &= E_{N+1,L+1} - E_{NL} \\ &= \left[B(L+1)((L+1)+1) + \left(N+1+\frac{1}{2}\right)hf \right] - \left[BL(L+1) + \left(N+\frac{1}{2}\right)hf \right] \\ &= \left[B(L+1)(L+2) + \left(1+N+\frac{1}{2}\right)hf \right] - \left[BL(L+1) + \left(N+\frac{1}{2}\right)hf \right] \\ &= \left[B(L+1)L + B(L+1)2 + hf + \left(N+\frac{1}{2}\right)hf \right] - \left[BL(L+1) + \left(N+\frac{1}{2}\right)hf \right] \\ &= [B(L+1)L + B(L+1)2 + hf] - [BL(L+1)] \\ &= 2B(L+1) + hf\end{aligned}$$

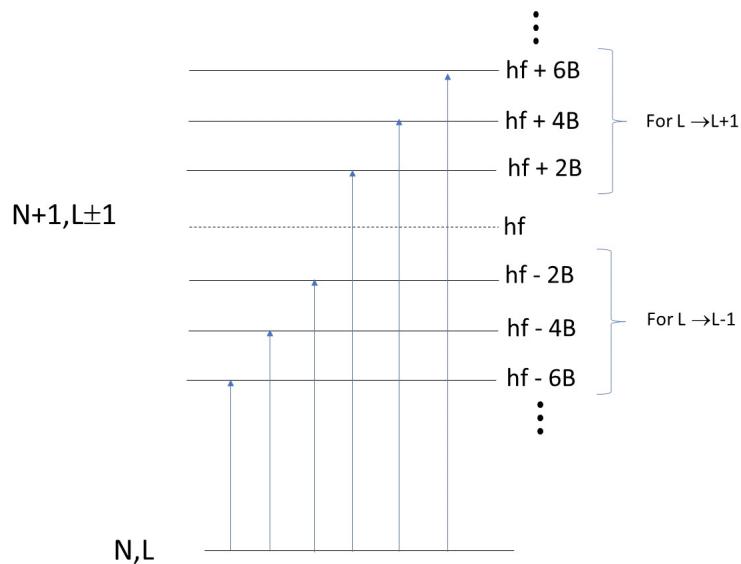
Now we need to do the $L-1$ case

$$\begin{aligned}\Delta E &= E_{N+1,L-1} - E_{NL} \\ &= \left[B(L-1)((L-1)+1) + \left(N+1+\frac{1}{2}\right)hf \right] - \left[BL(L+1) + \left(N+\frac{1}{2}\right)hf \right] \\ &= \left[B(L-1)(L) + \left(1+N+\frac{1}{2}\right)hf \right] - \left[BL(L+1) + \left(N+\frac{1}{2}\right)hf \right] \\ &= \left[BL^2 - BL + hf + \left(N+\frac{1}{2}\right)hf \right] - \left[(BL^2 + BL) + \left(N+\frac{1}{2}\right)hf \right] \\ &= [= -BL + hf] - [-BL] \\ &= -2BL + hf\end{aligned}$$

So for this transition we have two sets of allowed energies

$$\begin{aligned}\Delta E &= 2B(L+1) + hf \\ \Delta E &= -2BL + hf\end{aligned}$$

We get an energy diagram for the possible photon energies that looks like this



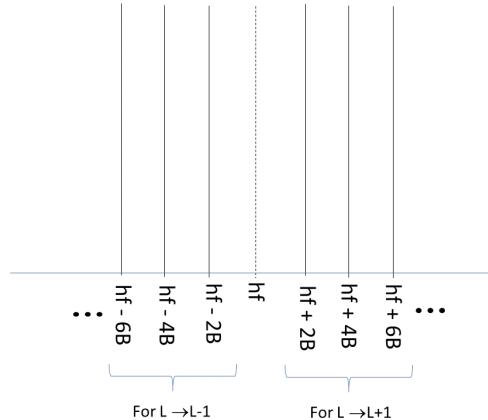
Notice that a transition from the N, L state to the $N + 1$ state with energy hf is not allowed. It obeys our selection rule on ΔN ,

$$\Delta N = +1$$

but not ΔL .

$$\Delta L = 0$$

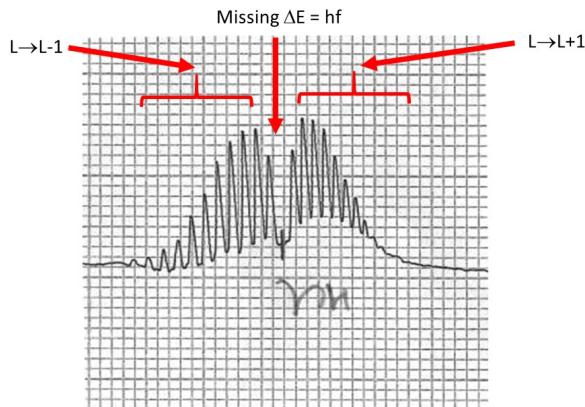
Spectroscopists tend to rotate this plot, because that is the way the spectra come out of spectrometers.



The energy values are given in terms of frequency by the spectrograph

$$f = \frac{\Delta E}{h}$$

Here is an example of a simple spectrograph of a molecular transition. Notice the many peaks on either side of the gap.



Spectrogram of a molecular transition. (Courtesy David-i98,
https://commons.wikimedia.org/wiki/File:HCl_rotational_spectrum.jpg, used with permission)

The gap is spacing is Δf and we have peaks at

$$\begin{aligned} f &= \frac{\Delta E}{h} = \frac{2B}{h} (L + 1) + f \\ f &= \frac{\Delta E}{h} = -hL + f \end{aligned}$$

Something is right about our simple approximate model for this molecular transition. We get the missing middle at the oscillator resonate frequency f . And we get a series of peaks on either side of the resonant frequency. But there are things our simple model didn't tell us. One is that the peaks aren't really equally spaced. This is because we assumed in the rotational discussion that the charges didn't move farther apart or closer together. But if we rotate a molecule with a stretchy bond, the distance between the two atoms can change as it rotates.

A more serious issue is that the peaks are not all the same height and seem to die off on either side. It is as though when we move away from $\Delta E = hf$ we get even less probability of that transition happening.

If we borrow from our future again we could say that the “population of a particular state” (meaning the likelihood that this particular transition will happen) is proportional to the Maxwell-Boltzmann distribution factor $e^{-\frac{E}{k_B T}}$. It's also true that with bigger L values we get more degenerate states. So the probability of a particular state being in the transition would be something like

$$P(E_{N,L}) = (2L + 1) e^{-\frac{E}{k_B T}}$$

where the $2L + 1$ is the angular momentum degeneracy. We can approximate this with

our equation for E

$$P(E_{N,L}) = (2L + 1) e^{-\frac{BL(L+1)+(N+\frac{1}{2})hf}{k_B T}}$$

and this explains our peak problem. For the first few levels the exponential factor is nearly 1. But as the values of L grow, the exponential term becomes smaller and smaller. If you are thinking this is great but where did the Maxwell-Boltzmann factor come from?—just hang on. We are headed there in our next lecture.

23 Statistical approach for many atoms

Fundamental Concepts in the Lecture

- Classical Statistics and Entropy
- Quantum Statistics
- Distribution Functions and Density of States Functions
- Gas of Particles Density of States
- Photon gas density of states

More than two atoms

We have been placing electrons in molecular states now for a while. But we had only really considered diatomic molecules. What happens if we have, say, a 200 kg ingot of gold? We can guess we might have a larger number of states.

Large chunks of sold metal might be too much of a jump for our first attempt at normal solid matter (solid state physics), but we want to be able to describe normal matter, so we should consider samples of more than single molecules. If we have millions of atoms, finding the states for each electron is at best tedious and probably just impossible to do. In PH123 we used statistical methods to form the branch of physics that we call *thermodynamics*, but sometimes this branch is called *statistical mechanics*. The statistical part of thermodynamics can be thought of as a subfield, in itself. So let's review a bit of statistics, and then apply statistics to our quantum model of molecules.

Classical Statistics

How many ways can I put balls in four boxes if only one ball fits in a box? Mostly the answer to this is that I don't care, I am a physicist and I don't get paid to put balls in

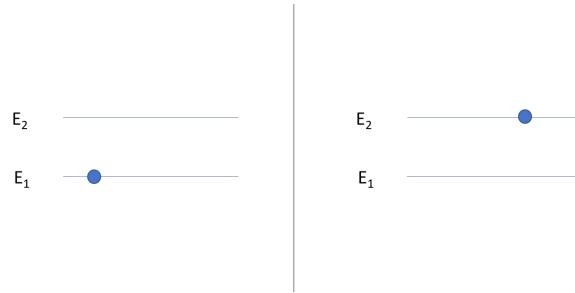
boxes. But suppose we want to put electrons into molecular orbital states. That makes the problem more interesting. So let's consider the easier ball and box problem and use it to work our way up to electrons.

In our ball and box problem, I can start with putting just one ball in one of the four boxes. There are four ways I could do that. There are six different ways I can put two balls in four boxes, and there are four ways to put three balls in four boxes, and one way to put four balls in four boxes and one way to put no balls in four boxes.

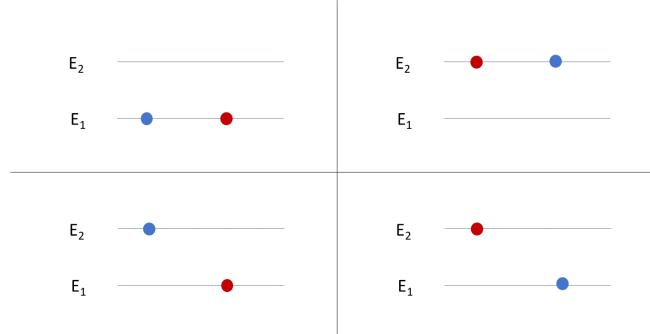
1	2	3	4
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	1	0	0
0	1	1	0
0	0	1	1
1	1	1	0
0	1	1	1
1	0	1	0
0	1	0	1
1	0	0	1
1	1	1	1
0	0	0	0
1	0	1	1
1	1	0	1

This is a good place to start thinking about lots of electrons in lots of molecular states for a small piece of a solid. But we know we need to go beyond just random placement. We need to care about how much energy we have for electrons.

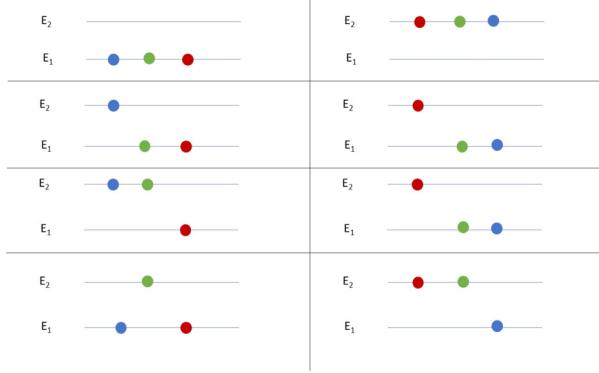
Suppose we have a system that has particles and two allowed energy levels (see figure below). If we are in an environment where we have sufficient thermal energy, we would expect that we would have a probability of $\frac{1}{2}$ that the particle will be in the state with E_1 and a probability of $\frac{1}{2}$ that the particle will be in the state with E_2 .



Now if I have two particles, the probability that two particles will be in the state E_1 will be $(\frac{1}{2})^2$ or $\frac{1}{4}$.



I can keep adding particles. For three particles the probability of being in the state E_1 is $(\frac{1}{2})^3$ or $\frac{1}{8}$.

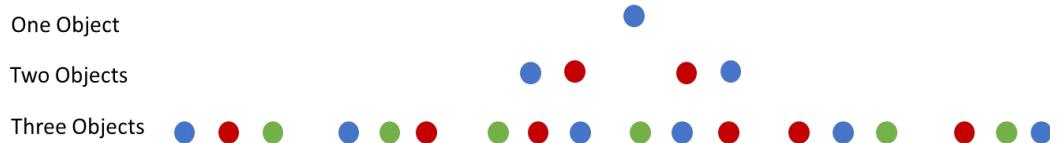


And for molecules we could have electrons as our particles and we have the possibility of having six to ten (or more) electrons in a molecular shell. But for a bulk metal, we could have many more electrons in a combined allowed state. So suppose we had 50 electrons. Then the probability of those 50 electrons all being in the E_1 state would be

$$\left(\frac{1}{2}\right)^{50} = \frac{1}{1125\,899\,906\,842\,624}$$

This is not very probable. Notice that we have marked our particles so we can tell them apart. This should worry you. But let's keep going with this assumption for now. From our math classes we "remember" that the number of ways to arrange N objects in a line is $N!$ So the number of ways to arrange one object is $1! = 1$ and the number of ways to arrange 2 objects in a line is $2! = 2$, and the number of ways to arrange three objects in a line is $3! = 6$

Lining up objects

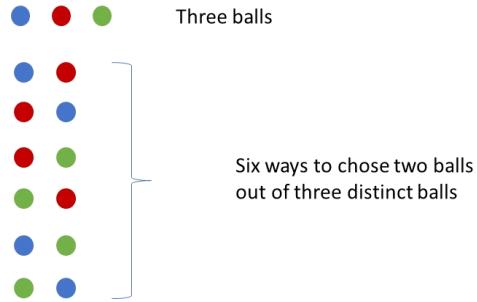


Knowing this we can see that the number of ways of selecting and arranging r objects taken from a group of N distinct objects is

$$\frac{N!}{(N-r)!}$$

For example, how many ways can we choose 2 particles from a group of 3 total particles.

$$\frac{3!}{(3-2)!} = 6$$

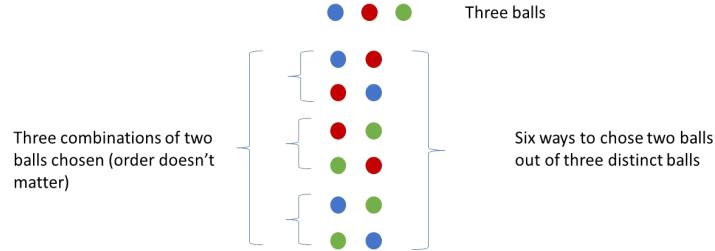


In this calculation the order of the balls was important. Blue red was different from red blue. But we expect that in choosing electrons the order won't matter. So we should consider how many ways we could choose r objects from N if the order isn't important. That is called a combination. So the number of combinations of N objects taken r at a time is

$$\frac{N!}{r!(N-r)!}$$

For our two balls chosen out of three case we get

$$\frac{3!}{2!(3-2)!} = 3$$



You “learned” all this at some time in a math class. But it might take thinking about this for a few minutes to remember why these equations work.⁹

We are going to go beyond this set of mathematical curiosities that work for picking numbered balls out of bins or people out of groups. We want to look at electrons in molecules and solids. So let’s consider a group of particles that can be in one of three energy states, but we have only two units of, say, thermal energy to spread among the particles. Further, say, we have four particles and we can still distinguish the particles. Then we might have states as shown in the following table.

	Microstates			
Macrostate	Particle 1	Particle 2	Particle 3	Particle 4
A	2	0	0	0
	0	2	0	0
	0	0	2	0
	0	0	0	2
B	1	1	0	0
	0	1	1	0
	0	0	1	1
	1	0	1	0
	0	1	0	1
	1	0	0	1

The states are grouped into two groups. One group has a particle alone in the $E_3 = 2u$ state. All the other particles have zero energy. The other group has the two units of energy spread over two particles with each of these particles carrying $1u$ of energy and the remaining two with zero energy. Because our energy is quantized, we can’t have all the particles, say, with half a unit of energy. We only get whole units. We call these groupings where the energy structure is similar *macrostates* and the individual ways we pull off the macrostate structure are called *microstates*.

⁹ If this is truly new, I recommend chapter 15 of Mary L. Boas’ book, *Mathematical Methods in the Physical Sciences*, 3rd Ed, Wiley, 2006. This is a good mathematical reference to have on hand.

We can do some predictive calculations based on just this information. Once again we assume that thermal or other processes make each microstate equally probable. Then if we wish to measure the energy of a particle and find that it has 2 units of energy and we repeat this experiment looking for particles with 2 units of energy we would expect to find such a particle about $4/40$ of the time. There are 4 microstates in which a particle can have energy of 2units. And in each of these microstates 1 particle has this energy, so we have 4 ways to get a particle with energy of 2units. And we have the 4 particles spread over 10 microstates so our analysis gave us $4 \times 10 = 40$ possible particle energy states. And only four of these have a particle energy of 2 units. So about $4/40 = 10\%$ of the time we would find one particle having 2units of energy. How about finding a particle with energy $E = 0$ units? We can count from the table and see that $24/40 = 0.6$ or 60% of the time.

Suppose we wish to know how many microstates we will have for a particular macrostate. We can say that for macrostate A we will choose 1 special particle out of the four to have 2 units of energy. Then (using our mathematical form that tells us how many combinations of N objects taken r at a time we can get) we should have

$$W_A = \frac{N!}{r!(N-r)!}$$

where we identify $N = 4$ particles and $r = 1$ with 2units of energy so

$$W_A = \frac{4!}{1!(4-1)!} = 4$$

microstates in macrostate A . For macrostate B we have $r = 2$ of the four particles with energy of 1u so we would expect

$$W_B = \frac{4!}{2!(4-2)!} = 6$$

microstates. But suppose we had 3 units of energy. We could distribute that energy in more ways among our four particles. We would expect a macrostate with one particle with 3 units of energy and there would be

$$W_A = \frac{4!}{1!(4-1)!} = 4$$

	Microstates			
Macrostate	Particle 1	Particle 2	Particle 3	Particle 4
A	3	0	0	0
	0	3	0	0
	0	0	3	0
	0	0	0	3

we could write this as

$$W_A = \frac{4!}{(1!)(3!)} = 4$$

and call N_3 the number of particles with 3u and N_2 the number of particles with 2u and

N_1 the number of particles with $1u$ and N_0 the number of particles with $0u$ of energy.

Then

$$W_A = \frac{4!}{(1!) (3!)} = \frac{N!}{N_3! N_0!}$$

and if we realize that $N_1! = 0! = 1$ and $N_2! = 0! = 1$ because we have no particles with $1u$ or $2u$ of energy, then we can write this as

$$W_A = \frac{4!}{(1!) (1) (1) (3!)} = \frac{N!}{N_3! N_2! N_1! N_0!}$$

where we just have to know N_0, N_1, N_2 , and N_3 for our particular case.

We expect that we could also make a macrostate where three particles could have one unit of energy

$$W_B = \frac{4!}{3! (4 - 3)!} = 4$$

with microstates

	Microstates			
Macrostate	Particle 1	Particle 2	Particle 3	Particle 4
B	1	1	1	0
	0	1	1	1
	1	0	1	1
	1	1	0	1

We could write this as

$$\begin{aligned} W_B &= \frac{N!}{(3!) (4 - 3)!} \\ &= \frac{N!}{N_1! N_0!} = 4 \end{aligned}$$

or even

$$W_B = \frac{N!}{N_3! N_2! N_1! N_0!} = 4$$

because $N_2 = N_3 = 0$, no particles have $E = 2u$ or $E = 3u$. And of course $N_1 = 3$ and $N_0 = 1$ in this case.

But we could also have one particle with 2 units and one particle with 1 unit. How many microstates would there be? Let's build the table and then see what the mathematical

function would be

	Microstates			
Macrostate	Particle 1	Particle 2	Particle 3	Particle 4
C	2	1	0	0
	2	0	1	0
	2	0	0	1
	1	2	0	0
	0	2	1	0
	0	2	0	1
	1	0	2	0
	0	1	2	0
	0	0	2	1
	1	0	0	2
	0	1	0	2
	0	0	1	2

We can use the new notation we developed in our first two cases to find the mathematical result. We would have $N_3 = 0$, $N_2 = 1$, $N_1 = 1$, and $N_0 = 2$ so we expect

$$W_C = \frac{N!}{N_3! N_2! N_1! N_0!} = \frac{4!}{(0!) (1!) (1!) (2!)} = 12$$

just as we see in the table.

In general the number of microstates in a macrostate is given by

$$W_r = \frac{N!}{N_r! \dots N_2! N_1! N_0!}$$

where r is the highest energy state occupied.

The total number of states would be

$$W_{total} = W_A + W_B + W_C = 20$$

But this should also be

$$W_{total} = \frac{4!}{(1!) (0!) (0!) (3!)} + \frac{4!}{(3!) (0!) (0!) (1!)} + \frac{4!}{(0!) (1!) (1!) (2!)} = 20$$

If we define $Q = 3u$ as our amount of energy the we could write this as

$$W_{total} = \frac{(N + Q - 1)!}{Q! (N - 1)!}$$

and we could check to make sure it works

$$W_{total} = \frac{(4 + 3 - 1)!}{3! (4 - 1)!} = 20$$

At any rate, we can use this to show the probability of a particle having a particular energy state. We did this before by counting the number of particles with a particular energy from a table and by dividing by the total number of possible states. Let's do that again, only using our mathematical forms. Let's define N_i as the number of particles

in each microstate (this is what we got from the table before). And let's define W_i as the number of microstates in each macrostate. This is also called the *multiplicity* of the macrostate. Then what we did before was to find the probability of a particle having a particular energy E

$$p(E) = \frac{\sum_i N_i W_i}{N \sum_i W_i}$$

Let's look at the denominator first to make sure we are achieving what we did before. It should be the number of particles multiplied by the number of microstates. This is the number of boxes in our tables for the A , B , and C macrostates combined. We can calculate this separately just so we can verify that it worked

$$\begin{aligned} N \sum_i W_i &= N(W_1 + W_2 + W_3) \\ &= N(W_A + W_B + W_C) \\ &= 4(20) \\ &= 80 \end{aligned}$$

Counting the boxes gives

$$16 + 16 + 48 = 80$$

so the denominator works. Now let's look at the numerator. $\sum_i N_i W_i$. If we want, say, the probability that a particle will be measured with an energy of $3u$ units, then we need N_i , the number of particles with energy of $3u$ in each microstate. We can see that $3u$ energies only occur in W_A which we would now label W_1 and in each of the 4 microstates one particle has $3u$ of energy, so $N_1 = 1$ and $W_1 = 4$. And for W_2 no particles have $3u$ units of energy so $N_2 = 0$ and likewise $N_3 = 0$ so the numerator would give

$$\sum_i N_i W_i = 4$$

so

$$\frac{\sum_i N_i W_i}{N \sum_i W_i} = \frac{4}{80} = 0.05$$

or about 5% of the time we would find a particle with $3u$ of energy. We can do this for all the possible values of E . For $2u$

$$N_1 = 0, W_1 = 4$$

$$N_2 = 0, W_2 = 4$$

$$N_3 = 1, W_3 = 12$$

so

$$\frac{\sum_i N_i W_i}{N \sum_i W_i} = \frac{12}{80} = 0.15$$

or about 15% of the time we would find a particle with $2u$ units of energy, and for one

unit of energy

$$N_1 = 0, W_1 = 4$$

$$N_2 = 3, W_2 = 4$$

$$N_3 = 2, W_3 = 12$$

$$\frac{\sum_i N_i W_i}{N \sum_i W_i} = \frac{24}{80} = 0.3$$

or 30% of the time we would find a particle with 1 unit of energy. We still need to determine how often we would find a particle with no energy.

$$N_1 = 3, W_1 = 4$$

$$N_2 = 1, W_2 = 4$$

$$N_3 = 2, W_3 = 12$$

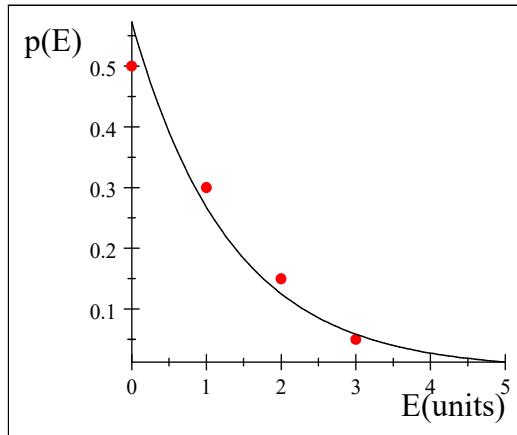
$$\frac{\sum_i N_i W_i}{N \sum_i W_i} = \frac{12}{80} + \frac{4}{80} + \frac{24}{80} = 0.5$$

or half the time we would find a particle with 0 u of energy. These must sum to 100%

$$50 + 30 + 15 + 5 = 100$$

and we could plot these probabilities

E (units)	$p(E)$
0	0.5
1	0.3
2	0.15
3	0.05



and a curve fit to the data gives

$$p(E) = 0.5727e^{-0.76E}$$

Note the exponential fall off of the probability with higher energy. This reminds us of the exponential term in the Maxwell-Boltzmann distribution, and indeed this is just the sort of analysis that produces that exponential. We expect collections of many atoms

to have an energy distribution that falls off exponentially with higher energy. But our analysis has been somewhat limited. We only allowed a little energy and a few states with a few particles. You can guess that this will be harder with very many particles. And that is what PH412 is for. So we will stop our classical analysis here, with just one little topic more, entropy.

Entropy

You may have had entropy defined in your PH123 class as

$$S \equiv k_B \ln (W) \quad (23.1)$$

The W in this equation is the same multiplicity that we have been using. We could consider a problem where we have three units of energy and that energy is all in one of the four particles. That would be macrostate W_A with its four microstates. Suppose we allow this system to change until one of the particles has two units and one has one unit of energy. That would be the W_C state. And we would have entropies

$$S_A \equiv k_B \ln (W_A)$$

and

$$S_C \equiv k_B \ln (W_C)$$

then the change in order is

$$\begin{aligned} \Delta S &= S_C - S_A \\ &= k_B \ln (W_C) - k_B \ln (W_A) \\ &= k_B \ln \left(\frac{W_C}{W_A} \right) \\ &= k_B \ln \left(\frac{12}{4} \right) \\ &= k_B \ln (3) \\ &= 1.0986 k_B \end{aligned}$$

which is larger than one. The entropy would increase. And we could say that it would be unusual for the W_C macrostate to evolve into the W_A macrostate because then ΔS would be negative.

Beginnings of Quantum Statistics

Much of what we have done for classical statistics is still true. But there are some differences that come because of quantization of energy and because of the nature of wave mechanics. Two major differences are:

1. *In quantum statistics like particles are not distinguishable.*

In our diagrams above, I color coded the particles so you could tell which is which. You might have guessed that this is not realistic. It is not possible to paint electrons different colors. And worse yet, electrons are waves. And it would, indeed, be futile to try to paint a wave. In classical statistics we could tell which microstate we had in macrostate *A* because we could tell the particles apart. But really that isn't possible for electrons. Because we can't distinguish the different microstates, we are forced to accept them all as one. The multiplicity of each macrostate becomes just 1. For our balls in boxes example, we would have just these possibilities because the other ways, say, to put one ball in one box look just like our first line no matter what box and ball we choose.

1	2	3	4
1	0	0	0
1	1	0	0
1	1	1	0
1	1	1	1
0	0	0	0

2. *In quantum statistics we have to obey the Pauli exclusion principle, so there are often a maximum number of wavicles that will fit in a particular state.*

So far we have only dealt with electron waves to any real extent. And electrons have spin states of $\pm\frac{1}{2}$. They must obey the Pauli exclusion principle. Let's consider our example with three units of energy. For electrons *only macrostate C* would be allowed because in macrostates *A* and *B* we would have three electrons with the same energy and therefore the same quantum numbers and that can't be.

But there are wavicles with spin states that are integers. Photons are one such wavicle. And when we do the lengthy calculations of their wave functions, we find that they don't obey the Pauli exclusion principle. We have really already acknowledged this when we did stimulated emission. The second photon produced in stimulated emission is exactly like the incoming photon. They would have the same quantum numbers. But because they don't have half integer spin, their wave functions are mutually compatible and they can just add up. And we easily get coherent light beams this way.

Let's try to recreate our graph of probability vs. energy for our 4 particle $3u$ of energy system for a quantum electron case. We only have macrostate C to use.

$$p(E) = \frac{\sum_i N_i W_i}{N \sum_i W_i}$$

and for $E = 0$ we would have, working from the table

$$p(0u) = \frac{24}{48} = 50\%$$

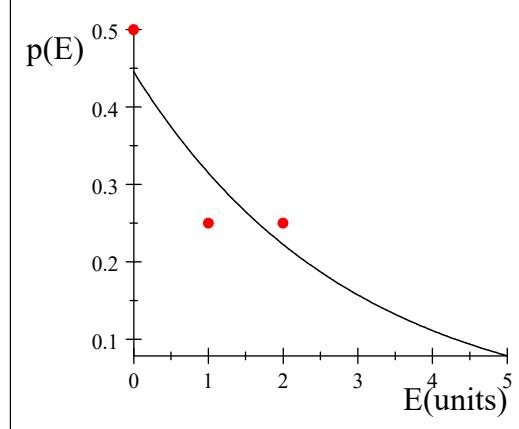
and for $E = 1u$

$$p(1u) = \frac{12}{48} = 25\%$$

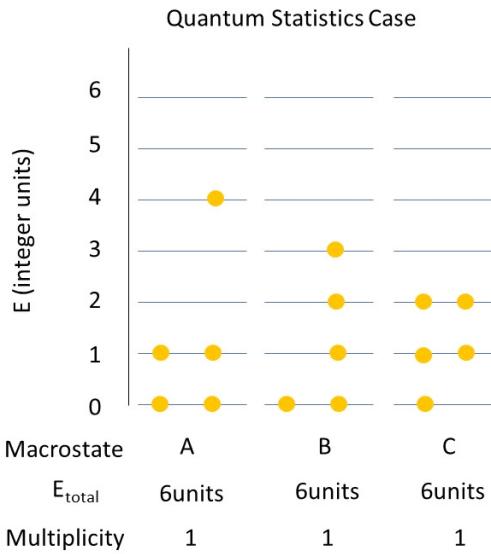
and for $E = 2u$

$$p(2u) = \frac{12}{48} = 25\%$$

so our graph looks like



and the exponential fit is quite unsatisfying. But with just three points it is hard to tell what is going on. Let's take a larger quantum case with, say, $6u$ of energy and 5 particles. Let's see how to arrange these with quantum statistics.



And the important thing is that these are the only possibilities. We can calculate our energy probabilities

$$p(0) = \frac{2+2+1}{(5)(3)} = 33\%$$

$$p(1) = \frac{2+1+2}{(5)(3)} = 33\%$$

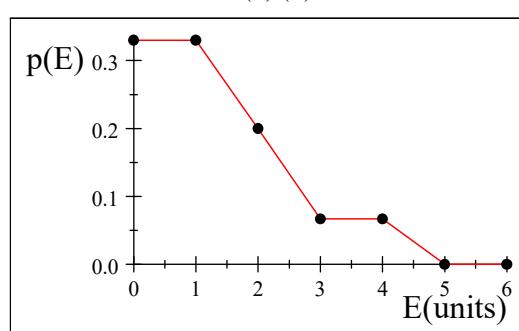
$$p(2) = \frac{0+1+2}{(5)(3)} = 20\%$$

$$p(3) = \frac{0+1+0}{(5)(3)} = 6.7\%$$

$$p(4) = \frac{1+0+0}{(5)(3)} = 6.7\%$$

$$p(5) = \frac{0+0+0}{(5)(3)} = 0\%$$

$$p(6) = \frac{0+0+0}{(5)(3)} = 0\%$$



And this is clearly not an exponential function.

Still, we are only up to six particles. That is a long way from a chunk of gold or the air in a classroom. We need to extend our analysis to much larger systems.

Distribution Functions and Density of States Functions

You might wonder if going forward is necessary, do large numbers of atoms exhibit any quantum behavior? We will find that even such normal everyday things like electric conductivity depend on quantum states of the metal atoms. So what we are doing is important to understanding the world we live in. It is so important that we have a whole class on this at the Senior level (PH412). But we can start here. So let's proceed.

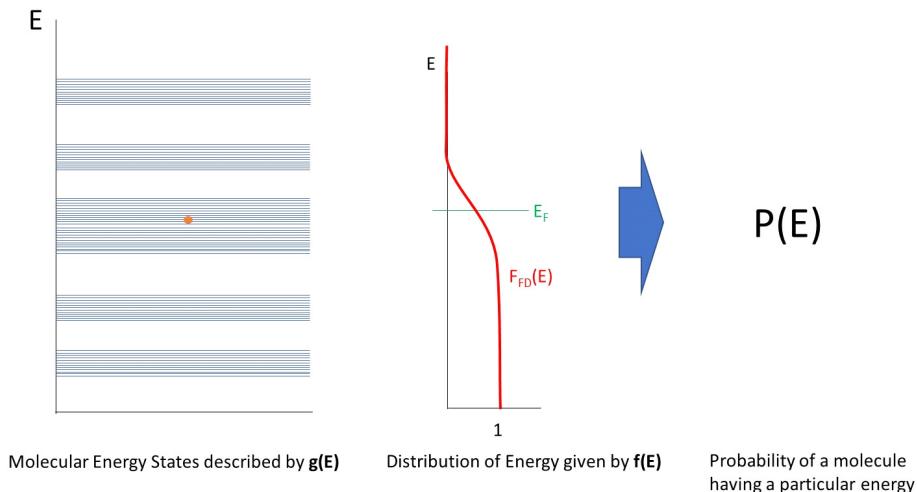
Let's give a name to the $p(E)$ vs. E graphs. The functions we see in this kind of graph is called a *distribution function*. We can generically represent this kind of function with the symbols

$$f(E)$$

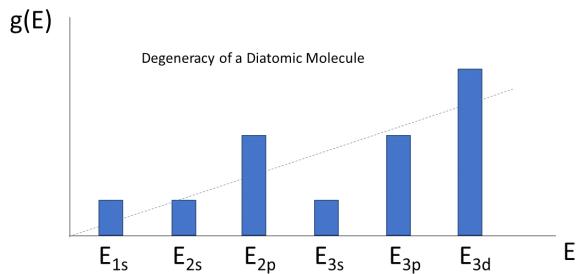
The Maxwell-Boltzmann function is just such a distribution function for classical systems.

So far our distribution functions have told us that it is more probable to find particles or wavicles in lower energy states. And the details of just what energy states are populated is important in understanding the parts of integrated circuits and glow paint, and so many other cool things. So we will need to find the distribution function for large groups of particles.

Our goal will be to find the probability of a particle (electron, atom, molecule, etc. depending on our situation) having a specific energy E . To find this we will calculate the number of particles that have a particular energy, E , and we will give that the symbol $N(E)$. And we will break this $N(E)$ into two parts. One part will be the distribution function that we just discussed. The distribution function tells us something about how much energy we have in the system and how spread out that energy is. Let's give it the symbol $f(E)$. The other part will be the places we can put that energy. We will give this piece the symbol $g(E)$ and we will call it the degeneracy function (because it includes the degeneracy of the energy states).



We know about molecules, so let's look at these two parts, $g(E)$ and $f(E)$ for an electron (our particle) in a diatomic molecule. We know that there are energy states for the electron. We have $1s$ and $2s$ and $2p$ and $3s$ states and so forth. The $g(E)$ function tells us about these states. For example for the $1s$ state energy would have $g(E_{1s}) \propto 4$ because in a diatomic molecule there the $1s$ state splits to include all four $1s$ electrons (two from each atom) so we have a degeneracy of 4 electrons at about the same energy. For the $2s$ state we also have a $g(E_{2s}) \propto 4$ but $g(E_{2p}) \propto 12$ and $g(E_{3p}) \propto 12$ and $g(E_{3d}) \propto 20$. Notice that the trend for $g(E)$ is to get bigger as the energy gets bigger.



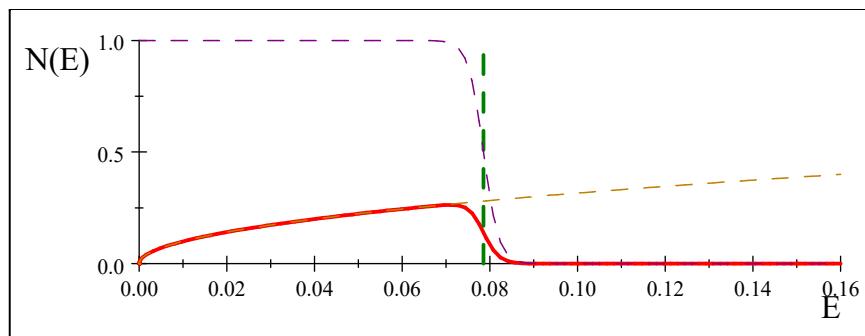
Notice also that we didn't use equal signs. The degeneracy function $g(E)$ is defined per unit volume. So it is proportional to the degeneracy, but not equal to it. We can write our $N(E)$ in terms of $g(E)$ and $f(E)$ as

$$N(E) = Vg(E)f(E)$$

to account for the per-unit-volume definition of $g(E)$.

If the $g(E)$ is about energy levels or where we can put energy, the distribution function $f(E)$ tells us which of these levels will be occupied. Notice in the figure the $f(E)$ shows that we have enough energy to fill the $1s$, the $2s$ and some of the $2p$ states. The $f(E)$ tells us which states have electrons with energy.

A single molecule is a very limited system. We want to calculate $N(E)$ for large systems, like containers of gas (think PH123) or silicon chips in computer electronics, and eventually for the whole universe. We expect our $N(E)$ to look something like this



where the orange dotted line is $g(E)$ and the purple dotted line is $f(E)$ and the red solid line is the product $Vg(E)f(E)$. The $g(E)$ function of this complicated system has become effectively continuous because there are so many states for this large system. Let's look at some physical systems in more detail and find $g(E)$ for those systems.

Dilute gases

Let's start with an easy case, a dilute gas, like a gas in one of the discharge tubes you have seen in PH123. To do this we will need to be able to find the number of available states with energy E . There are two ways we can add energy into the dilute gas. We can add energy to the electrons in the atoms (or molecules) or we can add kinetic energy (thermal energy) to the atoms (or molecules).

For the first part we might be able to just count the number of states as we have done before. The hydrogen atoms in our hydrogen gas discharge tubes are far enough apart that they don't interact and we can count the number of states with energy E_n by recalling that a state with quantum number n has possible l values up to $n - 1$ and for each l value it will have $2l + 1$ m_l values so it will have $l(2l + 1)$ substates. Substituting $l = n - 1$ gives on the order of $2n^2$ as the degeneracy of each n state.

We will have to work a bit harder to take into consideration the kinetic energy of the gas (coming up shortly)

The number of atoms in our dilute gas with a particular energy depends on which n state each atom has (with its degeneracy) and the distribution function that tells us how the energy is distributed in the gas. As we have seen distribution functions often favor lower energies, but the degeneracy favors higher energies. That is, there are more available states as n increases and the energy goes up. So there is a tug-o-war between the distribution function and the degeneracy. How many atoms will have a particular energy E will depend on the outcome of that tug-o-war.

We saw this in diatomic molecules as well. HCl is very popular so let's use it for our example. The degeneracy of each molecular energy state E_L marked with the molecular quantum number L is $2L + 1$ so the larger the energy the larger the number of available states to take that energy. But the energy distribution function for a gas of HCl would tend to favor lower energies.

Let's give a symbol for the other part of our tug-o-war, the degeneracy of the energy level E_n is written as

$$Vg(E) = d_n$$

and then we can write the number of particles that have a particular energy value E_n as

$$N_n = d_n f(E_n)$$

Like all probability functions, our distribution function must be normalized. But this time it must be normalized to the number of particles we have.

$$N = \sum_n N_n = \sum_n d_n f(E_n)$$

Or another way to say this is that all the particles must have some energy value so if we sum over all energy states we should find all the particles.

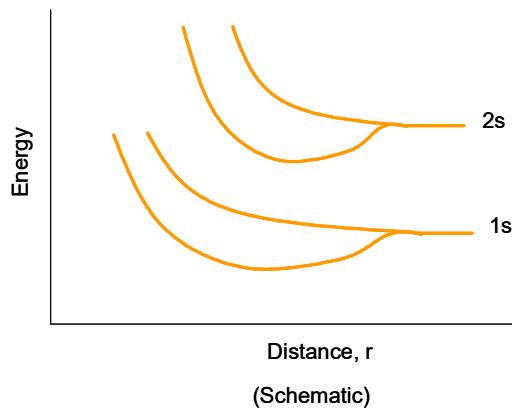
But we defined $g(E)$ per unit volume, so normally we would say

$$g(E) = \frac{d_n}{V}$$

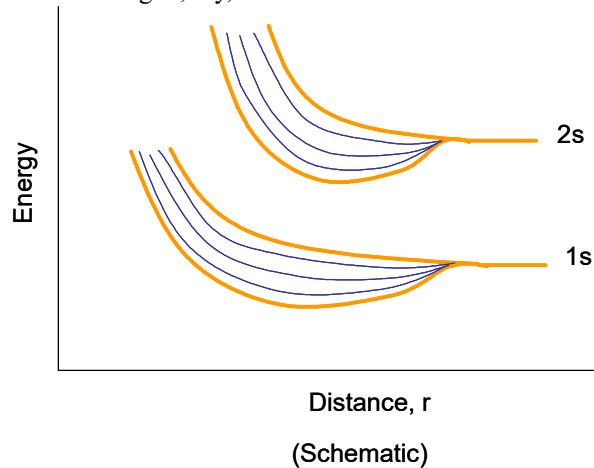
and

$$N = \sum_n Vg(E_n) f(E_n)$$

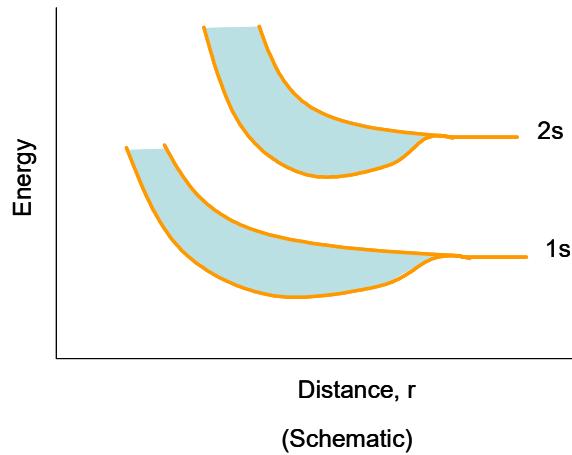
But not everything is a dilute gas of single atoms. And we know that if we bring atoms together their energy states split into molecular states. We did this with two atoms.



But what happens if we bring in, say, five atoms?



We get additional splitting of states where we once had two states we now have five different $1s$ molecular states each with a slightly different energy. What happens if we bring in many many atoms?



Now there are so many states that we just have a blue blur. A nearly continuous set of states in two bands. The states are really discrete, but in a calculus sense the ΔE between them is so small we would call it dE and we could mathematically treat these states as though they weren't quantized.

In this case we could replace summations with integration over dE .

$$N = \int dN$$

where our

$$dN = N(E) dE$$

We still need to do the same work. We need $N(E)$ to depend on the degeneracy of the states and we need a distribution function to know which states are filled. But we can write our degeneracy as a density, not a physical density, but a density of states.

$$N(E) = Vg(E)f(E)$$

where $f(E)$ is still the distribution function and $g(E)$ is the degeneracy per unit volume. This is a good choice because with many gas atoms with many electrons our degeneracy is so huge a degeneracy per volume is more wieldy. This also explains the V in our equation, it is the volume of our sample. With volumes involved, this feels like we are getting toward bulk material!

Our integrand is

$$dN = Vg(E)f(E)dE$$

and our integral to normalize our probability distribution function is

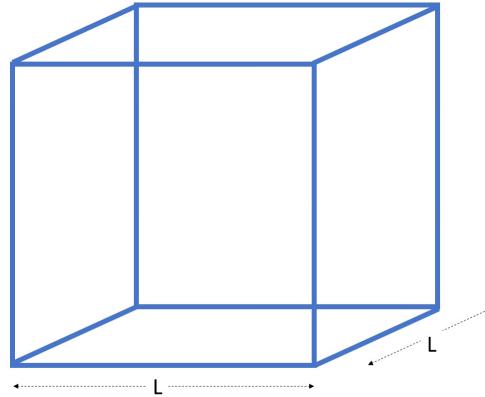
$$N = \int_0^\infty Vg(E)f(E)dE$$

If this discussion has not given you a firm friendly feeling about the function $g(E)$ it is not surprising. It will help to take a concrete example. Let's go back to a gas-like group

of particles because we understand somewhat what we expect to get there, then we can take on more difficult collections of particles once we understand our $g(E)$ function.

Gas of Particles Density of States

Let's assume a three dimensional infinite potential well.



Yes we have only done a one dimensional infinite potential well, but we can argue that we ought to be able to extend our analysis to a three dimensional case without doing all the math to solve the Schrödinger equation from scratch. In one dimension we have a wave function of the form

$$\psi(x) = A \sin kx$$

In three dimensions we expect to have waves in all three directions so

$$\psi(x, y, z) = A \sin(k_x x) \sin(k_y y) \sin(k_z z)$$

and we know from PH123 and our review that the standing waves have to fit in the box just right so we need even numbers of half wavelengths as our allowed states.

$$k_x = \frac{2\pi}{\lambda_x} = \frac{n_x \pi}{L}$$

where L is the size of the box. Let's make our box a perfect cube so each side has length L . Then

$$\begin{aligned} k_y &= \frac{n_y \pi}{L} \\ k_z &= \frac{n_z \pi}{L} \end{aligned}$$

and

$$\psi(x, y, z) = A \sin\left(\frac{n_x \pi}{L} x\right) \sin\left(\frac{n_y \pi}{L} y\right) \left(\sin\left(\frac{n_z \pi}{L} z\right)\right)$$

Let's take a semi-classical system for our first try, so we have kinetic energy

$$E = \frac{p^2}{2m}$$

but we can have momentum in all three directions so

$$E = \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2)$$

now lets assume we have wavicles so

$$p = \frac{h}{\lambda} = \frac{h}{\frac{2\pi}{k}} = \frac{h}{2\pi} k$$

and

$$p_x = \frac{h}{2\pi} \frac{n_x \pi}{L} = \frac{hn_x}{2L}$$

and likewise for p_y and p_z so that

$$\begin{aligned} E &= \frac{1}{2m} \left(\left(\frac{hn_x}{2L} \right)^2 + \left(\frac{hn_y}{2L} \right)^2 + \left(\frac{hn_z}{2L} \right)^2 \right) \\ &= \frac{h^2}{8mL^2} (n_x^2 + n_y^2 + n_z^2) \end{aligned}$$

and if we define

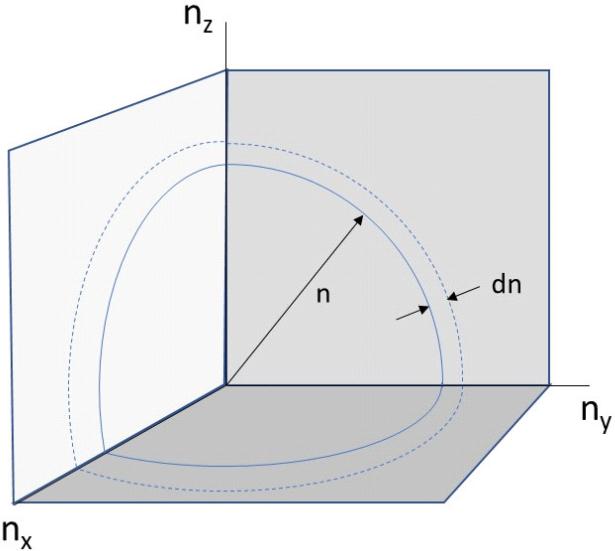
$$n^2 = n_x^2 + n_y^2 + n_z^2$$

then

$$E_n = \frac{h^2}{8mL^2} n^2$$

But we have done something really weird here. The quantity n likely won't be an integer. We have moved from having a few discrete states that are well separated in energy to having a combined set of states due to our particles being in a three dimensional system that will give us fractional n values.

To develop our equation we can do something even weirder. We could picture our n_x , n_y , and n_z as being coordinates on an imaginary n -space. This n -space isn't real, it is a mathematical trick. We want an equation for dE and $g(E)$ so we can integrate. We know $E \propto n^2$ and in this n -space n would be the radius of a sphere. I have drawn one quadrant of the sphere in the next figure.



We could define a spherical shell that would have the thickness dn . And since E_n is directly proportional to n^2 we can say that states with energies between E and $E + dE$ would have n values between n and $n + dn$. We have found a way to count all the states by defining a state-space labeled by our n values! This is just what we want for our $g(E) dE$ part of our integral. It should count all of the available energy states.

So we can integrate over dE just as we wanted to. There is one important restriction. Our n values can't be negative, So we only want the first quadrant of this state-space where n_x , n_y , and n_z are all positive. This gives part of the shell has a state-volume of

$$\frac{1}{8}4\pi n^2 dn$$

This amount of degeneracy is just due to kinetic energy of the gas particles. And if our particles in our gas have half integer spin we would have another degeneracy of $2s + 1$ because without an external magnetic field we get individual particles in any and all of the spin states. So our density of states function would be

$$g(n) dn = \frac{1}{8} \frac{2s+1}{V} 4\pi n^2 dn$$

We see the spin state factor in the $2s + 1$ but where do we get the V from? This is our box volume L^3 for our three dimensional infinite well and we wanted the density of states per unit volume.

$$g(n) dn = \frac{1}{8} \frac{2s+1}{L^3} 4\pi n^2 dn$$

I jumped a bit in writing this. Weren't we planning to have g be a function of E ? But we want the number of states per unit volume. This is a number. So if we realize that

E_n depends on n^2 directly then the number of states per volume we get by calculating $g(n)$ is the same number we would get by calculating $g(E)$ so we can just write

$$g(E) dE = g(n) dn$$

but we do have to write dn in terms of dE and n in terms of E so we can place them in our integral.

$$E = \frac{h^2}{8mL^2} n^2$$

so

$$n^2 = \frac{8mL^2}{h^2} E$$

or

$$n = \frac{L}{h} \sqrt{8mE}$$

Now we can take the derivative of n^2

$$2ndn = \frac{8mL^2}{h^2} dE$$

and substitute in for n

$$2 \left(\frac{L}{h} \sqrt{8mE} \right) dn = \frac{8mL^2}{h^2} dE$$

Solving for dn gives

$$\begin{aligned} dn &= \frac{8mL}{2\sqrt{8mE}h} dE \\ dn &= \frac{\sqrt{8mL}}{2\sqrt{E}h} dE \end{aligned}$$

and then we can write $g(E) dE$ by substituting in equations for n and dn into $g(n) dn$

$$g(E) dE = \frac{1}{8} \frac{2s+1}{L^3} 4\pi \left(\frac{8mL^2}{h^2} E \right) \left(\frac{\sqrt{8mL}}{2\sqrt{E}h} dE \right)$$

or

$$g(E) dE = (2s+1) 4\pi \left(\frac{m}{h^3} \sqrt{E} \right) \left(\sqrt{2m} dE \right)$$

or even

$$g(E) dE = 4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E} dE$$

and our density of states for our gas is

$$g(E) = 4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E}$$

This tells us how many states we have in our box per unit volume. To find out which states are occupied we would need to match this with a distribution function $f(E)$ and because we have many gas particles, we could, say, find the number of particles that have energy between 0 and E_o by integrating

$$N(0 : E_o) = \int_0^{E_o} vg(E) f(E) dE$$

and if we want the number of particles in the energy range $E_o, E_o + \Delta E$

$$N(0 : E_o + \Delta E) = \int_{E_o}^{E_o + \Delta E} vg(E) f(E) dE$$

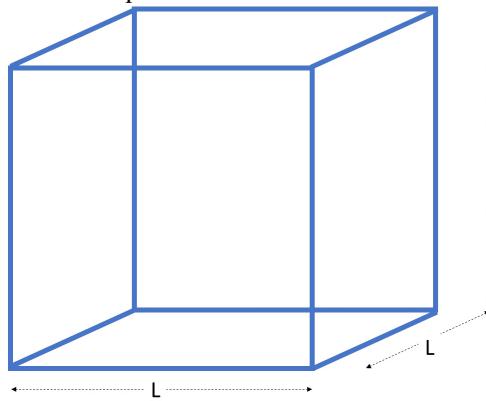
and if ΔE is very small we could even approximate this as

$$N(0 : E_o) = vg(E_o) f(E_o) dE$$

but we will need to figure out how to write $f(E)$ to do the calculations.

Photon gas density of states

There is another example of density of states that we might need later. So let's do it now for practice. We want the density of states function for a "gas" of photons confined in a reflective box with walls at temperature T .



Inside the box we have photons with energies from 0 to ∞ . And of course

$$f = \frac{E}{h}$$

so we get frequencies for 0 to ∞ . And we know

$$E = \frac{hc}{\lambda}$$

so we get

$$\lambda = \frac{hc}{E}$$

so our wavelengths go from ∞ to 0. And

$$E = pc$$

for photons so

$$E = c\sqrt{p_x^2 + p_y^2 + p_z^2}$$

We are going to get standing waves in all three directions in our box. The wave function must go to zero at the walls because they are reflective. For photons, this reflective box is an infinite potential well! That is nice, we can borrow from what we just did for the gas of particles.

$$p_x = \frac{hn_x}{2L}$$

and similarly for p_y, p_z . So

$$\begin{aligned} E &= c \sqrt{\left(\frac{hn_x}{2L}\right)^2 + \left(\frac{hn_y}{2L}\right)^2 + \left(\frac{hn_z}{2L}\right)^2} \\ &= \frac{ch}{2L} \sqrt{n_x^2 + n_y^2 + n_z^2} \end{aligned}$$

and this seems very familiar. We could write this as

$$E = \frac{ch}{2L} n$$

where

$$n^2 = n_x^2 + n_y^2 + n_z^2$$

just like before.

There are details of photons we haven't learned yet. We know they have $s = 1$. So we expect a factor of $2s + 1 = 3$ but photons are polarized perpendicular to their direction of travel so one of the spin components must be zero. Let's set $s_z = 0$ so we only get 2 instead of 3 for our spin state multiplicity. So let's assemble our $g(E)$. Starting with

$$g(n) dn = \frac{1}{8} \frac{2s+1}{V} 4\pi n^2 dn$$

because of photon polarization

$$g(n) dn = \frac{1}{8} \frac{2}{V} 4\pi n^2 dn$$

and using our energy as a function of n

$$E = \frac{ch}{2L} n$$

we can write n as a function of energy, so

$$n = \frac{2L}{ch} E$$

and

$$dn = \frac{2L}{ch} dE$$

so

$$\begin{aligned} g(E) dE &= \frac{1}{8} \frac{2}{L^3} 4\pi \left(\frac{2L}{ch} E\right)^2 \left(\frac{2L}{ch} dE\right) \\ g(E) dE &= \pi \left(\frac{2}{ch} E\right)^2 \left(\frac{2}{ch} dE\right) \\ g(E) dE &= \frac{8\pi}{h^3 c^3} E^2 dE \end{aligned}$$

so our density of states function is

$$g(E) = \frac{8\pi}{h^3 c^3} E^2$$

This is only half of what we need to find the number of particles with a particular energy. We still need probability density functions $f(E)$. We will take on this problem in our next Lecture.

24 Probability Distribution Functions and Quantum Statistics

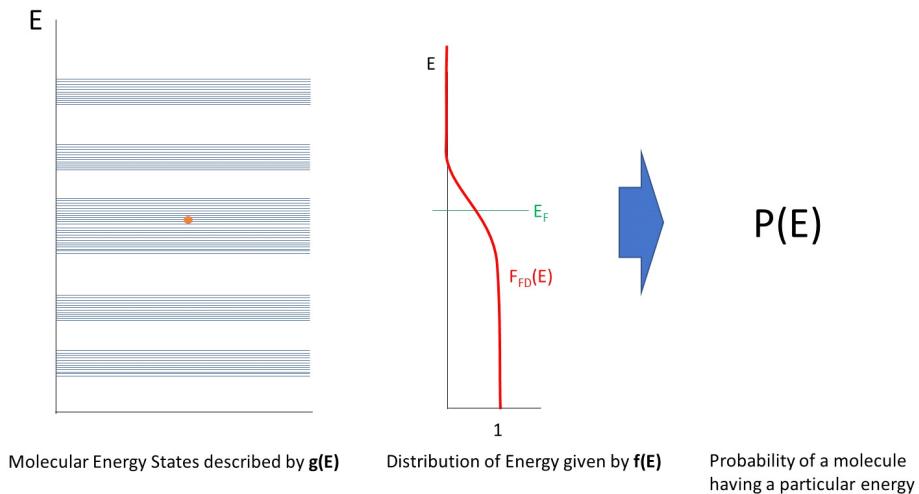
Fundamental Concepts in the Lecture

- Probability Distribution Functions
- Classical Maxwell-Boltzmann distribution function
- Doppler Broadening

In the last lecture we found out how to find the multiplicity $g(E)$ for a large system of particles (atoms, photons, etc.). But we also said that we wanted to find the probability of an event, like picking a particle with 2 units of energy out of a group of particles and that was found with

$$p(E) = \frac{\sum_i N_i W_i}{N \sum_i W_i}$$

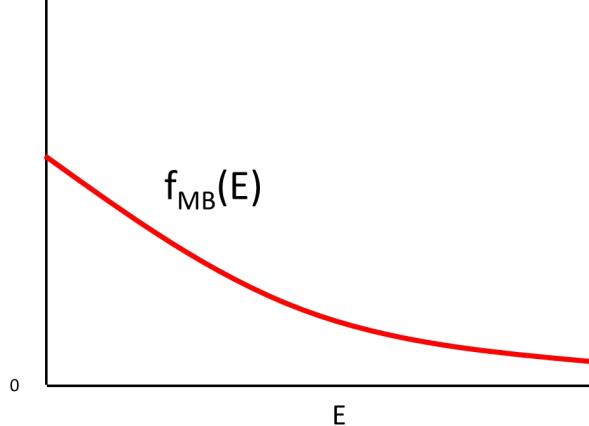
But for large systems, the energy states become close together and to a good approximation we can replace the summations with integrals. To do this we wanted to replace W_i , our multiplicity function that tells us how many ways we can put in particles with a particular pattern of energy, with $g(E)$ that gives the same information for systems with many energy states that are close together. Only $g(E)$ is a per volume quantity. We also want to replace our number of particles in a state with a particular energy, N_i , with the distribution function $f(E)$ so instead of calculating $\sum_i N_i W_i$ we would have $V g(E) f(E)$.



In the last lecture we looked at ways to find $g(E)$ for a system. Now we need to explore finding the distribution functions $f(E)$.

Distribution Functions

In a few paragraphs we will derive the Maxwell-Boltzmann distribution to see how we get a form for $f(E)$. The Maxwell-Boltzmann distribution function is the easiest because it is classical.



It has a maximum value at the intercept that is set by the number of particles we have. It says that as the energy goes up we expect fewer and fewer particles to have that energy. And this is generally our experience with classical systems. Balls roll down hill, things like to find the lowest energy state they can occupy.

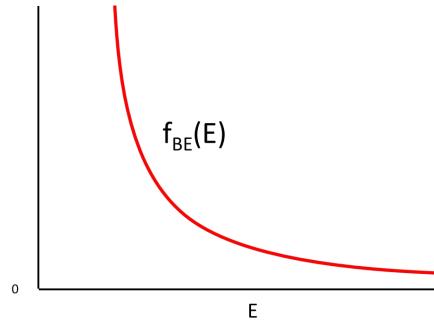


Figure 24.2.

But not all systems work this way. Quantum systems often very much don't work this way. And we will need distribution functions for the cases of quantum particles. We have two cases, one with wavicles with integral spin and another with wavicles that have half integer spin. Let's start with the integer spin case.

$$f_{BE}(E) = \frac{1}{A_{BE} e^{\frac{E}{k_B T}} - 1}$$

Let's give a name for particles with integer spin. We will call them *Bosons*. And the distribution function for Bosons is called the *Bose-Einstein distribution function* after the researchers that worked to describe it. The distribution function looks like this: Note that as $E \rightarrow 0$, $N(E)$ becomes infinite. This means that all the particles try to occupy the lowest energy state. This is not an easy condition to achieve. E must be very small, but when it does happen we call it a Bose-Einstein condensate. The word "condensate" refers to energy, it does not mean becoming a liquid. The particles "condense" to the lowest energy state. There must be quite a bit of math involved in finding the Bose-Einstein distribution function so let's postpone that for a minute (or until a junior level thermal class) and talk about our second case.

We know particles that have half half-integral spin. Electrons and protons both do. The distribution function for half integral spin particles looks like this

$$f_{FD} = \frac{1}{A_{FD} e^{\frac{E}{k_B T}} + 1}$$

This distribution function is called the *Fermi-Dirac distribution*, again after the researchers that developed it. Particles that have half-integer spin are called *Fermions*. The normalization constant A_{FD} isn't so constant. It has an energy dependence.

$$A_{FD} = e^{-\frac{E_F}{k_B T}}$$

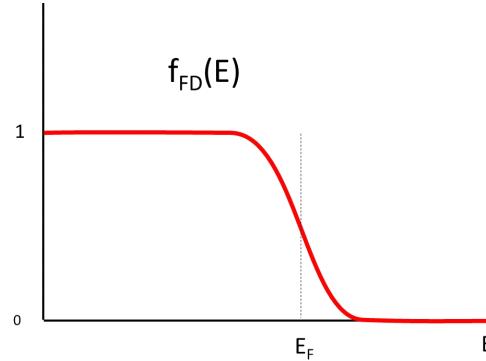
And even this is an approximation. But for the region where it holds true we can write

the Fermi-Dirac distribution as

$$\begin{aligned} f_{FD} &= \frac{1}{e^{-\frac{E_F}{k_B T}} e^{\frac{E}{k_B T}} + 1} \\ &= \frac{1}{e^{\frac{1}{k_B T}(E-E_F)} + 1} \end{aligned}$$

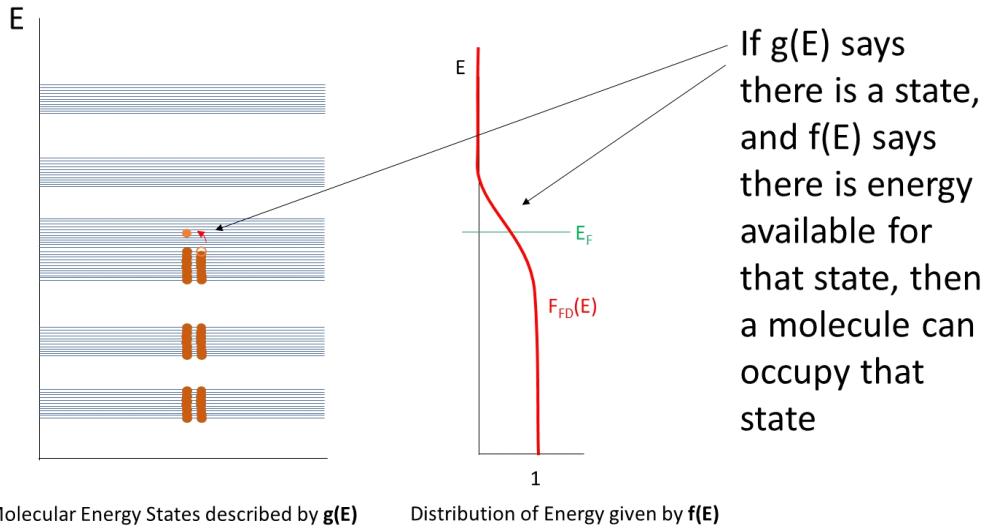
The energy E_F is called the Fermi energy. Again we will postpone the derivation.¹⁰

But a graph of the Fermi-Dirac distribution looks like this.



Note that at some point energies lower than the Fermi energy the distribution function $f_{FD}(E) = 1$. This means that all the allowed states with these energies are full. And a bit higher than the fermi energy all the allowed states are all empty. But right around the Fermi energy there is a transition. This means that, if there is an available energy state with that energy (think $g(E)$) then it could be occupied by a molecule. We can represent $g(E)$ with our usual energy diagram with occupied states designated with orange circles. We can match this with the Fermi-Dirac distribution function to show which states will be full.

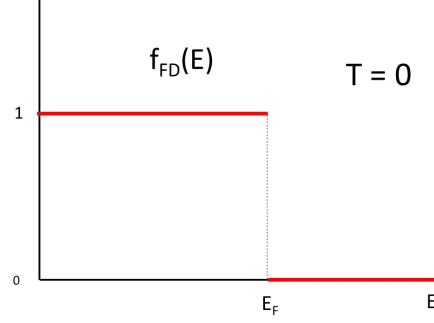
¹⁰ If you are curious, see Keith Stowe, **Introduction to Statistical Mechanics and Thermodynamics**, John Wiley & Sons, 1984. Both the Fermi-Dirac and Bose-Einstein distributions are derived in chapter 24.



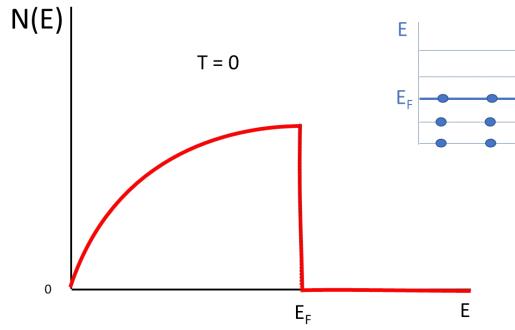
Molecular Energy States described by $g(E)$ Distribution of Energy given by $f(E)$

The lower states are all full where $f_{FD}(E) = 1$. But near $E = E_F$ we can see we have reached the upper shells for our material's atoms and the states start to be open. States directly above E_F will be occupied by electrons from states just below E_F . That is what the f_{FD} curve tells us in the region about E_F .

An important special case is if the temperature is right at absolute zero, $T = 0$ K, then the Fermi Dirac distribution transitions instantaneously from 1 to 0 right at E_F

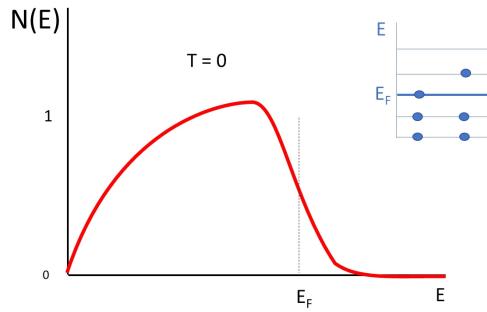


Of course we have just sketched $g(E)$ and $f(E)$ separately so far. We could combine them into a sketch of $N(E) = Vg(E)f(E)$. The degeneracy function $g(E)$ increases with E and $f_{GD}(E)$ is constant for a while then falls off quickly. Then $N(E)$ would look something like this for the $T = 0$ case



This is a ground state scenario where all the electrons are at the lowest energies possible (see the energy diagram insert in the last figure).

If the energy increases so that $T > 0$ then $N(E)$ changes to be more like



Notice that as the temperature goes up, our particles can move to a higher state.

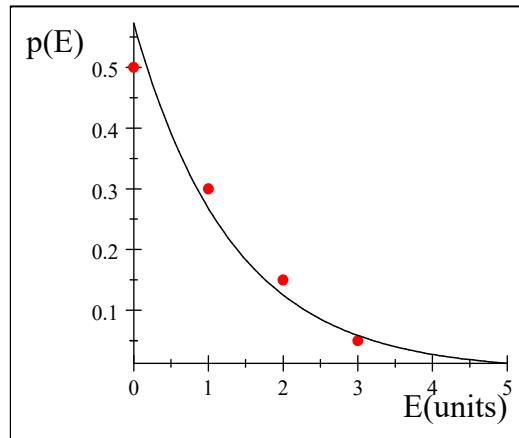
Classical Maxwell-Boltzmann distribution function

So now let's get one of these distribution functions. We will do the classical case and let our upper division quantum classes do the quantum cases. But this will get you ready for your upper division thermodynamics class where Maxwell-Boltzman will be used.

Last lecture we argued that a classical distribution function should look exponential in energy.

$$f(E) = Ae^{-bE}$$

where A and b are likely constants (in energy) and we even did an example



For classical systems like a gas of atoms it is traditional to write this slightly differently

$$f_{MB}(E) = A^{-1} e^{-E/(k_B T)}$$

where b has been written as one over Boltzmann's constant, $k_B = 8.617385 \times 10^{-5} \frac{\text{eV}}{\text{K}}$ multiplied by T , the temperature of the gas. This form is due to the work of Maxwell and Boltzmann (not surprising), thus the MB subscripts. The constant A^{-1} is our normalization constant that is chosen to ensure that the integral

$$\int dN = \int N(E) dE = \int V g(E) f(E) dE = N$$

so we don't miss any particles (or magically invent some that are not there by miscalculating). This is a normalization integral, and we know from a normalization integral we should be able to find the coefficient A^{-1} .

We know a way to write $g(E)$ for a gas from last lecture, That should apply to our case, since the Maxwell-Boltzmann distribution is for gasses.

$$\begin{aligned} \int N(E) dE &= \int V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E} \right) \left(A^{-1} e^{-E/(k_B T)} \right) dE \\ N &= VA^{-1} \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \right) \int_0^\infty E^{\frac{1}{2}} \left(e^{-E/(k_B T)} \right) dE \end{aligned}$$

We can use an integral table for this, but we need to look in the definite integral section.

My table gives

$$\int_0^\infty x^{\frac{1}{2}} (e^{-nx}) dx = \frac{1}{2n} \sqrt{\frac{\pi}{n}}$$

so for us $n = \frac{1}{k_B T}$ and our integral becomes

$$\int_0^\infty E^{\frac{1}{2}} \left(e^{-E/(k_B T)} \right) dE = \frac{1}{2 \left(\frac{1}{k_B T} \right)} \sqrt{\frac{\pi}{\left(\frac{1}{k_B T} \right)}} = \frac{\sqrt{\pi} (k_B T)^{\frac{3}{2}}}{2}$$

Putting this back in our normalization equation gives

$$N = VA^{-1} \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \right) \left(\frac{\sqrt{\pi}(k_B T)^{\frac{3}{2}}}{2} \right)$$

so we can find A^{-1}

$$\frac{N}{V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \right) \left(\frac{\sqrt{\pi}(k_B T)^{\frac{3}{2}}}{2} \right)} = A^{-1}$$

Which we can simplify a bit

$$\begin{aligned} A^{-1} &= \frac{Nh^3}{V (2(2s+1)\sqrt{2}) ((m\pi k_B T)^{\frac{3}{2}})} \\ A^{-1} &= \frac{Nh^3}{V (2s+1) ((2m\pi k_B T)^{\frac{3}{2}})} \\ A^{-1} &= \frac{N(hc)^3}{V (2s+1) c^3 ((2m\pi k_B T)^{\frac{3}{2}})} \\ A^{-1} &= \frac{N(hc)^3}{V (2s+1) ((2m\pi c^2 k_B T)^{\frac{3}{2}})} \end{aligned}$$

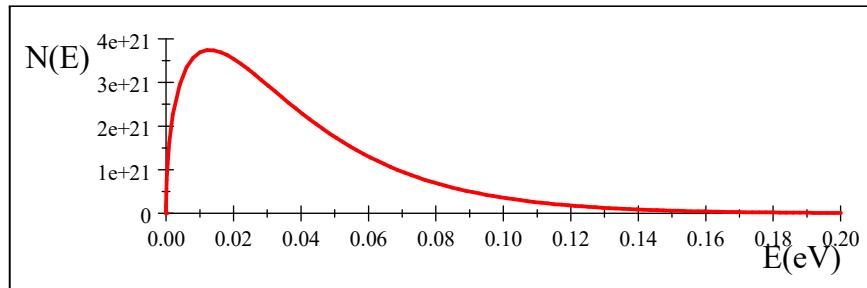
Then our energy distribution of gas molecules is

$$\begin{aligned} N(E) &= V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E} \right) \left(A^{-1} e^{-E/(k_B T)} \right) \\ &= V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E} \right) \left(\frac{N(hc)^3}{V (2s+1) ((2m\pi c^2 k_B T)^{\frac{3}{2}})} e^{-E/(k_B T)} \right) \end{aligned}$$

and many things cancel

$$\begin{aligned} N(E) &= V \left(4\pi \sqrt{2}(m)^{\frac{3}{2}} \sqrt{E} \right) \left(\frac{N(c)^3}{V ((2m\pi c^2 k_B T)^{\frac{3}{2}})} e^{-E/(k_B T)} \right) \\ &= (2\sqrt{E}) \left(\frac{N}{(\sqrt{\pi}(k_B T)^{\frac{3}{2}})} e^{-E/(k_B T)} \right) \\ &= \frac{2N}{\sqrt{\pi}(k_B T)^{\frac{3}{2}}} \sqrt{E} e^{-E/(k_B T)} \end{aligned}$$

and this is the Maxwell-Boltzmann distribution that we used earlier. For $N = 2 \times 10^{20}$ and $T = 300$ K this function looks like this



Note that even though this is for “classical” gases, we use spin state degeneracy in forming this distribution function! Quantum behavior is everywhere!

Molecular speeds

In PH123 we encountered the Maxwell-Boltzmann distribution as a distribution of molecular speeds. In our classical ideal gas model, all we had was kinetic energy so

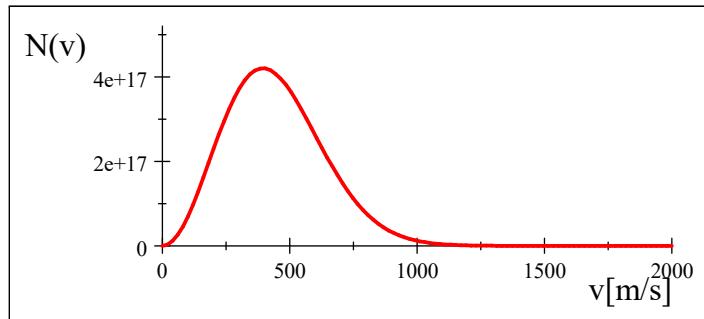
$$E = \frac{1}{2}mv^2$$

and we can put this into the Maxwell-Boltzmann distribution

Recall that our number of molecules is a number, it doesn’t matter how we count it so

$$\begin{aligned} N(v)dv &= N(E)dE \\ N(v) &= N(E) \frac{dE}{dv} \\ N(v) &= \frac{2N}{\sqrt{\pi}(k_B T)^{\frac{3}{2}}} \sqrt{E} e^{-E/(k_B T)} \frac{dE}{dv} \\ &= \frac{2N}{\sqrt{\pi}(k_B T)^{\frac{3}{2}}} \sqrt{\frac{1}{2}mv^2} e^{-(\frac{1}{2}mv^2)/(k_B T)} (mv) \\ &= N \sqrt{\frac{2}{\pi}} \left(\frac{m}{k_B T} \right)^{\frac{3}{2}} v^2 e^{-(mv^2)/(2k_B T)} \end{aligned}$$

This is called the Maxwell-Boltzmann velocity distribution. For $N = 2 \times 10^{20}$ and $T = 300$ K for diatomic oxygen this distribution looks like this.

Maxwell-Boltzman velocity distribution for O_2 at 300 K where $N = 2 \times 10^{20}$

Doppler Broadening

Let's use what we can do with this idea of finding the number of particles with a particular energy, or velocity, etc. Let's take on the problem of line spectra of a gas. We have dealt with the spectra from single atoms and single molecules. But in a gas the gas molecules move around. And we know about Doppler shift. We should worry. The line spectra frequencies must be affected by the motion of the molecules. Where we would have one very small ΔE given only by the uncertainty relation due to the Δt it takes for an electron to jump shells, we now expect to have a bigger uncertainty because some molecules will be moving away from us giving a smaller frequency and some will move toward us giving a higher frequency and some will be moving parallel to our view giving just the frequency

$$f = \frac{\Delta E}{h}$$

Because Doppler shift only depends on the component of the velocity along our view direction, this is really a one dimensional problem. So let's try a degeneracy function that looks like this

$$g(n_x) dn_x = \left(\frac{2s+1}{L} \right) (dn_x)$$

That is, we have allowed for spin angular momentum, and kinetic energy in the x direction as the two ways to give particles energy. We have L as the length of our one-dimensional system. In three dimensions, $g(E)$ is per unit volume, but this is a one dimensional case so our $g(E)$ will be per unit length, and that is where the L came from. L is the length of our collection of gas molecules. As usual, we need to go from $g(n_x) dn_x$ to $g(E) dE$. But because we only have kinetic energy to deal with it is actually a little more insightful to form $g(v) dv$

Again we have primarily kinetic energy for a group of gas molecules so

$$E = \frac{p^2}{2m}$$

and this is a one dimensional case so

$$E = \frac{p_x^2}{2m}$$

and from our previous gas problems we know that we get standing waves in our gas cavity and the energy states look like

$$E_n = \frac{n_x^2 h^2}{8mL^2}$$

and our gas molecules energy is kinetic energy is

$$\frac{(mv_x)^2}{2m} = \frac{n_x^2 h^2}{8mL^2}$$

so then

$$\begin{aligned} v_x^2 &= \frac{n_x^2 h^2}{4m^2 L^2} \\ v_x &= \sqrt{\frac{n_x^2 h^2}{4m^2 L^2}} \\ v_x &= \frac{n_x h}{2mL} \end{aligned}$$

The dv_x is just

$$dv_x = \frac{h}{2mL} dn_x$$

and solving for dn_x gives]

$$\frac{2mL}{h} dv_x = dn_x$$

so our degeneracy function is

$$\begin{aligned} g(v_x) dv_x &= \frac{2s+1}{L} \left(\frac{2mL}{h} dv_x \right) \\ &= \frac{2m(2s+1)}{h} dv_x \end{aligned}$$

and then our number of particles with speed v_x can be found

$$dN = N(v_x) dv_x = L g(v_x) f(v_x) dv_x$$

where we can use our Maxwell-Boltzmann velocity distribution again to find $f(v_x)$.

We have $g(v_x)$ so

$$dN = N(v_x) dv_x = L \left(\frac{2m(2s+1)}{h} \right) f(v_x) dv_x$$

We use the Maxwell-Boltzmann distribution because our gas molecules are non-relativistic and their energy of motion is kinetic, not quantized.

$$f(E) = A^{-1} e^{-\frac{E}{k_B T}}$$

then

$$dN = N(v_x) dv_x = L \left(\frac{2m(2s+1)}{h} \right) \left(A^{-1} e^{-\frac{E}{k_B T}} \right) dv_x$$

$$dN = N(v_x) dv_x = L \left(\frac{2m(2s+1)}{h} \right) \left(A^{-1} e^{-\frac{\frac{1}{2}mv_x^2}{k_B T}} \right) dv_x$$

$$dN = N(v_x) dv_x = L \left(\frac{2m(2s+1)}{h} \right) \left(A^{-1} e^{-\frac{mv_x^2}{2k_B T}} \right) dv_x$$

We will need to normalize to find A^{-1}

$$N = \int dN = \int_{-\infty}^{\infty} L \left(\frac{2m(2s+1)}{h} \right) \left(A^{-1} e^{-\frac{mv_x^2}{2k_B T}} \right) dv_x$$

where we need both positive and negative velocities so we have positive and negative limits of integration.

$$N = L \left(\frac{2m(2s+1)}{h} A^{-1} \right) \int_{-\infty}^{\infty} e^{-\frac{mv_x^2}{2k_B T}} dv_x$$

we can use the form

$$\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}$$

by defining

$$u^2 = \frac{mv_x^2}{2k_B T}$$

so

$$u = \sqrt{\frac{m}{2k_B T}} v_x$$

and

$$du = \sqrt{\frac{m}{2k_B T}} dv_x$$

Then dv_x is

$$du \sqrt{\frac{2k_B T}{m}} = dv_x$$

We can use these in our integral

$$N = L \left(\frac{2m(2s+1)}{h} A^{-1} \right) \int_{-\infty}^{\infty} e^{-u^2} \left(du \sqrt{\frac{2k_B T}{m}} \right)$$

$$N = L \left(\frac{2m(2s+1)}{h} A^{-1} \right) \sqrt{\frac{2k_B T}{m}} \int_{-\infty}^{\infty} e^{-u^2} du$$

$$N = L \left(\frac{2m(2s+1)}{h} A^{-1} \right) \sqrt{\frac{2k_B T}{m}} \sqrt{\pi}$$

which gives an expression for A^{-1}

$$\frac{Nh}{2Lm(2s+1)} \sqrt{\frac{m}{2\pi k_B T}} = A^{-1}$$

So our number of particles with speed v_x can be found from

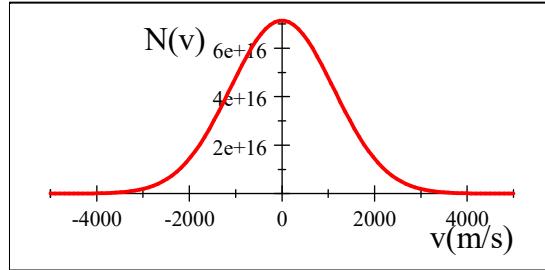
$$N(v_x) dv_x = L \left(\frac{2m(2s+1)}{h} \right) \left(\left(\frac{Nh}{2Lm(2s+1)} \sqrt{\frac{m}{2\pi k_B T}} \right) e^{-\frac{mv_x^2}{2k_B T}} \right) dv_x$$

$$= N \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{mv_x^2}{2k_B T}} dv_x$$

Which tells us that our velocity distribution function is

$$N(v_x) = N \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{mv_x^2}{2k_B T}}$$

which is a Gaussian function. For H₂ with $N = 2 \times 10^{20}$ and $T = 300\text{ K}$ we would have



We have found the distribution of molecules with each speed. But what we really want is to know what this does to the frequency of light emitted by the molecules. We want to know how many atoms will emit frequencies between f and $f + df$ so we want

$$dN = N(f) df$$

and now we need a little bit of relativity. We need the relativistic Doppler shift equations we derived earlier in this course. We got

$$f_B = f_A \frac{\sqrt{(1 - \frac{v_{BA}}{c})}}{\sqrt{(1 + \frac{v_{BA}}{c})}}$$

where our A frame is the frame of the moving molecule, and our B frame is the lab frame. We can use $v_x = v_{BA}$. Our speeds are small, so let's make some approximations

$$\begin{aligned} f_B &= f_A \frac{\sqrt{(1 - \frac{v_x}{c})}}{\sqrt{(1 + \frac{v_x}{c})}} \\ &= \frac{f_A \sqrt{(1 - \frac{v_x}{c})} \sqrt{(1 - \frac{v_x}{c})}}{\sqrt{(1 + \frac{v_x}{c})} \sqrt{(1 - \frac{v_x}{c})}} \\ &= f_A \frac{(1 - \frac{v_x}{c})}{\sqrt{(1 - \frac{v_x^2}{c^2})}} \end{aligned}$$

and since $v_x \ll c$

$$\begin{aligned} f_B &\approx f_A \frac{(1 - \frac{v_x}{c})}{\sqrt{(1 - 0)}} \\ &= f_A \left(1 - \frac{v_x}{c}\right) \end{aligned}$$

We can choose this to change variables

$$f_B = f_A \left(1 - \frac{v_x}{c}\right)$$

so

$$\begin{aligned}\frac{f_B}{f_A} &= \left(1 - \frac{v_x}{c}\right) \\ \frac{f_B}{f_A} - 1 &= -\frac{v_x}{c} \\ c \left(1 - \frac{f_B}{f_A}\right) &= v_x\end{aligned}$$

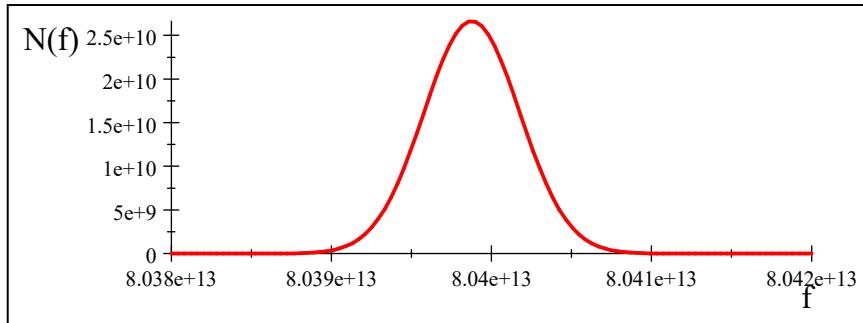
and

$$dv_x = \frac{c}{f_A} df_B$$

We, in the lab see f_B . Let's drop the subscript from this and just call it f , the frequency we see

$$\begin{aligned}N(f) df &= N \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{m(c(1-\frac{f}{f_A}))^2}{2k_B T}} \frac{c}{f_A} df \\ &= \frac{Nc}{f_A} \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{m(c(1-\frac{f}{f_A}))^2}{2k_B T}} df\end{aligned}$$

which is a Gaussian function. For H₂ with $N = 2 \times 10^{20}$, $T = 30000$ K, and $f_o = 8.0398 \times 10^{13}$ Hz we would have



and this shows us that the line indeed is broader because the molecules are moving. This is what a spectrograph will actually see.

How do we know when to use quantum statistics?

The general rule of thumb is that particles must be physically separated by more than a wavelength to be considered classical. If we consider thermal energy for our particles, then the amount of energy is about $k_B T$ and it will be mostly kinetic so if we set

$$\frac{p^2}{2m} = k_B T$$

then

$$p = \sqrt{2mk_B T}$$

and

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

so

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mk_B T}} = \frac{hc}{\sqrt{2mc^2 k_B T}}$$

The spacing between particles can be approximated by using the density

$$\rho = \frac{N}{V}$$

The spacing between particles is then roughly

$$d = \frac{1}{\sqrt[3]{\frac{N}{V}}}$$

We would consider the situation to be classical if

$$\frac{\lambda}{d} \ll 1$$

then

$$\frac{\frac{hc}{\sqrt{2mc^2 k_B T}}}{\sqrt[3]{\frac{N}{V}}} \ll 1$$

$$\frac{hc}{\sqrt{2mc^2 k_B T}} \left(\frac{N}{V} \right)^{\frac{1}{3}} \ll 1$$

Notice that for the Maxwell-Boltzmann distribution our amplitude would be

$$\begin{aligned} A^{-1} &= \frac{N(hc)^3}{V(2s+1) \left((2m\pi c^2 k_B T)^{\frac{3}{2}} \right)} = \frac{1}{(2s+1)\pi^{\frac{3}{2}}} \left(\frac{(hc) \left(\frac{N}{V} \right)^{\frac{1}{3}}}{\sqrt{2mc^2 k_B T}} \right)^3 \\ &= \frac{1}{(2s+1)\pi^{\frac{3}{2}}} \left(\frac{\lambda}{d} \right)^3 \end{aligned}$$

so if $\lambda/d \ll 1$ then

$$A^{-1} \ll 1$$

then we can calculate A^{-1} and test it to see if a classical treatment would be enough.

Examples of finding energy distribution functions

Let's try to use what we have done in finding $N(E)$ for some important cases.

Thermal radiation and the Plank Function

Let's go back to our photons in cavity. This case is a bit of a strange case. We expect

to use the Bose-Einstein distribution function because photons have integer spin. But there is a problem. The photons are particles that can be destroyed and recreated. Then in our radiator cavity, then N will constantly change. Then A_{BE} can be variable. We will start with being interested in the shape of $N(E)$ more than the numeric value. So let's set $A_{BE} = 1$ and we look at the shape only. We have

$$dN = N(E) dE = V g(E) f(E) dE = V \frac{8\pi}{(hc)^3} E^2 \frac{1}{e^{\frac{E}{k_B T}} - 1} dE$$

The photons will have different energies. The number of photons with energy between E and $E + dE$ would be $\int EdN$ but for small dE we could approximate this by just $EdN = EN(E) dE$. Let's find the energy density in the cavity.

$$\begin{aligned} u(E) dE &= \frac{EN(E) dE}{V} \\ &= \frac{8\pi}{(hc)^3} E^3 \frac{1}{e^{\frac{E}{k_B T}} - 1} dE \end{aligned}$$

and the total energy density is

$$\begin{aligned} U &= \int_0^\infty u(E) dE \\ &= \int_0^\infty \frac{8\pi}{(hc)^3} E^3 \frac{1}{e^{\frac{E}{k_B T}} - 1} dE \\ &= \frac{8\pi}{(hc)^3} \int_0^\infty \frac{E^3}{e^{\frac{E}{k_B T}} - 1} dE \end{aligned}$$

We can find from a table that

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15}$$

and we can use this with

$$\begin{aligned} x &= \frac{E}{k_B T} \\ x k_B T &= E \end{aligned}$$

and

$$\begin{aligned} dx &= \frac{dE}{k_B T} \\ k_B T dx &= dE \\ U &= \frac{8\pi}{(hc)^3} \int_0^\infty \frac{(x k_B T)^3}{e^x - 1} k_B T dx \\ U &= \frac{8\pi (k_B T)^4}{(hc)^3} \int_0^\infty \frac{x^3}{e^x - 1} dx \\ U &= \frac{8\pi (k_B T)^4}{(hc)^3} \left(\frac{\pi^4}{15} \right) \\ U &= \left(\frac{8\pi^5 (k_B)^4}{15 (hc)^3} \right) T^4 \end{aligned}$$

which is the Stefan-Boltzmann law. And if we remember

$$E = \frac{hc}{\lambda}$$

and

$$|dE| = \left| \frac{hc}{\lambda^2} d\lambda \right|$$

so we can recognize

$$\begin{aligned} u(E) dE &= \frac{8\pi}{(hc)^3} \left(\frac{hc}{\lambda} \right)^3 \frac{1}{e^{\frac{E}{k_B T}} - 1} \frac{hc}{\lambda^2} d\lambda \\ u(E) dE &= \frac{8\pi hc}{\lambda^5} \frac{1}{e^{\frac{E}{k_B T}} - 1} d\lambda \end{aligned}$$

And if you recall from PH220

$$I = u \frac{c}{4}$$

we have

$$I = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{\frac{E}{k_B T}} - 1}$$

which is the Planck function. Now we can see what was meant way back at the beginning of our course when we said that Plank envisioned the cavity radiator as not having energies of just any energy. The situation is that the photons reflect making standing waves. Those standing waves are quantized like all standing waves. The quantized energies that plank assumed just come from the allowed energies of the standing waves.

Bose-Einstein Condensation

Let's go back to the Bose-Enistein condensate. We have the Bose-Einstein distribution function

$$f_{BE} = \frac{1}{A_{BE} e^{\frac{E}{k_B T}} - 1}$$

and let's consider the density of states function for a gas

$$g(E) = 4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E}$$

Then

$$\begin{aligned} N(E) dE &= V g(E) f_{BE}(E) dE \\ &= V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E} \right) \left(\frac{1}{A_{BE} e^{\frac{E}{k_B T}} - 1} \right) dE \end{aligned}$$

But instead of using our normalization integral, let's find the maximum number of particles that we can include in volume V . We want to maximize this, so we need to

minimize the denominator. To do that we will make $A_{BE} \approx 1$ so that

$$N(E) dE = V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E} \right) \left(\frac{1}{(1)e^{\frac{E}{k_B T}} - 1} \right) dE$$

and we can integrate

$$\begin{aligned} N_{\max} &= \int_0^\infty V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E} \right) \left(\frac{1}{(1)e^{\frac{E}{k_B T}} - 1} \right) dE \\ &= V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \right) \int_0^\infty \sqrt{E} \left(\frac{1}{e^{\frac{E}{k_B T}} - 1} \right) dE \\ &= V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \right) \int_0^\infty \sqrt{E} \left(\frac{1}{e^{\frac{E}{k_B T}} - 1} \right) dE \end{aligned}$$

We have seen this before, let's do it numerically this time.

$$\int_0^\infty u^{\frac{1}{2}} \left(\frac{1}{e^u - 1} \right) du = 2.31516$$

so we want our integral to look like this. We are close. If we let

$$\begin{aligned} u &= \frac{E}{k_B T} \\ uk_B T &= E \\ du k_B T &= dE \end{aligned}$$

Then

$$\begin{aligned} N_{\max} &= V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \right) \int_0^\infty \sqrt{u k_B T} \left(\frac{1}{e^u - 1} \right) du k_B T \\ N_{\max} &= V \left(4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \right) (k_B T)^{\frac{3}{2}} \int_0^\infty \sqrt{u} \left(\frac{1}{e^u - 1} \right) du \\ N_{\max} &= V \left((2.31516) \pi (2s+1) \left(\frac{2mk_B T}{h^2} \right)^{\frac{3}{2}} \right) \end{aligned}$$

@@@@@ I get a different coefficient than Krane @@@@

This is the maximum that should be in a volume V . But we could increase N_{\max} by putting more particles into the volume, or by lowering the temperature. But something is wrong with our analysis. If this is the maximum, we should not be able to add to the maximum. And it takes a minute to figure out what our problem is. Let's look at the density of states again.

$$g(E) = 4\pi (2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E}$$

This says that if $E = 0$ then $g(0) = 0$. But there must be a ground state. So there must be a possibility of a forbidden transition to $E = 0$ and this transition isn't subject to our

maximum value of N . This is the “condensing” action, if the temperature is low enough the particles will transition to the $E = 0$ state and that state isn’t bound by our $g(E)$ so it doesn’t have a restriction on how many particles can be there. What we calculated in N_{\max} was the number of particles that are *not* in the ground state because the ground state wasn’t in our $g(E)$. The total number of particles has to be

$$N_{\text{total}} = N_0 + N_{\text{ex}}$$

where N_{ex} is the particles in the excited states, and N_{ex} is our N_{\max} .

We really should find the temperature at which this transition occurs. We can use what we found for N_{\max} , but we want the case when all the particles will just barely not be in the ground state so let

$$\left(\frac{N_{\text{total}}}{V(2.31516)\pi(2s+1)} \right)^{\frac{2}{3}} = \frac{2mk_B T}{h^2}$$

$$\left(\frac{N_{\text{total}}}{V(2.31516)\pi(2s+1)} \right)^{\frac{2}{3}} \frac{h^2}{2mk_B} = T_C$$

below T_C we will get particles starting to form a condensate.

25 More Probability Distribution Functions and Quantum Statistics

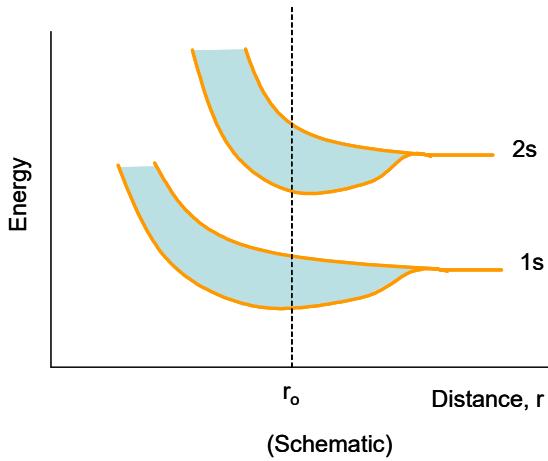
Fundamental Concepts in the Lecture

- Conduction in metals can be modeled as a “gas” of free electrons inside the pieces of metal.
- Heat capacity at low temperatures can be modeled using the Fermi-Dirac distribution
- Both use the idea of $N(E)dE = Vg(E)f(E)dE$

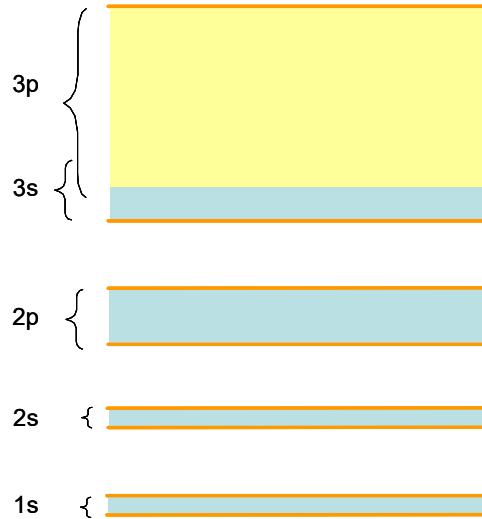
We found that we could use our description of the number of particles with a particular energy $N(E)$ as being the product of a degeneracy function $g(E)$ and a distribution function $f(E)$ such that $N(E)dE = Vg(E)f(E)dE$ allowed us to understand the Planck function and Bose-Einstein condensation. There are so many more phenomena that we can explain using statistical mechanics. This is so true that we have a whole class to deal with thermal energy for senior's in physics. For now, let's take on two more important effects, the conduction of electrons in metals, and heat capacities for solids (at least at low temperatures).

Free Electron Model of Metals

We are going to need to understand the quantum properties that make some elements conductive metals. We know already that metals have a loosely bound electrons in their highest electron energy level. We also know that as we bring atoms together the energy states split, making many states that are close together. This means the weakly bound electrons are allowed to take on energy and jump up into a vacant energy state. Here is our energy vs. separation distance graph again.



where r_o is the equilibrium distance. Since r_o is where the atoms usually are on this graph. We will not draw the whole diagram. We will instead just draw a slice of this last graph right around $r = r_o$. Here is an example.



And note that up in the $3s$ and $3p$ state region we have lots of states. But for many atoms these states are not full. We might have one $3s$ electron, but because the energy levels are so close together, that $3s$ electron can take on energy of almost any value in the *band* of energies from the bottom of the $3s$ region all the way to the top of the $3p$ region. And that energy *could* be kinetic energy. This means the electrons can move! This is what makes a metal a conductor.

Inside the metal, the outer electrons behave a lot like a confined gas. They are weakly

interacting with each other and strongly interact with the boundary of the metal. We can describe this situation with a Fermi-Dirac probability distribution

$$f_{FD} = \frac{1}{e^{\frac{(E-E_f)}{k_B T}} + 1}$$

and with a multiplicity function for individual electrons

$$g(n) dn = \frac{1}{8V} \frac{d}{dE} 4\pi n^2 dn$$

where d is the degeneracy for our “free” electrons (where we can put energy in the electrons, themselves) and the rest we know comes from forming standing waves in a container (where we can put energy by moving the electrons). Electrons can only have $\pm \frac{1}{2}$ so we would expect $d = 2$.

The electrons will be confined to our piece of metal. This is a little like a three dimensional infinite potential well. We expect that the electrons will act a lot like a “gas” confined to a potential well that they can’t escape. We can guess that the energy states would be like those from the infinite potential well.

$$E = \frac{\hbar^2 n^2}{8mL^2}$$

We once again we can write

$$n^2 = E \frac{8mL^2}{\hbar^2}$$

and, taking a derivative gives

$$2ndn = dE \frac{8mL^2}{\hbar^2}$$

We can write just n as

$$n = \sqrt{E \frac{8mL^2}{\hbar^2}}$$

so dn is

$$dn = \frac{1}{2} \left(E \frac{8mL^2}{\hbar^2} \right)^{-\frac{1}{2}} dE \frac{8mL^2}{\hbar^2}$$

which we can simplify

$$dn = \frac{1}{2} \frac{1}{\sqrt{E}} \left(\frac{8mL^2}{\hbar^2} \right)^{\frac{1}{2}} dE$$

so our degeneracy function would be

$$g(E) dE = \frac{1}{8V} 4\pi \left(E \frac{8mL^2}{\hbar^2} \right) \left(\frac{1}{2} \frac{1}{\sqrt{E}} \left(\frac{8mL^2}{\hbar^2} \right)^{\frac{1}{2}} dE \right)$$

and we want to simplify this a bit. We can combine the 8, m , and L parts

$$g(E) dE = \frac{1}{16V} \frac{d}{dE} 4\pi \sqrt{E} \left(\frac{8mL^2}{\hbar^2} \right)^{\frac{3}{2}} dE$$

or even take out the $8^{\frac{3}{2}}$

$$g(E) dE = \frac{1}{16V} \frac{d}{dE} 4\pi \sqrt{E} 16\sqrt{2} \left(\frac{mL^2}{\hbar^2} \right)^{\frac{3}{2}} dE$$

which cancels the 16

$$g(E) dE = \frac{d}{L^3} 4\pi\sqrt{E}\sqrt{2} \left(\frac{mL^2}{h^2} \right)^{\frac{3}{2}} dE$$

Applying the 3/2 power gives

$$g(E) dE = \frac{4\pi d\sqrt{E_n}\sqrt{2}m^{\frac{3}{2}}}{h^3} dE$$

and finally, putting in $d = 2$

$$g(E) dE = \frac{8\pi\sqrt{E}\sqrt{2}m^{\frac{3}{2}}}{h^3} dE$$

Note what we have done. We considered where we could put energy into our particles (electrons in this case, with spin states) and where we could put energy into the system by moving our particles (making standing waves). This is the same pattern we used for gases and photons, etc.

Combining our $g(E)$ with our $f(E)$ we would have a number of particles as a function of energy

$$N(E) dE = V \left(\frac{8\pi\sqrt{E}\sqrt{2}m^{\frac{3}{2}}}{h^3} \right) \left(\frac{1}{e^{\frac{(E-E_F)}{k_B T}} + 1} \right) dE$$

Let's put all the factors of energy in one place

$$N(E) dE = V \frac{8\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \frac{\sqrt{E}}{e^{\frac{1}{k_B T}(E-E_F)} + 1} dE$$

We need the value of E_F for the particular metal. We can find this by normalizing for the spacial case when $T = 0$. Then all the electrons will be in the lowest energy state we can get for our outer electrons. And since we said we would have one conduction electron in the highest energy orbital per atom, we know that N is equal to the number of atoms. The number of conduction electrons should be

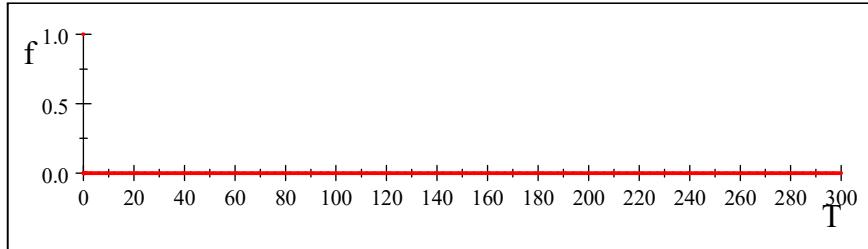
$$\begin{aligned} N &= \int_0^\infty V \frac{8\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \frac{\sqrt{E}}{e^{\frac{1}{k_B T}(E-E_F)} + 1} dE \\ &= \frac{8V\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \int_0^\infty \frac{\sqrt{E}}{e^{\frac{1}{k_B T}(E-E_F)} + 1} dE \end{aligned}$$

and we want to do this integral at $T = 0$.

Let's look at the Fermi-Dirac distribution as a function of T . Suppose we take $E - E_F = 2 \text{ eV}$, then

$$f(2 \text{ eV}) = \frac{1}{e^{\frac{1}{8.617385 \times 10^{-5} \text{ eV K}^{-1} T}(2 \text{ eV})} + 1}$$

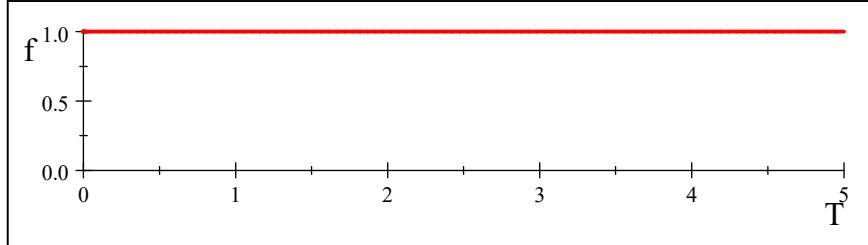
which looks like this



Now suppose we take $E - E_F = -2 \text{ eV}$ then

$$\frac{1}{e^{\frac{1}{8.617385 \times 10^{-5} \text{ eV K}^{-1} T} (-2 \text{ eV})} + 1}$$

which looks like this



so we see that if $E < E_F$ then $f_{FD} = 1$ and if $E > E_F$ then $f_{FD} = 0$ so we can write our integral just from 0 to E_F

$$\begin{aligned} N &= \frac{8V\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \int_0^{E_F} \sqrt{E} (1) dE \\ &= \frac{8\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \left(\frac{2}{3}E_F^{\frac{3}{2}} \right) \end{aligned}$$

then

$$\begin{aligned} N \frac{h^3}{8V\pi\sqrt{2}m^{\frac{3}{2}}} &= \frac{2}{3}E_F^{\frac{3}{2}} \\ N \frac{3h^3}{16\pi\sqrt{2}m^{\frac{3}{2}}} &= E_F^{\frac{3}{2}} \end{aligned}$$

Solving for E_F gives

$$\begin{aligned} E_F &= \left(N \frac{3h^3}{V\pi(8 \times 2)\sqrt{2}m^{\frac{3}{2}}} \right)^{\frac{2}{3}} \\ &= \frac{h^2}{2m} \left(\frac{3N}{8V\pi} \right)^{\frac{2}{3}} \\ &= \frac{h^2c^2}{2mc^2} \left(\frac{3N}{8V\pi} \right)^{\frac{2}{3}} \end{aligned}$$

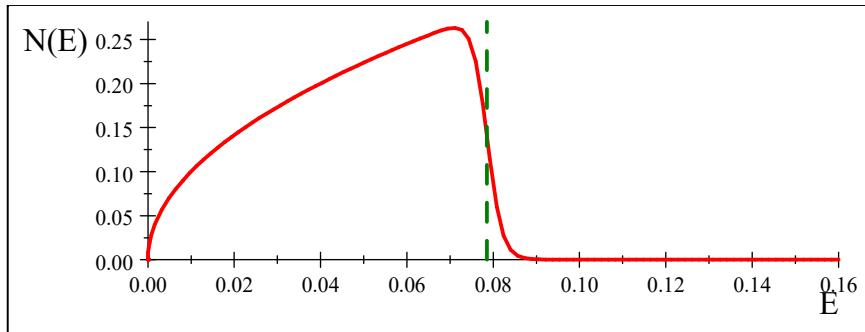
which is an expression for the Fermi energy.

Let's look at this for, say, $N = 1 \times 10^{20}$ electrons in a $V = 1 \text{ cm}^3$ piece of metal.

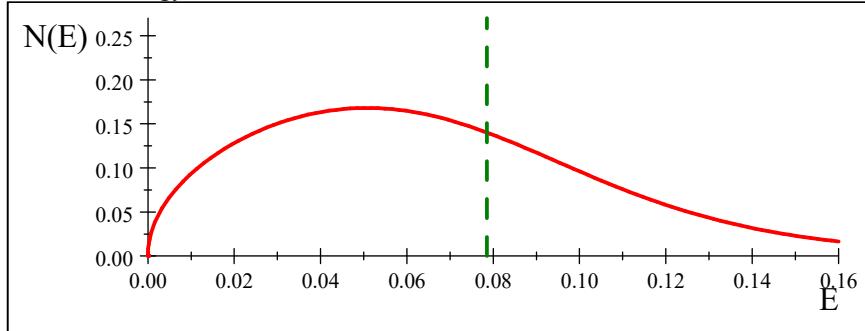
Recalling that $m = 0.51099906 \text{ MeV}/c^2$ for electrons, then

$$\begin{aligned} E_F &= \frac{(1239.8 \text{ eV nm})^2}{2(0.51099906 \text{ MeV})} \left(\frac{3(1 \times 10^{20})}{8(1 \text{ cm}^3)\pi} \right)^{\frac{2}{3}} \\ &= 7.8556 \times 10^{-2} \text{ eV} \end{aligned}$$

So for $T = 20$ we would get a graph of $N(E)$ that looks like this



The Fermi energy is indicated by a green dashed line. At $T = 300$ we can see that more electrons will be free to move by noting that the curve has a higher number on the right of the Fermi energy.



This, of course, means that the number had to go down to the left of the Fermi energy because we have a constant amount of electrons.

Heat capacity of Dilute Solutions of in ${}^4\text{He}$

As a final example, let's look at the origin of heat capacities. Remember that a heat capacity relates a change in temperature to a change in internal energy

$$Q = C\Delta T$$

or

$$C = \frac{\Delta E}{\Delta T}$$

with the limit for small changes dT of

$$C = \frac{dE}{dT}$$

So our strategy would be to find an expression for the energy and take a derivative with respect to temperature.

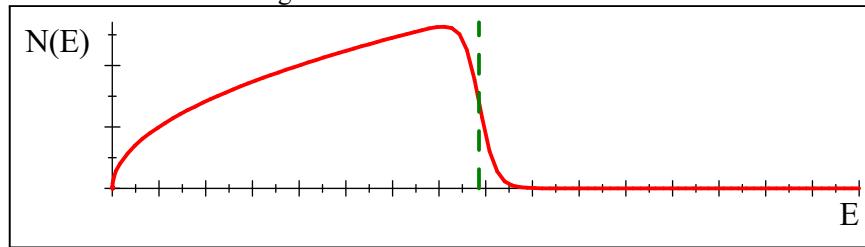
We will do more on this later, so for now let's do the heat capacity of dilute solutions of ^3He in ^4He . This is a pretty strange, specific example. But it is a cool one.

The isotope ^3He is rare, occurring at about 1 part in 10^6 . The two isotopes react chemically just the same, but we know ^4He has matching electrons, matching protons, and matching neutrons. Each has the possibility of $m_s = \pm \frac{1}{2}$. So as long as they are matched, the total spin is 0 in each case. So ^4He has integer spin and is a Boson. But if we remove one neutron, then the atom would have spin of $\pm \frac{1}{2}$ from the remaining neutron. And this would make ^3He a Fermion. And the Pauli exclusion principle would apply. Then if we cooled our helium sample, the ^4He would become a superfluid with all the ^4He atoms in the ground state at about $T = 2.17\text{ K}$. But the ^3He can't do this because of its half integer spin. So we would have a superfluid with ^3He atoms dispersed in it. This is like a very strange "gas" of ^3He . And we now have experience with very strange "gases."

This gas would follow Fermi-Dirac statistics because ^3He is a fermion. So we would expect

$$N(E) dE = V g(E) \left(\frac{1}{e^{\frac{(E-E_F)}{k_B T}} + 1} \right)$$

which should look something like this



as we have seen before.

First note that if $T = 0$ then

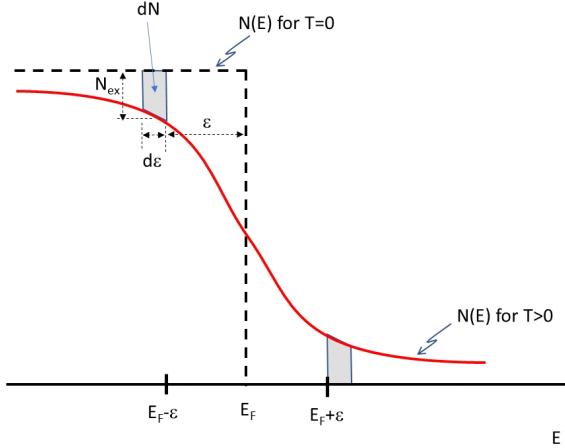
$$N(E) dE = V g(E) 1$$

so long as $E < E_F$ just as we found before. If we let the temperature rise to T then we

get

$$N(E) dE = V g(E) \frac{1}{e^{\frac{(E-E_f)}{k_B T}} + 1}$$

Both are plotted in the next figure which is like our previous graph, but expanded to show just the region around E_F .



Let's take a small number of ${}^3\text{He}$ atoms, dN that have an energy of about $E = E_F - \varepsilon$. Then

$$-\varepsilon = E - E_F$$

and we could write our equation for $T > 0$ as

$$N(E) dE = V g(E) \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1}$$

If we excite the atoms by changing the temperature we move the atoms from the region marked dN in the figure to the matching region at $E_F + \varepsilon$. The atoms would gain energy 2ε . We could approximate the energy gained by all the molecules in the shaded group as

$$dE = 2\varepsilon dN$$

The number of particles in the group is approximately $dN = N_{ex}(E) d\varepsilon$ where $N_{ex}(E)$ is the number of particles to move as a function of energy and we can find N_{ex} by taking the number of particles with a particular energy E at $T = 0$ (dashed squarish curve) and subtracting off the number of particles with energy E when the temperature

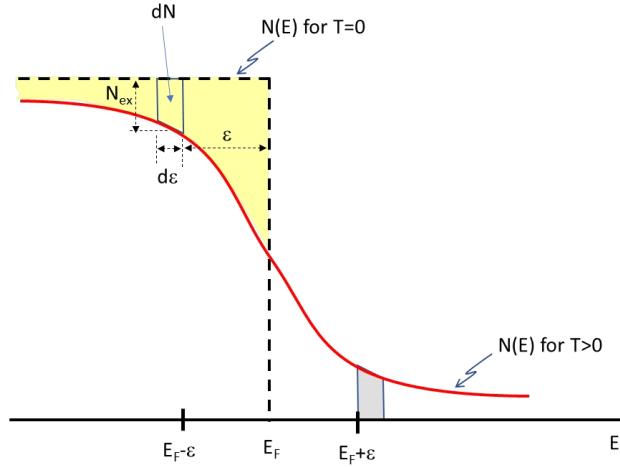
is T (red solid curve) right at $E = E_F - \varepsilon$

$$\begin{aligned} N_{ex} &= N(E_F - \varepsilon)_{T=0} - N(E_F - \varepsilon)_T \\ &= Vg(E_F - \varepsilon) 1 - Vg(E_F - \varepsilon) \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \\ &= Vg(E_F - \varepsilon) \left(1 - \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \right) \end{aligned}$$

To find the heat capacity we want to find dE and take a derivative with respect to T . So next let's find

$$\begin{aligned} dE &= 2\varepsilon dN \\ &= 2\varepsilon N_{ex}(\varepsilon) d\varepsilon \\ &= 2\varepsilon Vg(E_F - \varepsilon) \left(1 - \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \right) d\varepsilon \end{aligned}$$

and integrate it to find the energy of all the excited particles.



$$E_{ex} = \int_{E_F}^0 2\varepsilon Vg(E_F - \varepsilon) \left(1 - \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \right) d\varepsilon$$

This will be the entire yellow shaded area in the last figure. The limits of integration start at E_F and go to zero so we pick up the entire yellow area.

$$E_{ex} = 2\varepsilon V \int_{E_F}^0 g(E_F - \varepsilon) \left(1 - \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \right) d\varepsilon$$

We can approximate this by noting that $g(E_F - \varepsilon)$ varies much more slowly than f_{FD} so we could say that

$$g(E_F - \varepsilon) \approx g(E_F)$$

and take it out of the integral

$$E_{ex} = 2\varepsilon V g(E_F) \int_{E_F}^0 \left(1 - V \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \right) d\varepsilon$$

Now we want

$$C = \frac{dE_{ex}}{dT}$$

so then

$$C = \frac{d}{dT} 2Vg(E_F) \int_{E_F}^0 \varepsilon \left(1 - \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \right) d\varepsilon$$

and we can take the derivative inside the integral

$$\begin{aligned} C &= 2Vg(E_F) \int_{E_F}^0 \varepsilon \frac{d}{dT} \left(1 - \frac{1}{e^{\frac{(-\varepsilon)}{k_B T}} + 1} \right) d\varepsilon \\ &= 2Vg(E_F) \int_{E_F}^0 \varepsilon \left(\frac{-e^{\frac{(-\varepsilon)}{k_B T}} \frac{(-\varepsilon)}{k_B T^2}}{\left(e^{\frac{(-\varepsilon)}{k_B T}} + 1 \right)^2} \right) d\varepsilon \\ &= 2Vg(E_F) \frac{1}{k_B T^2} \int_{E_F}^0 \varepsilon \left(\frac{e^{\frac{(-\varepsilon)}{k_B T}}}{\left(e^{\frac{(-\varepsilon)}{k_B T}} + 1 \right)^2} \varepsilon \right) d\varepsilon \end{aligned}$$

which looks like a fun integral to do.

Let's do a u substitution with

$$\begin{aligned} u &= \frac{(-\varepsilon)}{k_B T} \\ \varepsilon &= -u k_B T \end{aligned}$$

then

$$\begin{aligned} du &= \frac{-1}{k_B T} d\varepsilon \\ d\varepsilon &= -k_B T du \\ C &= 2Vg(E_F) \frac{1}{k_B T^2} \int_{E_F}^0 \left(\frac{e^u}{(e^u + 1)^2} (-u k_B T)^2 \right) (-k_B T du) \\ &= 2Vg(E_F) \frac{(k_B T)^3}{k_B T^2} \int_{E_F}^0 \left(\frac{e^u}{(e^u + 1)^2} u^2 \right) du \end{aligned}$$

We can approximate again by noting that our integrand is non-zero only near E_F . Past E_F our yellow region has no area. So it won't hurt to integrate out further. So let's set the lower limit to ∞ . Then

$$C = 2Vg(E_F) \frac{(k_B T)^3}{k_B T^2} \int_{\infty}^0 \left(\frac{e^u}{(e^u + 1)^2} u^2 \right) du$$

Apparently a table gives

$$\int_{\infty}^0 \left(\frac{e^u}{(e^u + 1)^2} u^2 \right) du = \frac{\pi^2}{6}$$

so

$$C = 2Vg(E_F) \frac{(k_B T)^3}{k_B T^2} \frac{\pi^2}{6}$$

$$C = Vg(E_F) k_B^2 T \frac{\pi^2}{3}$$

We can borrow an earlier result for $g(E_F)$ for a gas

$$g(E_F) = 4\pi(2s+1) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E_F}$$

and we remember that for our 3He we can only have two spin states, $\pm 1/2$, so

$$g(E_F) = 4\pi(2) \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E_F}$$

We found earlier that

$$E_F = \frac{h^2 c^2}{2mc^2} \left(\frac{3N}{8V\pi} \right)^{\frac{2}{3}}$$

Solving for V gives

$$V = \frac{3N}{8\pi \left(\frac{2mE_F}{h^2} \right)^{\frac{3}{2}}}$$

so, putting in $g(E_F)$ and V our heat capacity becomes

$$\begin{aligned} C &= V \left(8\pi \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E_F} \right) k_B^2 T \frac{\pi^2}{3} \\ C &= \frac{3N}{8\pi \left(\frac{2mE_F}{h^2} \right)^{\frac{3}{2}}} \left(8\pi \frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E_F} \right) k_B^2 T \frac{\pi^2}{3} \\ C &= \frac{1}{\left(\frac{2mE_F}{h^2} \right)^{\frac{3}{2}}} \left(\frac{\sqrt{2}(m)^{\frac{3}{2}}}{h^3} \sqrt{E_F} \right) k_B^2 N T \pi^2 \\ C &= \frac{k_B^2 N T \pi^2}{2E_F} \end{aligned}$$

Note that we made many approximations, so this form won't work for all T . But for small T it does, well, okay.

We will return to heat capacities in the near future. But the idea is to show that we can get real physical results with our determination of $N(E)dE$. And with many particles we usually need to attack the problems this way.

26 Solids and Solid State Physics

Technically we started looking at solid state physics in our last lecture when we took on finding heat capacities. But liquid helium is not really a normal material and it's really not so solid. So in this lecture let's take our quantum model for atoms and try to explain the behavior of solids. This must work out for our quantum model to be correct. So this can be a further test of our quantum model for atoms. We will see that this works, but solids are complicated, so our analysis will be complicated too. A more complete picture of solid state physics is available in our solid state physics class.

Fundamental Concepts in the Lecture

- Ionic bonds create solids with specific crystal structures
- Covalent and metallic bonds form less geometrical bonds
- Quantized vibrations are called phonons
- Phonons solve our problems in forming heat capacities of solids

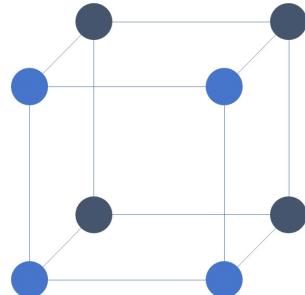
Structure and bonding of elemental solids

There are three types of elemental solids that come from our types of bonding. Ionic solids, covalent solids, and metallic solids.

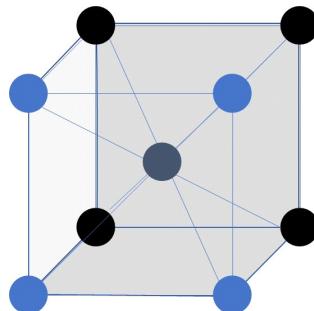
Ionic solids

Molecules that form ionic bonds are held together by Coulomb forces. These Coulomb forces can attract any other ion of the right charge. The ions can form into large structures. In such a large groups of ions the atoms are not really participating in molecular bonds (they don't have shared wavefunctions). They are just bound together electrically. The ions will try to form into the lowest energy state. This process builds

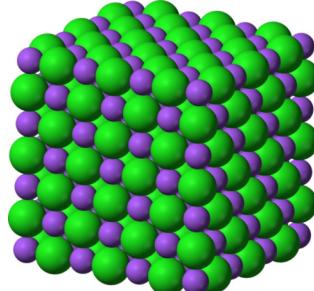
closely packed structures. Often they are in cubic patterns. One possibility is a simple cubic



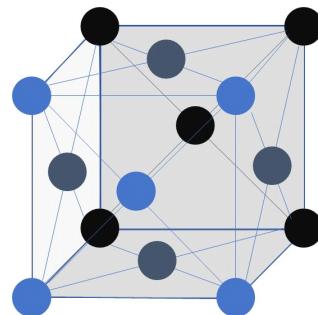
But many ionic solids pack in more ions. One way to do this is to put an ion in the center.



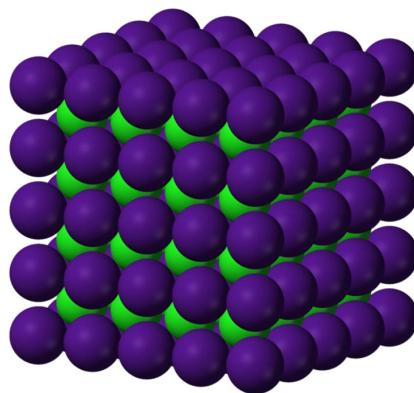
This is called body centered cubic structure or bcc. The solid looks like this



NaCl forms such a crystal. The small blue ions are the sodium, the larger green ions are the chlorine. There are other cubic structures that can form. If we add another ion to the face of each side of the cube we get a face-centered cubic structure or fcc.



CsCl forms such a solid structure



The large blue sphere's are the *Cs* and the *Cl* are the small green spheres.

Ionic solids are hard, and all the electrons are involved in the bonding and held in full shells, so the crystals are likely clear in the visible because visible photons can't move the electrons to cause absorption.

	NN Separation (nm)	Cohesive energy (kJ/mol)	n	Structure
NaCl	0.281	769	9	fcc
CsCl	0.356	657	10.5	bcc

We won't do a lot of calculations with solids, but let's try one. It is traditional to calculate the attractive Coulomb potential for a crystal. This is potential energy is just due the Coulomb force on one atom due to all the other atoms. The atoms are mostly neutral except for the ionic "shared" electrons. The sharing of electrons between two of the atoms creates separated charge. This separated charge is the reason the crystal sticks together. To calculate the potential energy we need to alternate positive and negative contributions. Each atom would have six neighboring atoms a distance r away.

And 12 next nearest neighbors, and so forth.

$$\begin{aligned} U_C &= \sum \frac{q_1 q_2}{4\pi\epsilon_0 r} = \frac{e^2}{4\pi\epsilon_0} \left(-6\frac{1}{r} + 12\frac{1}{r\sqrt{2}} - 8\frac{1}{r\sqrt{3}} + \dots \right) \\ &= \frac{e^2}{4\pi\epsilon_0 r} \left(-6 + \frac{12}{\sqrt{2}} - \frac{8}{\sqrt{3}} + \dots \right) \\ &= \frac{e^2}{4\pi\epsilon_0 r} (-\alpha) \end{aligned}$$

The factor α is purely geometric

$$\alpha = 6 - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} + \dots$$

and for the two cases

$$\alpha_{fcc} = 1.7476$$

$$\alpha_{bcc} = 1.7627$$

Of course there is also a repulsive force which is approximately due to the Pauli exclusion principle because we are dealing with electrons of spin $\pm 1/2$. Let's approximate this as

$$U_R = \frac{A}{R^n}$$

where the n is in the $8 - 10$ range. We won't derive this in this class.

Then the total potential energy would be

$$\begin{aligned} U &= U_C + U_R \\ &= \frac{e^2}{4\pi\epsilon_0 r} (-\alpha) + \frac{A}{r^n} \end{aligned}$$

Finding the minimum of the energy well can give a value for A

$$\begin{aligned} \frac{dU}{dr} &= \frac{1}{4\pi r^{n+2}\epsilon_0} (r^n \alpha e^2 - 4\pi A n r \epsilon_0) = 0 \\ r^n \alpha e^2 &= 4\pi A n r \epsilon_0 \\ \frac{r^n \alpha e^2}{4\pi n r \epsilon_0} &= A \\ \frac{r^{n-1} \alpha e^2}{4\pi n \epsilon_0} &= A \end{aligned}$$

The binding energy is

$$\begin{aligned}
 B &= -U(r_o) = -\frac{e^2}{4\pi\epsilon_0 r_o}(-\alpha) + \frac{\frac{r_o^{n-1}\alpha e^2}{4\pi n\epsilon_0}}{r^n} \\
 &= \frac{\alpha e^2}{4\pi\epsilon_0 r_o} - \frac{\frac{r_o^{n-1}\alpha e^2}{4\pi n\epsilon_0}}{r^n} \\
 &= \frac{\alpha e^2}{4\pi\epsilon_0 r_o} - \frac{r_o^{-1}\alpha e^2}{4\pi n\epsilon_0} \\
 &= \frac{\alpha e^2}{4\pi\epsilon_0 r_o} - \frac{\alpha e^2}{4\pi n\epsilon_0 r_o} \\
 &= \frac{\alpha e^2}{4\pi\epsilon_0 r_o} \left(1 - \frac{1}{n}\right)
 \end{aligned}$$

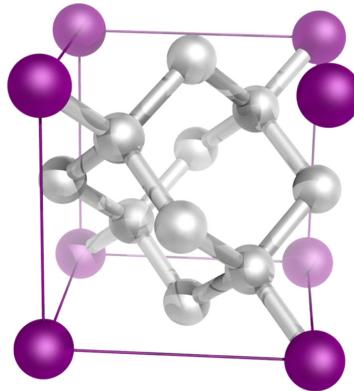
But we have an equal number of positive and negative ions in an ionic solid. So if we had a mole of negative ions we would have a mole of positive ions. That is a total of $2N_A$ ions per mole. Think if we had two ions bound together in the solid, then we would need the binding energy from the first ion interacting with all the other atoms. Same for the second. But in this we would include the interaction between the first and the second ion, and then the second and the first. But these are the same interaction, so we would be double counting. The cohesive energy is then

$$E_{coh} = \frac{1}{2}B(2N_A) = BN_A$$

We could calculate the cohesive energy for specific substances (which might happen in a homework problem).

Covalent solids

Covalent solids are not as orderly as ionic solids. This is because covalent bonds are directional, governed by the orbital structure and these solids are held together by actual bonds (not just Coulomb forces). Carbon forms this kind of bond with a tetrahedral shape, and that forms tetrahedral shaped solids.

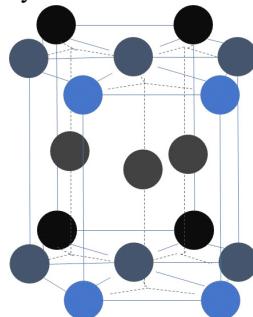


	NN Separation	Cohesive energy (kJ/mol)
C (diamond)	0.154	710

But of course, this is just one of many patterns possible governed by the orbital structure of atoms.

Metallic Bonds

Metallic bonds are different because they are not based on molecular bonds at all. The atoms of the metal share the valence electrons to form our “gas” of electrons. The metal is held together because the positive ions are attracted to the electron gas. Metals often have a close packed hexagonal crystal form.



Molecular Solids

We saw that for ionic solids we started with diatomic molecules that were polar due

to sharing of electrons. But we could use more complicated molecules that have separated charge due to their bonds to form solids. For example some molecules have a permanent dipole moment. Water is one of these. The shared electrons spend most of their time with the oxygen atom. The oxygen is then negatively charged and the hydrogen are left positively charged. The molecules are held together by Coulomb forces between the dipoles. These Coulomb forces are known as *van der Waals forces*.

With this as a basis, we will take on finding the heat capacities of these solids.

Heat Capacity of Solids

We talked about heat capacities of ^3He gas in ^4He superfluid. If you are at CERN you are very concerned about this because you want all of your He to be a superfluid, and the ^3He is not. But otherwise, this might not be a very common concern. But everyday phenomena from beach breezes (I wish this was an everyday phenomena at BYU-I) to using hot pads while cooking, to city climate warming all depend on the heat capacity of solid matter. Back in PH123 we studied molar heat capacities for gases.

$$C = \frac{\Delta E_{int}}{n\Delta T}$$

We can try starting with this to build a more complete model this for solids. And it is good to do this problem because it will lead us to a better quantum approach. So think of an atom in a crystal matrix. The atom doesn't have the ability to move much because it is bonded to the atoms around it. You could mentally envision the atom suspended in place by springs. It can vibrate on the springs. But it is restricted to a small region around its equilibrium location. We can think of this vibration as happening in any or all of the three directions x , y , and z , and in each of these directions we have two ways to have energy, kinetic and potential energy. So all told there are six ways to store energy, two for each of the three directions. This is 6 degrees of freedom.

And from PH123 we know we get $\frac{1}{2}k_B T$ of energy for each degree of freedom so our internal energy is

$$\begin{aligned} E_{int} &= 6 \times \frac{1}{2}k_B T \\ &= 3k_B T \end{aligned}$$

If we have a whole mole of atoms in this situation we would have a total energy of

$$E_{int} = N_A 3k_B T$$

and we might remember from PH123 that

$$Nk_B = nR$$

so

$$E_{int} = 3nRT$$

and our molar heat capacity would be

$$\begin{aligned} C &= \frac{\Delta E_{int}}{n\Delta T} \\ &= \frac{3nRT}{n\Delta T} \\ &= 3R \\ &= 24.944 \frac{\text{J}}{\text{K mol}} \end{aligned}$$

and we already see a problem. This is a constant! From what we did before at low temperatures we expect there must be some change in heat capacity with temperature. We saw this for gasses. But in this formulation there are no jumps to higher rotational or vibrational states. We just have this number. This approach is called the Dulong-Petit method after the researchers that came up with it. But it is not very good. Here are some experimental measurements at room temperature where it is not always terrible.

Substance	$T = 293 \text{ K}$	$T = 100 \text{ K}$
Aluminum	24.3	12.8
Copper	24.4	16.0
Iron	25.1	12.0
Gold	25.4	21.1
Lead	26.5	23.8

But at $T = 100 \text{ K}$ the numbers are much worse. There must be a temperature

dependence that we have missed.

In an attempt to fix this we could try what we did for ${}^3\text{He}$ in ${}^4\text{He}$. The approach should work for any fermions. And the in metals we know we have a “gas” of electrons. Perhaps the thermal properties of solids are solely due to this “gas” of electrons. We know the result from what we did before.

$$C = \frac{k_B^2 NT \pi^2}{2E_F}$$

We could write this as

$$C = \frac{\pi^2}{2} \frac{nRk_B T}{E_F}$$

or if we choose $n = 1$

$$C = \frac{\pi^2}{2} \frac{Rk_B T}{E_F}$$

Let's try this for, say, a mole of copper. We need E_F

$$E_F = \frac{(hc)^2}{2mc^2} \left(\frac{3N}{8V\pi} \right)^{\frac{2}{3}}$$

We will need N/V which we can get from the density of copper

$$\rho = \frac{M}{V} = 8960.0 \frac{\text{kg}}{\text{m}^3}$$

and each copper atom weights

$$12 \text{ u} = 1.9926 \times 10^{-26} \text{ kg}$$

The mass of the sample M divided by the mass of an atom should give the number of atoms

$$\frac{N}{V} = \frac{M}{Vm_{Cu}} = \frac{\rho}{m_{Cu}} = \frac{8960.0 \frac{\text{kg}}{\text{m}^3}}{1.9926 \times 10^{-26} \text{ kg}} = \frac{4.4966 \times 10^{29}}{\text{m}^3}$$

This is the number density of atoms. If each atom contributes one electron then it is also the number density of electrons. We get a Fermi energy of

$$E_F = \frac{(hc)^2}{2mc^2} \left(\frac{3}{8\pi} \frac{N}{V} \right)^{\frac{2}{3}}$$

$$E_F = \frac{(1239.8 \text{ eV nm})^2}{2(0.511 \text{ MeV})} \left(\frac{3}{8\pi} \frac{4.4966 \times 10^{29}}{\text{m}^3} \right)^{\frac{2}{3}} = 21.401 \text{ eV}$$

Then our heat capacity would be

$$C = \frac{\pi^2}{2} \frac{Rk_B T}{E_F}$$

$$C = \frac{\pi^2}{2} \frac{(8.314510 \text{ J mol}^{-1} \text{ K}^{-1})(8.617385 \times 10^{-5} \text{ eV K}^{-1})(100 \text{ K})}{21.401 \text{ eV}}$$

$$= 1.6522 \times 10^{-4} T \frac{\text{J}}{\text{K}^2 \text{ mol}}$$

This is temperature dependant, which is good. But if we calculate C at $T = 100K$

$$C = 1.6522 \times 10^{-2} \frac{\text{J}}{\text{K mol}}$$

Our value is no where near the experimental value. This function for C is actually not too bad near 0 K. But we have clearly missed something. Thermal conduction is not due solely to free electrons.

Einstein solved this problem. His approach was to look at the atomic oscillations as another wave function. That wave function could be quantized like any wave function. Or in other words the vibrations could be looked at as like standing waves that only allow certain frequencies and therefore have certain energies. Electromagnetic waves we now know have a quanta of a photon. Mechanical waves can also be quantized and have a quanta. That quanta is called a *phonon*. the phonon has energy

$$E = \hbar\omega$$

This is not too strange. We already found that in atoms and molecules the angular

momentum was quantized and that gave us space quantization. This is like an extended form for the motion of] not just the electrons, but the whole atoms. Because the atoms oscillate in the crystal lattice, we can model them as quantum oscillators.

We already studied the quantum oscillator and we found that it gave energy values of

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right)$$

Now we can look at this another way. Our oscillator energies are integer increments, or we have an increment of a phonon.

Einstein postulated that phonons obey Bose-Einstein statistics. So he chose

$$f = f_{BE} = \frac{1}{e^{\frac{E}{k_B T}} - 1}$$

and he reasoned that one mole of the solid would have N_A atoms, and each atom can oscillate along the three axes, so there could be $3N_A$ oscillators in a mole of the solid.

Then

$$g(E) = \frac{3N_A}{V}$$

so

$$N(E) = V \left(\frac{3N_A}{V} \right) \left(\frac{1}{e^{\frac{E}{k_B T}} - 1} \right)$$

but each phonon has energy $\hbar\omega$ so

$$N(\hbar\omega) = 3N_A \frac{1}{e^{\frac{\hbar\omega}{k_B T}} - 1}$$

The total number of oscillators would seem to be the integral of this

$$N = \int_0^\infty (3N_A) \left(\frac{1}{e^{\frac{E}{k_B T}} - 1} \right) dE$$

But we don't have continuous energy! We need to evaluate this *just* at $E = \hbar\omega$. So we have only

$$N = (3N_A) \left(\frac{1}{e^{\frac{\hbar\omega}{k_B T}} - 1} \right)$$

for the lowest state. The total energy would be the number of phonons times the phonon energy

$$E_{total} = EN = \hbar\omega N = \frac{3N_A \hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1}$$

which is what we needed to find C . We can now take a derivative

$$C = \frac{dE}{dT} = 3N_A \hbar\omega \frac{e^{\frac{\hbar\omega}{k_B T}} \frac{\hbar\omega}{k_B T^2}}{\left(e^{\frac{\hbar\omega}{k_B T}} - 1 \right)^2}$$

Einstein defined a term $T_E = \frac{\hbar\omega}{k_B}$ called the *Einstein temperature* so we could write our

heat capacity as

$$C = \frac{dE}{dT} = 3N_A\hbar\omega \frac{\hbar\omega}{k_B T^2} \frac{e^{\frac{\hbar\omega}{k_B T}}}{\left(e^{\frac{\hbar\omega}{k_B T}} - 1\right)^2}$$

$$C = \frac{dE}{dT} = 3N_A\hbar\omega \frac{T_E}{T^2} \frac{e^{\frac{\hbar\omega}{k_B T}}}{\left(e^{\frac{\hbar\omega}{k_B T}} - 1\right)^2}$$

Multiplying by k_B/k_B

$$C = 3N_A \frac{k_B}{k_B} \hbar\omega \left(\frac{T_E}{T^2}\right) \frac{e^{\frac{T_E}{T}}}{\left(e^{\frac{T_E}{T}} - 1\right)^2}$$

or

$$C = 3k_B N_A \left(\frac{T_E^2}{T^2}\right) \frac{e^{\frac{T_E}{T}}}{\left(e^{\frac{T_E}{T}} - 1\right)^2}$$

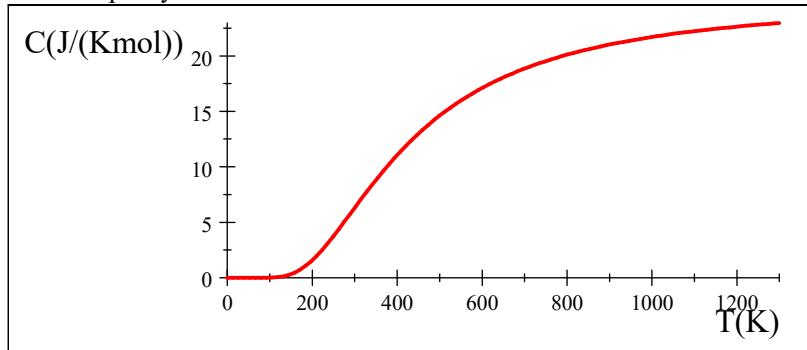
to put this in terms of R

$$C = 3(1)R \left(\frac{T_E}{T}\right)^2 \frac{e^{\frac{T_E}{T}}}{\left(e^{\frac{T_E}{T}} - 1\right)^2}$$

where the (1) is a reminder that we have one mole of atoms in our sample. Einstein tried this on diamond. He found from data that $T_E = 1300$ K so then

$$\begin{aligned} \omega &= \frac{k_B T}{\hbar} = \left(\frac{(8.617385 \times 10^{-5} \text{ eV K}^{-1})(1300 \text{ K})}{6.5821220 \times 10^{-16} \text{ eV s}} \right) \\ &= 1.702 \times 10^{14} \text{ Hz} \end{aligned}$$

and our heat capacity looks like this



which is fantastic improvement over the Dulong-Petit model. Of course the assumption that all phonons have the same energy is too restrictive and the curve is wrong near $T = 0$.

Debye relaxed this restriction on phonon energy by using the same approach we used

for a “photon gas” in a cavity. For low temperatures Debye got

$$C = \frac{12\pi^4}{5} R \left(\frac{T}{T_D} \right)^3$$

where T_D is a material dependent parameter called the *Debye temperature*.

This does better but is still not enough. At low temperatures it turns out that both the electron heat capacity we calculated before and the phonon heat capacity we just found are important. So we expect the actual heat capacity to be of the form

$$C = aT + bT^3$$

So we can see that we have used our technique of finding the number of particles with a particular energy for both the electrons and the atoms separately in this case. Clearly this is a powerful technique, but a complicated one.

But we have strayed from our everyday objects again. The Einstein method is really quite good for normal earth surface temperatures.

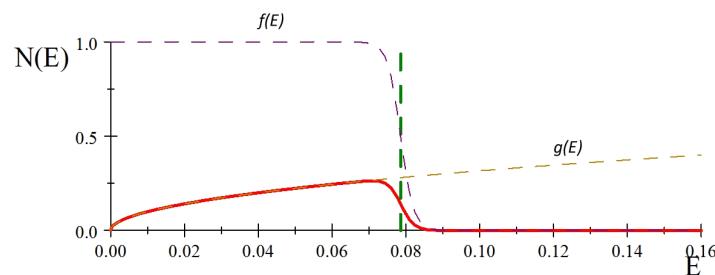
Quantum Conduction

We are alternating between a theory of conduction and a theory of thermal heat capacity. There is a reason these are tied together. Both electrical conduction and thermal conduction in metals are accomplished at least in part by the semi-free electron gas.

We have already found that for our electron gas we have

$$N(E) dE = V \left(\frac{8\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \sqrt{E} \right) \left(\frac{1}{e^{\frac{1}{k_B T}(E-E_F)} + 1} \right) dE$$

and we plotted this to find the red curve shown below.



The dashed curves are the Fermi-Dirac distribution

$$f(E) = \left(\frac{1}{e^{\frac{1}{k_B T}(E-E_F)} + 1} \right)$$

and the multiplicity function

$$g(E) = \left(\frac{8\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \sqrt{E} \right)$$

and we found the Fermi energy for the special case of $T = 0$

$$E_F = \frac{h^2 c^2}{2mc^2} \left(\frac{3N}{8V\pi} \right)^{\frac{2}{3}}$$

but we didn't find the Fermi energy for higher temperatures. We would have to integrate

$$N = V \left(\frac{8\pi\sqrt{2}m^{\frac{3}{2}}}{h^3} \right) \int_0^\infty \left(\frac{\sqrt{E}}{e^{\frac{1}{k_B T}(E-E_F)} + 1} \right) dE$$

but this is done numerically. We won't do it ourselves in this class. But the answer is an approximate (it's a numerical solution)

$$E_F(T) = E_F(0) \left(1 - \frac{\pi^2}{12} \left(\frac{k_B T}{E_F(0)} \right)^2 \right)$$

So this is where we are with our quantum theory of electrical conduction. We ended with a numeric solution.

Before we go on, let's remember a few things from PH220 and use them to calculate a classical estimate for the conductivity. The current density

$$\mathbf{j} = \sigma \mathbf{E}$$

and for the electrons in the metal we have a force

$$\mathbf{F} = -e \mathbf{E}$$

and the electrons should have an acceleration of

$$\mathbf{a} = \frac{-e \mathbf{E}}{m}$$

but we know that the electrons don't accelerate to the end of the piece of metal. In PH220 we learned that there is a constant drift velocity v_d

$$\mathbf{v}_d = \frac{-e \mathbf{E}}{m} \tau$$

where τ is the average time between collisions between the free electrons and the atoms. And that is the key, the electrons do accelerate in between collisions with the atoms. But when the electrons do collide with the atoms of the conductor, they are slowed. There is a net force due to the collisions much like a drag force, and when the Coulomb force and the drag force are equal we have a kind of terminal velocity for the electrons that is our v_d .

We can write the current density is

$$\begin{aligned}\mathbf{j} &= -ne\mathbf{v}_d \\ &= n \frac{e^2 \tau \mathbf{E}}{m}\end{aligned}$$

where n is the electron number density, so the conductivity σ is

$$\sigma = n \frac{e^2 \tau}{m}$$

where the mean time between collisions is given by the

$$\tau = \frac{\ell}{v_{ave}}$$

and where ℓ is the mean free path and v_{ave} is the average speed of the electrons. If we treat this classically we could take our thermally based average velocity from thermodynamics.

$$v_{ave} = \sqrt{\frac{3k_B T}{m}}$$

Let's apply this to copper and see what $\sigma_{classical}$ would be. For $T = 293\text{ K}$

$$\begin{aligned}v_{ave} &= \sqrt{\frac{3(1.3806568 \times 10^{-23} \text{ J K}^{-1})(293\text{ K})}{9.1093897 \times 10^{-31} \text{ kg}}} \\ &= 1.1542 \times 10^5 \frac{\text{m}}{\text{s}}\end{aligned}$$

then if we take ℓ to be the spacing between atoms in a copper lattice

$$\ell = 0.256\text{ nm}$$

so

$$\tau = \frac{\ell}{v_d} = \frac{0.256\text{ nm}}{1.1542 \times 10^5 \frac{\text{m}}{\text{s}}} = 2.218 \times 10^{-15}\text{ s}$$

We need the number density n

$$n = \frac{\rho N_A}{M} = \frac{\left(8960 \frac{\text{kg}}{\text{m}^3}\right) \left(6.02 \times 10^{23} \frac{1}{\text{mol}}\right)}{0.0635 \frac{\text{kg}}{\text{mol}}} = 8.4944 \times 10^{28} \frac{1}{\text{m}^3}$$

and finally for the conductivity we get

$$\begin{aligned}\sigma &= n \frac{e^2 \tau}{m} \\ &= \left(8.4944 \times 10^{28} \frac{1}{\text{m}^3}\right) \frac{(1.60217733 \times 10^{-19} \text{ C})^2 (2.218 \times 10^{-15} \text{ s})}{9.1093897 \times 10^{-31} \text{ kg}} \\ &= 5.3092 \times 10^6 \frac{1}{\Omega \text{ m}}\end{aligned}$$

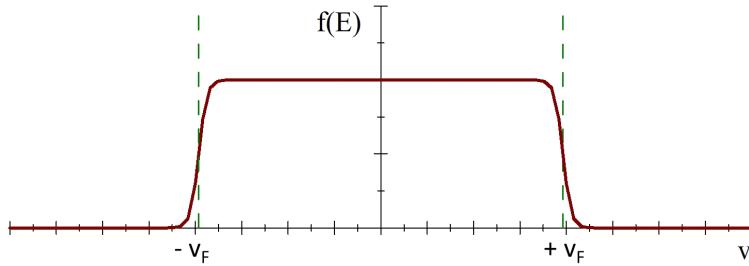
The measured value is

$$\sigma_{measured} = 5.96 \times 10^7 \frac{1}{\Omega \text{ m}}$$

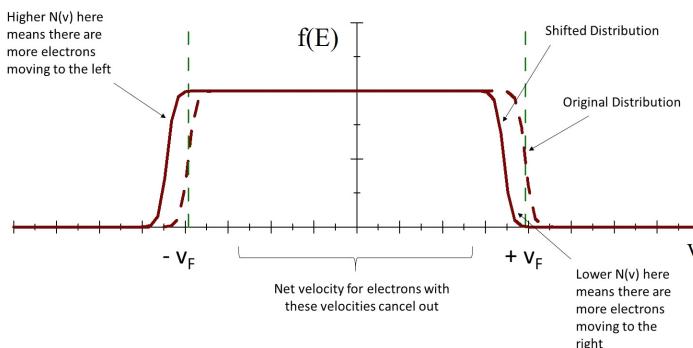
We should expect that the classical value would not be exactly right, but we are more than an order of magnitude off. This seems very bad. We need a quantum theory of conduction.

Fermi-Dirac velocity distribution and electrical conduction

For our electrons, the velocity is related to kinetic energy, so we should be able to express our Fermi-Dirac probability distribution in terms of velocities. Here is what it would look like.



Notice that the distribution includes negative velocities. That is as it should be. Our electrons can go either direction down a wire. Also notice that the distribution is 1 in the center and falls to zero at a velocity that represents the speed at the Fermi energy. We could call this the Fermi velocity. Now suppose that we apply an external electric field in the positive direction. This adds in potential energy, and we obtain a drift velocity. This shifts the distribution by v_d .



On the left side of the distribution we now have a larger number of electrons with velocities going the opposite direction of the field. On the right side we see we have fewer electrons going in the field direction. Only electrons that are in the region right around the Fermi velocity are affected. Thus we could say that only electrons with

$$v \approx v_F$$

are free to move. Then we could write

$$\sigma = \frac{ne^2\tau}{m} = \frac{ne^2\ell}{mv_d} \approx \frac{ne^2\ell}{mv_F}$$

which is a (very) rough approximation (v is really not v_F , in fact, it must be larger than v_F on average, but it is not too different...). It is not exactly right, but it would be more correct than what we have done before. Sadly, this approximation is not enough to fix our order of magnitude difference in the measured and predicted σ (in fact, it makes things worse!). But we are not done, the electrons are waves, so we need to consider wave behavior in calculating our mean free path.

Let's look at this backwards. Let's find out what the mean free path would need to be in order for our model to match the measurement. We could solve for ℓ

$$\begin{aligned}\sigma &= \frac{ne^2\ell}{mv_F} \\ \frac{\sigma mv_F}{ne^2} &= \ell\end{aligned}$$

We will need the Fermi velocity, and let's keep using copper so we know that

$$E_F = 7.05 \text{ eV}$$

then

$$\begin{aligned}v_F &= \left(\frac{2E_F}{m} \right)^{\frac{1}{2}} = \left(\frac{2(7.05 \text{ eV})}{9.1093897 \times 10^{-31} \text{ kg}} \right)^{\frac{1}{2}} \\ &= 1.5748 \times 10^6 \frac{\text{m}}{\text{s}}\end{aligned}$$

so that

$$\begin{aligned}\ell &= \frac{\sigma mv_F}{ne^2} \\ &= \frac{(5.96 \times 10^7 \frac{1}{\Omega \text{m}})(9.1093897 \times 10^{-31} \text{ kg})(1.5748 \times 10^6 \frac{\text{m}}{\text{s}})}{(8.4944 \times 10^{28} \frac{1}{\text{m}^3})(1.60217733 \times 10^{-19} \text{ C})^2} \\ &= 3.9211 \times 10^{-8} \text{ m} \\ &= 39.2 \text{ nm}\end{aligned}$$

Copper atoms in a solid piece of copper are about 0.256 nm apart. And we can see that the mean free path that works is 153.13 bigger than the atomic spacing! It is easy to imagine an electron missing an atom or two, but 153 seems ridiculous. What is happening?

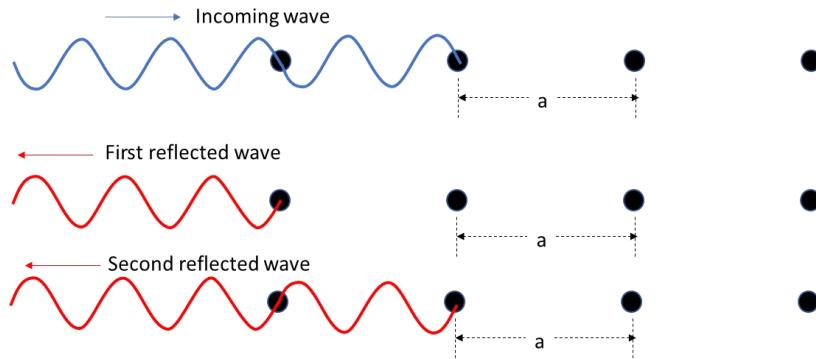
The key is to consider that the electrons are waves. Like light photons can go through a crystal lattice and ignore the atoms, the extended wave functions of the electrons in the metal can mostly wave around the lattice atoms of the copper. If the lattice were perfect, the electrons would travel very long distances (from the atomic perspective) before being affected by an atom. Disturbances in the lattice shorten these distances. The disturbances come from two different sources. One is impurities in the material, and the other is thermal motion of the lattice atoms. The first just breaks up the lattice causing scattering of the electron waves. The second moves the lattice atoms randomly

also upsetting the electrons waves.

These lattice vibrations are just the same vibrations we looked at when we were working to find a quantum theory for heat capacities. we know the lattice vibrations will be quantized. and we call these quantized vibrations phonons.

We can use a one dimensional model to get some idea of how this works. Of course it can only be an approximation, but let's save the complete three dimensional problem for a Solid State Physics Class.

In our one dimensional problem we have incoming electron waves. These electrons can reflect (scatter) off of lattice atoms.



In the figure we have a reflection from the first atom, and a second reflection from the second atom. Then the wave going to the left is a combination of two waves, and we can have constructive and destructive interference! In the case shown, the atom spacing, a , is just right to cause constructive interference. The path difference would be $2a$, so we would expect constructive interference when

$$2a = n\lambda \quad n = 1, 2, 3, \dots$$

Clearly this can only happen for a wavelength that just matches. The wave number

$$k = \frac{2\pi}{\lambda}$$

would then be

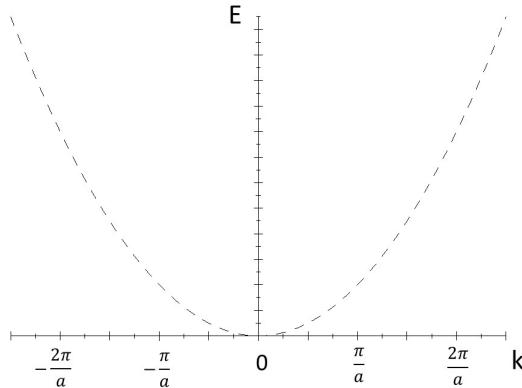
$$k = \frac{2\pi}{\frac{2a}{n}} = \frac{n\pi}{a}$$

and for an electron wave

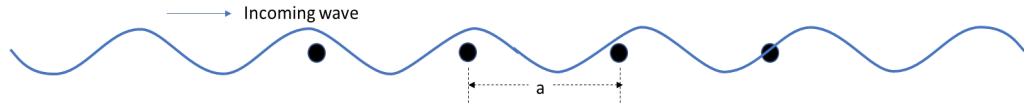
$$E = \frac{p^2}{2m} = \frac{h^2 k^2}{2m}$$

If it weren't for the reflection off of the lattice atoms, there would be no restriction on k and the electrons could have any energy. Once again it is the standing waves that make

a difference. Notice that our energy depends on k^2 . This gives a potential well that is a parabola in k



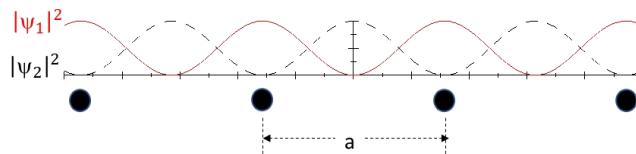
But what happens to wavelengths that don't fit our criteria for constructive interference?



They act as free particles missing all the atoms. But with the ones that do reflect, we are removing energy from the wave that won't travel along the conductor. Instead the energy at these wavelengths will be stored in the standing waves. And because k is quantized, so will be E for these wavelengths.

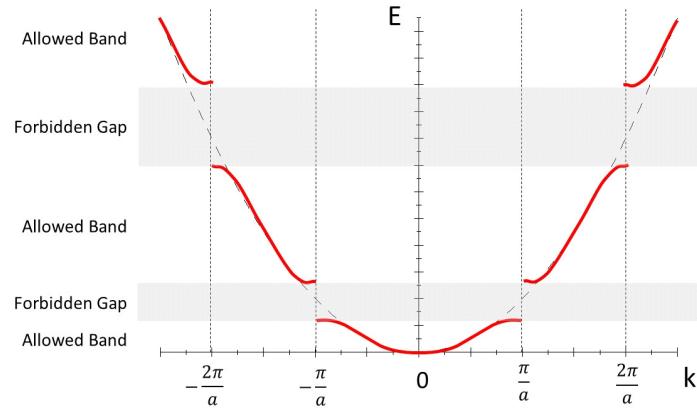
$$E = \frac{\hbar^2}{2m} \left(\frac{n\pi}{a} \right)^2$$

Since our idea of quantization came from constructive and destructive interference, this shouldn't be too much of a surprise. But this means that there are some electron wavelengths that act as free particles and some that don't. For the particles that are reflected, we could still have a phase change on reflection or not, so we have two possibilities ψ_1 and ψ_2 that could have different phases. We could draw the probability density functions for these two cases.



The ψ_1 case has a larger probability of detection right around the lattice ions. So it will feel the Coulomb force of the ions more strongly, and will therefore have a lower energy. The ψ_2 case has a higher probability of being detected in between the ions (dashed line), and therefore will have a higher energy. and the electrons with

wavelengths far from $\lambda = 2a/n$ will have an energy in between these two. We can draw what would happen as a graph with the electrons energies that can travel through the lattice unhindered in red. Note there is the sort of curling of the allowed state that we would expect from Fermi-Dirac statistics.

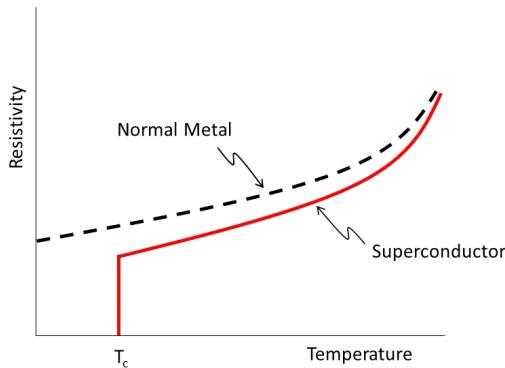


So we get some electrons traveling with some energies but some energies are not allowed because of the interactions of the electron waves with the atoms. What we learned in introductory electrodynamics (PH220) was true, but not the whole picture!

But we will leave the more exact calculation in three dimensions for a future solid state class. Yet, our quantum electron gas model of conduction has taken us pretty far, and even the weird energy gaps in conduction electrons are experimentally observed. This is a great success for quantum mechanics. But before we leave conduction we should consider superconductors.

Superconductivity

It was a surprise when superconductivity was first discovered. The situation is as shown in the next figure.

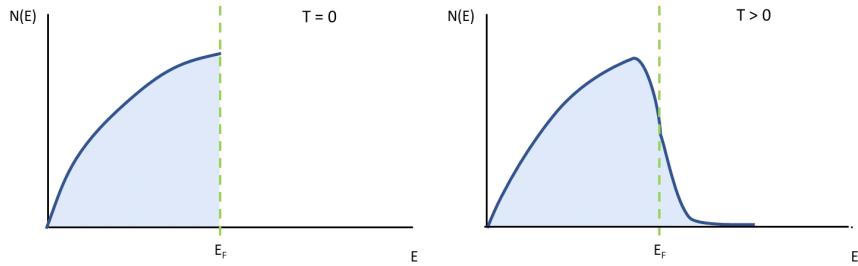


We expect the resistance of materials to lower with temperature. With lower temperature the lattice vibrations are less, so the interruption of the electron waves is less. But for some materials the resistance goes to zero below a material dependent temperature T_C . It doesn't drop to near zero. It goes to zero. There is no resistance. There is no heating of the superconductor because of the current flowing through the superconductor because there is no resistance at all. Normal metals have a power loss that goes like I^2R . But for superconductors $R = 0$. No heating. How does this work?

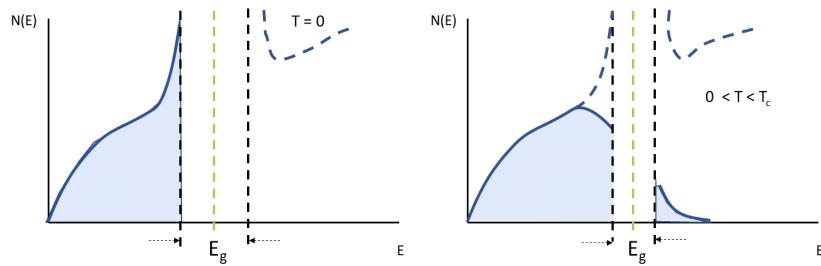
Well a quick look at materials that are superconducting tells us that we need more than just our electron gas model. Copper, for example, is not superconducting at any temperature, nor is gold nor silver. So the best conductors at normal temperatures are never superconducting at low temperatures.

In fact, the model for superconducting is that we don't want the electrons to bypass the lattice atoms. Instead we want the electrons to cause a disturbance in the lattice sort of like a boat makes a wake in water. The lattice atoms begin to vibrate, and these vibrations propagate through the material and interact with another electron. The two electrons can become coupled through the lattice vibrations (a little like two people jumping on a trampoline can enhance each other's motion through the material of the trampoline). These two electrons now move in tandem. And because they move in tandem they will have a spin that is integer! They are no longer restricted by Pauli exclusion, and therefore are not subject to Fermi-Dirac statistics. Let's use our population diagrams to show what would happen.

In a normal metal, the electrons do obey Fermi-Dirac statistics. We get a population graph like the left hand side of the following figure for $T = 0$



and when $T > 0$ we get the right hand graph. We see that we have some electrons with energy above E_F with non zero probability. But for a superconductor the electrons form pairs and this changes our graphs. At $T = 0$ all the electrons form pairs and the situation is as shown in the left hand side of the next figure.



So the shape of the curve is no longer the same. Notice that the electron pairs have lowered the energy. This is like a filled shell having a lower energy in an atom. So we see the electrons all have energy lower than the Fermi energy. When the temperature is above T_c no pairs form. But when the temperature goes below T_c , the electron pairs start to form and as the electrons form pairs they drop down in energy. Then the energy gap around E_F starts to form. The gap energy is roughly given by

$$E_g = 3.53k_B T$$

so it is a function of temperature. As the temperature lowers, more pairs form, and the gap widens. Each pair makes two fermions into a pair with spin 1 which changes their wave functions in such a way that they miss the atoms. And this seems to be superconductivity.

Much research has been and continues to be done to find semiconductors with higher T_c . Ideally we would find something with T_c at or higher than liquid nitrogen temperatures. But so far the records is about 130 K.

27 Examples of Solid State Physics, and Introduction to the Nucleus

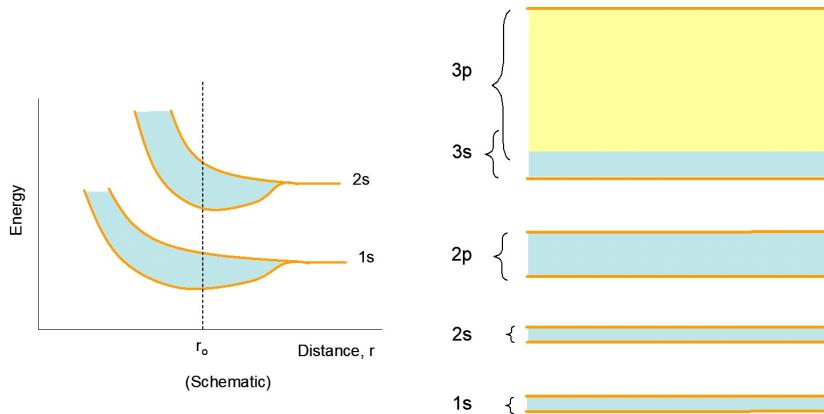
11.7-12.2

Fundamental Concepts in the Lecture

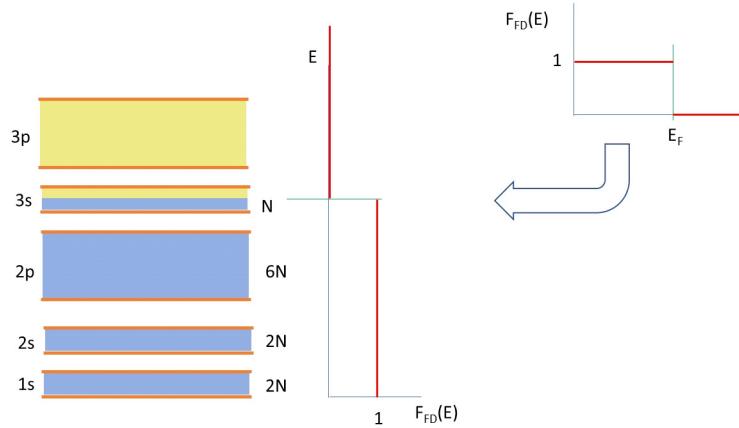
- Band Theory in solids and semiconductors
- Diamagnetism and paramagnetism
- Structure and Properties of the Nucleus
- Binding Energy and Nuclear Forces

Semiconductors

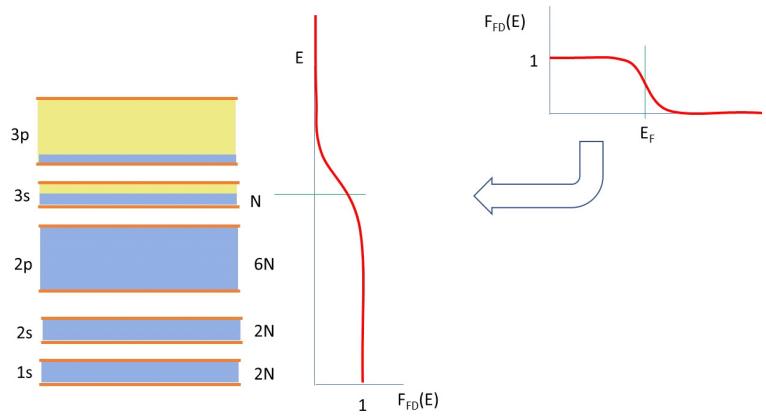
We have dealt with conductors. Let's look at what makes something an insulator. To do this, let's use our band model and of course Fermi-Dirac statistics. Our band model tells us that as we bring atoms together their electrical orbits split into many states due to the Pauli exclusion principle into many molecular orbitals, and then into solid energy states.



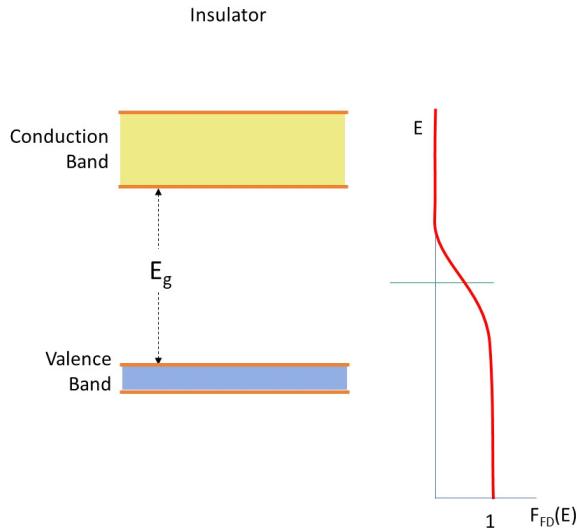
If we have a band that isn't full, then the electrons can take on energy and that could be kinetic energy. This is a conductor. Let's add to our energy graph a rotated graph of the Fermi-Dirac distribution function. In the next graph, drawn for $T = 0$, we can see that the energy states that are full (blue) are the ones below the Fermi energy. The ones that are empty are those above the Fermi energy.



If we allow the temperature to be above zero ($T > 0$) then the extra thermal energy will allow some electrons to move up into higher shells. The Fermi-Dirac distribution shows this as well.



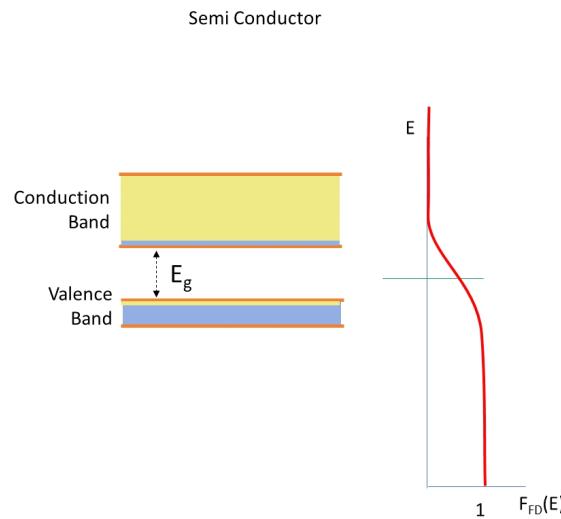
But we could have the case where the gap energy E_g is large compared to the thermal energy (on the order of $k_B T$). Notice that in this case the Fermi-Dirac distribution has gone to zero before we reach an allowed state in the upper band. So no electrons are likely to be in the upper state beyond the gap energy.



Let's call the upper band in such a situation the *conduction band* and the lower, filled band the *valence band*. The situation in our last figure is an insulator. It would take a lot of energy to move an electron from the valence band to the conduction band. It is possible, but normal thermal energy in a day-to-day environment won't do it. So this material would be a good choice to coat electrical wires. Note that given sufficient

energy, all materials are conductors. Which is why we must be careful around high voltage electrical systems .

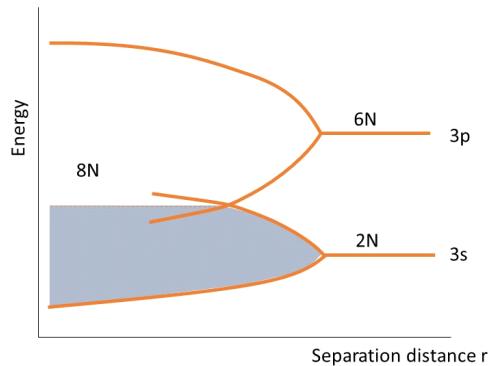
But what if the situation is the same, the valence band is full, the conduction band is empty, but the gap energy is not so big?



In this case the gap energy might be on the order of $k_B T$ and at normal temperatures the Fermi-Dirac distribution tells us that some electrons will be able to use the thermal energy to move from the valence band to the conduction band. And once there they are free to take on kinetic energy (subject to all the things we have already discussed about conduction).

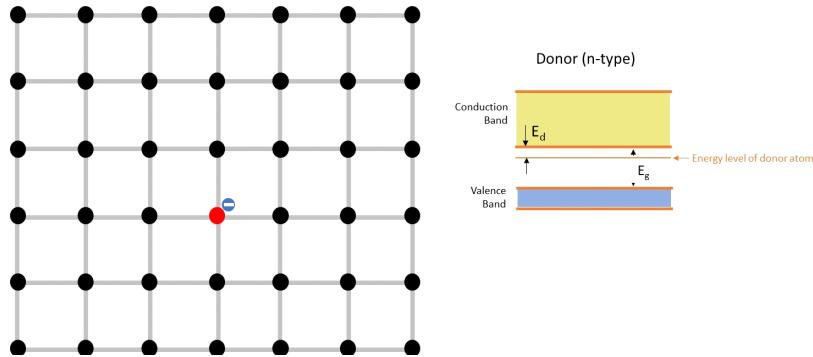
Many materials naturally do this. And we have generated more by building crystal lattices with carefully placed impurities to modify the gap separation.

Note that band theory is an extension of our model beyond the atomic model. From the atomic structure of magnesium $1s^2 2s^2 2p^6 3s^2$ would not be a conductor. It's outer shell is full. But band theory tells us that if we push many magnesium atoms together their states will split and form bands. And a good numerical calculation will show that the $3p$ and $3s$ state's bands overlap. That allows the $3s$ electrons plenty of energy states to use to take on kinetic energy.



P-type and n-type semiconductors

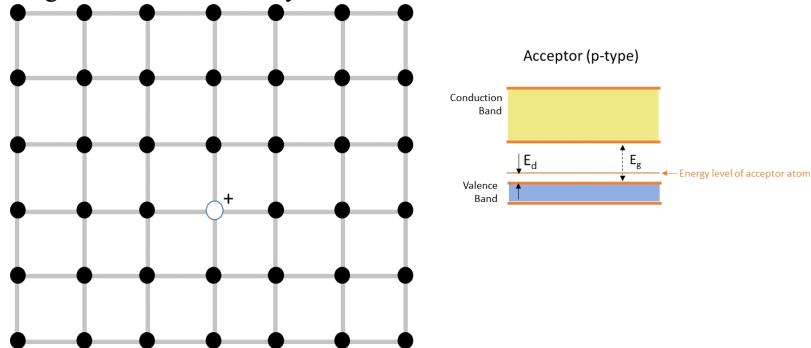
Although we now understand the basics of semiconductors, there are some important details to understanding how electronic systems work that we need yet to see. So far we have only considered pure semiconductors. But suppose we put in on atom of the wrong kind on purpose. Say we have silicon which has four electrons available for sharing and each *Si* atom is sharing with four neighbors. But suppose we put in an arsenic atom that has five valence electrons. This extra electron isn't participating in the bonding, so it is relatively easy to break free.



Since there is an extra negative charge in this type of semiconductor let's call this an *n*-type. The material is electrically neutral—all the electrons are matched with protons. But one electron is easy to remove due to the impurity. And we say that the impurity is a *donor* because it can allow an electron to be broken free.

Of course we could go the other way. We could replace one of the Si atoms with

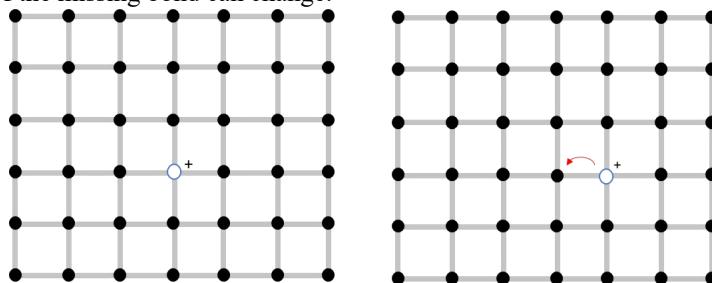
something like Gallium that only has three valence electrons.



All three electrons would participate in lattice bonds. But there is one bond missing from the structure. The structure would like to have another electron forming that final bond. That would likely be a lower energy state if the final bond were filled. So it is a little like the structure wants another electron, or a little like it has extra positive charge in the lattice because of the lack of a fourth valence electron. Let's call this type of semiconductor a *p*-type. We say the impurity is an *acceptor* because it would like an extra electron in its energy levels.

Both situations where impurities are added to semiconductors are called *doped*. But in thinking about this remember that we put neutral atoms into the lattice, so there is no net charge. Just *n*-type semiconductors have an electron that is more free because it doesn't participate in a bond and *p*-type semiconductors have one place in the lattice where there is a missing bond. It is the energy of a filled bond state that will make the difference, not an overall net charge.

A detail that we will need is that when the *p*-type semiconductors are formed, the location of the missing bond can change.



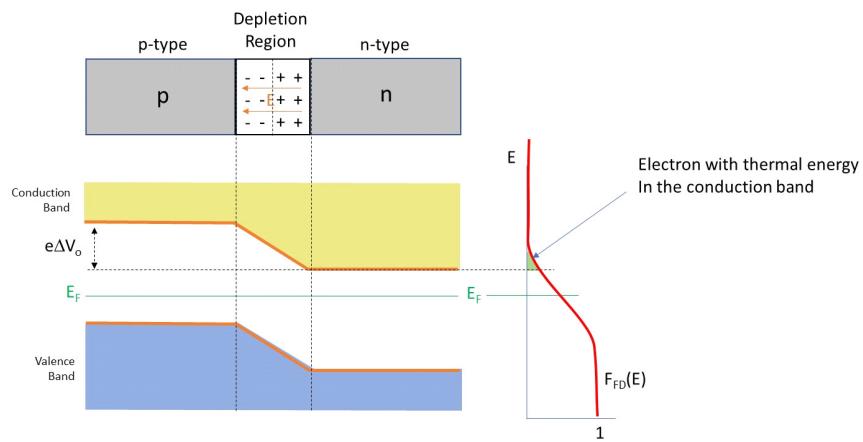
An electron can move over from a neighboring atom and provide the missing bond for the acceptor atom. In effect, the "hole" in the bond structure would move when this happens. And that is just what we say, we say that there is a new kind of charge carrier, a *hole* that effectively has a positive charge because it is formed by the lattice structure

wanting an electron so it can reach a lower energy state.

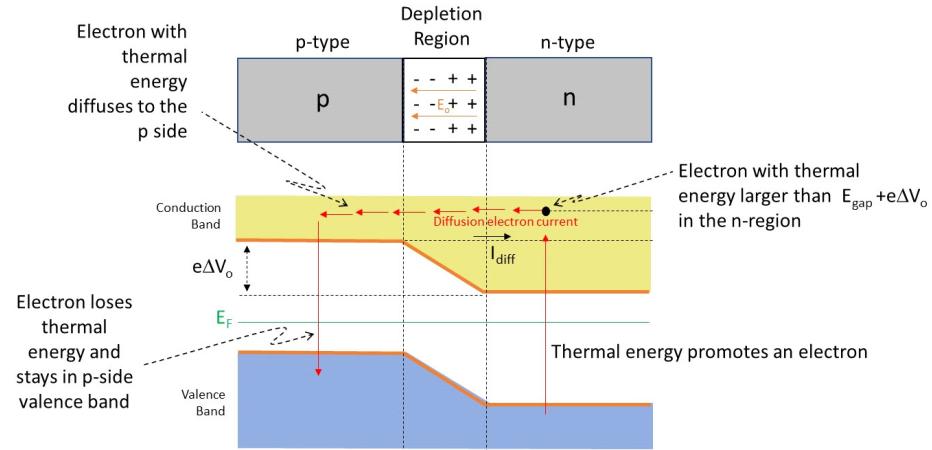
These different types of semiconductors become useful when we attach a *p*-type semiconductor to an *n*-type semiconductor.



There is a potential energy just because in the *p*-type semiconductor the lattice would like another bond at the acceptor so the lattice can have a lower energy.

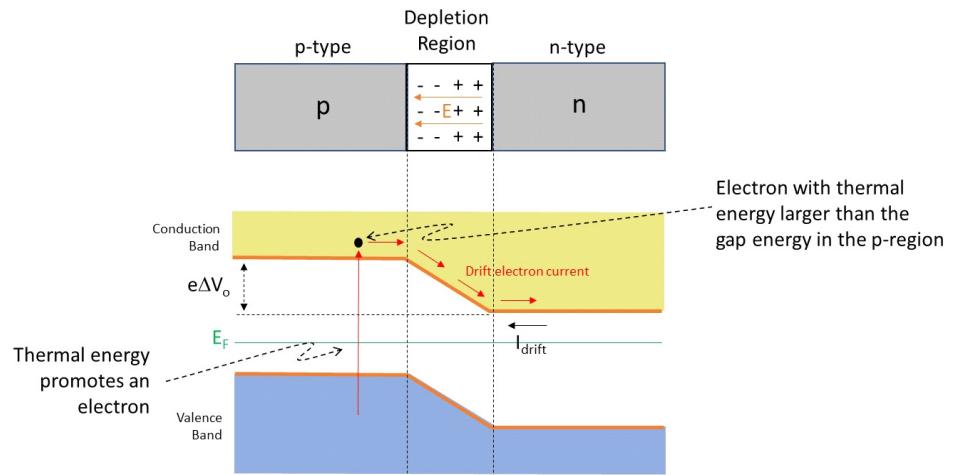


On the *n*-type side, we have electrons that are relatively free to move, so some will move from the *n* side to the *p* side to fill the vacant bonds. There has to be enough of a tail in the Fermi-Dirac distribution to make this happen or the electrons won't be able to take on kinetic energy to make the move. But if there is enough of a tail in the distribution (that is, if there is enough available kinetic energy) a few donor electrons in the *n*-type semiconductor move to the *p*-type semiconductor filling the "holes" in the acceptor states. This makes the edge part of the *p*-type region negatively charged (remember both regions are neutral to begin with), and leaves the edge region of the *n*-type region positively charged because it lost electrons. This sets up a separation of charge and that means we will have an electric field. This electric field will eventually stop the flow of electrons from the *n*-type region to the *p*-type region. This region in the middle with an electric field is called the *depletion zone* because it is, in a sense, out of charge carriers and no more will enter the region without some effort.



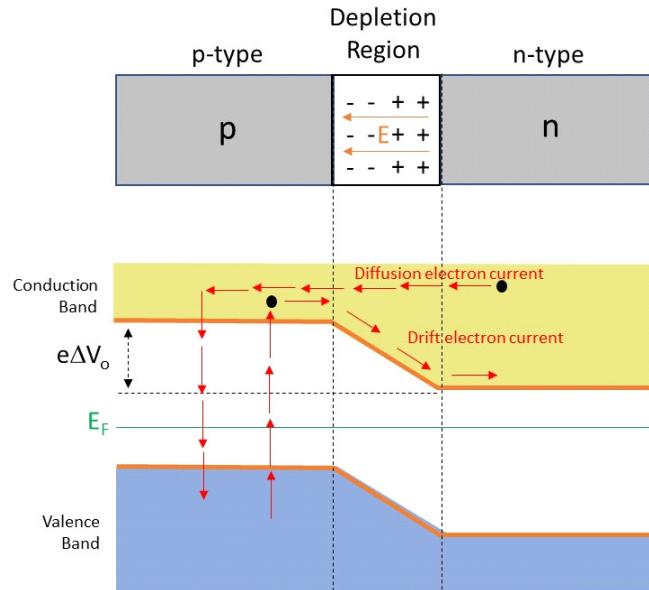
The weak current formed by electrons moving to the *p* side is called the *diffusion current* and in the usual confusing way we label currents, the electrons go from the *n* side to the *p* side, but the current goes the other way because we always label currents as though positive charges moved the opposite direction the negative charges actually went.

Of course we have thermal energy on the *p* side as well and if there is enough of a tail in the Fermi-Dirac distribution on the *p* side, then some electrons will gain enough energy to be freed, and the electric field that the depletion current formed will accelerate these newly freed electrons over to the *n* side.

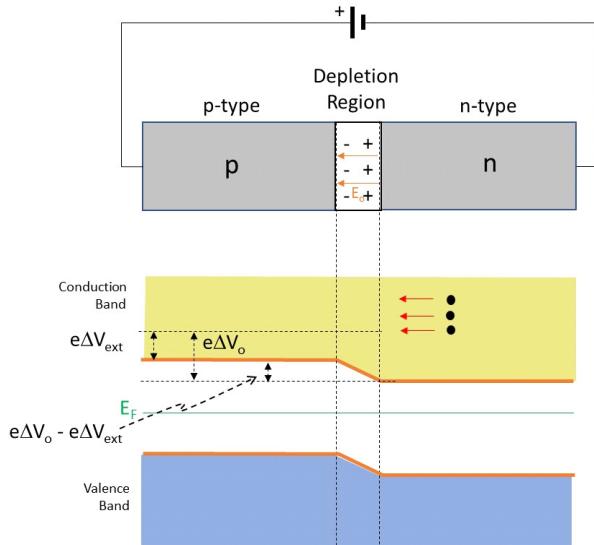


This is called the *drift electron current* but also in our normal strange way of labeling currents we would say there is a current I_{drift} going the opposite direction of the drift electron current. The net effect of these two currents is zero because in steady state, the

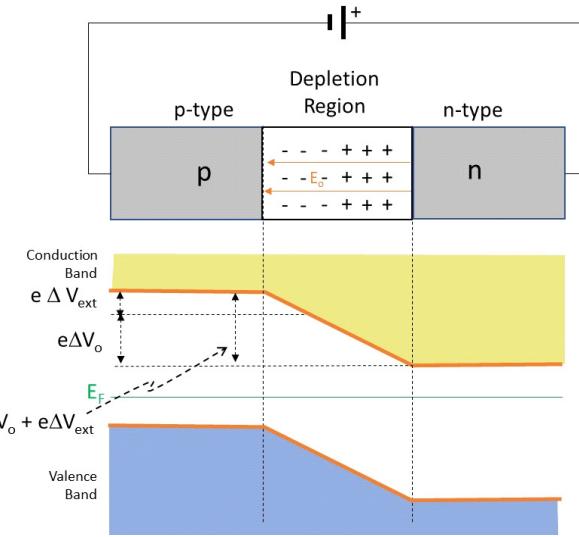
drift and diffusion currents will be the same, but in opposite directions.



We could cause an imbalance in the net current so that charge would flow through this *pn* junction. To do this, let's decrease the energy difference in the conduction band on the *p* side. We can do this by connecting a battery with the positive terminal on our *p* side and the negative terminal on the *n* side. The positive connection to the battery can remove some of the electrons from the *p* side of the depletion zone and the negative connection can provide some electrons to remove positive holes on the *n* side of the depletion zone. This narrows the depletion zone and therefore reduces the electric field strength in the zone. And it decreases the electric potential across the depletion zone. Then more electrons in the *n* side would have enough energy to participate in the diffusion current. Then $I_{diff} > I_{drift}$ and we get a net current.

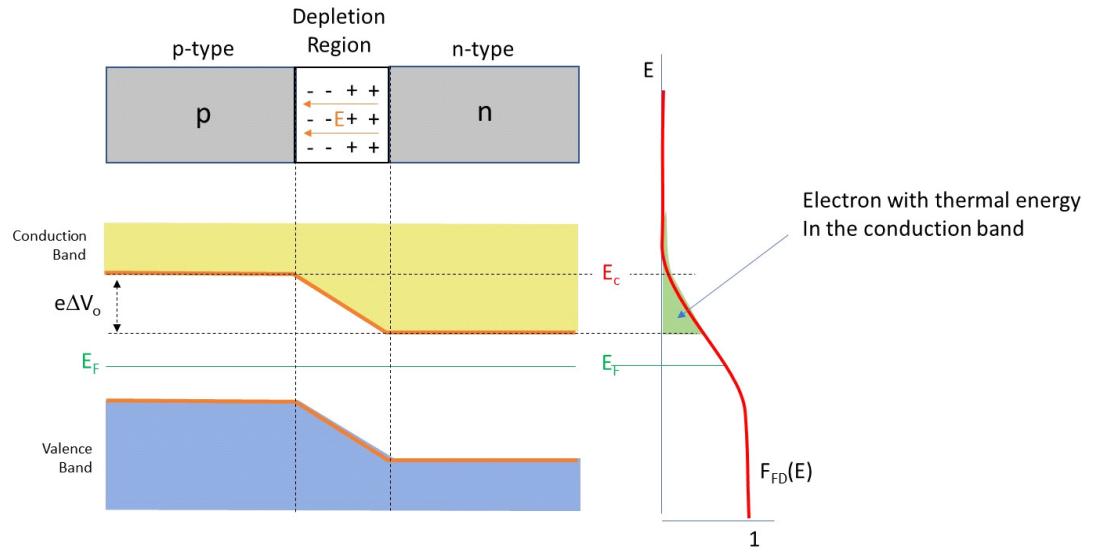


Notice in the diagram that the valence band energy difference between the *n* side and the *p* side of the junction is lower by an amount ΔV_{ext} provided by the battery, and notice that now many more electrons in the *n* side can move to the *p* side through diffusion. This configuration of a *pn* junction and a battery is called a *forward bias* and it allows current to flow through the junction. Notice that not only did the energy level change in a forward bias situation, but the depletion zone narrowed further helping the current flow. Of course if we hook the battery up the other way so it is a *reverse bias* the current will have a much harder time flowing. The reverse battery voltage adds to the potential energy difference in the valence zone between the *p* and *n* sides. If the battery voltage is large enough, the diffusion current will stop all together, and the diffusion zone will widen further stopping the diffusion current.



Under these circumstances, all we have is the small drift current that goes in the opposite direction.

We can approximate the drift and diffusion currents using our statistical physics ideas. I have repeated the next diagram to remind us of the shape of the tail. We are looking at the part in green. We know we need Fermi-Dirac distributions, but we are on the very tail end of the distribution so we can approximate the tail with an exponential.



In the p side, we have

$$N_p = \zeta e^{-(E_c - E_F)/k_b T}$$

where ζ is a scaling factor of proportionality and E_c is the conduction electron energy.

In the n side

$$N_n = \zeta e^{-(E_c - e\Delta V_{ext} - E_F)/k_b T}$$

and

$$I_{diff} \propto N_n$$

$$I_{drift} \propto N_p$$

and the total current is

$$\begin{aligned} I &= I_{diff} - I_{drift} \propto N_n - N_p \\ &= \zeta e^{-(E_c - e\Delta V_{ext} - E_F)/k_b T} - \zeta e^{-(E_c - E_F)/k_b T} \\ &= \zeta e^{-(E_c - E_F)/k_b T} \left(e^{(-e\Delta V_{ext})/k_b T} - 1 \right) \\ &= \zeta e^{-(E_c - E_F)/k_b T} \left(e^{(e\Delta V_{ext})/k_b T} - 1 \right) \end{aligned}$$

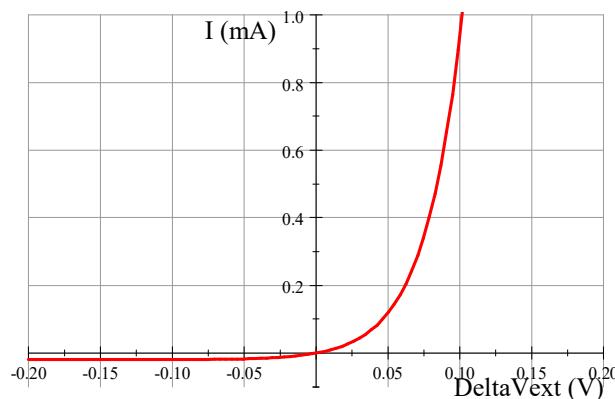
which we can write as

$$I = I_o \left(e^{\frac{e\Delta V}{k_b T}} - 1 \right)$$

where

$$I_o = \zeta e^{-(E_c - E_F)/k_b T}$$

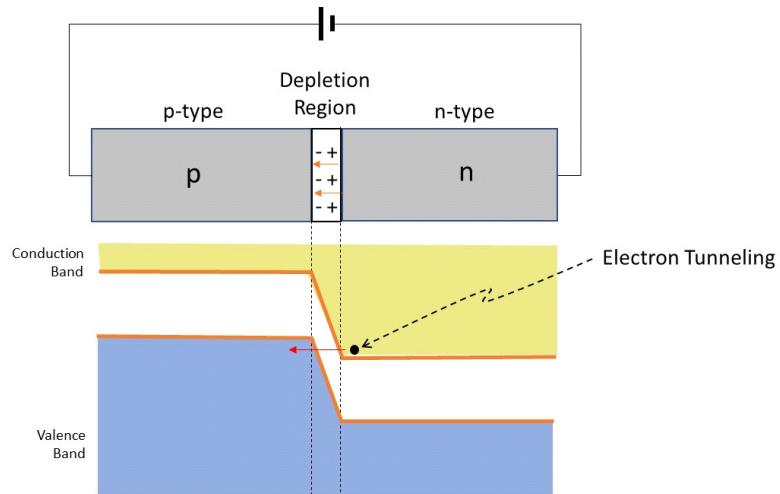
which looks like this.



Notice that this is essentially a one-way current valve. If you hook up the pn junction in reverse bias mode, almost no current flows, but in forward bias mode the current can flow. This pn junction is given the electronics name of *diode* and it is used in devices that convert alternating current to direct current because it only allows current one direction.

Tunneling diodes

There are other diode designs that are of interest. One such design is to make the depletion zone very thin by increasing the doping.

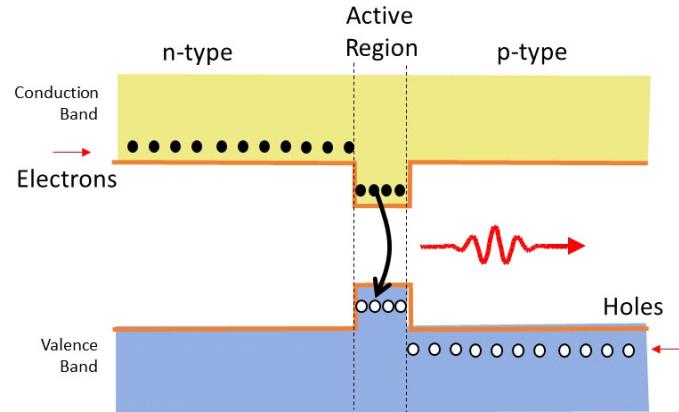


The heavy doping plus a strong forward bias makes the *n* side conduction band energies overlap with the *p* side valence band energies. Then electrons participating in the diffusion current are faced with a narrow barrier. And we know about barriers. Because the electrons are waves, if the barrier is narrow enough it the electrons will tunnel through to the *p* side valence band. This configuration is called a *tunneling diode*.

Photo diodes

Another common form of a diode is the light emitting diode or LED. If you have taken PH250 you have several of these in your Arduino kit and probably made them blink for your first lab. These diodes use the electrical potential supplied by the battery to excite electrons into the conduction band. But the diode is constructed so part of the conduction band is closer to the valence band at one point. And at that point the electron can jump down to the valence band in the same way an electron in an atomic excited state transitions to a lower shell.

Photo Diode



And just like when an atomic transition happens, a photon is emitted. The photon will have energy $E = hf$ that matches the energy of the electron transition. So if we can dope the diode to adjust the energy difference in the active region of the diode, we can make any color of photon we want.

Just like with atomic transitions, we could work this process backwards. We could have a photon absorbed, and use such a diode as a light detector.

Quantum Magnetic Materials

Let's go back to the idea of magnetism. We found that for atoms, the magnetic moment was primarily due to the spin of the electrons. But what happens when we put these atoms into solids? Each atom has its own magnetic moment, and the material will have an overall magnetic moment that is the sum of all the atomic magnetic moments.

$$\mathbf{M} = \frac{\sum_{i=1}^N \boldsymbol{\mu}_i}{V}$$

The magnetic moments usually are randomly aligned, so \mathbf{M} is usually zero. But if we put the sample in an external magnetic field, the individual magnetic moments tend to align. Different substances are more or less able to have all their dipole moments align. How well each substance is able to have it's magnetic dipole moments align with the external field is called the magnetic susceptibility, χ

$$\mu_o \mathbf{M} = \chi \mathbf{B}_{ext}$$

Diamagnetism

Some materials (e.g. noble gases) have paired electrons. The atomic orbitals are distorted by a magnetic field as we have seen before. But with all the electrons paired, they immediately relax when the external magnetic field is turned off. These materials have a magnetic moment opposite the direction of the external field. This *diamagnetic* effect is usually very weak.

Paramagnetism

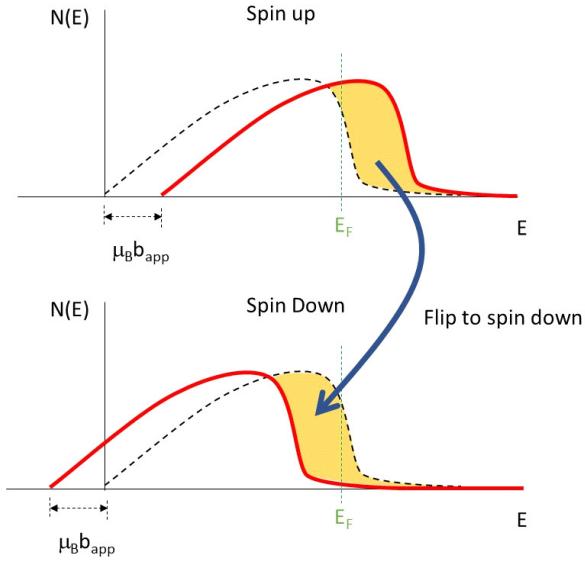
Other materials have their magnetic moments align with the external field. These are called *paramagnetic*. To form a theory for paramagnetic materials let's start with an electron gas in a metal. Recall that electrons have spin angular momentum, so they have a magnetic moment. The energy in the electrons in the gas would be

$$E = -\mu_s \cdot \mathbf{B}_{app}$$

where $\mu_s = -(e/m)\mathbf{s}$ is the spin magnetic moment of the electrons. Let's assume an applied magnetic field is in the z -direction. Then

$$\begin{aligned} E &= -\left(\frac{e}{m}\right)\mathbf{s} \cdot \mathbf{B}_{app} \\ &= -\left(\frac{e}{m}\right)s_z B_{app} \\ &= -\left(\frac{e}{m}\right)m_s \hbar B_{app} \\ &= \pm \mu_B B_{app} \end{aligned}$$

where $m_s = \pm \frac{1}{2}$. The electrons with $m_s = +\frac{1}{2}$ will gain energy when they align with the field, and the electrons with $m_s = -\frac{1}{2}$ will lose energy. The situation is as described in the next figure.



The spin up electrons's distribution is shifted up in energy and the spin down electrons's distribution is shifted down. But, the electrons are able to interact. So the higher energy electrons (orange area in the upper part of the figure) will flip to the lower energy state, leaving a surplus of spin down electrons. The strip on the graph that represents the change is roughly

$$\Delta E = 2\mu_B B_{app}$$

and the number of electrons in this strip would be

$$\Delta N = \frac{1}{2} V g(E) f_{FD}(E) \Delta E$$

where we recognize the Fermi-Dirac distribution and the multiplicity function $g(E)$. And the $1/2$ comes from the fact that we start with half of the electrons in the spin up state and half in the spin down state. Usually ΔE is very small, so the drawing exaggerates the effect. If we concentrate on very low temperatures, near $T = 0$, then we know the Fermi-Dirac distribution goes to 1 for $E < E_F$

$$\begin{aligned}\Delta N &= \frac{1}{2} V g(E) (1) \Delta E \\ &= \frac{1}{2} V g(E) 2\mu_B B_{app} \\ &= V g(E) \mu_B B_{app}\end{aligned}$$

and the switch of the electrons is all near the fermi energy so we can approximate with

$$\Delta N \approx V g(E_F) \mu_B B_{app}$$

The z component of the magnetization would be

$$M_z = \frac{1}{V} \sum_i \mu_{iz}$$

and if we take the case where all the atomic magnetic moments are aligned this is just

$$M_z = \frac{1}{V} \mu_B \Delta N$$

and we can write the susceptibility as

$$\begin{aligned} \chi &= \frac{\mu_o M}{B_{app}} = \frac{\mu_o \frac{1}{V} \mu_B \Delta N}{B_{app}} \\ &= \frac{\mu_o \frac{1}{V} \mu_B V g(E_F) \mu_B B_{app}}{B_{app}} \\ &= \mu_o \mu_B^2 g(E_F) \end{aligned}$$

Pauli estimated $g(E_F)$ in this case to be

$$g(E_F) = \frac{3N}{2VE_F}$$

and with this assumption we have

$$\chi = \frac{3\mu_o \mu_B^2}{2E_F} \frac{N}{V}$$

where we need to be careful because the number density N/V is the density of electrons and individual atoms might give more than one valence electron into the electron gas. This equation is called the *Pauli susceptibility*.

This only works for substances where saying the valence electrons form gas is a good approximation. Sodium is one such material and the calculated susceptibility is quoted as 8.2×10^{-6} .

But not every substance is well modeled by a lattice of ions with a gas of valence electrons. Other substances would have an overall spin angular momentum. Let's give this the symbol J . It would have a z -component

$$J_z = m_J \hbar$$

with m_J values

$$m_J = 0, \pm 1, \pm 2, \dots \pm J$$

for a given value of J there are $2J + 1$ possible m_J values. If we define a new factor, g , that ranges from 0 to 1 that depends on the particular combination of spin states that make J , then the effective magnetic moment for the substance would be

$$\mu = -g\mu_B \mathbf{J}$$

$$\mu_z = -g\mu_B m_J$$

and then the energy of interaction between the applied field and effective magnetic

moment would be

$$E = -\mu \cdot \mathbf{B} = g\mu_B m_J B_{app}$$

And the m_J states are not degenerate so we get the number of atoms in each magnetic substate to be

$$N_{mJ} = \frac{1}{A} e^{-\frac{E}{k_B T}} = \frac{1}{A} e^{-\frac{g\mu_B m_J B_{app}}{k_B T}}$$

where we are using the Maxwell-Boltzmann distribution. These are not continuous states, so we need to sum them rather than integrate. We need to find the normalization constant, A^{-1} . If we sum over all the states, we should get all the atoms.

$$N = \sum_{m_J=-J}^{+J} N_{mJ} = A^{-1} \sum_{m_J=-J}^{+J} e^{-g\mu_B m_J B_{app}/k_B T}$$

so

$$A^{-1} = \frac{N}{\sum_{m_J=-J}^{+J} e^{-g\mu_B m_J B_{app}/k_B T}}$$

so then

$$N_{mJ} = \frac{N}{\sum_{m_J=-J}^{+J} e^{-g\mu_B m_J B_{app}/k_B T}} e^{-\frac{g\mu_B m_J B_{app}}{k_B T}}$$

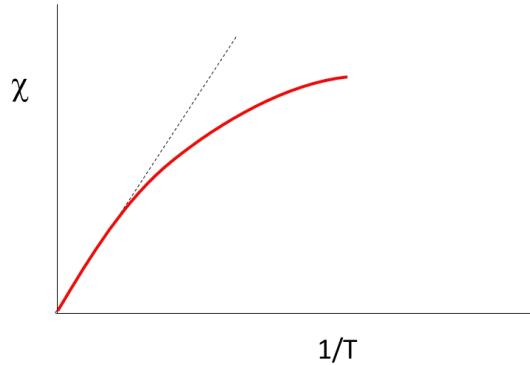
and then the z -component of the magnetization is the average magnetic moment per unit volume, and now we can find that

$$\begin{aligned} M_z &= \frac{1}{V} \sum_{\text{all atoms}} \mu_z = \frac{1}{V} \sum_{m_J=-J}^{+J} N_{mJ} g\mu_B m_J \\ &= \frac{1}{V} \sum_{m_J=-J}^{+J} \frac{N}{\sum_{m_J=-J}^{+J} e^{-g\mu_B m_J B_{app}/k_B T}} e^{-\frac{g\mu_B m_J B_{app}}{k_B T}} g\mu_B m_J \\ &= \frac{\frac{N}{V} g\mu_B \sum_{m_J=-J}^{+J} m_J e^{-\frac{g\mu_B m_J B_{app}}{k_B T}}}{\sum_{m_J=-J}^{+J} e^{-g\mu_B m_J B_{app}/k_B T}} \end{aligned}$$

and if our B_{app} is also in the z direction, then

$$\chi = \frac{\mu_o M_z}{B_{app}} = \frac{\mu_o \frac{N}{V} g\mu_B \sum_{m_J=-J}^{+J} m_J e^{-\frac{g\mu_B m_J B_{app}}{k_B T}}}{B_{app} \sum_{m_J=-J}^{+J} e^{-g\mu_B m_J B_{app}/k_B T}}$$

This tells us that paramagnetism is temperature dependent. It is traditional to plot χ as a function of $1/T$ which puts high temperatures on the left and low temperatures on the right.



For high temperatures we get a nice linear behavior. Believe it or not, this is just what our equation predicts. We can take T as large then terms like $e^{-\frac{g\mu_B m_J B_{app}}{k_B T}}$ are like $e^x \approx 1 + x$ for small x so

$$\chi = \frac{\mu_o \frac{N}{V} g \mu_B \sum_{m_J=-J}^{+J} m_J \left(1 - \frac{g\mu_B m_J B_{app}}{k_B T}\right)}{B_{app} \sum_{m_J=-J}^{+J} \left(1 - \frac{g\mu_B m_J B_{app}}{k_B T}\right)}$$

or

$$\chi = \frac{\mu_o \frac{N}{V} g \mu_B \sum_{m_J=-J}^{+J} \left(m_J - \frac{g\mu_B B_{app}}{k_B T} m_J^2\right)}{B_{app} \sum_{m_J=-J}^{+J} \left(1 - \frac{g\mu_B B_{app}}{k_B T} m_J\right)}$$

and it is easier to look at some of these sums on their own.

$$\sum_{m_J=-J}^{+J} 1 = 2J + 1$$

and

$$\sum_{m_J=-J}^{+J} m_J = 0$$

and finally

$$\sum_{m_J=-J}^{+J} m_J^2 = \frac{1}{3} J (J + 1) (2J + 1)$$

and we can write our equation as

$$\begin{aligned} \chi &= \frac{\mu_o \frac{N}{V} g \mu_B \left[\sum_{m_J=-J}^{+J} m_J - \frac{g\mu_B B_{app}}{k_B T} \sum_{m_J=-J}^{+J} m_J^2 \right]}{B_{app} \left[\sum_{m_J=-J}^{+J} (1) - \frac{g\mu_B B_{app}}{k_B T} \sum_{m_J=-J}^{+J} (m_J) \right]} \\ \chi &= \frac{\mu_o \frac{N}{V} g \mu_B \left[0 - \frac{g\mu_B B_{app}}{k_B T} \frac{1}{3} J (J + 1) (2J + 1) \right]}{B_{app} [2J + 1 - 0]} \\ \chi &= \frac{-\mu_o \frac{N}{V} g \mu_B \left[\frac{g\mu_B B_{app}}{k_B T} \frac{1}{3} J (J + 1) (2J + 1) \right]}{B_{app} [2J + 1]} \\ \chi &= \frac{-\mu_o N g^2 \mu_B^2 [J (J + 1)]}{3 k_B T V} \end{aligned}$$

which sure enough is linear in $1/T$. We did make quite a few assumptions and

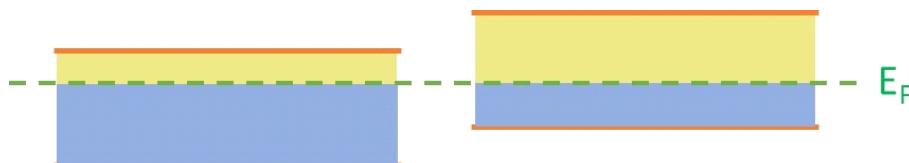
approximations, but we can see that our ideas about molecular spin states causing paramagnetism seem to be useful in explaining the phenomena.

Ferromagnetism

Neither of these methods of inducing magnetism are the basis of permanent magnets. Permanent magnets happen because of molecular orbital states that depend on electron spin. These states depend on the inter-atomic spacing. If the spacing is just right, we get permanent magnets. We can get a small idea of how this works by using our band model. For iron, when the magnetic interaction is in place each atom contributes 3 electrons to the molecular states, and the spin up and spin down states are at the same energy.



But if the magnetic interaction is in place, the molecular structure changes and each atom gives 4 electrons with spin up and 2 with spin down.



which provides a net magnetic moment.

It will take a bit more (in another class) to finish how this works.

We have studied the electron orbitals in an atom and found that chemistry resulted. We also found an explanation for the atomic spectra. It is time to take on the nucleus of the atom and see what wonders are waiting there.

Matter and Energy

To study the nucleus, we need to use Einstein's relationship between matter and

energy.

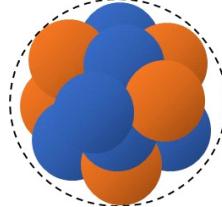
$$E_{rest} = mc^2 \quad (27.1)$$

Let's review what this means. This equation tells us that mass and energy are really the same thing. They are related through a constant of proportionality, but really they are the same stuff. When we talk of mass, we call it the *rest energy*, that is, the amount of energy an atom has at rest (no kinetic energy).

Nature of the Nucleus

From your high school chemistry classes, you know that the nucleus has two different types of particles. They are *protons* and *neutrons*. Generically we refer to both as *nucleons*. The protons are positively charged, the neutrons are negatively charged. These nucleons exist in close packed form at nearly the same density in all nuclei. We use some symbols to describe the parts of the nucleus. The first, Z , we already used in our model of the atom. It is the atomic number. It tells us the number of protons. The others are N , the number of neutrons, and A the *mass number* which is really the number of neutrons and protons.

The nucleons can be roughly modeled as spheres. In the nucleus, they are close packed, like nesting ping-pong balls.



Each chemical element has a unique number of protons, but it may have different numbers of neutrons. As we have already found in our quantum atom, the chemical properties of an atom come from the electron orbital structure. And an atom has the same number of electrons as protons. So we expect the chemical properties to vary by Z and to mostly not depend on N . So carbon with 12 neutrons and carbon with 13 neutrons is still carbon. We call atoms of the same element with different N values *isotopes*.

You may have heard of one of the isotopes of carbon which has 14 nucleons. This isotope is quite famous because it is radioactive and is somewhat prevalent. It is used in

dating artifacts up to about 1000 years old.

We have a standard notation for a nucleus

$$_Z^A X_N \quad (27.2)$$

where X is the chemical symbol for the atom, and we already know A , Z , and N . Note that the N is redundant and is almost always left off. But in the full notation our isotope of carbon with 14 nucleons would be written as

$$_6^{14} C_8$$

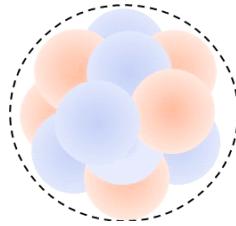
Normal carbon is written as

$$_6^{12} C_6 \quad (27.3)$$

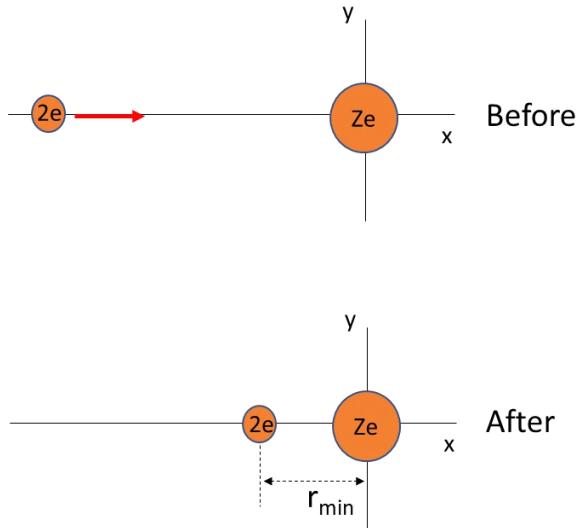
to represent carbon with 12 neutrons. And we can see that really the Z subscript is also redundant because the chemical symbol tells you what the number of protons will be. So you will often see our carbon isotopes written as ^{14}C or ^{12}C . Note that the subscripts are on the other side of the element name so we won't forget this is nuclear physics and not chemistry!

Nuclear Size and shape

But so far we have found that our high school physics was a rough approximation to our current models. We could guess that a quantum view of the nucleus will need to fill in some missing details. Nucleons are really not little hard balls. They are wavicles. So our last figure might not be so accurate. Nucleons are fuzzy in size because they are, of course, waves.



Rutherford found the approximate size of the atom. We did this problem earlier but let's review.



Rutherford shot α particles toward gold foil. By using the conservation of energy, he found when the α particles turned around. That would have to be at about the size of the nucleus, or at least we know the nucleus would have to be smaller than that distance. We start the alpha particle with an initial kinetic energy

$$K_i = \frac{1}{2}mv_i^2 \quad (27.4)$$

and let's start the α particle far enough away that the potential energy is essentially zero. When the α makes its closest approach, the potential energy will be maximum and the kinetic energy zero. The potential energy for the Coulomb force will be

$$U_f = k_e \frac{q_1 q_2}{r_f} \quad (27.5)$$

so then

$$\frac{1}{2}mv_i^2 = k_e \frac{q_1 q_2}{r_f} \quad (27.6)$$

We can identify q_1 for the α particle, and q_2 for the gold atom

$$\begin{aligned} \frac{1}{2}mv_i^2 &= k_e \frac{(2e)(Ze)}{r_{\min}} \\ r_{\min} &= k_e \frac{(2e)(Ze)}{\frac{1}{2}mv_i^2} \end{aligned} \quad (27.7)$$

where the r_{\min} is the distance of closest approach. If we solve for r_{\min}

$$r_{\min} = k_e \frac{(2e)(Ze)}{\frac{1}{2}mv_i^2} \quad (27.8)$$

This gives a very small number

$$r_{\min} = 3.2 \times 10^{-14} \text{ m} \quad (27.9)$$

We often use *femptometers* ($1 \text{ fm} = 10^{-15} \text{ m}$) as our unit for such small distances. This would be

$$r_{\min} = 32 \text{ fm}$$

We get the impression that the nucleus is roughly spherical. We can give an approximate equation for the radius that follows Rutherford's findings

$$r = r_o A^{\frac{1}{3}} \quad (27.10)$$

where $r_o = 1.2 \times 10^{-15} \text{ m}$ and where r is the radius for any atom we choose. A is the atomic mass number.

The number of nucleons per unit volume seems to be approximately constant, so we could say

$$\frac{N_{\text{nucleons}}}{V_{\text{nucleus}}} = \frac{A}{\frac{4}{3}\pi r^3} \simeq \text{constant.}$$

The density of the nucleus is fairly high. We can give an good estimate of this. A times the mass of a nucleon would be the mass of the nucleus. Since protons and neutrons have about the same mass, let's just use the mass of a proton in our estimate. For ^{12}C we would have

$$\begin{aligned} \rho &= \frac{N_{\text{nucleons}} m_{\text{nucleon}}}{V_{\text{nucleus}}} \approx \frac{Am_p}{\frac{4}{3}\pi r^3} \\ &= \frac{Am_p}{\frac{4}{3}\pi (r_o A^{\frac{1}{3}})^3} \\ &= \frac{12 (1.6726231 \times 10^{-27} \text{ kg})}{\frac{4}{3}\pi ((1.2 \times 10^{-15} \text{ m}) (12)^{\frac{1}{3}})^3} \\ &= 2.3108 \times 10^{17} \frac{\text{kg}}{\text{m}^3} \end{aligned}$$

which is very dense! We describe the nucleon mix in a nucleus with the term *nuclear matter* and this nuclear matter has very high densities.

28 Nuclear Binding Energy and Radioactivity

12.3-12.6

Fundamental Concepts in the Lecture

- Radioactivity
- Alpha Decay
- Beta decay
- Gamma Decay

Now that we know a little about the nucleus, it is time to ask about how nuclei are held together. From the start, we know that the positively charged protons will have a strong electrical repulsion force. We can draw on our models of electron states in atoms and our understanding of thermal statistical mechanics to build intuition into the dynamics of nucleons.

Binding Energy

The mass of the nucleus is *less than* the sum of the masses of the nucleon's from which it is made! This seems a little strange. But think of our atom with electrons. We found that when an electron approaches a proton and it falls into the potential well we get a release of a photon with energy of about 13.6 eV. The newly formed hydrogen atom would still have a proton and an electron, but it would have 13.6 eV worth of mass less than the two combined masses before they were bound together. And mass is a form of energy

$$E_{rest} = mc^2$$

and the bound hydrogen atom has less energy than the separate proton and electron, so it has less mass than a separate proton and electron. The missing mass left in the form of the energy of the photon. This is something we discovered in this class some time

ago. You might object. Isn't it really just that we have changed the potential energy

$$U = -\frac{1}{4\pi\epsilon_0} \frac{e^2}{r}$$

by making r much smaller? The answer is, of course, yes! But this change in potential energy is a change in... energy, and the mass of the hydrogen atom system is formed from the rest energy of the hydrogen atom system. So if we change the rest energy we change the mass. Internal potential energy of a system on the quantum scale is expressed as part of the system mass.

But we should ask if this would be true for the nucleus. If we bring a proton and a neutron together to make a nucleus (2H) then there must be a binding energy holding the proton and neutron together. Would the mass of the nucleus be less than the mass of a separate proton and a separate neutron?

It takes energy to remove nucleons from the nucleus. This energy to remove a nucleon is the binding energy of a nucleon. We lose this energy when the nucleons are brought together. The energy will leave in the form of photons. A photon is released, a gamma ray in this case. So the nuclear mass will be less than the mass of the separate nucleons. Let's do the numbers for 2H . The binding energy would be

$$\begin{aligned} B &= \text{mass of the separate nucleons} - \text{mass of the nucleolus} \\ &= (m_n c^2 + m_p c^2) - m_D c^2 \end{aligned}$$

where m_D is the nuclear mass of 2H (because 2H is called "deuterium"). We can play a clever trick to make this calculation easier. Note that the rest mass energy of the 2H atom includes an electron mass. So

$$m_{[{}^2H]} c^2 = m_D c^2 + m_e c^2 - 13.6 \text{ eV}$$

where of course the -13.6 eV is the binding energy that holds the electron bound to the hydrogen atom. We could do this for 1H as well

$$m_{[{}^1H]} c^2 = m_p c^2 + m_e c^2 - 13.6 \text{ eV}$$

Now if we add the rest mass energy of 1H to the rest mass energy of the neutron we get

$$m_{[{}^1H]} c^2 + m_n c^2 = (m_p c^2 + m_n c^2) + m_e c^2 - 13.6 \text{ eV}$$

and the part in parenthesis shows up in our nuclear binding equation. Let's solve for that part in parenthesis and substitute it back into our binding energy equation

$$m_{({}^1H)} c^2 + m_n c^2 - m_e c^2 + 13.6 \text{ eV} = (m_p c^2 + m_n c^2)$$

so

$$B = (m_n c^2 + m_p c^2) - m_D c^2$$

becomes

$$B = (m_{[{}^1H]} c^2 + m_n c^2 - m_e c^2 + 13.6 \text{ eV}) - m_D c^2$$

We could rearrange this as

$$B = m_{[{}^1H]} c^2 + m_n c^2 - (m_D c^2 + m_e c^2 - 13.6 \text{ eV})$$

and now we recognize that the part in parenthesis is the rest energy of a 2H atom (including the electron and it's binding energy).

$$B = m_{[{}^1H]} c^2 + m_n c^2 - m_{[{}^2H]} c^2$$

This might not be exciting at first, but there are many collected tables of the masses of whole atoms, but few tables of masses of nuclei. With what we have done we recognize that we can put in the masses of the full atoms, knowing that under the hood the electron masses and binding energies will cancel out. Let's look at our 2H binding energy one more time and write out the atomic masses so we can see the electron rest mass energy and binding energy cancel explicitly.

$$\begin{aligned} B &= (m_p c^2 + m_e c^2 - 13.6 \text{ eV}) + m_n c^2 - (m_D c^2 + m_e c^2 - 13.6 \text{ eV}) \\ &= m_p c^2 + m_n c^2 - (m_D c^2) \end{aligned}$$

just as it should.

We did this with the simplest of nuclei. But it works for any nucleus. We could write this in general for an atom,

$$B = N m_n c^2 + Z m_p c^2 - m_{[A]} c^2$$

where N is the number of neutrons, Z is the number of protons, and A is the atomic mass, so it tells us which atomic nuclei we have. But We could use the atomic masses to write this as

$$B = N m_n c^2 + Z m_{[{}^1H]} c^2 - m_{[{}^AX_N]} c^2$$

where we used the letter X as a generic atomic symbol (like C or carbon or Na for sodium). Just as in our 2H case the electron contributions cancel out leaving us with our nuclear binding energy. Let's try this for a few atoms.

For iron, ${}^{56}_{26}Fe_{30}$, we have $N = 30$, $Z = 26$, and $A = 56$. So we can find a binding energy of

$$B = 30 m_n c^2 + 26 m_{[{}^1H]} c^2 - m_{[{}^{56}_{26}Fe_{30}]} c^2$$

The masses are (from the table in our book¹¹)

$$\begin{aligned}m_n &= 1.0086649160 \text{ u} \\m_{[{}^1H]} &= 1.0078250 \text{ u} \\m_{[{}^{56}_{26}Fe_{30}]} &= 55.934937 \text{ u}\end{aligned}$$

and we know the conversion

$$1 \text{ u} = 931.49406 \frac{\text{MeV}}{c^2}$$

so the rest energies are

$$\begin{aligned}E_n &= 1.0086649160 \text{ u} \frac{931.49406 \frac{\text{MeV}}{c^2}}{1 \text{ u}} \\&= 939.57 \frac{\text{MeV}}{c^2} \\E_{[{}^{56}_{26}Fe_{30}]} &= 55.934937 \text{ u} \frac{931.49406 \frac{\text{MeV}}{c^2}}{1 \text{ u}} \\&= 52103. \frac{\text{MeV}}{c^2} \\E_{[{}^1H]} &= 1.0078250 \text{ u} \frac{931.49406 \frac{\text{MeV}}{c^2}}{1 \text{ u}} \\&= 938.78 \frac{\text{MeV}}{c^2}\end{aligned}$$

Then our binding energy is

$$\begin{aligned}B &= 30(939.56538 \text{ MeV}) + 26(938.78 \text{ MeV}) - 52103. \text{ MeV} \\&= 492.24 \text{ MeV}\end{aligned}$$

but because we did two unit conversions from u to MeV/c^2 it is easier to do everything in terms of u and convert at the end.

$$\begin{aligned}B &= (30(1.0086649160 \text{ u}) c^2 + 26(1.0078250 \text{ u}) c^2 - (55.934937 \text{ u}) c^2) \frac{931.49406 \frac{\text{MeV}}{c^2}}{\text{u}} \\&= 492.26 \text{ MeV}\end{aligned}$$

This is the total binding energy holding all the nucleons together. We could get an estimate of how much binding energy an individual nucleon would have by calculating the average binding energy per nucleon

$$\frac{B}{A} = \frac{492.26 \text{ MeV}}{56} = 8.7904 \frac{\text{MeV}}{\text{nucleon}}$$

Let's do this again for a larger atom, say, uranium ${}^{235}_{92}U_{146}$

¹¹ I am still working on this table. But the most current data on atomic masses is from the US National Institute of Standards (NIST). <https://www.nist.gov/pml/atomic-weights-and-isotopic-compositions-relative-atomic-masses>

$$B = 146m_n c^2 + 92m_{[{}^1H]} c^2 - m_{[{}^{235}_{92}U_{146}]} c^2$$

and from our table.

$$\begin{aligned} m_n &= 1.0086649160 \text{ u} \\ m_{[{}^1H]} &= 1.0078250 \text{ u} \\ m_{[{}^{235}_{92}U_{146}]} &= 235.043930 \text{ u} \end{aligned}$$

The our binding energy is

$$\begin{aligned} B &= [143(1.0086649160 \text{ u}) c^2 + 92(1.0078250 \text{ u}) c^2 - (235.043930 \text{ u}) c^2] \frac{931.49406 \frac{\text{MeV}}{c^2}}{\text{u}} \\ &= 1783.9 \text{ MeV} \end{aligned}$$

with the binding energy per nucleon being

$$\frac{B}{A} = \frac{1783.9 \text{ MeV}}{235} = 7.5911 \frac{\text{MeV}}{\text{nucleon}}$$

Notice that the binding energy per nucleon for iron is bigger than for uranium. How about for something like Lithium, 3_3Li_3 ?

$$B = 3m_n c^2 + 3m_{[{}^1H]} c^2 - m_{[{}^3_3Li_3]} c^2$$

and

$$\begin{aligned} m_n &= 1.0086649160 \text{ u} \\ m_{[{}^1H]} &= 1.0078250 \text{ u} \\ m_{[{}^3_3Li_3]} &= 6.015123 \text{ u} \end{aligned}$$

which gives a binding energy of

$$\begin{aligned} B &= [3(1.0086649160 \text{ u}) c^2 + 3(1.0078250 \text{ u}) c^2 - (6.015123 \text{ u}) c^2] \frac{931.49406 \frac{\text{MeV}}{c^2}}{\text{u}} \\ &= 31.994 \text{ MeV} \end{aligned}$$

with the binding energy per nucleon being

$$\frac{B}{A} = \frac{31.994 \text{ MeV}}{6} = 5.3323 \frac{\text{MeV}}{\text{nucleon}}$$

It appears that iron, ${}^{56}_{26}Fe_{30}$, has a higher binding energy than both 3_3Li_3 and ${}^{235}_{92}U_{146}$.

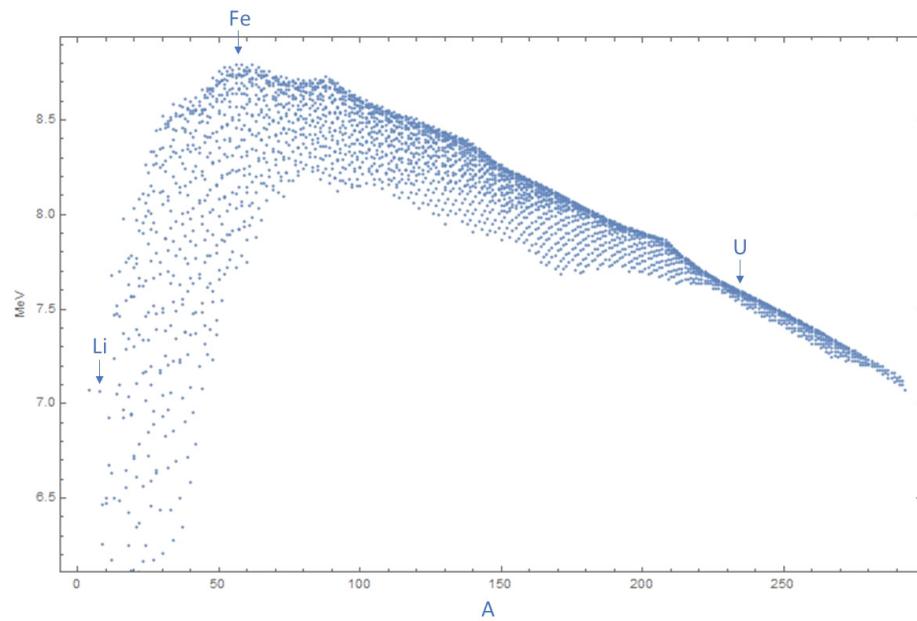
Nuclear Stability

You might be wondering why the protons agree to sit calmly in the nucleus next to each other. They do have positive charges. There must be some force that is stronger than the Coulomb force that keeps them in place! That is exactly what we believe happens. It is called the *nuclear* force, but this force has to be very strong to keep the protons together, stronger than the Coulomb force. So most often the nuclear force is called the

strong force. It drops dramatically as a function of distance. After about 2 fm it is very weak, but at closer distances it is much stronger.

For small atomic numbers, even numbers of protons and neutrons make stable atoms, but for larger atoms it seems to be better to have more neutrons than protons. For large atoms, the Coulomb force will be large, so spreading out the protons helps to reduce the repulsion. Thus, for large atoms, more neutrons make the atom more stable.

You might guess that there is a limit to all this. At some point the Coulomb force would just be larger. You would be right. $Z = 83$ seems to be the limit for stable atoms. In the next figure the binding energy per nucleon of all isotopes of all atoms are show. Lithium, iron, and uranium are called out on the graph.

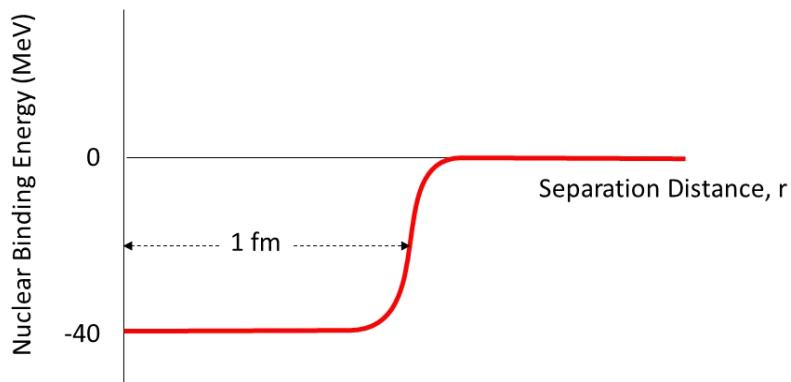


We can see that atoms with the highest binding energies per nucleon are right around iron with iron having the highest binding energy. That means that iron is the most stable, holding its nucleons the tightest. Then, there are two ways to get energy from the nuclear reactions. We could liberate energy by splitting heavy nuclei into smaller, more stable products. This is what we call nuclear fission. The other route would be to take smaller nuclei and combine them to form larger, more stable nuclei. This is what we call nuclear fusion. Both can provide a release of energy as the product nuclei move toward the more stable region near iron.

Nuclear energy states

Let's return to our discussion of the strong force. We expect the strong force to form a potential energy well. Then, as we bring in a new nucleon, we would expect the quantum well to have many allowed energy states like the Coulomb force well has for electrons in the orbitals. For deuterium, there doesn't seem to be any excited states. If we take a neutron and bring it into a proton, it falls all the way to a ground state emitting just one photon of energy 2.224 MeV. In nuclei with more nucleons, we do see energy levels.

Deuterium studies do teach us some characteristics of the strong force. The strong force is very short range. It is limited to distances of about 1 fm.



The size of a proton is on the order of a femtometer and so is the size of a neutron, so this means the strong force can only tie nearest neighbor nucleons together. This is a little like a crystal, where the atoms are usually bound to their nearest neighbor. The strong force appears to be the same for proton-proton bonds, and neutrons-bonds, and proton-neutron bonds.

The strong force is usually modeled as an attractive exchange force. But it is more complicated than, say, the Coulomb force. And this takes some explaining. What happens in this model is that one of the nucleons emits a particle to which it and other near by nucleons are attracted. Thinking of 2H we would have a proton and a neutron. If the neutron emitted such a particle then both the proton and the neutron would be attracted to the particle. It's like having a pillow fight, and both you and your opponent grab the same pillow and pull on it. You and your opponent would move together because you are both moving toward the particle. The proton and neutron would appear to be attracted to each other in this way. they are both attracted to the mitigating *exchange particle*. But, you may object, wouldn't constantly emitting particle reduce

the rest energy of the nucleons?

Let's consider our uncertainty principle again.

$$\Delta E \Delta t \approx \hbar$$

or

$$\Delta t \approx \frac{\hbar}{\Delta E}$$

and if the exchange particle has mass m we need energy $\Delta E = mc^2$ to create the particle. Then if the particle exists for less than about

$$\Delta t \approx \frac{\hbar}{mc^2}$$

to within what we can know, we have not violated conservation of energy. But this says that the exchange particle must have a limited lifetime. The particle can exist, borrowing an amount of energy ΔE so long as within the time Δt that energy is returned to the system as the particle rest mass energy mc^2 when the particle is destroyed. Under this condition energy conservation is not violated.

This may sound crazy, but remember that our “particles” are not really “particles” but waves. No one is surprised at the beach when occasionally there is a bigger wave.¹² And that is what we have going on here. And the energy in our nuclear wavicles is spread out physically and temporally. We are really exchanging potential energy between two nucleon waves with an intermediate exchange wave. A sort of “splash in the middle.”

We could see how far this exchange waveicle would go, assuming its speed is not relativistic

$$\begin{aligned}\Delta x &= c\Delta t \\ &= \frac{c\hbar}{mc^2}\end{aligned}$$

which gives a limit on the range of our strong force. The strong force can't exist when the particle ceases to exist. So if the nucleons move beyond this distance the strong force will be essentially zero and the nucleons won't seem to attract each other. This is how it is that the strong force only attracts nearest neighbor nucleons.

But we should ask if we can identify this exchange particle. We can't open up a nucleus and peer inside with a microscope. And the exchange particles don't travel outside the nucleus because of their limited lifetime. If the particle left the nucleus we would violate conservation of energy. But suppose we provide the energy ΔE by sending in a

¹² Well, they may be surprised, and that is the stuff of funny YouTube videos. But no one finds it strange.

photon. Then the particle could be released without violating conservation of energy. When this experiment is tried, we get particles with $mc^2 = 140 \text{ MeV}$. These particles are given the name π mesons. Our model would predict

$$\begin{aligned} mc^2 &= \frac{c\hbar}{\Delta x} \\ &\approx \frac{197.326972 \text{ MeV fm}}{1 \text{ fm}} \\ &\approx 200 \text{ MeV} \end{aligned}$$

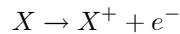
which is a rough estimate but matches the rest energy of a π meson fairly well. We will return to exchange particles and the strong force when we study quantum chromodynamics later in this course.

Separation energy

Armed with a model for the strong force, let's look at the energy required to remove the least tightly bound nucleon from the nucleus. This is like removing an upper electron from an atomic electron state. We know how to do this with electrons. We called this the ionization energy and it is given by

$$\begin{aligned} E_i &= m_{X^+}c^2 + m_e c^2 - m_X c^2 \\ &= (m_{X^+} + m_e - m_X) c^2 \end{aligned}$$

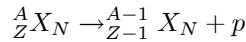
where X is a generic element symbol. We write the reaction that creates the ion as



We want to do the same sort of thing for a nuclear separation energy

$$S_p = \left(m_{[Z-1]X_N} + m_{[1]H} - m_{[Z]X_N} \right) c^2$$

where we have used atomic masses because the electrons will once again cancel out and we have tables of atomic masses. This is at least a very good approximation. We could write the reaction as



where we assumed that the most loosely held nucleon is a proton which makes sense due to the Coulomb repulsion for small nuclei. But for larger nuclei it could be a neutron.

$$S_n = \left(m_{[Z]^{A-1}X_{N-1}} + m_n - m_{[Z]X_N} \right) c^2$$

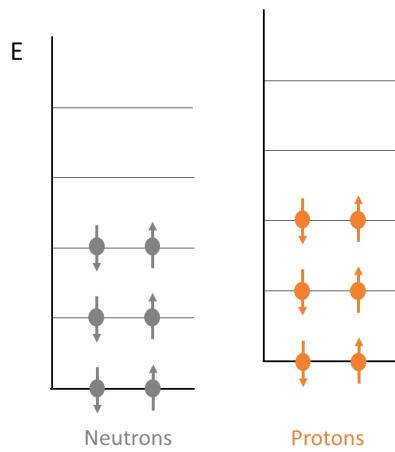
$${}_{Z}^{A}X_N \rightarrow {}_{Z}^{A-1}X_{N-1} + n$$

By inserting gamma rays on the order of $5 - 10 \text{ MeV}$ we can remove a nucleon, so the separation energies should be on this order.

But if we can remove a nucleon by inserting it's separation energy, this reminds us so much of removing an electron by inserting it's ionization energy. We could even call the most loosely held nucleon a *valence* nucleon. We would expect excited states for nucleons.

Nuclear energy states

We studied deuterium and found we could separate the two nucleons if we inserted the right amount of energy. But it turns out that with just two nucleons we only found a ground state and separate particles. In our analysis we didn't see discrete energy states in between these two extremes like we found for an electron in hydrogen. But if we allow more nucleons the situation is more complicated. Three nucleons is a multi-body problem and we would have to approach it numerically. We might gain some insight by looking at a one-dimensional infinite potential well for nucleons. We would expect we would get a number of energy states E_n as we did for electrons in atoms. And because neutrons and protons are Fermions, we would expect to have two neutrons or two protons in each state. The potential energy for neutrons is different than for protons because protons have Coulomb forces. So neutrons and protons would have different energy levels. Schematically¹³ this could be visualized like this



These energy levels will lead us into a theory for radioactive decay. Which will study in

¹³ Remember that "schematically" means we are drawing like an artist. We didn't calculate the energy states and plot them. The figure is just to get an idea of what it might look like.

our next lecture.

29 Radioactive Decay

Fundamental Concepts in the Lecture

- α , β , and γ decay
- Half-Life and Rate of Decay
- Calculations Involving Decay Rates and Half-Life

Radioactivity was discovered, like many things, by accident. In 1896 Becquerel found that uranium salt would darken photographic paper. This launched a field of study that has, at its head, one of the first recorded great women scientists, Marie Curie.

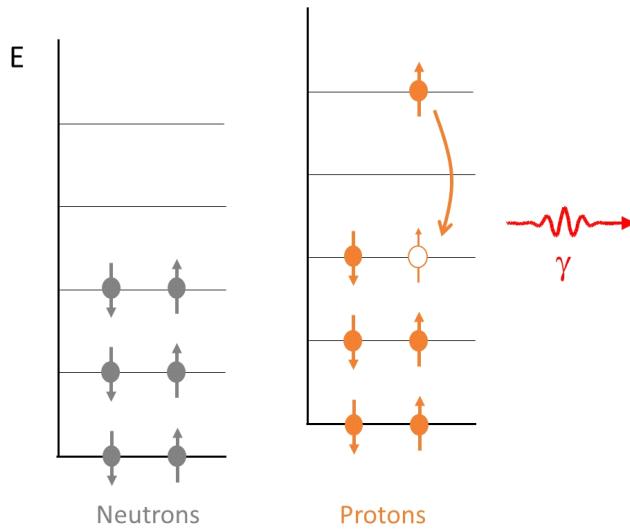
Radioactive decay basics

Question 30.4 a-c

Question 30.6 a-c

We now know from our study of binding energy per nucleon that radioactivity is the decay of large nuclei that are unstable. There are three basic types of nuclear decay. we can describe all of them with reference to our nuclear energy level diagrams.

1. Gamma (γ) emission: the emission of a high energy photon. This is the easiest to understand. A nucleon in an excited state drops down into a lower state.

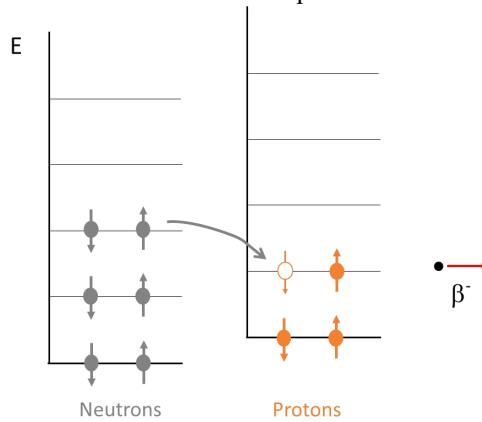


From what we know of electron energy levels we would expect a photon to be produced with

$$\Delta E = hf$$

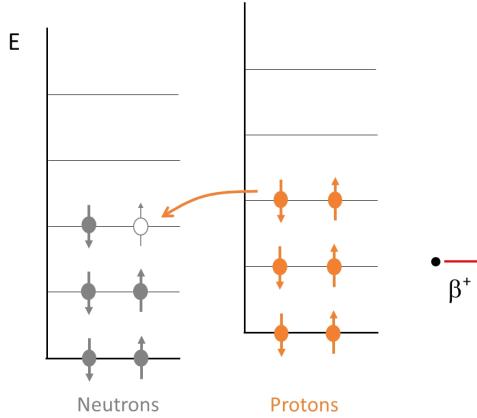
The existence of gamma emission is good evidence for nuclear excited states.

2. Beta (β) emission: The emission of an electron or its antiparticle, the positron. This is more complicated because it involves one nucleon changing into a different type of nucleon. For example a neutron could become a proton.



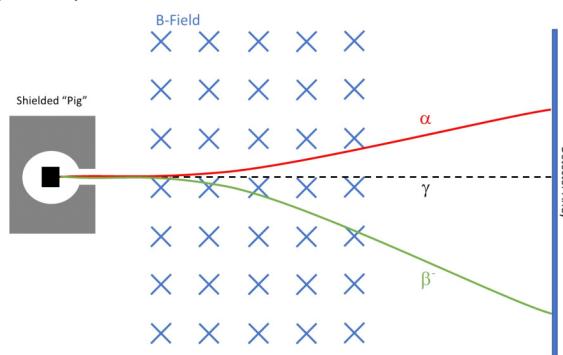
But to maintain conservation of electrical charge this would mean a negative charge equal to the new proton charge must be created, and this is what happens. A new electron is created. But electrons don't interact by the strong force. So the electron will leave the nucleus. And thus an electron is emitted. Because early researchers didn't realize that β particles were electrons they gave them the symbol β to distinguish be-

tween this kind of emission and γ emission (and α emission). The process we described can work the other way as well. We could convert a proton into a neutron. But we need to remove the positive proton charge to do this. So an anti-electron, a positron, is produced.



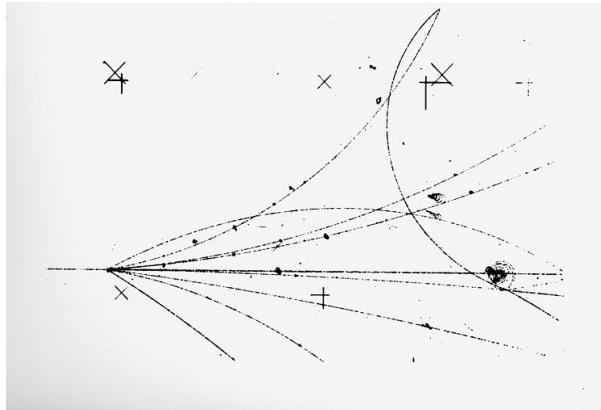
3. Alpha (α) emission: The emission of a ${}^4_2\text{He}$ nuclei. This one is not at all obvious. We will study this in more detail later, but under some circumstances it is more probable that a group of two protons and two neutrons will leave the parent nucleus than a single proton or neutron leaving. When this happens the resulting emission was called an α particle by early researchers. Two protons and two neutrons make a helium nucleus, and helium nuclei are still referred to as α particles.

Often β or α decay will leave the parent nucleus with some nucleons in excited states. So often α or β decay are accompanied by γ emission. We can actually observe these particles leaving nuclei. We use a device like the one shown below



You would expect α particles to be positively charged, and electrons to be negatively charged. So if they travel through a magnetic field as shown, they will deflect different

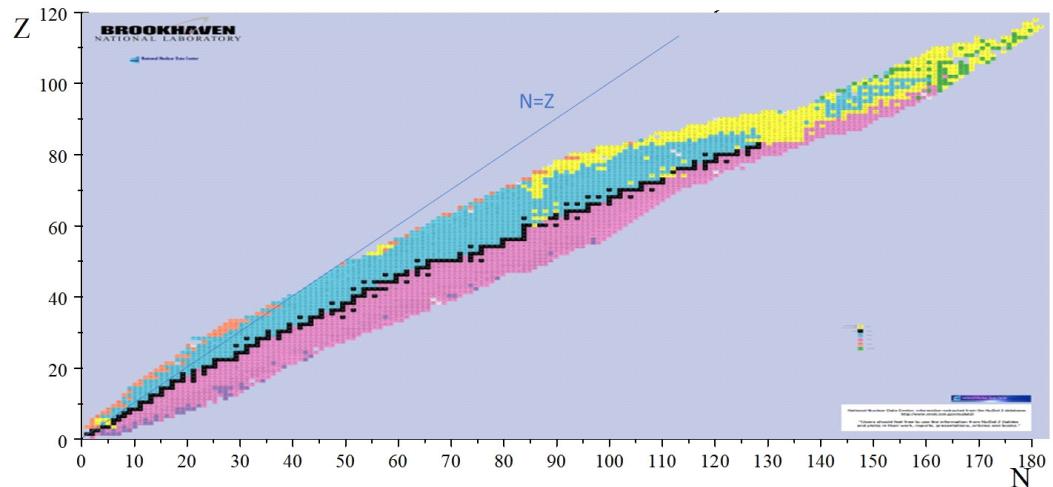
directions. γ particles will not deflect at all. We use a cool device to detect these particles. We make a cooled chamber and fill it with alcohol vapor (or similar gas that can supersaturate). The chamber is placed in the path of the particles. Then, when the particles cross the chamber, they make condensation trails like jet airplanes. We can see the path they take!



One of the first photographs taken with the thirty-inch bubble chamber at NAI, June 15, 1972. A two-hundred GeV proton enters the chamber and interacts with liquid hydrogen. The resulting collision produces a spectacular event with ten visible nuclear fragments emerging. The tracks are nearly thirty inches long. Photo Courtesy US Department of Energy.

Since γ rays have shorter wavelengths and higher energies than X-rays, you might guess that they are hard to stop. They can penetrate through lead. The heavier electrons (β particles) can penetrate an aluminum plate, and the α particles have trouble going through a piece of paper.

In the next figure is a chart taken from Brookhaven National Laboratory that shows all the different combinations of neutron and protons that we know occur. Each different combination is called a *nucleoid*. The little boxes show different nucleoids, but the boxes are so small you can't read that there are tiny words (no doubt not reproduced either by the printer or your screen and maybe not even by my software). But notice there is a color coding. The black boxes represent stable nuclei. The yellow is for α decay, light blue is β^+ decay and pink is β^- decay.

Table of the nucleotides (live version at <https://www.nndc.bnl.gov/nudat2/>)

Here is a small section of the chart expanded so we can see the individual nucleotides boxes and that there are more details in our color scheme.

^{234}Pu 8.8 H $\alpha: 94.00\%$ $\alpha: 6.00\%$	^{235}Pu 25.3 M $\epsilon: 100.00\%$ $\alpha: 2.8E-3\%$	^{236}Pu 2.858 Y $\alpha: 100.00\%$ $SF: 1.9E-7\%$	^{237}Pu 45.64 D $\epsilon: 100.00\%$ $\alpha: 4.2E-3\%$	^{238}Pu 87.7 Y $\alpha: 100.00\%$ $SF: 1.9E-7\%$	^{239}Pu 24110 Y $\alpha: 100.00\%$ $SF: 3.E-10\%$	^{240}Pu 6561 Y $\alpha: 100.00\%$ $SF: 5.7E-6\%$	^{241}Pu 14.329 Y $\beta: 100.00\%$ $\alpha: 2.5E-3\%$	^{242}Pu 3.75E+5 Y $\alpha: 100.00\%$ $SF: 5.5E-4\%$
^{233}Np 36.2 M $\epsilon: 100.00\%$ $\alpha: 1.0E-3\%$	^{234}Np 4.4 D $\epsilon: 100.00\%$	^{235}Np 396.1 D $\epsilon: 100.00\%$ $\alpha: 2.6E-3\%$	^{236}Np 153E+3 Y $\epsilon: 86.30\%$ $\beta: 13.50\%$	^{237}Np 2.144E+6 Y $\alpha: 100.00\%$ $SF \leq 2E-10\%$	^{238}Np 2.117 D $\beta: 100.00\%$	^{239}Np 2.366 D $\beta: 100.00\%$	^{240}Np 61.9 M $\beta: 100.00\%$	^{241}Np 13.9 M $\beta: 100.00\%$
^{232}U 68.9 Y $\alpha: 100.00\%$ $24\text{Ne}: 9E-10\%$	^{233}U 1.592E+5 Y $\alpha: 100.00\%$ $24\text{Ne}: 9E-10\%$	^{234}U 2.455E+5 Y 0.0054% $\alpha: 100.00\%$ $SF: 1.6E-3\%$	^{235}U 7.04E+8 Y 0.7204% $\alpha: 100.00\%$ $SF: 7.0E-9\%$	^{236}U 2.342E7 Y $\alpha: 100.00\%$ $SF: 9.4E-8\%$	^{237}U 6.75 D $\beta: 100.00\%$	^{238}U 4.468E9 Y 99.2742% $\alpha: 100.00\%$ $SF: 5.4E-5\%$	^{239}U 23.45 M $\beta: 100.00\%$	^{240}U 14.1 H $\beta: 100.00\%$
^{231}Pa 3.276E+4 Y $\alpha: 100.00\%$ $SF < 3E-10\%$	^{232}Pa 1.32 D $\beta: 100.00\%$ ϵ	^{233}Pa 26.975 D $\beta: 100.00\%$	^{234}Pa 6.70 H $\beta: 100.00\%$	^{235}Pa 24.4 M $\beta: 100.00\%$	^{236}Pa 9.1 M $\beta: 100.00\%$	^{237}Pa 8.7 M $\beta: 100.00\%$	^{238}Pa 2.28 M $\beta: 100.00\%$	^{239}Pa 1.8 H $\beta: 100.00\%$
^{230}Th 7.54E+4 Y $\alpha: 100.00\%$ $24\text{Ne}: 6E-11$ 0%	^{231}Th 25.52 H $\beta: 100.00\%$ $\alpha: 4E-11\%$	^{232}Th 1.40E+10 Y 100% $\alpha: 100.00\%$ $SF: 1.1E-3\%$	^{233}Th 21.83 M $\beta: 100.00\%$	^{234}Th 24.10 D $\beta: 100.00\%$	^{235}Th 7.1 M $\beta: 100.00\%$	^{236}Th 37.3 M $\beta: 100.00\%$	^{237}Th 4.7 M $\beta: 100.00\%$	^{238}Th 9.4 M $\beta: 100.00\%$

Zoomed in view of the Table of the nucleotides (live version at

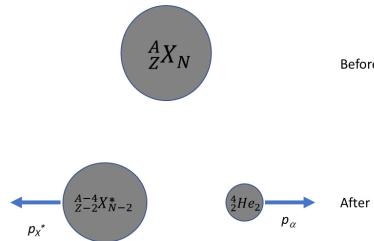
<https://www.nndc.bnl.gov/nudat2/>)

Near the middle of the figure you can see ^{235}U with a black box to note it's long half life (discussed below). It is interesting that so many nucleotides are unstable. Let's look at the decay processes in detail.

Details of the decay processes

Knowing the basics of how nuclei decay, let's now take a deeper look at each of the decay processes.

Alpha decay



Let's take a generic nucleus called X and create a new nucleus called Y . The equation for producing Y from X by α -decay is



We call the original nucleus the *parent nucleus*. That would be $\frac{A}{Z}X$. We call the new nucleus the *daughter nucleus*.¹⁴ That would be $\frac{A-4}{Z-2}Y$.

Let's take a real example.



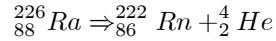
This is almost like chemistry! All we have to do is to balance the equation so the mass numbers, A , and the atomic numbers Z add up to the same thing on each side of the reaction!

There is a small catch, the parent nucleus must have a larger mass than the combined mass of the daughter and the α particle. The lost mass is converted into the kinetic energy of the daughter particles (and sometimes other forms of energy).

The α -particle will have the higher kinetic energy because its mass is smaller. This is like an explosion problem from PH121!

¹⁴ "Daughter nucleus" is what we find in the literature. I may call this a "child nucleus" to be less gender specific, although the use of "daughter" was in part to balance gender references in physics. This part of language has become difficult to navigate. Please be patient while society settles on new rules.

Let's find the amount of energy liberated in this decay.



We start by finding the mass of the parent nucleus. We look at a table of isotopes to find this

$$m_{Ra} = 226.025403 \text{ u} \quad (29.3)$$

and then we sum the masses of the daughter products (the *Ra* nucleus and α particle)

$$m_{Rn} = 222.017570 \text{ u} \quad (29.4)$$

$$m_\alpha = 4.002603 \text{ u} \quad (29.5)$$

so

$$m_{Rn} + m_\alpha = 222.017570 \text{ u} + 4.002603 \text{ u} \quad (29.6)$$

$$= 226.020173 \text{ u} \quad (29.7)$$

then the lost mass is

$$\Delta m = m_{Ra} - (m_{Rn} + m_\alpha) \quad (29.8)$$

$$= 226.025403 \text{ u} - (226.020173 \text{ u}) \quad (29.9)$$

$$= 0.00523 \text{ u} \quad (29.10)$$

$$= 4.8717 \text{ MeV}/c^2 \quad (29.11)$$

This change in the rest mass energy is called the *Q value* and we can write our process as

$$Q = (m_{Ra} - (m_{Rn} + m_\alpha)) c^2$$

We will often find problems like this that ask us to find the binding energy or the energy released. You should be familiar with this process. The α -decay is a form of *spontaneous decay*. Remember we studied spontaneous emission of light. That was emission that happened on its own without something causing it to happen. The word spontaneous means the same thing here.

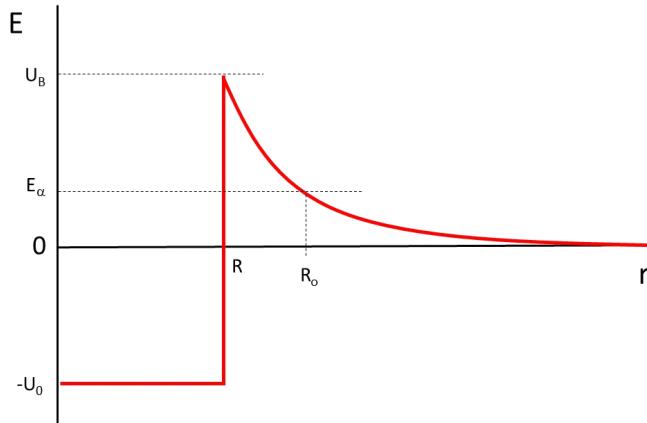
Quantum Tunneling and α -Decay

But how would an α particle just leave a nucleus? We know that there would be a Coulomb potential of

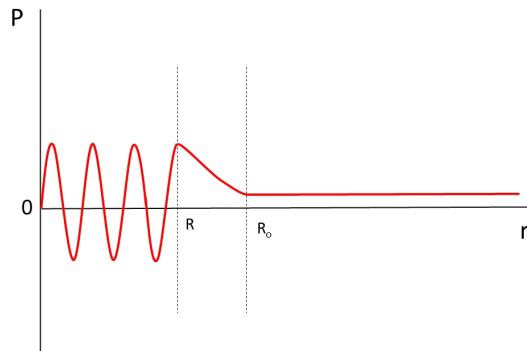
$$\begin{aligned} U_B &= \frac{1}{4\pi\epsilon_o} \frac{q_1 q_2}{r} \\ &= \frac{1}{4\pi\epsilon_o} \frac{2e^2 (Z-2)}{r} \end{aligned}$$

and we know that there is a strong force potential. We don't have a simple mathematical

form for the strong force potential¹⁵. But the combined potential looks something like this



Less than a separation distance (R) the potential is large and negative due to the strong force. At R the potential jumps up and is positive (repulsive) preventing the α particle from leaving. This is an energy barrier that the α particle must overcome to leave the nucleus. Classically it doesn't have enough energy. However, we know about barriers and wave functions. We know that there is a chance of the α particle tunneling because it really is a waveicle. And that is what happens in α decay. We could plot the α wave function, $\psi(r)$ squared (to get the probability) for this set of boundaries. The result would look something like this.



The *probably per unit time*, λ , of actually detecting an α particle outside the atom depends on both the probability of the α particle penetrating the barrier and the number

¹⁵ We won't give the numeric solution in this class, we save that for a graduate level nuclear physics or quantum chromodynamics class.

of times per second that the α particle/wave strikes the barrier.

Beta decay

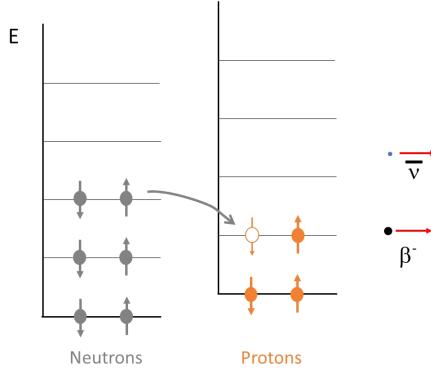
Beta decay can emit either electrons or positrons. We write the electron symbolically as e^- and the positron symbolically as e^+ . The electron or positron is emitted because a neutron decays into a proton plus a beta particle. We can write this as



but this is not what we actually see. If we just had the creation of a β particle all the β particles would have exactly the same energy (like atomic transitions create exactly the same wavelength photon). But that is not what we see. The reason is that in β decay there is a *third* particle created, and that particle will have both energy and momentum. The particle is very small, small enough that there is still debate about its mass. It is called a *neutrino* and it is given the symbol ν (which is not a “v,” it is the Greek letter nu, pronounced “new”). So what we have is the decay process



where $\bar{\nu}$ is actually an antineutrino, the neutrino’s antiparticle (like the positron is the antiparticle for the electron).



Note that all this happens in a nucleus, the mass number will not change, there are the same number of nucleons (protons and neutrons). But the atomic number will change. Chemically, the atom will act differently. Let’s try one

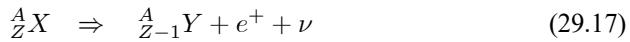
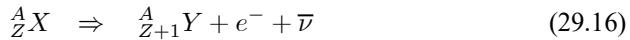


This is the famous carbon-14 that is used in dating things. The mass number did not

change, but the atomic number did. We can run the process backwards



so there are two types of beta decay processes. Generically we can write



with a minimum Q value of

$$Q = \left(m({}_{\bar{Z}}^A X) - m({}_{\bar{Z}+1}^A Y) \right) c^2$$

where we have ignored the neutrino mass because it is very small and the electron is taken care of by using atomic masses. For β^+ decay we would have

$$Q = \left(m({}_{\bar{Z}}^A X) - m({}_{\bar{Z}-1}^A Y) \right) c^2$$

But let's think about this a little further. For a single neutron decay, we get a Q value of

$$Q = (m_n - (m_p + m_e + m_{\bar{\nu}})) c^2$$

where in practice $m_{\bar{\nu}}$ is small enough that it is ignored, but we could have kinetic energy of the products! This kinetic energy must come from our released Q energy. Then for a neutron decay in a nucleus we would get a Q value of

$$Q = E_{\bar{\nu}} + K_e + K_Y$$

where $E_{\bar{\nu}}$ is both the kinetic energy and rest energy of the neutrino and K_e is the kinetic energy of the electron and K_Y is the kinetic energy (recoil) of the nucleus.

Of course momentum must be conserved, so finding the exact values for the kinetic energy is a small amount of work.

Gamma Decay

As you can imagine, if we remove part of the nucleus, or change a neutron into a proton, the potential energy of the nucleus will change. The remaining nucleons won't be in the right positions to be in the new well's ground state. The nucleons will eventually drop down to a lower energy state by emitting a photon. The photon will have a very short wavelength and a very high energy. It will be a γ ray.

Often the mechanism for getting the nucleon into a higher energy state is through beta decay. For example when Boron-12 decays into carbon-12



but the ${}_6^{12} C$ has a nucleon in an excited state. We usually write this as



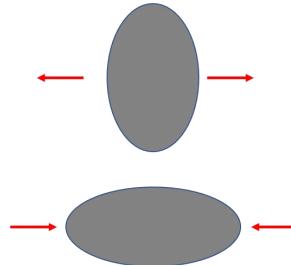
to show that the carbon is excited. Then we have a second reaction



The values of γ energy are more complicated to calculate than we have said so far. For small nuclei we have to take into account the recoil of the nucleus that happens when the γ particle is created.

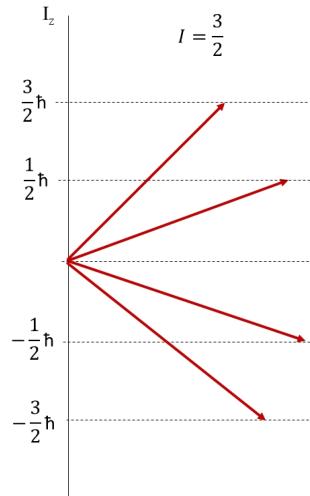
$$E_\gamma = E_i - E_f - K_{\text{recoil}}$$

And there are other ways to excite a nucleus. We could take the whole nucleus and force it to vibrate. The nucleons are only nearest neighbor attracted, so the nucleus is deformable. The nucleus, as a whole can exhibit oscillation modes. Water drops can do this when hit by an acoustic wave. You can imagine a water droplet oscillating between two elongated spheroidal shapes.



Our nuclei can do this as well.

Nuclei can also rotate. And that would mean we have angular momentum, and a possible magnetic moment. Nuclear rotations must be quantized just like electron angular momentum. So it will have $\sqrt{I(I+1)}\hbar$ rotational states where I is a new quantum number. If $I = 3/2$ we would have the following possible orientations for I



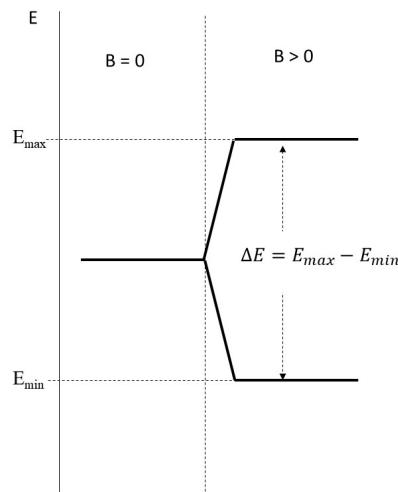
and the magnetic moment would be μ measured in units of the *nuclear magneton*

$$\mu_n = \frac{e\hbar}{2m_p}$$

where m_p is the proton mass. If we put our nucleus in a magnetic field, like all objects with magnetic moments we get a potential energy

$$U_{nuc} = -\mu \cdot \mathbf{B}_{ext}$$

and our nucleus will precess as it spins. The precession frequency is called the *Larmor precessional frequency* and it is given the symbol ω_L . This precession frequency is proportional to the external field. If the nuclear magnetic moment lines up with the field (as closely as quantum mechanics allows) there is a minimum energy, and when it is anti-aligned there is a maximum when it is aligned.



We can use this to image inside of things like people. The person is placed inside a large coil that makes a specially varying magnetic field. The protons in the hydrogen atoms in the body will precess at different frequencies in different parts of the body, and in this way the different parts of the body are spatially coded. Adding in a mechanism to flip the direction of the protons, we can cause a detectable signal. And this is the basis for an MRI machine.

Conservation of energy and momentum and charge

In radioactive decays we need to obey the conservation laws that we have learned over the last few years. Conservation of energy is really what we were talking about when we defined Q values. For

$$\begin{aligned} X &\rightarrow Y + x \\ Q &= (m_X - (m_Y + m_x)) c^2 \end{aligned}$$

Momentum must be conserved. In the rest frame for the parent particle, the momentum of the products would be

$$\mathbf{p}_Y + \mathbf{p}_x = 0$$

We also must have conservation of angular momentum. So the parent angular momentum must equal the product combined angular momentum. An example would be a neutron decay. Neutron decay into just a proton and an electron is forbidden because

$$\begin{aligned} m_{s,p} &= \pm \frac{1}{2} \\ m_{s,e} &= \pm \frac{1}{2} \end{aligned}$$

so the total spin angular momentum would be either 0 or 1, but the neutron has spin $\pm \frac{1}{2}$. So angular momentum would not be conserved in this case. We need a neutrino with spin $\pm \frac{1}{2}$ to make the reaction work.

We also have to conserve charge. We learned this in introductory electrodynamics. But here it is more serious. A proton could capture an electron with the result being a neutron, so far as conservation of charge is concerned. The net charge did not change. But a proton cannot capture an positron to become a neutron, because a neutron has no charge and the parent particles would have charge $2e$.

We also have a new conservation law. It is the law of conservation of nucleon number. In a nucleus we can create π mesons or other particles, but not nucleons. So in nuclear

reactions, the total nucleon number won't change. We can convert neutrons to protons and protons to neutrons, but we can't create new nucleons out of energy (that is a job for particle physics). The proof of this is beyond this course (stay tuned for Nuclear and Particle Physics in your junior or senior year).

Mathematically describing radioactive decay

Let's take a sample of radioactive material that has N radioactive nuclei at $t = 0$. We would like to know how many nuclei decay in a time dt . We will call this number that decay dN . Experiment shows that these quantities are related by

$$\frac{dN}{dt} \propto N \quad (29.21)$$

Notice that this is not an equality, we will do the normal physics thing to make it an equality. We will find a constant that contains the material properties of the particular nuclei buried in it.

$$\frac{dN}{dt} = -\lambda N = -a \quad (29.22)$$

NOTE: this is NOT a wavelength! We are reusing the letter λ with a totally new meaning, so be careful. The quantity a is called the *activity* and it is the number of decays per second.

This λ is called the *decay constant*. We describe $\frac{dN}{dt}$ as a *decay rate*. We sometimes call the decay rate R for "rate" but when we do this we usually define R to be positive.

$$R = \left| \frac{dN}{dt} \right| \quad (29.23)$$

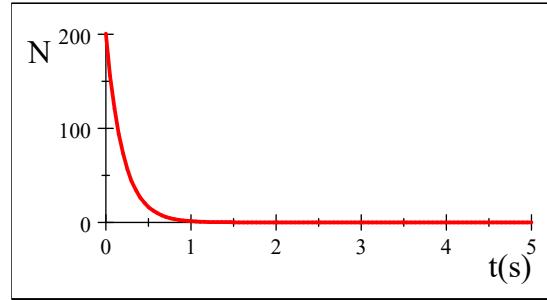
Our decay rate equation is a differential equation in N

$$\frac{dN}{dt} = -\lambda N \quad (29.24)$$

we need a solution, and it looks like we need a function that is very like it's own derivative. Exponentials come to mind.

$$N = N_0 e^{-\lambda t} \quad (29.25)$$

This is an *exponential decay*, much like a wave function tunneling within a barrier. The amplitude is going to die off to zero.



Let's try it, first find the derivative

$$\frac{dN}{dt} = N_o (-\lambda) e^{-\lambda t}$$

so then

$$\frac{dN}{dt} = -\lambda N \quad (29.26)$$

becomes

$$N_o (-\lambda) e^{-\lambda t} = -\lambda (N_o e^{-\lambda t}) \quad (29.27)$$

after substitution which clearly works. The coefficient N_o must be the original number of radioactive atoms.

We could multiply our solution by λ on both sides

$$N = N_o e^{-\lambda t} \quad (29.28)$$

$$\lambda N = \lambda N_o e^{-\lambda t} \quad (29.29)$$

$$-a = -a_o e^{-\lambda t} \quad (29.30)$$

$$a = a_o e^{-\lambda t} \quad (29.31)$$

which gives us an equation for the activity as well.

But we don't usually hear about a material's decay constant. Usually we describe the material by saying how long it takes for half of the nuclei to decay. This is called the *half life* of the material. We often write the half life with the symbol $T_{1/2}$ or τ . I will use $T_{1/2}$. We can give an equation for the half life by considering the time finding N such that n half lives have passed.

$$N = N_o \left(\frac{1}{2}\right)^n \quad (29.32)$$

Then we could find the time it takes for n half lives. Let's choose $n = 1$ then to find the

time for one half life.

$$\begin{aligned} N &= N_o e^{-\lambda t} \\ N_o \left(\frac{1}{2}\right)^n &= N_o e^{-\lambda T_{1/2}} \\ \left(\frac{1}{2}\right)^1 &= e^{-\lambda T_{1/2}} \\ \ln\left(\frac{1}{2}\right) &= -\lambda T_{1/2} \\ T_{1/2} &= -\frac{1}{\lambda} \ln\left(\frac{1}{2}\right) \\ T_{1/2} &= \frac{1}{\lambda} \ln(2) \end{aligned}$$

We did this for $n = 1$, but we could have n be anything. Usually we use the letter n to mean an integer, but it can be any fraction as well. To find n , take

$$n = \frac{t}{T_{1/2}} \quad (29.33)$$

Units of decay

We have some units to describe radioactive decay. Since we are talking about a rate of decay we expect units like other rates we have used. Let's remind ourselves of our first rate we studied, speed.

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

which had units of meters/second. We expect a similar type of unit, something per second. We will take *decays* per second as the basis of our units. Madam Curie and others did not start knowing λ for their studies, so she picked a convenient number of decays per second to use as a unit.

$$1 \text{ Ci} = 3.7 \times 10^{10} \text{ decays/second} \quad (29.34)$$

This may seem like a lot, but some materials decay quite rapidly, and this unit makes sense for them.

The SI unit for decay rate is the Becquerel

$$1 \text{ Bq} = 1 \text{ decay/second} \quad (29.35)$$

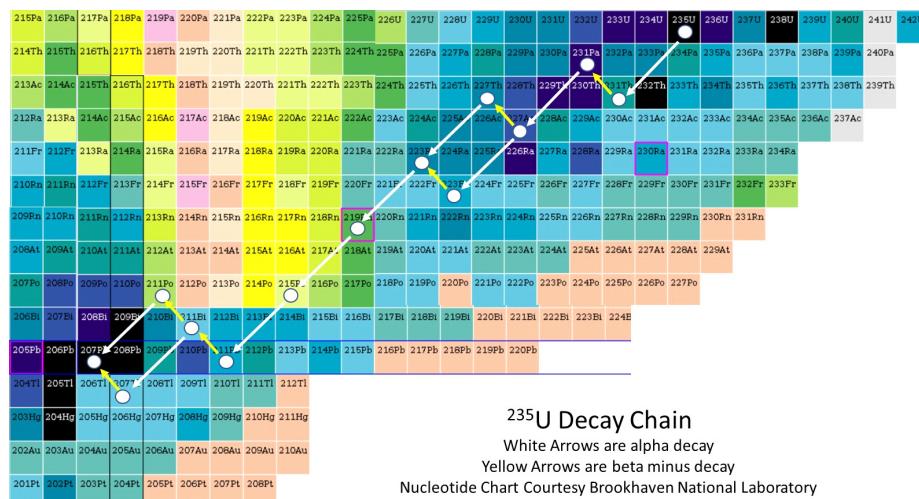
which seems to make more sense, but is not used much in the US.

Natural Reactions

People today often forget that radioactivity was discovered by accident. It is a naturally occurring phenomena. In fact, in Africa's rift valley, the naturally occurring uranium concentration was so high that for many years the whole valley operated as a low level nuclear reactor! This was before Africa had a human population, so there were fewer anti-nuke protests.

When the Earth was formed there were likely many radioactive elements. But only a few radioactive nucleotides have half lives that are longer than the Earth is old. So only a few natural radioactive elements remain. The process of decaying from one radioactive nuclei to another until the process produces a stable nucleus is called a *decay chain*. The decay chains for the few natural radioactive elements that remain are called the Uranium, Actinium and Thorium chains. They start with ^{238}U for the Uranium chain, ^{235}U for the Actinium chain, and ^{237}Th for the Thorium chain.

The Actinium decay chain starting with ^{235}U is shown below superimposed on the chart of the nucleotides.



With this type of plot, we can follow the parent and daughter nuclei through the entire series or chain of reactions until a stable nucleus is produced. You can see that this decay chain is a series of alpha and beta decays. And we can see that alpha decays are a diagonal line downward and to the left that skips one row of nucleotides and betta

decays are diagonal upward to the left one nucleotide.

As interesting as it is to find natural radioactivity, we can benefit from the careful creation of artificial radioactive nucleotides and from artificial radioactive reactions. Let's take on this topic in the next lecture.

30 Artificial Nuclear Reactions

12.10 - 13.2

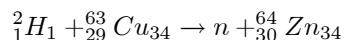
Fundamental Concepts in the Lecture

- Particle capture as a way to induce nuclear reactions
- Cyclotrons
- Van de Graaff accelerators
- Nuclear spectroscopy
- Nuclear Cross Sections

Knowing more about the nucleus, we could understand natural radioactivity. But knowing that nuclear reactions happen, we can force such reactions and learn more about the nucleus by watching what happens. In this lecture we contemplate shooting protons or neutrons at a nucleus to see what happens.

Producing nuclear reactions

We can force a nuclear reaction by attacking the nucleus with another particle. For example we could bombard copper with deuterium nuclei.



We force a collision by accelerating one particle (${}_1^2H_1$) and directing it at a nucleus (${}_{29}^{63}Cu_{34}$) to see what happens. In this case there is a product particle (n) and a product nucleus (${}_{30}^{64}Zn_{34}$). Generically we could write this as

$$x + X \rightarrow y + Y$$

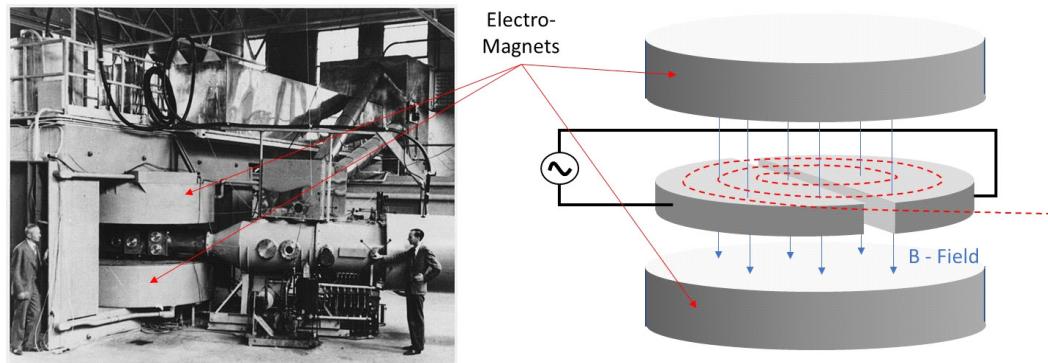
where the capital letters represent nuclei and the small letters represent particles. It is probably good to remember that all of these things are waves. But in this case they approximately act like particles.

You might wonder how we could get a hydrogen nucleus to accelerate. There are many

ways, but two popular devices are the cyclotron and the van de Graaff accelerator.

Cyclotrons

In the next figure you can see an early cyclotron (1939).



Photograph shows the 60-inch cyclotron at the University of California Lawrence Radiation Laboratory, Berkeley, in August, 1939. The metal frame at left is the machine's huge electromagnet. The flat vacuum chamber in which the particles are accelerated is located in the narrow space between the magnet's 60 inch (152 centimeter, 5 foot) pole pieces. The beamline which analyses the particles resulting from the collisions is at right.

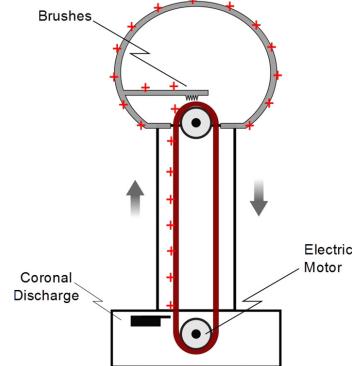
The cyclotron uses powerful magnets to turn charged particles. In the schematic drawing part of the figure, you can see the path of a charged particle. The magnetic fields make it turn because

$$F_B = q\mathbf{v} \times \mathbf{B}$$

The center of the cyclotron has large *D* shaped areas where the particle moves. These *D*s are connected to an alternating voltage source. So at any given time one of the *D*s is positive and one is negative. There is a gap in between the *D*s and because the *D*s are charged there is a potential difference between the two *D*s. So when the charged particle crosses the gap it is accelerated. By the time the charged particle gets to the gap again, the polarity of the *D*s has changed so the particle is accelerated again. As the particle accelerates, it spirals outward until it leaves the cyclotron and can impact a nuclear target.

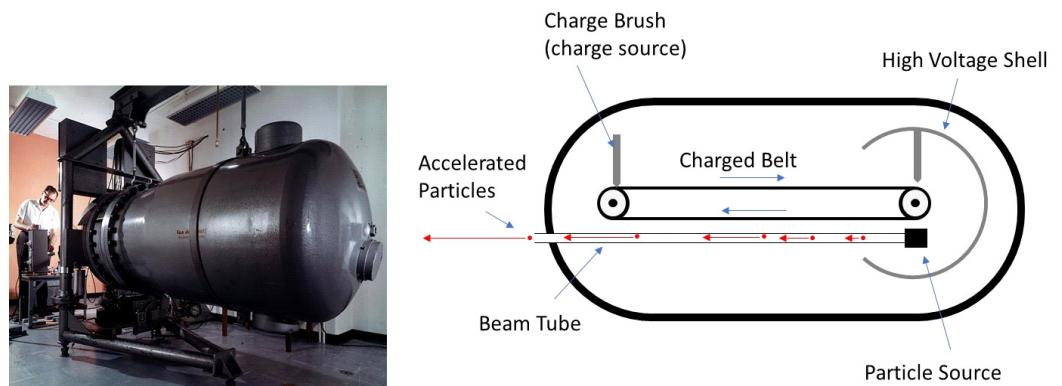
Van de Graaff accelerators

Another method of accelerating charged particles is the van de Graaff accelerator. If you were paying attention back in your introductory electrodynamics class (PH220 at BYUI) you would remember the van de Graaff generator.



Usually a professor uses one of these devices in class as a demonstration. The device has a large metal ball on top of a tube. The tube has a rubber belt in it. A device in the base of the generator sprays charge onto the belt. because the belt is an insulator, the charge is stuck in place as the belt moves upward. At the top of the tube, inside the large metal ball, there are a second set of brushes that collect the charge. The charge is motivated to leave the belt because of Coulomb repulsion. The brush is attached to the metal ball, so the metal ball get's charged. This device can't make large currents, but it can create large electric potentials between the base and the ball (ours easily produces 30000 V). And that large electric potential could accelerate charged particles.

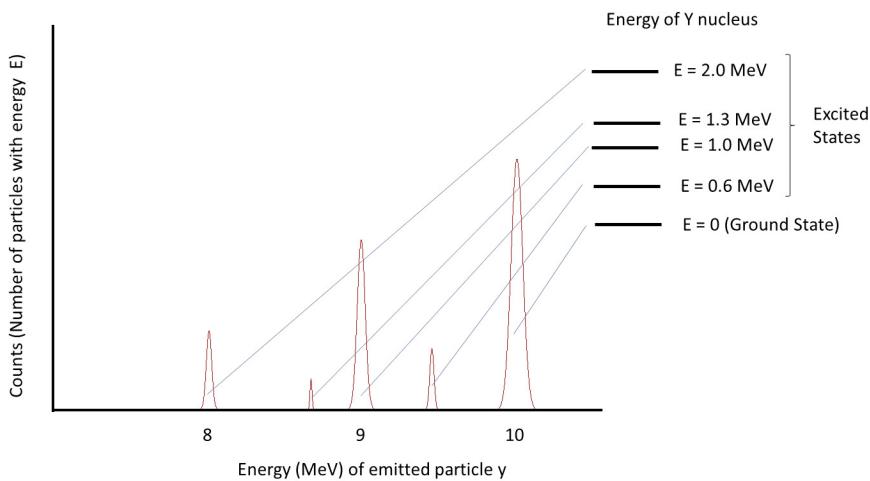
In a van de Graaff accelerator an insulating belt is used in the same way to separate charge.



Van de Graaff Accelerator (Photo courtesy US Department of Energy)

The belt is usually not vertical in this case. But it works the same way. Charge is sprayed on the belt and the belt moves the charge to a large shell on one side of the accelerator. The shell can reach high voltages as the belt transfers more and more charge to it. If a radioactive isotope is placed in the accelerator at the high voltage end, then charged decay products can be accelerated down the beam tube and out toward a target nucleus.

But what would we see from such an experiment? Well, a particle detector might measure the γ (particle) product and its energy. The particle counter could count the number of product particles as a function of energy and form an energy spectrum graph like this



There is a reason that we choose to measure the γ particle. Usually the Y nucleus is still buried in a sample of material. We couldn't get to it without cutting into the sample. And that would take time. The collision takes on the order of 10^{-20} s so we wouldn't have time to observe the Y nucleus until it came to rest in the material lattice. But we can see the γ particles because they escape the sample. Usually we measure two things about γ . The particle energy (that is where the peaks show up on the energy spectrum graph) and the reaction probability. That is how often a particular reaction happens, and that shows up as the height of the peak on the graph.

By making these measurements we can find the ground state and the excited states of the nucleus. Suppose we have x (the incoming particle) with an energy of about 10 MeV. If we measure the energy of our γ particle to be, say 8 MeV, and we assume that the number of nucleons does not change (so the before and after rest mass energies

are the same) then the Y product nucleus must be left in its excited state with about 2 MeV of energy.

$$10 \text{ MeV} = 8 \text{ MeV} + 2 \text{ MeV}$$

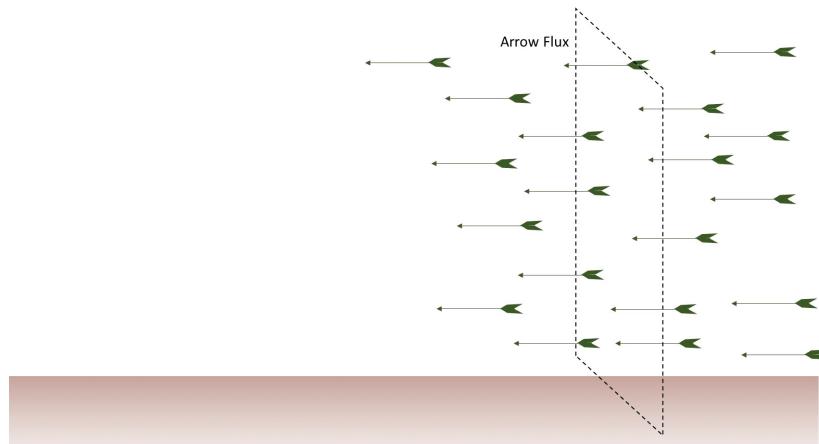
That would be the highest nuclear *energy state* represented on the graph (see the little energy state graph in the upper right hand corner). The other states can be found by looking at the peaks and doing the conservation of energy calculations. And in this way we have learned something about the nuclear energy states even though we can't calculate them. Energy spectra have been collected for many different nuclei under neutron or proton bombardment. Using these standard tables, you can excite a nucleus using a bombardment, and collect its energy spectrum. By matching that spectrum to the tables, you can identify an element in your sample. Though it is often just as good to look at the secondary γ spectrum from your sample because bombardment will leave the Y nuclei in the sample in an excited state.

The height of the peaks is proportional to the probability that a particular reaction will happen. The vertical axes on our graph is the number of times y is found with a particular energy. So if we take the total number of reactions (counts) for a given energy and divide by the total number of reactions we would get the probability of that particular reaction with that particular energy happening. So the height gives us a measure of probability of that particular reaction that is nicely visual.

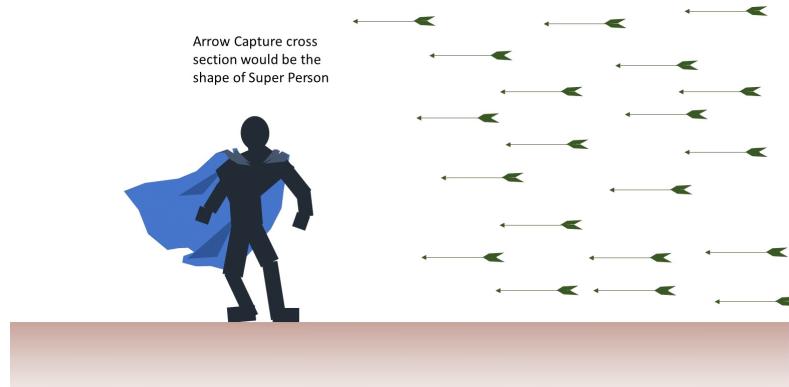
Reaction probabilities are most often expressed in a new term, the cross section

Cross Sections

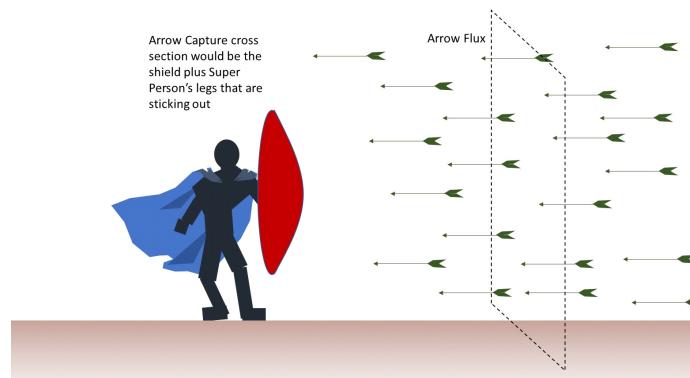
The idea of a cross section is easy. For example, if we took a broom handle and cut it in half we could look at the end of one of the halves and see a circle. That circle has an area and that area is the physical cross section of the broom handle. A cross section is an area. But suppose we had a flux of something, say, arrows from one of our favorite comic book characters, the Green Lawn Dart!



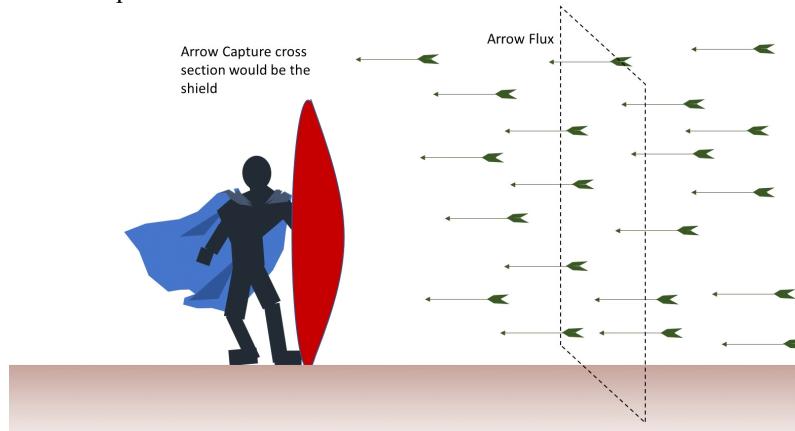
A flux is a number of something that penetrates an area per unit time. This is close to the definition we gave to flux back in introductory electrodynamics (PH220) but there we were dealing with field fluxes (how much electric field penetrated an area) and our fields didn't change in time. But if we are shooting arrows (or protons) the arrows will pass through the area and now it is meaningful to ask how many arrows (or protons) pass through the area per unit of time. We will use the same symbol we used in the past for flux, Φ , but remember that now it is a number per area *per time*. So we have our arrow flux. Suppose that flux is incident on another comic book hero, Super Person.



It would be a bad day for Super Person. Super Person would have arrows on half of Super Person's body. The arrows would hit in an area that would be the size and shape of a Super Person shadow with the light coming from the arrow direction. That area would be Super Person's cross section. But suppose our intrepid hero has a shield.

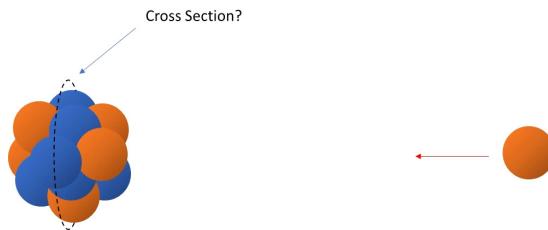


Super Person could increase his cross sectional area with this shield. The arrows would hit in an area that would be the circular cross section of the shield plus the cross sectional area of Super Person's boots. A better shield would include the boot area.



Not that now not only is Super Person safer, but Super Person's cross sectional area is larger than the actual cross sectional area of Super Person's body. And this cross section would give a measure of how likely it is that the Super Person system (Super person plus shield) would interact with an arrow. The larger the area, the bigger the probability of some part of Super Person getting hit (hopefully all on the shield).

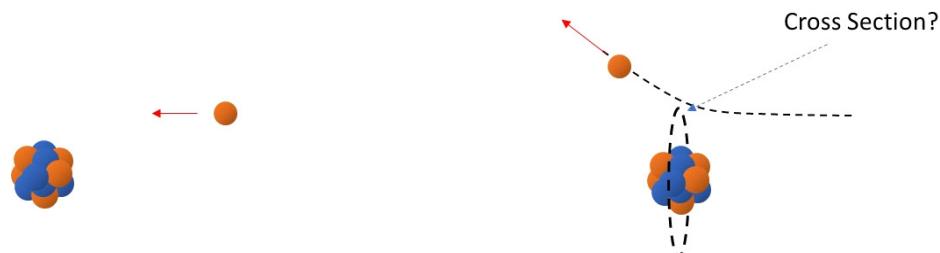
We could take this analogy and apply it to nuclear bombardment. Suppose we take a proton (x) and send it flying into a nucleus (X) what would be the cross section for nucleus X ?



We could guess that it should be something like the physical cross section of the nucleus. But let's adjust our experiment so the proton doesn't come centered on the nucleus. We have done the centered problem and we know the proton bounces back.



But if it comes in off-center something like this might happen.

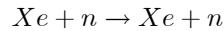


It looks like the X nucleus has had an effect that goes beyond its physical dimensions (whatever that means for a wave based thing). And that is true. The protons in the nucleus will have an effect on the incoming x proton that extends beyond the physical dimension of the nucleus due to Coulomb force. So the nuclear cross section might be larger than the actual nucleus. Remember that we want a probability of an interaction. And in our off-center case the x and X did interact. The proton didn't just pass by unaffected.

Nuclear cross sections should be on the order of the area of the nucleus, which is an area of about 10^{-28} m^2 . This is typical for a midrange atom. And nuclear physicists gave this area a name, the *barn*

$$1b = 10^{-28} \text{ m}^2$$

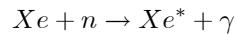
and most of the time cross sections are about this size. But sometimes something changes, and that tells us much about the nucleus. For example



is an inelastic scattering with a cross section

$$\sigma = 4b$$

but if the neutron is captured, then

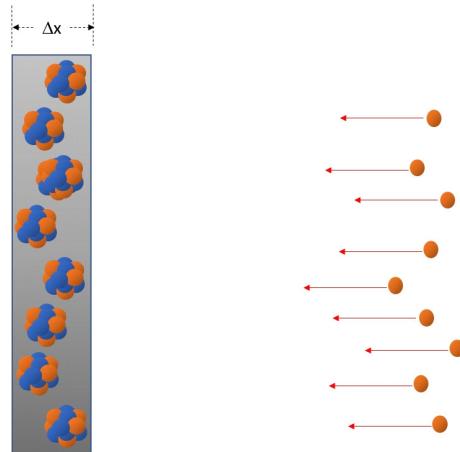


and the cross section for this interaction is

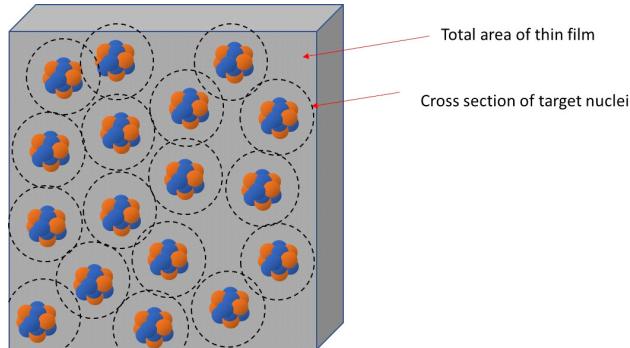
$$\sigma = 10^6 b$$

which says that neutron is much more likely to be captured by the Xe atom for some reason (like Super Person's shield might be more likely to "capture" arrows than to deflect them). And by measuring the two cross sections, we have learned something about Xe nuclei.

Like Rutherford, modern physicists like to do bombardment experiments with thin film targets.



This makes the math easier because if we look at the target sideways we see that to a fair approximation the probability of hitting a target nuclei is like $N\sigma$ where N is the number of target nuclei in the film.



It would be much harder if the target nuclei cross sections overlapped. If the total thin film area is S , then for this condition the reaction probability would be proportional to the number of incoming particles, I_o and the ratio of the number of particles times a cross section to the total area of the film $N\sigma/S$ or

$$R = \frac{\sigma N}{S} I_o$$

This accounts for the parts of the film that don't have a target nuclei in them (the blank spots in our last figure).

We could write I_o as a flux

$$I_o = \Phi S$$

then

$$R = \Phi \sigma N$$

and we are used to finding the number of atoms in a pure sample

$$N = \frac{m N_A}{M}$$

where m is the nuclear mass of the target and M is the molar mass. Then for a pure sample

$$R = \Phi \sigma \frac{m N_A}{M}$$

It is worth remembering that flux has units of particles per area *per second* and cross section has units of area per nucleus per incident particle. So $\Phi \sigma$ has units

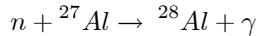
$$\frac{\text{particles}}{\text{At}} \frac{A}{\text{nucleus} \times \text{particle}} = \frac{1}{t \times \text{nucleus}}$$

and $\phi \sigma N$ then would be just inverse time. That is, reactions per time.

Let's take an example.

Suppose we have a sample ($m = 8.1 \times 10^{-5} \text{ kg}$) of aluminium foil (^{28}Al) that is rolled thin enough to assume it is about one atom thick (this would be hard to do!). And

suppose the ^{28}Al nuclei undergo neutron capture



with a measured cross section of

$$\sigma = 2.0 \times 10^{-31} b = 2.0 \times 10^{-31} \text{ m}^2$$

and the incoming neutron flux is

$$\Phi = 5.0 \times 10^{12} \frac{\text{neutrons}}{\text{cm}^2 \text{s}}$$

Find the reaction rate R .

We know that

$$R = \Phi \sigma N = \Phi \sigma \frac{m N_A}{M}$$

so we will need the molar mass of Al

$$M = 27 \frac{\text{g}}{\text{mol}}$$

then

$$\begin{aligned} R &= \Phi \sigma \frac{m N_A}{M} \\ &= \left(5.0 \times 10^{12} \frac{1}{\text{cm}^2 \text{s}} \right) (2.0 \times 10^{-31} \text{ m}^2) \frac{(8.1 \times 10^{-5} \text{ kg}) (6.02 \times 10^{23} \frac{1}{\text{mol}})}{27 \frac{\text{g}}{\text{mol}}} \\ &= 1.806 \times 10^7 \frac{1}{\text{s}} \end{aligned}$$

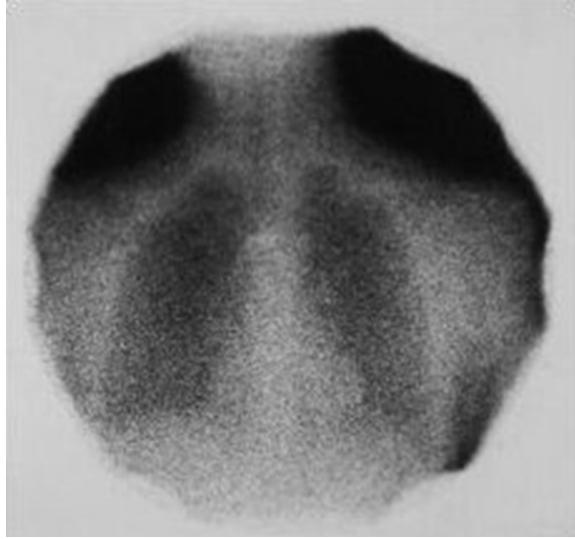
or about 1.8×10^7 reactions per second.

You might object saying that we can't always prepare a sample to be a thin foil, and that is true. If the target atom is only one component of the sample and only a small component, we can still use our fairly simple formula because the target nuclei won't overlap as we look through the sample. But if the target nuclei will overlap, then the geometry considerations are more difficult and it is time for a solid state physics or junior level nuclear physics class.

Radioisotope production

We use short half life radioactive materials in medicine. An example would be as a tracer that allows doctors to image a patient by making some part of their body radioactive. Technetium-99m is ^{99}Tc but in an excited state that is metastable, $^{99}Tc^*$. It has a half life of about 6 hours. So if it is inserted into a part of a patient's body, that part will emit gamma rays that can be detected with a gamma ray camera. If the tracer is placed in the blood stream, we could follow the blood trough the patient's body with the gamma camera. We wouldn't want the person to have radioactivity going on in

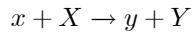
them for a long time, so the 6 hour half life is a good thing (the decay products may be radioactive, but with long half lives and lower energy β^- emission).



“A PA transmission scan of my chest acquired using an uncollimated gamma camera and an 18.5 MBq point source of ^{99}mTc at a focal distance of two metres, on 18 Jan 1984 during the installation of a GE gamma camera at the Meath Hospital, Dublin. Generated by Kieran Maher by scanning an original transparency film. marz 04:32, 25 October 2006 (UTC).” This is not the most detailed gamma image from ^{99}Tc but it is one that Kieran Maher kindly let me use)

But how can we have a radioactive element with a half life of only 6 hours when the Earth is so much older? Clearly this cannot be an element left over from the creation of the Earth. And this is true, no short half life radioactive isotopes can exist without more modern creation. We build such radioactive isotopes through artificial nuclear reactions.

Suppose we want to create a product isotope Y with a short half life. We have the reaction



but the half life of Y is short enough that there might be significant decay during production while x is bombarding X .

Let R be the constant rate at which we produce Y by x bombardment. We expect

$$R = \Phi\sigma N$$

in a time dt we would expect to produce Rdt new Y nuclei. The number of Y nuclei that decay in dt is $\lambda N dt$. So the change in the number of Y nuclei has two parts, the part where we are making new Y nuclei and the part where Y nuclei decay

$$dN = Rdt - \lambda N dt$$

which we can write as

$$\frac{dN}{dt} = R - \lambda N$$

and now all we have to do is to guess an answer. We remember that with no production we would have $N = N_o e^{-\lambda t}$. So we expect to have a term in $e^{-\lambda t}$ but we need something else. Let's try

$$N = \frac{R}{\lambda} (1 - e^{-\lambda t})$$

The derivative is

$$\frac{dN}{dt} = 0 + \frac{R}{\lambda} e^{-\lambda t} (-\lambda) = Re^{-t\lambda}$$

then

$$\frac{dN}{dt} = R - \lambda N$$

becomes

$$\begin{aligned} Re^{-t\lambda} &= R - \lambda \frac{R}{\lambda} (1 - e^{-\lambda t}) \\ Re^{-t\lambda} &= R - R (1 - e^{-\lambda t}) \\ Re^{-t\lambda} &= R - (R - Re^{-\lambda t}) \\ Re^{-t\lambda} &= Re^{-\lambda t} \end{aligned}$$

which is true. So this guessed solution works.

The activity would be λN

$$a = \lambda N = R (1 - e^{-\lambda t})$$

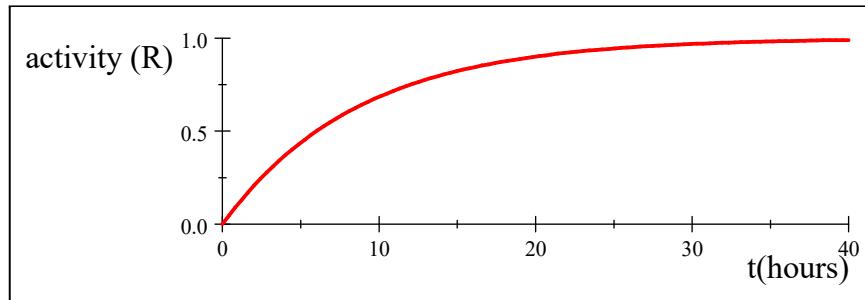
We should look at this before we try an example. When $t = 0$ we have

$$a = R (1 - e^{-\lambda(0)}) = 0$$

no activity since we have not produced any radioactive Y nuclei yet. If we take $t \gg T_{1/2}$ we would have

$$a = R (1 - e^{-\lambda(\infty)}) = R$$

And we have reached an equilibrium. We have an activity that is equal to the reaction rate. For an isotope with a 6 hour half life the graph looks like



Let's try an example of isotope production. Thirty milligrams of gold are exposed to a neutron flux of 3.0×10^{12} neutrons/(cm² s) for 60 s. The neutron capture cross section of gold is 99b. Find the activity of ¹⁹⁸Au.

We have to use the tables from the back of our book or the NIST website and Brookhaven National Laboratory here. We know

$$\sigma = 99b = 99 \times 10^{-28} \text{ m}^2$$

$$\Phi = 3.0 \times 10^{12} \text{ neutrons/}(\text{cm}^2 \text{ s})$$

$$\Delta t = 60 \text{ s}$$

$$m = 30 \text{ mg}$$

We look up

$$A = 197$$

$$T_{1/2} = 2.70 \text{ d} = 2.3328 \times 10^5 \text{ s}$$

$$M = 197 \frac{\text{g}}{\text{mol}}$$

We also know

$$R = \Phi \sigma N = \Phi \sigma \frac{m N_A}{M}$$

and for the case of production

$$a = R (1 - e^{-\lambda t})$$

but we realize we have a problem. Our Δt isn't much larger than $T_{1/2}$ nor is it zero. In fact, $\Delta t \ll T_{1/2}$ but not zero. We didn't look at that case. We can get a good approximation under these circumstances if we approximate $e^{-\lambda t}$.

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

then

$$e^{-\lambda t} = 1 - \lambda t + \frac{(-\lambda t)^2}{2!} + \frac{(-\lambda t)^3}{3!} + \dots$$

and since

$$\lambda = \frac{\ln 2}{T_{1/2}}$$

$$e^{-\lambda t} = e^{-\frac{\ln 2}{T_{1/2}} t} = 1 - \frac{\ln 2}{T_{1/2}} t + \frac{\left(\frac{-\ln 2}{T_{1/2}} t\right)^2}{2!} + \frac{\left(\frac{-\ln 2}{T_{1/2}} t\right)^3}{3!} + \dots$$

and we know that

$$\frac{t}{T_{1/2}}$$

will be a very small number so

$$\left(\frac{t}{T_{1/2}}\right)^2$$

will be even smaller and

$$\left(\frac{t}{T_{1/2}}\right)^3$$

is frighteningly smaller. We could approximate

$$e^{-\lambda t} \approx 1 - \lambda t$$

without too much loss of accuracy. Then

$$a = R(1 - e^{-\lambda t}) \approx R(1 - 1 - \lambda t) = R\lambda t$$

so our activity for very small Δt values is linear in t . We need R and λ to finish the problem

$$R = \Phi \sigma \frac{mN_A}{M}$$

$$R = \left(3.0 \times 10^{12} \frac{1}{\text{cm}^2 \text{s}}\right) (99 \times 10^{-28} \text{ m}^2) \frac{(30 \text{ mg})(6.02 \times 10^{23} \frac{1}{\text{mol}})}{197 \frac{\text{g}}{\text{mol}}}$$

$$= \frac{2.7228 \times 10^{10}}{\text{s}}$$

and

$$\lambda = \frac{\ln 2}{T_{1/2}}$$

$$\lambda = \frac{\ln 2}{2.3328 \times 10^5 \text{ s}} = \frac{2.9713 \times 10^{-6}}{\text{s}}$$

then if we define $t_i = 0$ so $\Delta t = t$

$$a = R\lambda t$$

$$= \left(\frac{2.7228 \times 10^{10}}{\text{s}}\right) \left(\frac{2.9713 \times 10^{-6}}{\text{s}}\right) (60 \text{ s})$$

$$= 4.8542 \times 10^6 \frac{1}{\text{s}}$$

$$= 4.8542 \times 10^6 \text{ Bq}$$

506 Chapter 30 Artificial Nuclear Reactions

This is cool and useful. Medical physicists do this kind of work in major hospitals. We have several recent graduates that have followed this career path and love their jobs. Perhaps this is something to consider.

But we haven't gotten to energy production yet. Let's take that up in our next lecture.

31 Nuclear Energy Reactors

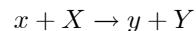
13.3-13.5

Fundamental Concepts in the Lecture

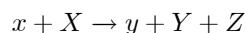
- Nuclear deformations and oscillations
- Slow nuclear reactions
- Threshold kinetic energies

Nuclear Fission

So far we have studied nuclear reactions where a particle comes in and affects a nucleus, and then a particle and a new nucleus comes out.

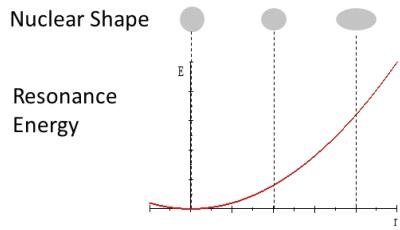


But there is also the possibility that when x hits X the X nucleus becomes unstable in a bigger way. Then we might have

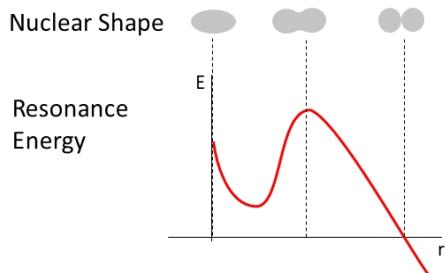


where both Y and Z are smaller nuclei. How can this happen?

Because the strong force is a nearest-neighbor type force, we can send nuclei into an oscillatory resonance. A fairly spherical nucleus will elongate in resonance. We could roughly model this like a harmonic oscillation. The energy for such an oscillator should be proportional to the separation distance squared. So as we stretch the nucleon mass, the energy should increase (because we have done work on the nucleus).

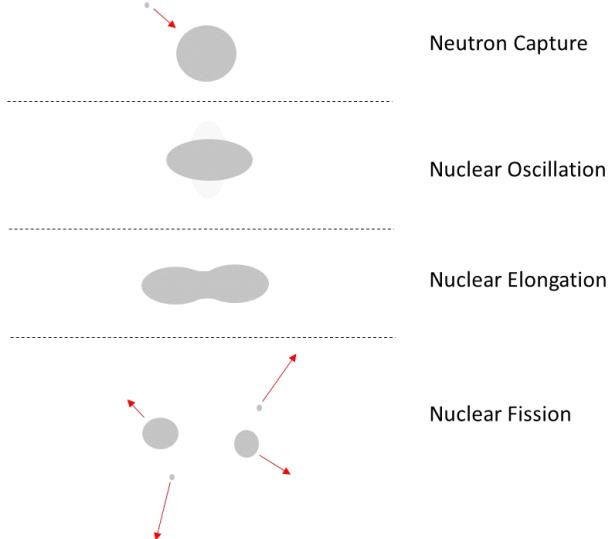


But not all nuclei are spherical. Some are elongated to begin with. Those nuclei experience a greater effect from oscillation because the nuclear force is weaker because along one axis the nucleons are farther apart.



So we still get a dip with a minimum for oscillation, but at some distance the potential energy begins to diminish as the nucleus flattens and then gets thinner in the middle. Fewer and fewer nucleons are participating in strong force reactions in that middle section. At some distance, the parent nucleus breaks apart and the two sides form new smaller nuclei.

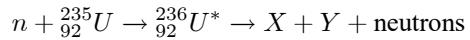
The overall process might look something like this



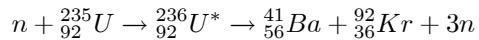
A neutron strikes a large nucleus. The nucleus captures the neutron, but this causes waves in the nucleons. The oscillating nucleus can elongate and reduce the strong force in the middle leading to the splitting of the nucleus into two smaller nuclei. This is called *nuclear fission*.

Slow nuclear reactions

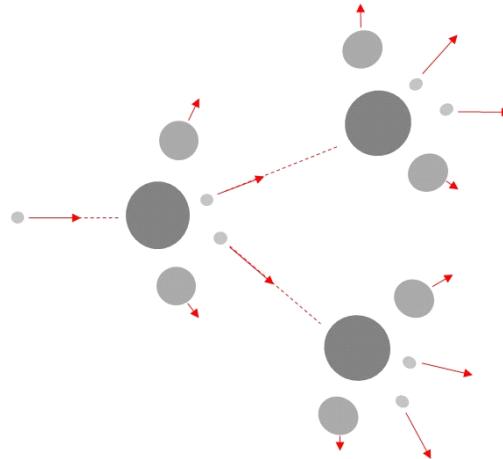
The nuclear fission process isn't as exact and simple as the reactions we have studied up till now. One large nucleus might split into many different combinations of two product nuclei with a number of free neutrons produced as well. We could write this as



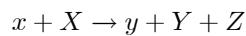
The new nuclei, X and Y are called fission fragments. One possibility is



If the reaction creates more neutrons, these neutrons can go on to cause additional reactions.



A chain reaction can form. The fission fragments have positive charge. So they repel each other. So the reaction will result in kinetic energy of the fission fragments and some kinetic energy of the neutrons. We can find out how much energy is released by finding the Q value of the reaction. Symbolically for the reaction



we would have

$$\begin{aligned} Q &= (m_i - m_f) c^2 \\ &= (m_x + m_X - m_y - m_Y - m_Z) c^2 \end{aligned}$$

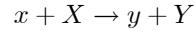
The kinetic energy of the products share the Q energy

$$Q = K_y + K_Y + K_Z - K_x$$

where I have assumed the original nucleus, X , was not moving in our lab frame.

It might be better to do this in the center of mass frame. In that frame it looks like both x and X are moving together and in the limiting case the y and Y would seem to stay at the collision location after the reaction. This would be the minimum case where we have enough energy to just make the reaction happen, but no extra for kinetic energy of the products y and Y . We call this a *threshold* reaction. If you are a nuclear physicist and want this reaction to happen, you must provide enough initial kinetic energy to make it happen (K_{th}) and the way to do this is to shoot in the x particle in the lab frame with the necessary $K_x = K_{th}$. Let's find K_{th} for such a reaction.

Let's assume that our fission products don't move at relativistic speeds. and let's take the simple reaction



Then, thinking of our incoming x and our stationary X . The center of mass would move with speed

$$v = \frac{m_x v_x + m_X v_X}{m_x + m_X} = \frac{m_x v_x}{m_x + m_X}$$

The we would have a center of mass frame speed for our threshold case (where y and Y are just produced, but not moving)

$$\begin{aligned} v_{x,cm} &= v_x - v \\ v_{X,cm} &= -v \\ v_{y,cm} &= v_{Y,cm} = 0 \end{aligned}$$

And we can write conservation of energy (that includes rest energy)

$$\begin{aligned} E_i &= E_f \\ \frac{1}{2} m_x (v_x - v)^2 + \frac{1}{2} m_X (-v)^2 + m_x c^2 + m_X c^2 &= m_y c^2 + m_Y c^2 \\ \frac{1}{2} m_x \left(v_x - \frac{m_x v_x}{m_x + m_X} \right)^2 + \frac{1}{2} m_X \left(\frac{m_x v_x}{m_x + m_X} \right)^2 &= m_y c^2 + m_Y c^2 - (m_x c^2 + m_X c^2) \end{aligned}$$

We recognize the right hand side as our Q value for the reaction

$$\frac{1}{2} m_x v_x^2 \left(1 - \frac{m_x}{m_x + m_X} \right)^2 + \frac{1}{2} m_X v_x^2 \left(\frac{m_x}{m_x + m_X} \right)^2 = -Q$$

$$\begin{aligned}
& \frac{1}{2}v_x^2 \left(m_x \left(1 - \frac{m_x}{m_x + m_X} \right)^2 + m_X \left(\frac{m_x}{m_x + m_X} \right)^2 \right) = -Q \\
& \frac{1}{2}v_x^2 \left(m_x \left(\frac{m_x + m_X}{m_x + m_X} - \frac{m_x}{m_x + m_X} \right)^2 + m_X \left(\frac{m_x}{m_x + m_X} \right)^2 \right) = -Q \\
& \frac{1}{2}v_x^2 \left(m_x \left(\frac{m_X}{m_x + m_X} \right)^2 + m_X \left(\frac{m_x}{m_x + m_X} \right)^2 \right) = -Q \\
& \frac{1}{2}v_x^2 \left(m_x \frac{m_X^2}{(m_x + m_X)^2} + m_X \frac{m_x^2}{(m_x + m_X)^2} \right) = -Q \\
& \frac{1}{2}v_x^2 \left(\frac{m_x m_X^2}{(m_x + m_X)^2} + \frac{m_X m_x^2}{(m_x + m_X)^2} \right) = -Q \\
& \frac{1}{2}v_x^2 \left(\frac{m_x m_X^2 + m_X m_x^2}{(m_x + m_X)^2} \right) = -Q \\
& \frac{1}{2}v_x^2 \left(\frac{m_x m_X (m_X + m_x)}{(m_x + m_X)^2} \right) = -Q \\
& \frac{1}{2}v_x^2 \left(\frac{m_x m_X}{(m_x + m_X)} \right) = -Q
\end{aligned}$$

Now let's take out m_x

$$\frac{1}{2}m_x v_x^2 \left(\frac{m_X}{(m_x + m_X)} \right) = -Q$$

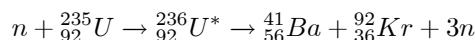
and identify $K_{th} = \frac{1}{2}m_x v_x^2$ so

$$K_{th} \left(\frac{1}{\left(\frac{m_x}{m_X} + 1 \right)} \right) = -Q$$

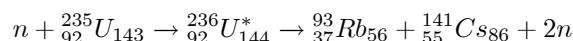
and finally

$$K_{th} = -Q \left(1 + \frac{m_x}{m_X} \right)$$

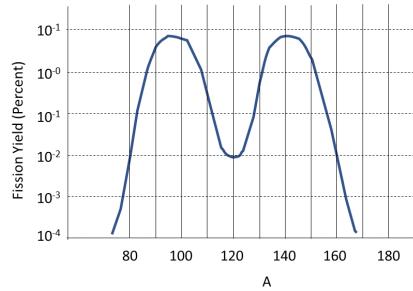
So we need at least K_{th} to make the nuclear reaction happen. In our $^{235}_{92}U$ example, the n in the left hand side of



would carry this kinetic energy. But this is not the only possible set of products Another possible uranium 235 reaction would be.

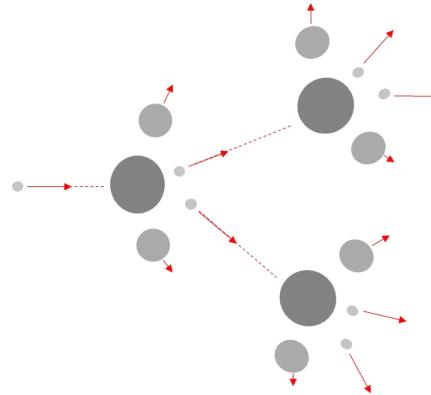


The percentage of fission fragments by atomic mass number is given in the next figure.



The distribution is bimodal and that makes sense. We expect to see the $^{235}\text{U}_{143}$ split into two nuclei. But notice the broad peaks. We can get many many different fission reactions from the same input particle and nucleus.

Also notice that the reactions produce extra neutrons. These come from the product nuclei being ripped from the parent. These product neutrons can hit other target nucleus and cause other nuclear reactions.

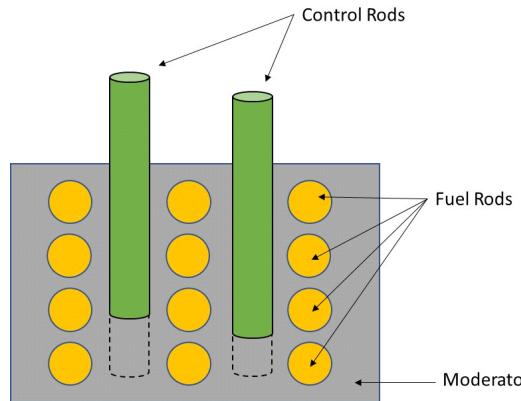


If we start with $K > K_{th}$ we get kinetic energy of the products, and the kinetic energy of the products becomes thermal energy of the material and this type of chain reaction can get out of control quickly. The result would generally be bad, from a melting of the radioactive material to a nuclear explosion like in nuclear bombs.

Control rods

To control such a chain reaction we can introduce a material that captures neutrons but does not produce a nuclear chain reaction. Cadmium has a large neutron absorption cross section and is often used in reactors. In one design for a nuclear reactor the target nuclei, say, ^{235}U is placed in rods in the reactor. These *fuel rods* contain the ^{235}U and

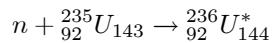
cladding to keep the radioactive material contained.¹⁶



Rods of control material are included to slow down the production of new neutrons so the reactor doesn't overheat and melt. The control rods are designed to be raised or lowered to allow more or less neutron production so they can slow down or speed up the nuclear reaction.

Enrichment

To make a safe reactor, we want a nuclear chain reaction that produces one new reaction for every reaction. If the chain produces even a fraction more than a one for one chain, the reaction will go out of control. If it produces even a little less than one for one, then the reaction will die out. Both are bad for energy production (but one is far worse than the other!). The number of reactions that result from the products of the previous reaction is called the *reproduction constant*. To understand how to control the reaction we need to look at the specific reaction we are to use. So suppose it is

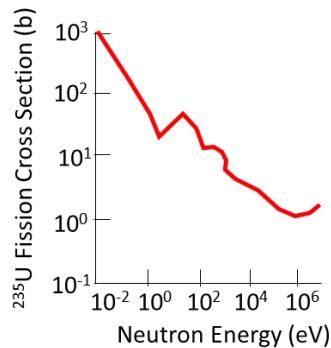


Natural uranium only has about 0.7% ${}_{92}^{235}\text{U}_{143}$ and the rest ${}_{92}^{238}\text{U}_{146}$ and that ${}_{92}^{238}\text{U}_{146}$ won't participate in our chain reaction. This means the neutrons that should keep our reaction going will likely be captured by ${}_{92}^{238}\text{U}_{146}$ and the reaction will die out. One solution is to artificially increase the percentage of ${}_{92}^{235}\text{U}_{143}$ to somewhere in between 3% to 5%. Then there is enough ${}_{92}^{235}\text{U}_{143}$ to sustain a reaction. This process of increasing the amount of fissile material (${}^{235}\text{U}$ in our case) is known as *enrichment*. A richer fuel mix would need less control. But there is another complicating factor.

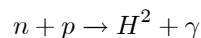
¹⁶ INL and our own Grain Boundary Group study the physics of nuclear fuel rod design.

Moderators

We can enrich less if we can make it more likely that the neutrons will be captured. We can look at the neutron capture cross section. We find that it is not constant, but is different for different initial kinetic energies of the incoming neutrons.



Notice that the higher the neutron energy the *lower* the cross section. We could get more reactions if we slowed down the neutrons. And we know how to do that, we let them collide with something that won't absorb them. The collision will transfer some of the neutron's energy, slowing them down. Water is a good choice because the neutrons tend to strike the hydrogen atoms in the water molecules. But normal hydrogen can also capture the neutrons



It is more efficient if we use water made with deuterium instead of hydrogen. Then the cross section for neutron capture is very low. Some reactor designs use more enrichment, some choose deuterium based "heavy water."

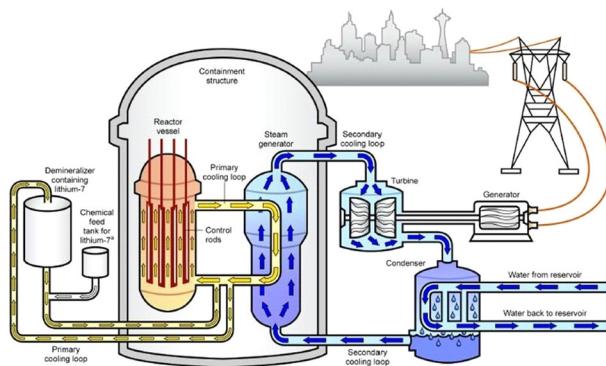
Natural control through delayed reactions

Some decay products have longer half lives than others. This is a benefit to reactor design. ^{93}Rb is one such fission product from ^{235}U . It has a half life of 6 seconds. This isn't the most probable decay route, but remember we need our reaction to produce exactly one more reaction to keep the chain going. The control rods in a reactor can be adjusted, but not infinitely quickly. If we plan for quick reactions to keep our reactor at just under one reaction per reaction, then the slower but less probable reactions can make up the difference. And we can move control rods within a 6 second window. This

makes fine control of a uranium reactor possible.

Power reactors

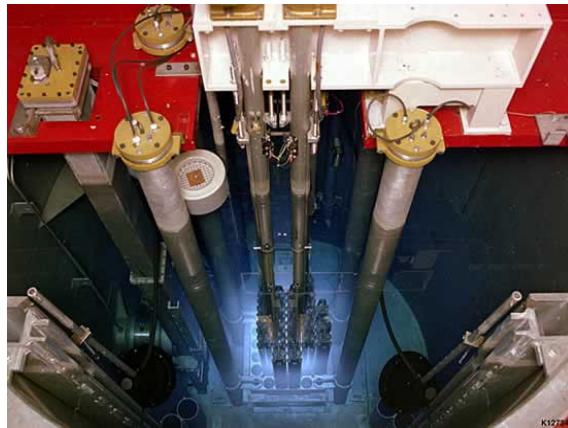
The next figure is a schematic of a power generating reactor.



Pressurized Water Reactor (Image courtesy the US Government Accountability Office)

This is just one of many designs. This design is called a *pressurized water reactor* because it uses pressurized water to transfer thermal energy. The pressurized water (yellow arrows in the figure) is heated by the reactor. That pressurized water in turn heats non-pressurized water which is turned into steam and drives turbines to generate electrical power. The water that passes through the reactor might be activated by the neutrons from the nuclear reactions. The pressurized water reactor ensures no radioactive water is involved with the steam turbines so the process is much safer.

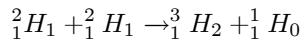
The next figure shows a reactor at INL. This is a research reactor, not designed for energy production. We can see many of the same design items as in the energy producing reactors. There are fuel and control rods (both vertical in this reactor) and water as a moderator. Often this kind of reactor also used carbon rods as a moderator.



Fusion

Ages ago (a few lectures ago) we said iron was the most stable element and we could get energy from the nucleus in two ways. We could take large nuclei and split them, allowing the nucleons in the product nuclei to achieve lower energy states. Or we could combine smaller nuclei, once again allowing the nucleons in the combined nucleus to have a lower energy state. Both methods release energy. We now know the first method is called Fission. The second is called fusion.

An example reaction would be



We take two deuterium nuclei and combine to get tritium and a normal hydrogen nuclei (a proton). If we calculated the Q value for this we would get about 4 MeV which doesn't sound like much but if we compare energy per nucleon it is about the same as fission.

The trick in doing this is to squeeze the deuterium nuclei close enough that the strong force can act. But of course the deuterium nuclei repel each other because of the Coulomb force. The deuterium radius is about

$$\begin{aligned} r &= r_o A^{\frac{1}{3}} = 1.2 \text{ fm} (2)^{\frac{1}{3}} \\ &= 1.5119 \times 10^{-15} \text{ m} \end{aligned}$$

The electrostatic potential energy for two deuterium nuclei that just touch ($r = 3 \text{ fm}$) is

$$\begin{aligned} U(3 \text{ fm}) &= \frac{1}{4\pi (8.854187817 \times 10^{-12} \text{ F m}^{-1})} \frac{(1.60217733 \times 10^{-19} \text{ C})^2}{3 \times 10^{-15} \text{ m}} \\ &= 4.7999 \times 10^5 \\ &= 0.5 \text{ MeV} \end{aligned}$$

so if we supply about 0.5 MeV to one of the deuterium nuclei, we should have a reaction in which about $4 \text{ MeV} + 0.5 \text{ MeV} = 4.5 \text{ MeV}$ of energy is released. Of course, you have to use energy to accelerate the deuterium nucleus, and you must keep this reaction going to be practical. Particle accelerators are not good approaches to this.

A different approach would be to heat deuterium gas until each nucleus has 0.25 MeV of energy. Then the collisions between the deuterium atoms would produce fusion. But to do this for a large number of atoms requires temperatures on the order of 10^9 K . If you are not doing this in a star, this is difficult to achieve.

Fusion reactors

Another approach is to switch to a different reaction. There are several promising fusion reactions that might produce energy. Some of these are

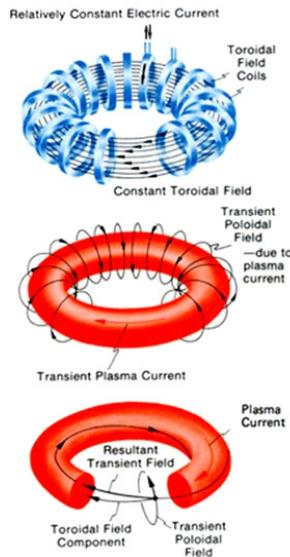
Reaction	Q (MeV)
${}^2H + {}^2H \rightarrow {}^3H + p$	4.0
${}^2H + {}^2H \rightarrow {}^3He + n$	3.3
${}^2H + {}^3H \rightarrow {}^3He + n$	17.6

The last seems most promising because it yields the largest release of energy. It is called the D-T reaction for deuterium and tritium.

There are two promising designs for fusion reactors. One method uses high temperatures (only on the order of 10^8 K) combined with high density to push the nuclei together. The other uses powerful lasers to force the nuclei together. In both cases, the trick is to not spend too much energy causing the reaction so that there is more energy left over after the reaction.

The first kind of fusion reactor needs to hold the nuclei at a high temperature and in close proximity. And the nuclei need to be kept in those conditions for a long time. The temperatures are too high to use a physical bottle. So one method of providing long confinement times is to use a magnetic field. The standard configuration is a toroid. And the kind of reactor that uses such a magnetic field confinement is called a Tokamak

(Russian acronym for “toroidal magnetic chamber”).

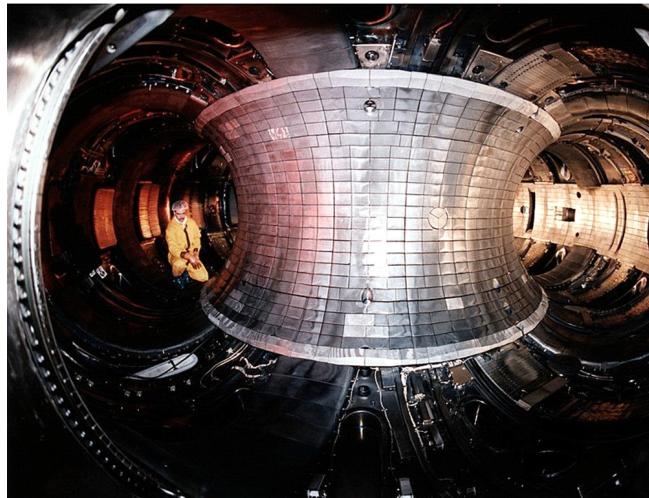


Tokamak magnetic fields, image courtesy United States Department of Energy

To achieve a net energy, the reactor needs to heat the gas to a plasma with an amount of power proportional to the number density, n , of atoms. The power from the reaction is proportional to $n^2\tau$ where τ is the confinement time. To achieve net energy, the product

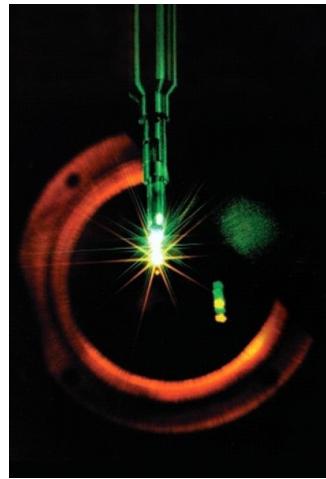
$$n\tau > 10^{20} \text{ s m}^{-3}$$

which is known as Lawson's criterion. The International Thermonuclear Experimental reactor (ITER) in France is projected to surpass the break even point where net energy is produced. A research group at China's Experimental Advanced Superconducting Tokamak (EAST) announced last year that they had achieved $1.6 \times 10^7 \text{ K}$ which is a new record. They are also close to sustained fusion.

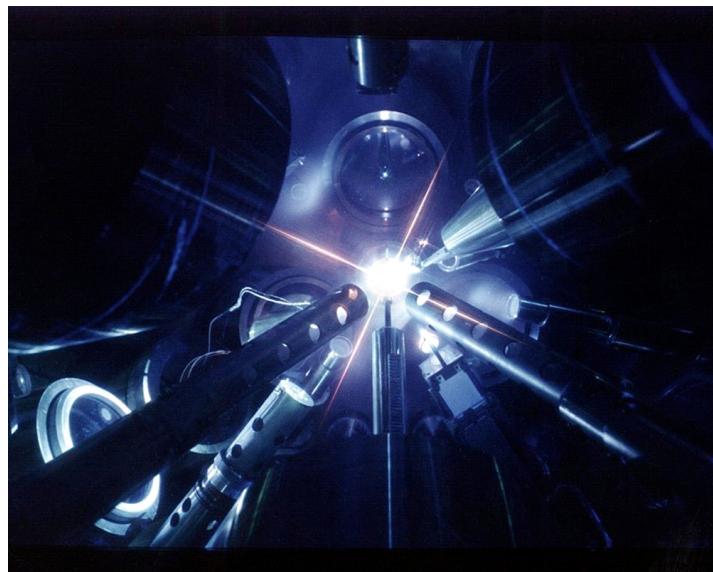


Tokamak Reactor Interior at Princeton Plasma Physics Labroarory (US Department of Energy)

The second of our methods for producing fusion is called *Inertial confinement*. The idea is to fulfill the Lawson criterion by making n large so τ does not have to be so long. This is done by taking D-T fuel confined in a small pellet and hitting the pellet with intense laser beams from all directions. Recall that these laser beams will have momentum. So the pellet is vaporized and the atoms are ionized and pushed together. The resulting plasma can have densities up to $10^{29} / \text{m}^3$. Usually short laser pulses are used, on the order of a nanosecond. In the next figure you can see the flash of light from a pellet being irradiated.



Implosion of a fusion microcapsule on the NOVA laser system. This miniature “star” was created in the Nova laser target chamber as 300 trillion watts of power hit a 0.5-millimeter-diameter target capsule containing deuterium–tritium fuel. (Lawrence Livermore National Laboratory)
Here is another setup with several laser beams visible



View inside the OMEGA chamber at the moment of operation of the 24-beam, 12-terawatt OMEGA laser (U.S. Department of Energy)

And finally a design for a reactor.



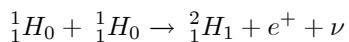
Image showing the arrangement of the later versions of the MEF or LIFE.1 power plant. The grey boxes arranged in groups at the top and bottom (just visible, shaded) are the laser beam-in-a-box systems. Their light, in blue, shines through the optical system into the yellow colored target chamber in the center. The machinery on the left pumps liquid lithium or fibre into the target chamber walls to cool the reactor, extract energy for generation, and produce tritium.

(Lawrence Livermore National Laboratory)

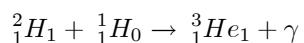
But no research group has produced a sustained fusion reaction that provides more output energy than was required to make the reaction happen. Several research groups seem so close, but we are not there yet.

Fusion in stars

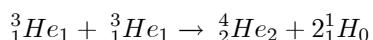
But what if you do have a star? The temperatures are high, the pressures are high, Lawson's criterion would be much easier to satisfy. This doesn't help us make reactors on Earth, but it does provide us with energy from the Sun. The fusion reaction in stars is really a chain of reactions.



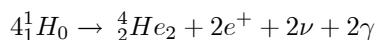
which yields deuterium. Then



giving us helium. This reaction must happen twice before the next reaction can happen.



the net process is

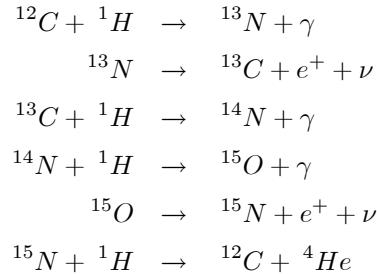


The Q value is a little bit hard to calculate. Our trick of using atomic masses is not as

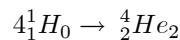
straight forward here. We need four electrons to be added to the left hand side to make the hydrogen nuclei into atoms. That means we need four electrons on the right side as well. Two would go with the helium nucleus to make a helium atom. The remaining two would combine with the positrons making more gamma rays. So we could write the Q equation as

$$\begin{aligned} Q &= (m_i - m_f) c^2 \\ &\quad (4m_{(1H)} - m_{(4He)}) c^2 \\ &= ((4(1.007825 \text{ u}) - 4.002603 \text{ u})) c^2 \left(\frac{931.5 \text{ MeV}/c^2}{\text{u}} \right) \\ &= 26.731 \text{ MeV} \end{aligned}$$

This is likely the principal source of solar energy. But with such enormous densities ($\rho = 1.5 \times 10^5 \text{ kg/m}^3$) and temperatures (10^7 K) as are found in the center of the Sun, we can have larger elements fuse. An example is the carbon cycle



The net process is



and we end up with the same amount of carbon as we started with. But the carbon is a catalyst that allows the reaction to happen faster. A new star would have to synthesize carbon before this process could start. This process likely dominates with star core temperatures above $20 \times 10^6 \text{ K}$, much hotter than our Sun. Let's continue to look at making atoms in stars in our next lecture, and that will lead us to a taste of particle physics.

32 Nucleosynthesis and the Particle Zoo

13.6 - 14.2

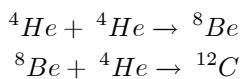
Fundamental Concepts in the Lecture

- How new elements are produced in stars
- CAT Scans, MRI
- The four fundamental forces
- Classification Schemes, Leptons, Hadrons

In our last lecture we hinted about something about the formation of atoms in stars. We learned that large stars probably use the carbon cycle to turn hydrogen into helium. But where did the carbon come from?

Making the atoms

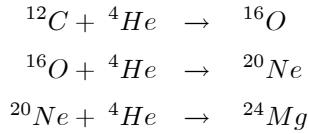
As stars age they will use up their hydrogen. This leaves an abundance of helium but without the nuclear fusion the star collapses inward due to gravitation. The rushing inward of the star matter raises the temperature of the matter to about 10^8 K (recall that thermal energy is strongly related to kinetic energy). The higher kinetic energy of the atoms will drive atoms into atoms. And some will have enough energy to overcome the Coulomb repulsion between the nuclei. And this starts a second set of nuclear fusion reactions.



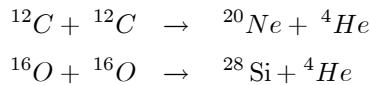
and this is where the carbon comes from to start the carbon cycle. This reaction has a Q value of 92keV. Carbon is not produced quickly because the 8Be has a half life on the order of 10 fs, so the second reaction is not very probable. But stars are big and hot so

at 10^8 K there will be some 8Be all the time and the $^8Be + ^4He$ reaction has a large cross section so carbon is slowly produced.

Once there is enough ^{12}C a new fusion chain can start

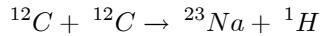


All of these reactions contribute energy to the star and therefore raise the star's temperature. When the temperature gets high enough the kinetic energy of the atoms is large enough to fuse carbon and oxygen atoms



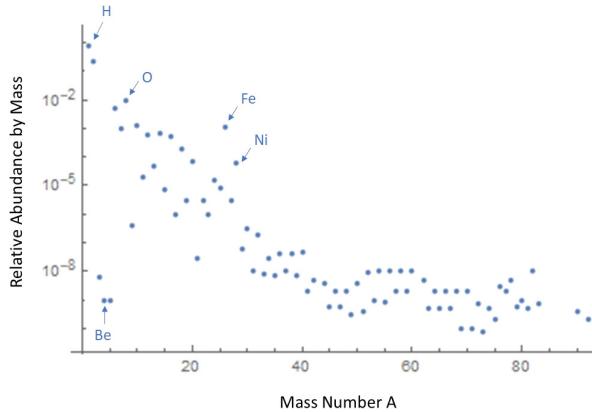
There are many more such process leading to eventually produce iron. Noting the pattern of these processes we expect to find in stars many light even Z elements, and much less abundant odd Z elements. Notice also that we skipped Li , Be , and B so these elements should be less abundant. And eventually we should see lots of Fe because it is so stable.

We will get some of the odd Z elements through less probable reactions, for example

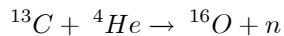


produces sodium.

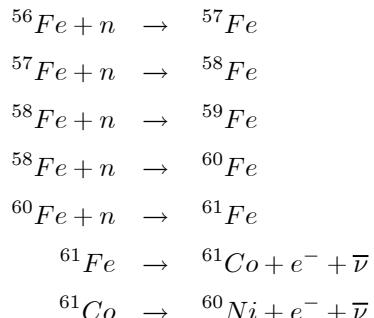
We can plot the relative abundance of different elements and find that there are indeed quite a lot of hydrogen and helium atoms in the universe, fewer Li , Be , and B , and lots of Fe .



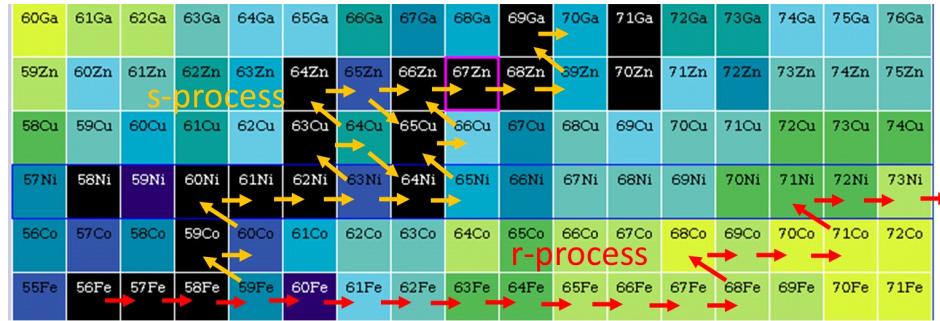
Beyond iron we need to have free neutrons available to create the neutron heavy nuclei. Without the extra neutrons the heavy nuclei above iron are not stable. Reactions like



could contribute to large nuclei production. With free neutrons we could have a chain like this



Many of the nuclei in this chain are unstable but with half lives that allow the chain to function in a star. Here are two example processes for generating heavy nuclei. The *s*-process works by beta decay and it is slower (hence the *s*). The rapid *r*-process is quicker but goes quickly away from the stable region where the *s*-process sticks close to the stable region. The *r*-process is likely to happen in supernova explosions.



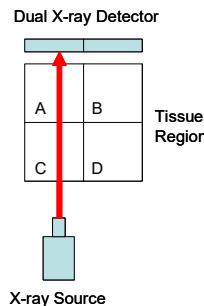
Only a few possibilities are shown in the figure. There are many such *r* and *s*-processes.

Applications of Nuclear Physics

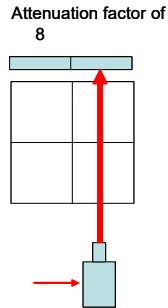
Here are a few more practical applications of nuclear physics

CAT Scans

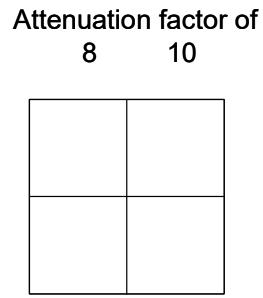
A computed axial tomography (CAT) scan finds tissues of different composition by sending X-rays through the body at many different angles. As an idea of how this works, let's take a region of the body and mentally divide it up into four smaller volumes. I will use my X-ray source and detectors to find out how many of the X-rays are absorbed. I will use generic units so we won't spend time worrying about the actual doses of radiation. We just need to know that big numbers tell us there has been more absorption.



I move the x-ray source, and have a detector array so I can see what the combined absorption was for subregion A and C together. I now move the source so I can find the absorption of subregion D and B together.

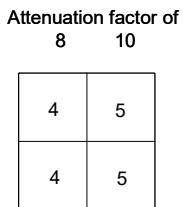


The results are as shown

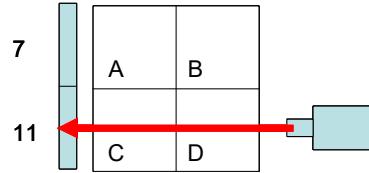
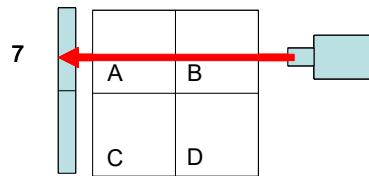


Result of first scan

for A and C together I have an attenuation of 8 generic units, for B and D together I have 10 generic units. I guess that half the attenuation of the combined A and C scan is due to the material in each subregion equally



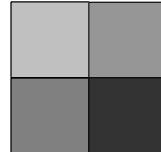
Likewise I assume half the attenuation from regions B and D together is due to each cell, B and D equally. It is a little like Suduko puzzles.



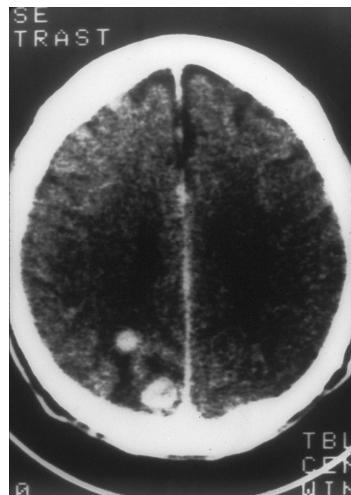
Now I turn my X-ray source and detector array. I scan A and B together and C and D together giving two new combined attenuation values.. The first tells me that I over estimated. My first guess would have given an attenuation factor for the sideways scan of 9 for A and B together. So I reduce each guess by one. Likewise, the bottom row, C and D would have given an attenuation factor of 9, but I got 11 with the sideways scan. I will change my guesses for C and D by adding one to each. My guesses now look like

3	4
5	6

I convert my guesses into a grayscale image



which tells us instantly that subregion D has the most dense material. The images are easier to see and can be quite detailed if many more than just two different directions are taken to form the image.



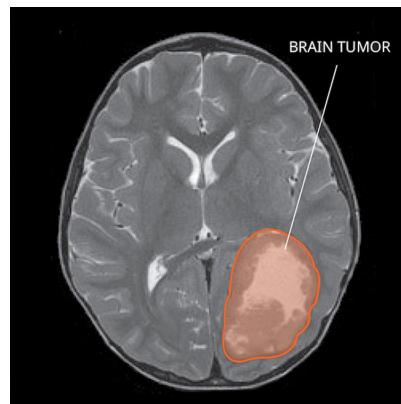
CAT scan of a human brain of a patient with AIDS (Image courtesy of the National Cancer Institute)

MRI

We already discussed the MRI technique. It really has nothing to do with a nuclear reaction , but it does deal with the nucleus. We already found that MRIs come from placing the person in an magnetic field that makes the nuclei align their spins. Then, when microwaves hit the atoms, nuclei with just the right spin energy, the nuclei flip their spin state. This causes the microwave radiation to be absorbed, and it is this absorption that we use as the MRI signal.

Originally, this technique was called Nuclear Magnetic Resonance, but the word “nuclear” scared patients, so the name was changed to MRI.

Let me emphasize this. A MRI is far less damaging to tissue than a radiation based imaging technique. MRI does use microwaves (like your cell phone) but the risk is very much less that with a CAT scan.



Magnetic Resonance Imaging of Primitive Neuroectodermal Tumor of the Central Nervous System. (Image courtesy of the National Institute of Health)

Fundamental forces

We found that matter is made of atoms, and atoms are made of electron orbitals and nuclei, and nuclei are made of proton and neutron wave functions. But what are electrons, protons and neutrons made from? And we found that the strong force was mediated by mesons, what is a meson and how does this work? To answer these questions we need to go back to the fundamental forces that hold the universe together.

In the last four physics classes (including this one) you have learned about four basic force types.

Fundamental Forces
Gravitation
Electric
Magnetic
Strong

Usually we combine the Coulomb and magnetic forces and just call them the electromagnetic force. From our introduction to electrodynamics (PH220) we can see why. Electric and magnetic fields are dependent on our reference frame, so what is an electric field in one reference frame is a magnetic field in another frame. Electrical and magnetic forces are really part of the same electromagnetic field interaction.

Fundamental Forces
Gravitation
Electromagnetic
Strong

Note that there is an open spot in our table. There is actually one more force. It is a nuclear force like the strong force, but, well, it is weaker—about 10^{-6} times the strong force magnitude. Thus the name. The *weak force* is a short-range force. It is responsible for instability that causes beta decay. The weak force doesn't play much of a role in holding nuclei together. So we haven't needed it up till now. The weak force has an interaction range that is about 0.001 fm, so it is even more short range than the strong force. Really it only acts inside a nucleon. Our fundamental forces are

Fundamental Forces
Gravitation
Weak
Electromagnetic
Strong

Particles can interact through any of these forces. We can describe the four fundamental forces by their range, relative strength, characteristic time, and the particles that use this force to interact.

Force Type	Range	Relative Strength	Typical Lifetime	Particles acted on by force	Mediating particles
Gravitational	∞	10^{-38}	?	all particles	None needed
Electromagnetic	∞	10^{-2}	10^{-14} to 10^{-20} s	charged particles	Virtual Photon
Weak	10^{-3} fm	10^{-7}	10^{-8} to 10^{-13} s	quarks, leptons	W^\pm, Z^0 bosons
Strong	1 fm	1	$< 10^{-20}$	quarks, hadrons	Gluon

Some of the words in this table are strange still. We will get to them shortly. It is a little bit wrong to think of the weak and strong force going to zero on the order of a femtometer. Really they approach zero asymptotically. But they approach zero so much faster than the gravitational and electromagnetic forces that this is practically correct.

The last column of the table may be strange. It talks about a mediating particle. We spoke of this back when we first encountered the strong force. The strong force is well modeled by assuming a nucleon emits a particle, and that all nucleons are attracted to that particle. The nucleons are not so much attracted to each other, but are attracted to the mediating particles. So two nucleons will move closer to a mediating particle, and therefore move closer to each other. For the strong force the mediating particles

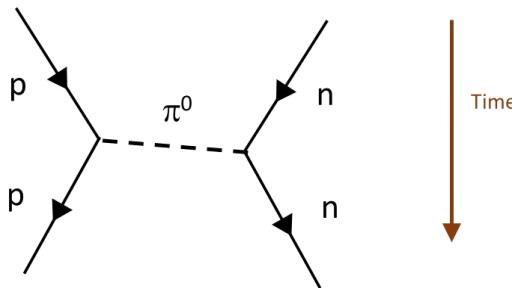
have been named *gluons*. It turns out that quantum electrodynamics models the electromagnetic field force as also mediated by a particle, and an obvious choice would be the quantized bit of field we know as a photon. The weak force also has mediating particles.

You might ask if gravitational forces have a mediating particle. And the answer depends on how you view gravity. Einstein's general theory of relativity (something we have not gotten to yet) would say no mediating particle is needed. But Einstein's theory is not complete, it has singularities in the center of black holes, for example. Quantum gravity theories seek to fix the holes in general relativity and to merge quantum mechanics and general relativity. There are competing theories (quantum loop gravity, modified gravity, etc.) and some of these postulate a particle that has been called the *graviton*. No evidence for this particle exists yet. But from theory we might have some idea of what its properties would be if it did exist. Here is a table of mediating particle characteristics.

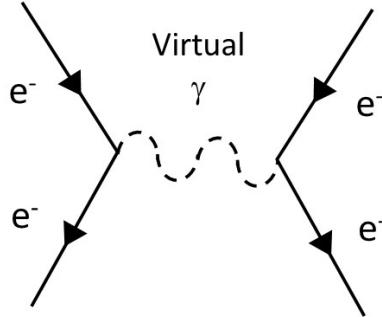
Force Type	Mediating Particle	Symbol	Charge	Spin	Rest Energy
Gravitational	Graviton?	g	0	2	0
Electromagnetic	Photon	γ	± 1	1	0
Weak	W^\pm Z^0	W^\pm, Z^0	0	1	80.4 GeV 91.2 GeV
Strong	Gluon		0	1	0

Note that for the strong force we found before that a π meson might be a good candidate for the gluon. A better theoretical fit is the virtual π meson, and that will need to be something we explain a bit further as we go.

Richard Feynman developed a way to draw particle interactions to show the workings of the exchange particles. Here is such a diagram for the strong force interaction between a proton and a neutron.



In the diagram you can see a proton approaching a neutron. And at some point an exchange particle, π^0 , is emitted. Notice that time moves from the top to the bottom of a Feynman diagram. Here is another example for the electromagnetic force.



Antimatter

We already have found that electrons have anti-particles, positrons. The existence of the positron was predicted by relativistic quantum theory (Dirac). This was my last class in graduate physics. It took the professor about two weeks of class time to do a problem. We won't do any in this class (sigh). But the results of the theory predicts that every particle will have an anti particle. This has been great for the science fiction writers, but despite this, it seems to be true. In nuclear reactions we used neutrinos and antineutrinos (some theories allow particles to be their own anti-particle, the neutrino may be one of these).

The positron and electron are good examples of the strange things that happen with antimatter.

$$e^- + e^+ \rightarrow \gamma_1 + \gamma_2$$

if we allow a positron and an electron to collide, they destroy each other, converting their rest mass (and any kinetic energy) to photons with an energy of 0.511 MeV. A way to think about this might be to envision the free positron and the free electron as still being waves (they are) with wave functions. But these wave functions can interfere with each other to the point of destroying their wave functions. The energy that was in these wave packets must be released. And the easiest way to release energy is to create photons. This will happen for protons and antiprotons as well

$$p + \bar{p} \rightarrow \gamma_1 + \gamma_2$$

with a photon energy of 938 MeV. We call such reactions *annihilation reactions*. This is the Star Trek matter-antimatter annihilation that you see in SiFi literature. SiFi might

exaggerate what we can do with annihilation, but the reactions are reality. PET scanners use this. They use a positron emitting substance (beta decay) in a patient, usually mixed in a glucose solution. Then the gamma emission due to the positrons annihilating with electrons can be detected. Since the brain uses more glucose when it is active, this is one way we have mapped the brain to the actions or functions for which each part of the brain takes control.

A high energy gamma ray can strike a nucleus and convert its energy into mass. Because charge must be concerned, we expect either a neutral result or a pair of particles, one of which is positive and one of which is negative. Experiment tells us that we get an electron and a positron. We create mass by converting the gamma ray's energy into particles. We often call this *pair production*.

Classification Schemes

We know about protons, neutrons, electrons, neutrinos, and π mesons. There are many more particles. Before we study individual particles and their properties, let's group them into categories based on those properties.

Leptons

The word leptons comes from the greek word for “small” (think of leprechauns). We have not been able to split apart this kind of particle, so we don't know of any internal structure. Since (we think) leptons are unique and different with no internal parts,¹⁷ we must conserve something leptonish in reactions. We assign a lepton number to each lepton like a charge. We demand lepton number to be conserved like charge. This means the wave functions must have a particular characteristic that would have a lepton quantum number come out in the solutions to the Schrödinger equation. Here are the leptons

¹⁷ By “no internal parts” of course we mean that their intrinsic wave functions are simple and not a combination of other wave functions that we recognize as independent particles. Protons, however, do seem to be a combination of three quark wave functions.

Leptons	Symbol	Lepton Number	Charge (e)	Spin	Mean Life (s)	Anti-particle	Lepton Number	Rest Energy (MeV)	Decay Products
Electron	e^-	$L_e = +1$	-1	$\frac{1}{2}$	∞	e^+	$L_e = -1$	0.511	
Electron-Neutrino	ν_e	$L_e = +1$	0	$\frac{1}{2}$	∞	$\bar{\nu}_e$	$L_e = -1$	really small	
Muon	μ^-	$L_\mu = +1$	-1	$\frac{1}{2}$	2.2×10^{-6}	μ^+	$L_\mu = -1$	105.7	$e^- + \bar{\nu}_e + \nu_\mu$
Muon-Neutrino	ν_μ	$L_\mu = +1$	0	$\frac{1}{2}$	∞	$\bar{\nu}_\mu$	$L_\mu = -1$	really small	
Tau	τ^-	$L_\tau = +1$	-1	$\frac{1}{2}$	2.9×10^{-13}	τ^+	$L_\tau = -1$	1784	$\mu^- + \bar{\nu}_\mu + \nu_\tau$
Tau-Neutrino	ν_τ	$L_\tau = +1$	0	$\frac{1}{2}$	∞	$\bar{\nu}_\tau$	$L_\tau = -1$	really small	

Notice that all leptons have spin of $\frac{1}{2}$. Leptons interact using the electromagnetic or weak forces.

Hadrons

There are two subclasses for Hadrons. Baryons and mesons. We have talked recently about mesons. Baryons are more “normal” particles like protons and neutrons. The Baryons have a baryon-ness that we also say is conserved like a charge. The Baryons interact through the strong force.

Baryons	Symbol	Baryon Number	Particle Charge (e)	Spin	Strange-ness	Mean Life (s)	Anti-particle e	Baryon Number	Rest Energy (MeV)	Decay Products
Proton	p	+1	+1	$\frac{1}{2}$	0	∞	\bar{p}	-1	938.3	None
Neutron	n	+1	0	$\frac{1}{2}$	0	886	\bar{n}	-1	939.6	$-p + e^- + \bar{\nu}_e$
lambda	Λ^0	+1	0	$\frac{1}{2}$	-1	2.6×10^{-10}	$\bar{\Lambda}^0$	-1	1115.6	$p + \pi$
Sigma	Σ^+	+1	+1	$\frac{1}{2}$	-1	8.0×10^{-11}	Σ^+	-1	1189.4	$p + \pi^0$
	Σ^0	+1	0	$\frac{1}{2}$	-1	7.4×10^{-20}	$\bar{\Sigma}^0$	-1	1192.5	$\Lambda^0 + \gamma$
	Σ^-	+1	-1	$\frac{1}{2}$	-1	1.5×10^{-10}	Σ^-	-1	1197.3	$n + \pi^-$
Xi	Ξ^0	+1	0	$\frac{1}{2}$	-2	2.9×10^{-10}	$\bar{\Xi}^0$	-1	1315	$\Lambda^0 + \pi^0$
	Ξ^-	+1	-1	$\frac{1}{2}$	-2	1.6×10^{-10}	$\bar{\Xi}^-$	-1	1321	$\Lambda^0 + \pi^-$
	Δ^*		+2, +1, 0, -1	$\frac{3}{2}$	0	5.6×10^{-24}	$\bar{\Delta}^*$		1232	$p + \pi$
	Σ^*		+1, 0, -1	$\frac{3}{2}$	-1	1.8×10^{-23}	$\bar{\Sigma}^*$		1385	$\Lambda^0 + \pi$
	Ξ^*		-1, 0	$\frac{3}{2}$	-2	7.2×10^{-23}	$\bar{\Xi}^*$		1533	$\Xi + \pi$
Omega	Ω^-	+1	-1	$\frac{3}{2}$	-3	8.2×10^{-11}	$\bar{\Omega}^-$	-1	1672	$\Lambda^0 + K^-$

Here are the mesons

Mesons	Symbol	Particle Charge (e)	Spin	Strange-ness	Mean Life (s)	Anti-particle	Rest Energy(MeV)	Decay Products
Pion	π^+	+1	0	0	2.6×10^{-8}	π^-	139.6	$\mu^+ + \nu_\mu$
	π^0	0	0	0	8.4×10^{-17}	π^0 (self)	135.0	$\gamma + \gamma$
Kaon	K^+	+1	0	+1	1.2×10^{-8}	K^-	493.7	$\mu^+ + \nu_\mu$
	K_S^0	0	0	+1	0.89×10^{-10}	K_S^0	497.7	$\pi^+ + \pi^-$
	K_L^0	0	0	+1	5.2×10^{-8}	K_L^0	497.7	$\pi^\pm + e^\mp + \bar{\nu}_e$
Eta	η	0	0	0	5.1×10^{-19}	η (self)	548.8	$\gamma + \gamma$
	η'	0	0	0	3.2×10^{-21}	η' (self)	958	$\eta + \pi^+ + \pi^-$
	ρ^+	+1	1	0	4.4×10^{-24}	ρ^-	775	$\pi^+ + \pi^0$
	D^+	+1	0	0	1.0×10^{-12}	D^-	1869	$K^- + \pi^+ + \pi^-$
	J/ψ	0	1	0	7.1×10^{-21}	J/ψ (self)	3097	$e^+ + e^-$
	B^+	+1	0	0	1.6×10^{-12}	B^-	5279	$D^- + \pi^+ + \pi^-$
	Υ	0	1	0	1.2×10^{-20}	Υ (self)	9460	$e^+ + e^-$

For hadrons, it was noticed that some of the particles are only produced in pairs. They also seemed to experience weak force decay (e.g. beta decay) even though they were particles that usually are governed by the strong force. Since this is strange, the property that allowed these behaviors is called *strangeness*. We give them a strangeness number that must be conserved, S .

We can use the leptonness, baryonness, and strangeness to predict the outcome of particle interactions. We will do this in the next lecture.

33 Particle Conservation Laws and particle detectors

4.3 - 14.6

Fundamental Concepts in the Lecture

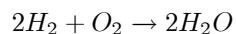
- Conservation of Lepton Number
- Conservation of Baryon Number
- Detecting particles: Geiger tubes, Diodes, Scintillators, PMT, Cloud and Vapor chambers
- Energy and momentum in particle interactions

We introduced the classifications of particles in our last lecture. Now let's look at those classifications and the interactions between particles of different classification. We will also need to know a little bit about detectors that we can use to verify particle experiment predictions.

Conservation laws

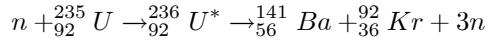
The classification system we introduced in our last lecture is more than just a nice way to see the different kinds of particles. The classification system is designed to help with a new set of conservation rules that come from experiment and theory. Like the conservation rules we have had so far, selection rules tell us what reactions are likely to take place without solving a very large differential equation. This was true for conservation of energy and momentum. But we found more conservation laws as we went.

For example, we had a conservation law that said in chemical reactions we couldn't destroy electrons. So if I have



you would know that there are ten electrons in four hydrogen and two oxygen atoms and ten electrons in two water molecules. No electrons created or destroyed. Electrons are leptons. We will find that there is something conserved in lepton interactions.

In nuclear reactions we found a conservation of nucleons. So in

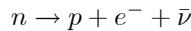


we can count up all the neutrons and protons and we should get the same number of nucleons on both sides. Nucleons are baryons. There seems to be something conserved in a baryon reaction.

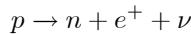
In each case we could introduce a new quantum number to allow us to figure out what reactions are more likely. We did this with electrons states. Our n , l , m_l , and m_s really stood for different wave functions. We need something like this for leptons and baryons. The details of the math are the part of physics called quantum electrodynamics or QED. QED is a graduate level physics topic (and really fascinating). For now we will introduce and use the selection rules to find what reactions are possible. This is a little like using n , l , m_l , and m_s without ever seeing the Schrödinger Equation or the solutions for hydrogen. But we can get a good conceptual idea of what happens this way, so let's get started.

Lepton Number

Let's look at a typical lepton interaction



Another is



But it never happens that in the first interaction we get a neutrino instead of an antineutrino even though they have the same very small mass. Conservation of energy would allow it, but something else makes the reaction not happen. Something inherent to the lepton type of wave functions. We can assign a quantum number to leptons that shows if a reaction is probable or not. In our lepton table there were actually three new lepton quantum numbers L_e , L_μ and L_τ . Let's see how these work in our example interactions. For the first reaction we mentioned above

	n	\rightarrow	p	$+$	e^-	$+$	$\bar{\nu}$
L_e	0	\rightarrow	0	$+$	1	$+$	-1

The neutron and proton are baryons, so their lepton number has to be zero (no lepton-ness). But the electron and the antineutrino are leptons and now we see that to balance the lepton number we do need an antineutrino, not a neutrino. Likewise for the second reaction

	p	\rightarrow	n	$+$	e^+	$+$	ν
L_e	0	\rightarrow	0	$+$	-1	$+$	1

and L_e balances, on both sides of the reaction. We could write similar reactions involving L_μ and L_τ .

	ν_μ	$+$	n	\rightarrow	μ^-	$+$	p
L_μ	1	$+$	0	\rightarrow	1	$+$	0

or even

	μ^-	\rightarrow	e^-	+	$\bar{\nu}_e$	+	ν_μ
L_e	0	\rightarrow	1	+	-1	+	0
L_μ	1	\rightarrow	0	+	0	+	1

Of course our use of lepton quantum numbers obscures a lot of math. But it does allow us to see what would come from individual particle reactions easily.

Baryon Number

In our baryon classification table there is a baryon quantum number listed for each baryon. We already know this conservation rule because we used it before, but we called it conservation of atomic number or conservation of nucleons. But by careful observation we can assign specific baryon numbers to baryons and then use this to predict the outcome of interactions. Let's look at one of our previous examples. We now know the proton and neutron should have had baryon numbers

	n	\rightarrow	p	+	e^-	+	$\bar{\nu}$
L_e	0	\rightarrow	0	+	1	+	-1
B	1	\rightarrow	1	+	0	+	0

For this reaction we have not employed a conservation rule for all the particles. We need to do this for every interaction in order to predict outcomes accurately.

Perhaps the most famous (or at least the very cool) baryon interaction is

$$p + p \rightarrow p + p + p + \bar{p}$$

which is the reaction that produces an antiproton. We could write this with the baryon numbers to see why this must be the right interaction.

	p	+	p	\rightarrow	p	+	p	+	p	+	\bar{p}
B	1	+	1	\rightarrow	1	+	1	+	1	+	-1

what this means is that we accelerate one (or both) of the initial protons in an accelerator and direct them to have a collision. The kinetic energy of the collision is converted into a proton and an antiproton. CERN is one laboratory that can produce antiprotons and the first place to make antihydrogen (an antiproton with a positron orbital cloud).



CERN Low Energy Antiproton Ring (LEAR) where anti hydrogen was first produced.

Strangeness number

Lepton number and baryon number proved to be great predictors of which interactions would happen, but this didn't work for mesons. This is one of the reasons for the meson category. Mesons are frustrating because they don't seem limited in interactions like baryons and leptons so that the interactions

$$\begin{aligned} p + p &\rightarrow p + n + \pi^+ \\ p + p &\rightarrow p + p + \pi^0 \\ p + p &\rightarrow p + n + \pi^+ + \pi^0 \\ p + p &\rightarrow p + p + \pi^+ + \pi^- \end{aligned}$$

are all possible! But there are some mesons that seemed to act strangely. For example the reaction

$$p + p \rightarrow p + n + K^+$$

seems like it should happen but it doesn't. If the energy of the protons is higher we can get

$$p + p \rightarrow p + n + K^+ + K^-$$

It looks like we can only make Kaons in pairs. We can assign another quantum number to mesons to describe this restriction.

$$p + p \rightarrow p + n + K^+ + K^-$$

	p	$+$	p	\rightarrow	p	$+$	n	$+$	K^+	$+$	K^-
S	0	+	0	\rightarrow	0	+	0	+	1	+	-1

This strange behavior was given the name *strangeness* and the quantum number is the strangeness quantum number. All Hadrons can have strangeness. For strangeness, there

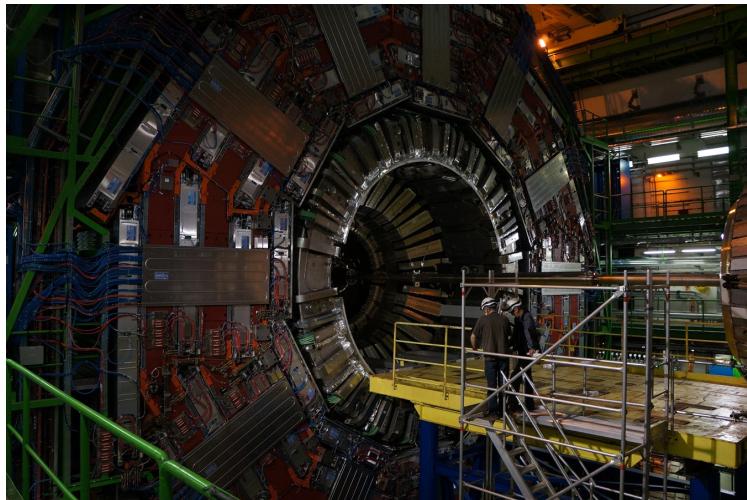
is a further restriction, the weak interaction can only change strangeness by one unit.

$$\Delta S = \pm 1 \quad (\text{Weak interactions})$$

which of course is a new selection rule!

Detecting particles

The next figure is the CERN Compact Muon Solenoid (CMS) detector. The CMS detector is a 15 m high, 21 m long “compact” detector specifically designed to detect muons. The detector system includes huge liquid helium cooled magnets to make charge particle trajectories curve. It has layers of detectors that track the path of the particles through the detector system. It is an engineering marvel designed to do exacting scientific tests. The CMS is pictured below during routine maintenance. Notice the size of the workers compared to the size of the huge cylindrical CMS system.



Superconducting Compact Muon Solenoid (CMS) at CERN. The detector system is two stories tall and normally kept at 3K. In this photo the detector is separated into two halves (one side is not visible in this photo) for maintenance.

But just how would such a detector system work? Let's look at the basic strategies for detecting particles.

Basic detection strategies

We can't see or hear particles. We don't feel electric or magnetic fields. So the question should be raised, just how do we know all these particles exist? Here are some detection strategies used in particle physics.

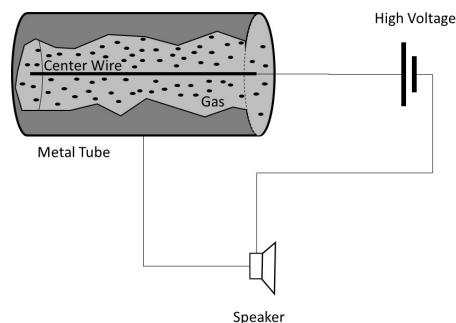
We will look at smaller versions of particle detectors, but the principles used in simple hand held detectors are the same as those used in monster detectors like the CMS. Our

detection strategies will take advantage of the energy of the particles that come from our reactions to move charges on collision. This is similar to photoelectric effect.

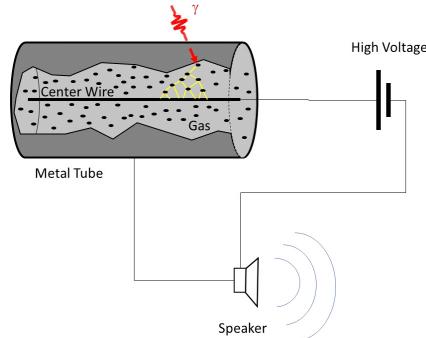
Geiger counter



The Geiger counter will count radiation that passes through it's tube. You have seen these on SiFi shows, no doubt. We know enough physics to see how it works. We have a wire at a high positive potential compared to the metal tube that surrounds it. The tube is filled with a gas.



As radiation enters the tube, it ionizes the gas by knocking an electron off a gas atom. The electrons from the atom accelerates toward the wire. On the way it hits other atoms and knocks off more electrons. This is called and electron “avalanche.” The avalanche is a burst of current (it is moving charge) which can trigger a counter or speaker to let you know ionizing radiation passed through the tube.



This avalanche is the “click” you hear when a Geiger counter detects radiation.

Semiconductor diode

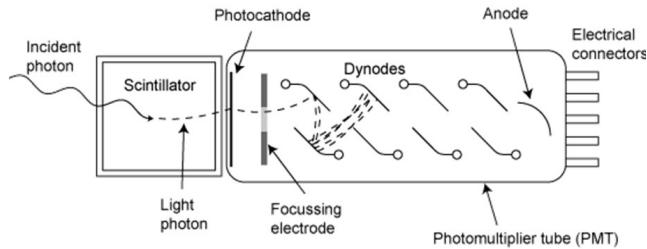
This is just a *p-n* junction. When ionizing radiation hits within the depletion region, it separates an electron from its atom. We know this now as an electron-hole pair. The electron and hole will move in the depletion zone field, creating a small current pulse that can drive a counter or speaker. This is very like the photodiodes we used in lab, only we now have small particles that can remove electrons as well as photons.

scintillation counter

These are fun. Scintillation crystals glow when struck with radiation. The radiation excites an atom, and then the atom returns to the ground state. This sometimes is complicated, because the atom is often in a crystal, so crystal excited states are involved. But the photon is detectable, using photomultiplier (next section) or avalanche photodiodes (not discussed here).

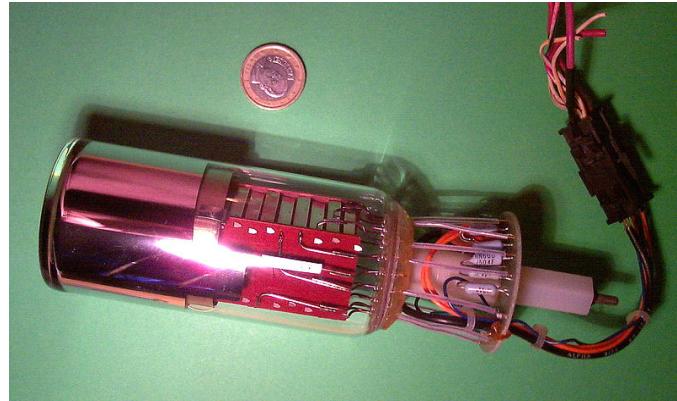
Photomultiplier

Photomultipliers are another tube device like the Geiger tube. The difference is that a photomultiplier has several dynodes, extra electric connections at different potentials along the tube. It starts with a scintillation crystal that emits a photon. The photon excites an electron using the photoelectric effect. This electron causes an avalanche at each dynode. The avalanche grows to the point at which it can be detected. Photomultipliers can be very sensitive. They can detect individual photons from the scintillation crystal, so they can detect single particles.



Photomultiplier with a scintillation crystal (Image courtesy Colin Eberhardt)

The tube photomultipliers are now being replaced with silicon based photomultiplier technology. But the older tube based version are in some ways more rugged. The new devices are lower power, and not made of glass tubes, so they have some important advantages.



A Photomultiplier Tube. Photo courtesy Antonio Pedreira.
The image intensifiers used in night vision goggles work on the same principle as photomultipliers.

Track detectors (cloud, bubble and, wire chambers)

These are cool. The bubble chamber uses a liquid near its boiling point. As a charged particle passes through, it ionizes atoms breaking their bonds to the rest of the fluid. This leaves a bubble trail. For both cloud chambers and bubble chambers you have to photograph the trail quickly because it will dissipate. More on this later. The reverse situation of a vapor cooled to super saturation also works. The charged particles ionize the vapor by collision, forming nucleation centers. So you get a particle vapor trail through the vapor. Here is a picture of a bubble chamber from CERN.



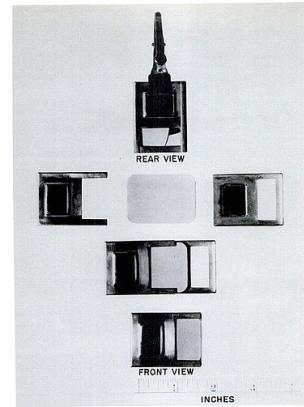
But a picture of a chamber in operation is better and we will get a few such pictures in the next section.

Wire track detectors

The wire track detectors use thousands of wires in two grids, one perpendicular to the other. As the charged particle passes a wire, it creates ions in the gas surrounding the wire grid. The electrons collect on the nearest wire in each direction. This is sort of a very fancy Geiger tube. Since there are two directions of wires, we can use the two wires that feel a current to give an x and a y location for the particle.

Film Badges

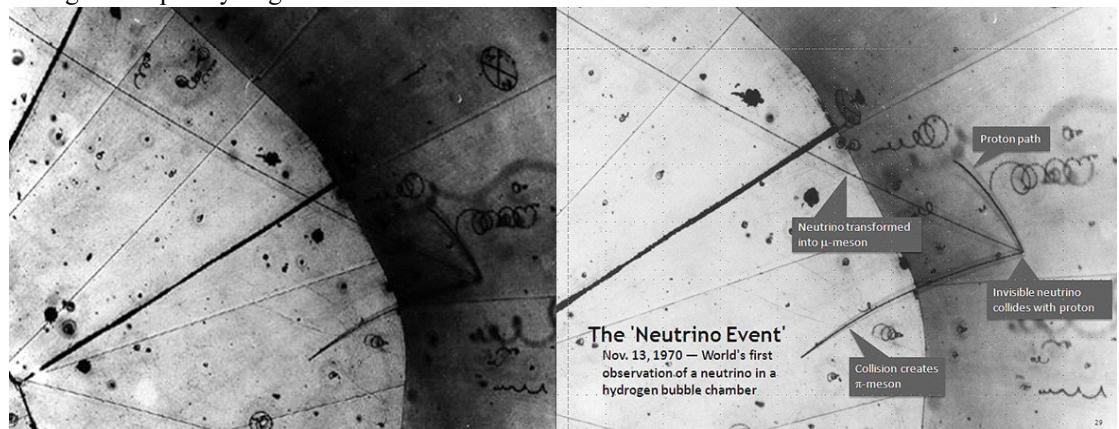
A simple radiation detector is photographic film, like the kind used in old cameras. Some of you may remember that the x-ray machines at airports can hurt your vacation photos if you are still using film-based cameras. Many of us have had x-rays taken of our bones. We know film reacts to x-rays and other ionizing radiation. Many people who work with radioactive materials wear a “film badge” that simply contain a piece of film. Periodically the badge is submitted to the rad-safety people that service the business. They develop the film. The darkness or lightness of the film tells how much radiation the badge has experienced, and therefore, how much exposure the person wearing the badge has had.



Older design of a Film Badge. You can see the Front View, the Back view, and the badge disassembled in the middle.

Detection strategies

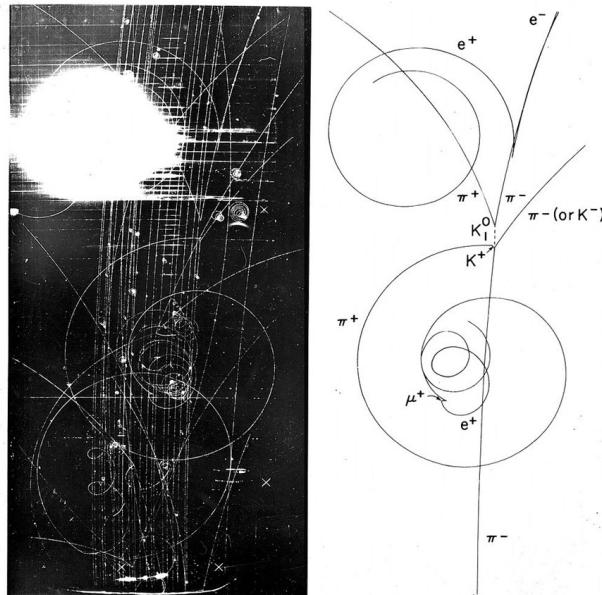
To detect particles we need not only a detector that can stop them and determine the energy they had, but we need to see the path they take so we can determine life of the particle and, with a magnetic field, determine charge. The first detectors to do all this were called bubble chambers. In the picture below you see a bubble chamber tank filled with liquid hydrogen. Charged particles cause tiny bubbles to form as they pass through the liquid hydrogen.



The first use of a hydrogen bubble chamber to detect neutrinos, on November 13, 1970. A neutrino hit a proton in a hydrogen atom. The collision occurred at the point where three tracks emanate on the right of the photograph. Image courtesy of Argonne National Laboratory.

Here is another example of a bubble chamber trail showing several reactions. The

positive charged particles make circular paths.

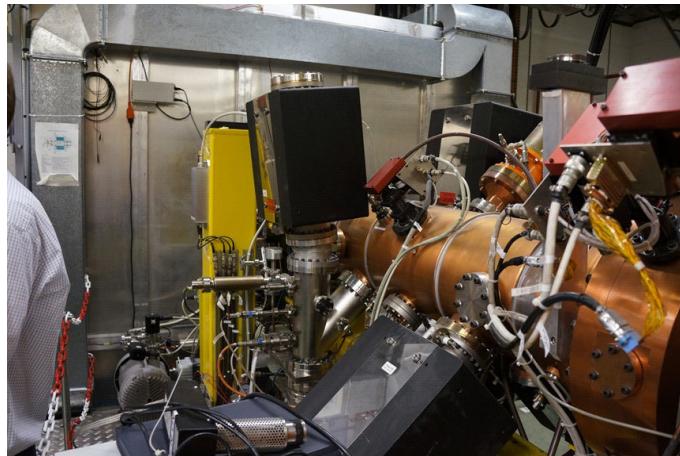


Photograph of bubble traces with drawing of traces from a specific particle interaction. Neutral or uncharged particles do not leave tracks. (Courtesy United States Department of Energy)

A typical particle experiment accelerates a charged particle like protons and collides them into a target material. From that interaction many particles are formed and the scientist can select a secondary particle type and form it into a beam which is directed into a detector like a bubble chamber.

Another typical experiment uses a radioactive source to produce particles, and then accelerates those particles with a linear accelerator (like our van de Graaff accelerator we studied earlier).

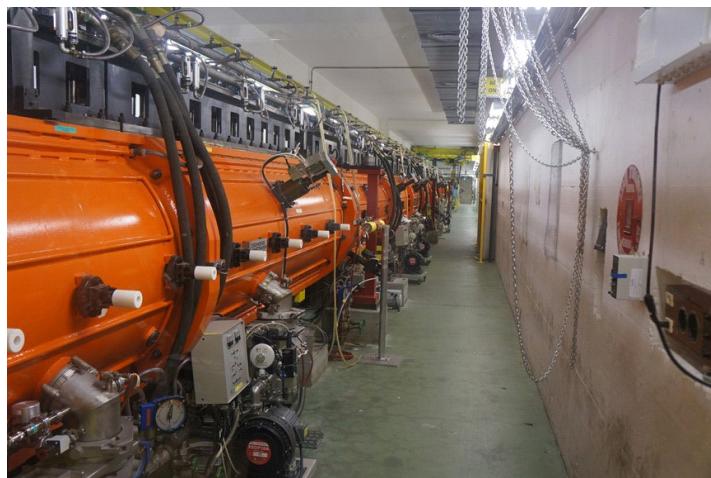
The next figure is a photograph of the CERN Low Energy Antiproton Ring (LEAR) initial proton source.



The protons are accelerated in a linear accelerator



The orange accelerator stages each have an electric potential difference so each produces acceleration.



From here the protons are often turned by a powerful magnet. In this case the LEAR uses magnets (large orange corner pieces) to direct charged particles in the square ring.



And somewhere along the path there will be detectors to study either the secondary particles themselves, or their collision products.



You might notice that our secondary particles have traveled some distance. If the particle travels about 10 m it would take about 10^{-7} s with the particle going near the speed of light. It would seem that CERN made the ring too big! But we have to include the effect of time dilation because in most particle experiments the particles are going at speeds that are significant fractions of the speed of light.

Energy-momentum in particle experiments

In analyzing particle interactions we use all the physics we have learned. We need conservation of energy (calculated as a Q value) and conservation of momentum and all of our quantum conservation laws. But usually in such experiments the particles are going at speeds that are significant fractions of the speed of light. This *high energy physics* requires us to consider relativistic energy momentum. Let's try an example

Suppose we have an Λ^0 particle¹⁸ and it is at rest and decays

$$\Lambda^0 \rightarrow p + \pi^-$$

Let's find the kinetic energy of each product particle.

We can use conservation of energy to find the Q value

$$Q = (m_i - m_f) c^2$$

and now we need our tables to find the masses

$$\begin{aligned} m_\Lambda &= 1115.6 \text{ MeV}/c^2 \\ m_p &= 938.3 \text{ MeV}/c^2 \\ m_\pi &= 139.6 \text{ MeV}/c^2 \end{aligned}$$

¹⁸ Λ is the capital λ so it is pronounced "lambda."

Then

$$\begin{aligned} Q &= 1115.6 \text{ MeV} - 938.3 \text{ MeV} - 139.6 \text{ MeV} \\ &= 37.7 \text{ MeV} \end{aligned}$$

and we can say that

$$K_p + K_\pi - K_\Lambda = Q$$

We know that $K_\Lambda = 0$ but we don't know how much of the 37.7 MeV belongs to the proton and how much belongs to the π^- .

$$K_p + K_\pi = Q = 37.7 \text{ MeV}$$

We do know that relativistic kinetic energy is given by

$$K = \sqrt{(pc)^2 + (mc^2)^2} - mc^2$$

which we can use for both product particles

$$Q = K_p + K_\pi = \sqrt{(p_p c)^2 + (m_p c^2)^2} - m_p c^2 + \sqrt{(p_\pi c)^2 + (m_\pi c^2)^2} - m_\pi c^2$$

and we also know from conservation of momentum in the center of mass frame

$$p_\Lambda = p_p - p_\pi = 0$$

so

$$p_p = p_\pi$$

then

$$\begin{aligned} Q &= \sqrt{(p_p c)^2 + (m_p c^2)^2} - m_p c^2 + \sqrt{(p_\pi c)^2 + (m_\pi c^2)^2} - m_\pi c^2 \\ Q + m_p c^2 + m_\pi c^2 &= \sqrt{(p_p c)^2 + (m_p c^2)^2} + \sqrt{(p_\pi c)^2 + (m_\pi c^2)^2} \end{aligned}$$

Most textbooks tell us that we could solve this with "a little algebra." And that is true, just like there is a little water in Lake Eyre. I decided to do this numerically

$$Q + 938.3 \text{ MeV} + 139.6 \text{ MeV} = \sqrt{(p_p c)^2 + (938.3 \text{ MeV})^2} + \sqrt{(p_p c)^2 + (139.6 \text{ MeV})^2}$$

$$\begin{aligned} 37.7 \text{ MeV} + 938.3 \text{ MeV} + 139.6 \text{ MeV} &= \sqrt{(p_p c)^2 + (938.3 \text{ MeV})^2} + \sqrt{(p_p c)^2 + (139.6 \text{ MeV})^2} \\ 1115.6 \text{ MeV} &= \sqrt{(p_p c)^2 + (938.3 \text{ MeV})^2} + \sqrt{(p_p c)^2 + (139.6 \text{ MeV})^2} \end{aligned}$$

We want p_p in terms of MeV/c so we can solve for x where $x = pc$

$$\begin{aligned} 1115.6 \text{ MeV} &= \sqrt{(x)^2 + (938.3 \text{ MeV})^2} + \sqrt{(x)^2 + (139.6 \text{ MeV})^2} \\ x &= p_p c = 100.38 \text{ MeV} \end{aligned}$$

then

$$K_p = \sqrt{(100.38 \text{ MeV})^2 + (938.3 \text{ MeV})^2} - 938.3 \text{ MeV} = 5.3541 \text{ MeV}$$

$$K_\pi = \sqrt{(100.38 \text{ MeV})^2 + (139.6 \text{ MeV})^2} - 139.6 \text{ MeV} = 32.343 \text{ MeV}$$

We will continue looking at energy and momentum in particle interactions in our next lecture.

34 Quantum Chromo Dynamics in a Nut Shell

14.7 -14.8

Fundamental Concepts in the Lecture

- Strange Particles? Charm? Maybe a New Model is Needed!
- Quarks
- The “Standard Model”: Quantum Chromodynamics (QCD) and the Electroweak Theory
- Grand Unification Theories
- Strings and Supersymmetry

Momentum and Energy in particle interactions

In particle physics today there are two standard techniques for producing collisions and their products. Along the way we have talked about both techniques.

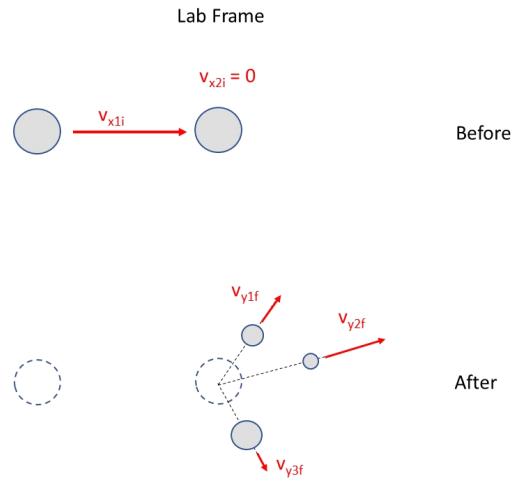
One is to have a stationary target made of some material and to accelerate charged particles toward that target. In such a reaction so long as Q is negative, the reaction has a possibility of occurring due to conservation of energy. If we have the reaction

$$x_1 + x_2 \rightarrow y_1 + y_2 + y_3 + \dots$$

then

$$Q = (m_{x_1} + m_{x_2} - m_{y_1} - m_{y_2} - m_{y_3} - \dots) c^2$$

gives us the change in rest energy. But in the lab frame for this experiment we know that x_2 is stationary and x_1 is not. So there is some initial linear momentum.

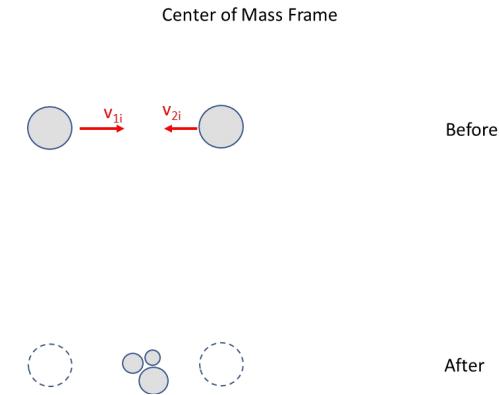


And we know that momentum is conserved so

$$\mathbf{p}_{x_1} + \mathbf{p}_{x_2} = \mathbf{p}_{y_1} + \mathbf{p}_{y_2} + \mathbf{p}_{y_3} + \dots$$

which tells us that the product particles must have some kinetic energy or they would have no momentum and we would break conservation of momentum. So we need Q to be large enough to not only allow the product particles to form, but also to supply them with some kinetic energy such that we conserve momentum. And of course x_1 has some initial kinetic energy to contribute to the conservation of energy equation.

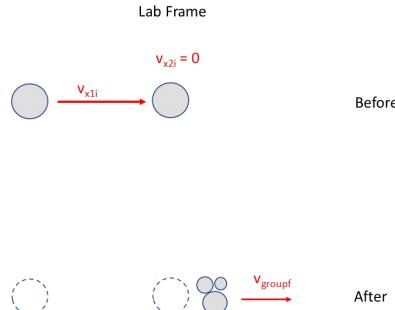
The second particle physics technique is to accelerate two charged particles and to have them collide. In this case the initial momentum can be zero, so the product momentum can be zero and we would not need to use any of the rest mass energy to provide kinetic energy for the product particles.



In either case if Q is negative, we need to add energy into the reaction to make it go. This is done in the form of initial kinetic energy of one or both initial particles. Then there must be a threshold kinetic energy below which the reaction won't happen. This is why we build huge supercolliders. We need to give the particles enough kinetic

energy so that the theoretical reactions would be possible, and then we try it and see if the reaction happens in real experimental life.

The most efficient way to distribute the final momentum to the group of product particles is to have them move off together with their velocities equal in the x -direction.



If we say the group of product particles has a combined mass M then we can use relativistic energy and momentum. The energy conservation equation is (for the lab frame)

$$\sqrt{(p_{x1}c)^2 + (m_{x1}c^2)^2} + \sqrt{(p_{x2}c)^2 + (m_{x2}c^2)^2} = \sqrt{(p_Mc)^2 + (Mc^2)^2}$$

and we know $p_{x2i} = 0$ in the lab frame so

$$\sqrt{(p_{x1}c)^2 + (m_{x1}c^2)^2} + (m_{x2}c^2) = \sqrt{(p_Mc)^2 + (Mc^2)^2}$$

We can square both sides

$$\left(\sqrt{(p_{x1}c)^2 + (m_{x1}c^2)^2} + (m_{x2}c^2) \right)^2 = (p_Mc)^2 + (Mc^2)^2$$

and write out the left hand side

$$(p_{x1}c)^2 + (m_{x1}c^2)^2 + 2\sqrt{(p_{x1}c)^2 + (m_{x1}c^2)^2} (m_{x2}c^2) + (m_{x2}c^2)^2 = (p_Mc)^2 + (Mc^2)^2$$

Solving for the term with the square root gives

$$\sqrt{(p_{x1}c)^2 + (m_{x1}c^2)^2} = \frac{(p_Mc)^2 + (Mc^2)^2 - (m_{x2}c^2)^2 - (p_{x1}c)^2 - (m_{x1}c^2)^2}{2(m_{x2}c^2)}$$

and we know from conservation of momentum that

$$p_{x1} = p_M$$

so then

$$\sqrt{(p_{x1}c)^2 + (m_{x1}c^2)^2} = \frac{(p_{x1}c)^2 + (Mc^2)^2 - (m_{x2}c^2)^2 - (p_{x1}c)^2 - (m_{x1}c^2)^2}{2(m_{x2}c^2)}$$

or, after canceling the $(p_{x1}c)^2$ terms,

$$\sqrt{(p_{x1}c)^2 + (m_{x1}c^2)^2} = \frac{(Mc^2)^2 - (m_{x2}c^2)^2 - (m_{x1}c^2)^2}{2(m_{x2}c^2)}$$

And we recognize the left hand side as part of the relativistic kinetic energy so

$$\begin{aligned} K_{x_1} &= \sqrt{(p_{x_1}c)^2 + (m_{x_1}c^2)^2} - m_{x_1}c^2 \\ &= \frac{+ (Mc^2)^2 - (m_{x_2}c^2)^2 - (m_{x_1}c^2)^2}{2(m_{x_2}c^2)} - m_{x_1}c^2 \end{aligned}$$

and we need a common denominator

$$\begin{aligned} K_{x_1} &= \frac{-(m_{x_1}c^2)^2 + (Mc^2)^2 - (m_{x_2}c^2)^2}{2(m_{x_2}c^2)} - \frac{m_{x_1}c^2(2(m_{x_2}c^2))}{2(m_{x_2}c^2)} \\ K_{x_1} &= \frac{-(m_{x_1}c^2)^2 + (Mc^2)^2 - (m_{x_2}c^2)^2 - 2(m_{x_1}c^2)(m_{x_2}c^2)}{2(m_{x_2}c^2)} \end{aligned}$$

The numerator can be written in a tricky way.

$$M^2c^4 - m_{2x}^2c^4 - 2m_{2x}m_xc^4 - m_x^2c^4 = (Mc^2 - m_{x_1}c^2 - m_{x_2}c^2)(Mc^2 + m_{x_1}c^2 + m_{x_2}c^2)$$

so then

$$K_{x_1} = \frac{(Mc^2 - m_{x_1}c^2 - m_{x_2}c^2)(Mc^2 + m_{x_1}c^2 + m_{x_2}c^2)}{2(m_{x_2}c^2)}$$

and we recognize the first term in the numerator as Q and the second term as the sum of the masses times c^2

$$\begin{aligned} K_{x_1} &= \frac{-Q(M + m_{x_1} + m_{x_2})c^2}{2(m_{x_2}c^2)} \\ &= \frac{Q((m_{y_1} + m_{y_2} + m_{y_3} + \dots) + m_{x_1} + m_{x_2})c^2}{2(m_{x_2}c^2)} \end{aligned}$$

which gives us

$$K_{th} = -Q \frac{\text{total mass of all particles involved}}{2(\text{mass of target particle})}$$

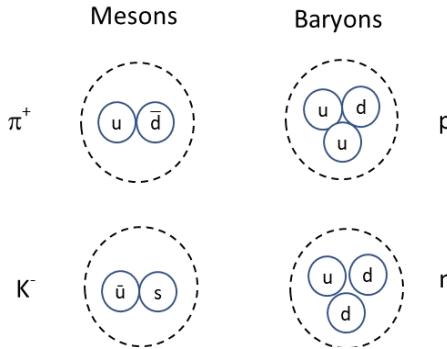
And this would be the kinetic energy required to make the reaction take place.

We are now going to leave particle physics as a broad topic, and concentrate on a special part of particle physics that deals with making up hadrons. To do this we need to introduce a new particle, the *quark*.

Quarks

Leptons seem to be truly fundamental particles, but we know that some baryons decay (like the neutron). We assume that they must have internal structure if they decay. So what are they built from?

The answer is that there is some wavicle more fundamental than hadrons, and hadrons are made of those fundamental wavicles. Here is an example.



Notice in the figure the baryon is made of three quarks with different properties. The original quark theory included three types of quarks. All quark names are weird. I guess that the particle physicists expressed the difficulty with visualizing these ultra-small wavicles and their behaviors with the names they gave them. These three quarks are called the *up*, *down*, and *strange* quarks. Since these early days, new properties of quarks were discovered, and new odd names resulted. Now we have properties of not only *topness*, *bottomness*, and *strangeness*, but we have properties of *charm* and *color*. These are just names, though. Of course the color of a particle that is much smaller than any wavelength of visible light is not something that even makes sense. Here are some properties of the quarks

Quark Name	Symbol	Spin	Charge	Baryon Number	Strangeness	Charm	Bottomness	Topness
Up	u	$\frac{1}{2}$	$\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	0
Down	d	$\frac{1}{2}$	$-\frac{1}{3}e$	$\frac{1}{3}$	0	0	0	0
Strange	s	$\frac{1}{2}$	$-\frac{1}{3}e$	$\frac{1}{3}$	-1	0	0	0
Charmed	c	$\frac{1}{2}$	$\frac{2}{3}e$	$\frac{1}{3}$	0	+1	0	0
Bottom	b	$\frac{1}{2}$	$-\frac{1}{3}e$	$\frac{1}{3}$	0	0	+1	0
Top	t	$\frac{1}{2}$	$\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	+1

and the antiquarks.

Antiquark Name	Symbol	Spin	Charge	Baryon Number	Strangeness	Charm	Bottomness	Topness
Anti-up	\bar{u}	$\frac{1}{2}$	$-\frac{2}{3}e$	$-\frac{1}{3}$	0	0	0	0
Anti-down	\bar{d}	$\frac{1}{2}$	$\frac{1}{3}e$	$-\frac{1}{3}$	0	0	0	0
Anti-strange	\bar{s}	$\frac{1}{2}$	$\frac{1}{3}e$	$-\frac{1}{3}$	+1	0	0	0
Anti-charmed	\bar{c}	$\frac{1}{2}$	$-\frac{2}{3}e$	$-\frac{1}{3}$	0	-1	0	0
Anti-bottom	\bar{b}	$\frac{1}{2}$	$\frac{1}{3}e$	$-\frac{1}{3}$	0	0	-1	0
Anti-top	\bar{t}	$\frac{1}{2}$	$-\frac{2}{3}e$	$-\frac{1}{3}$	0	0	0	-1

Here are a few recipes for making larger particles out of quarks. In the recipes, the up

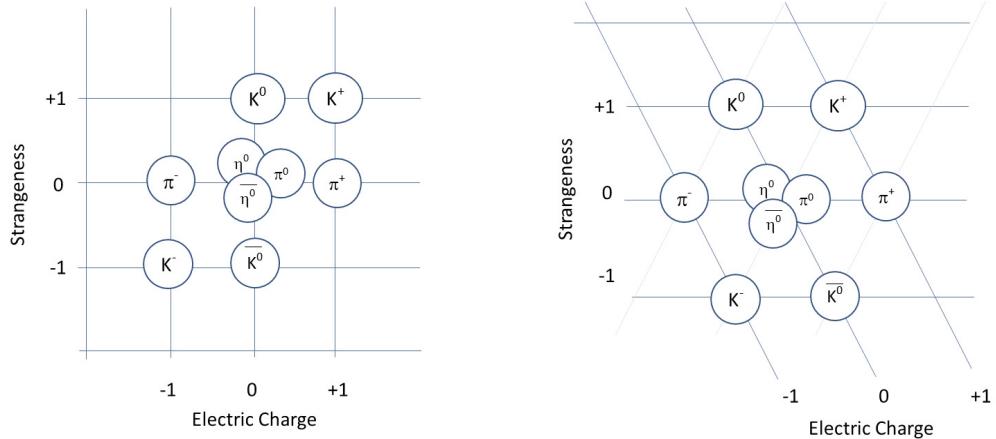
quark is given the symbol u , the down, d , and the strange s .

Particle	Quark Combination
p	uud
n	udd
π^+	$\bar{d}u$
π^-	$\bar{u}d$

This is all part of a complicated mathematical theory called Quantum Chromo Dynamics or (QCD). It is a field that takes some of the largest experiments (Fermi lab or CERN) and is at the cutting edge of physics.

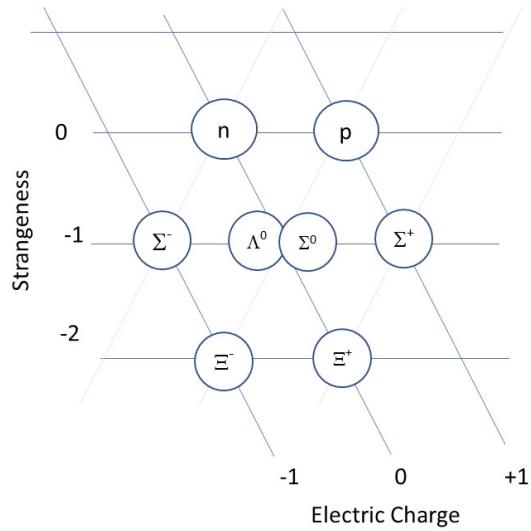
But let's back up and see how the quark model was able to describe the particles we know to exist from our particle physics experiments.

Two physicists (Gell-Mann and Zweig) independently used geometry to solve this problem of how hadrons are built. They constructed charts like the one below.

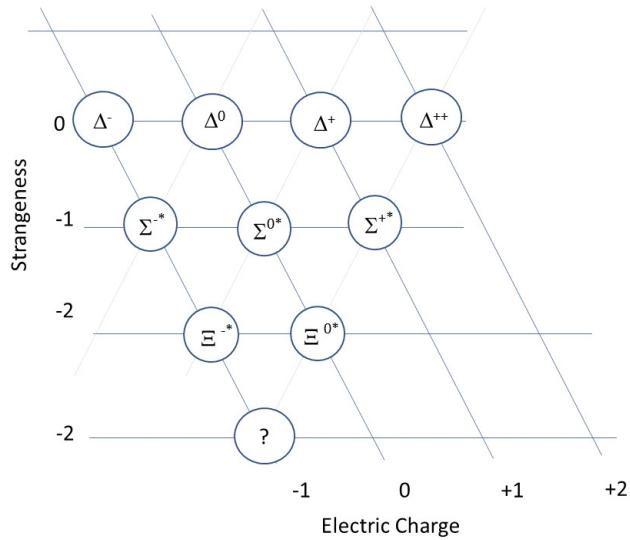


Notice that the charts are the same, but one has slanted axes. Gell-Mann liked the slanted axes and we will use his axis structure. When the mesons were plotted in this fashion, a clear pattern showed up (for Gell-Mann it was a hexagon).

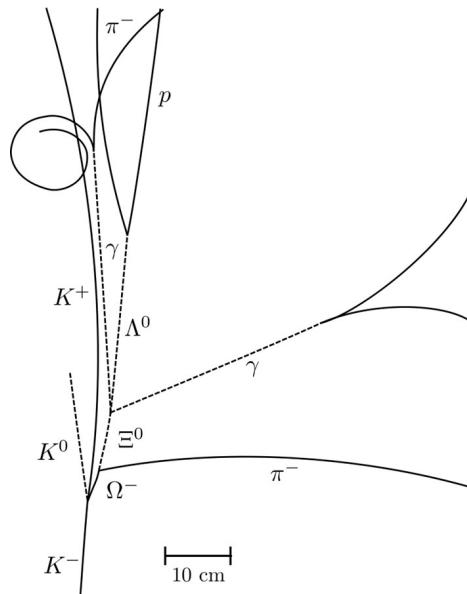
The Baryons follow the same basic structure



Gell-Mann gave the structure a very far-Eastern name, the *eight fold way*. This structure was useful in understanding the underlying wave functions. For example if we plot the high mass spin 3/2 baryons that were known at the time Gell-Mann was working you get this triangular pattern.

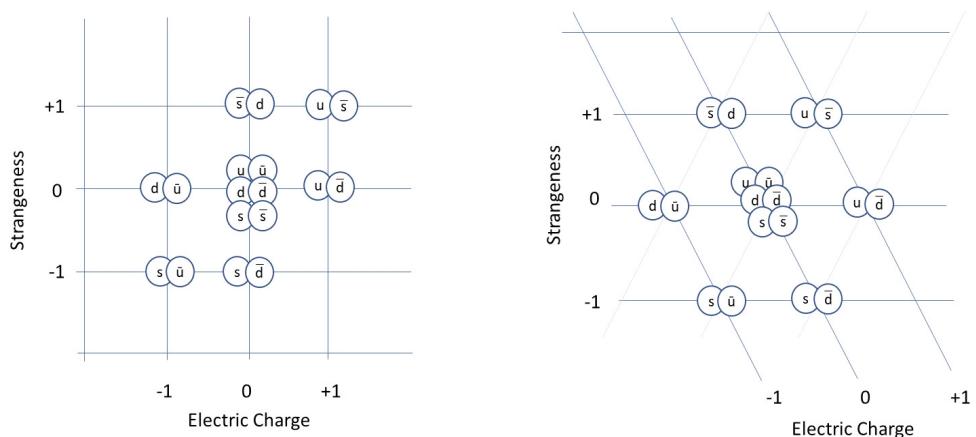


But the pattern was missing the tip of the triangle. Gell-Mann could predict that there would be a new particle discovered with the characteristics of this vertex in the graph. He called it the Ω^- particle and it was later discovered in an experiment at Brookhaven National Laboratory.



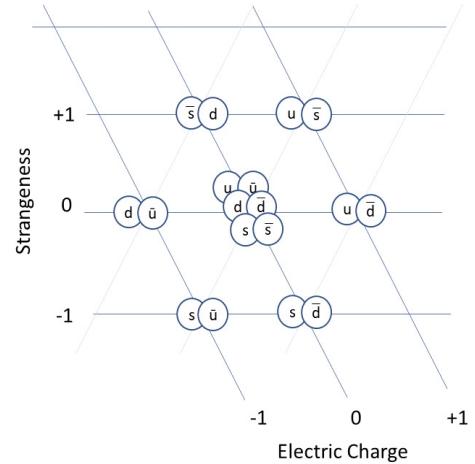
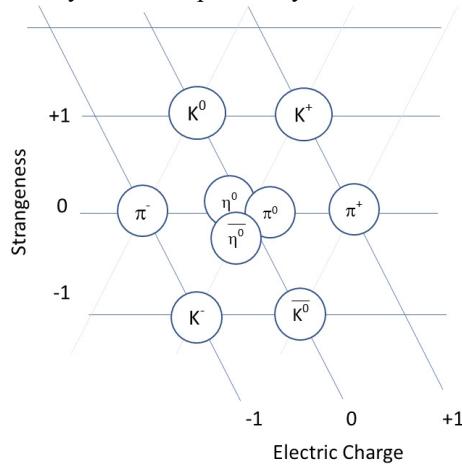
Traced bubble chamber trails from the Brookhaven National Laboratory experiment that found the Ω^- particle.

But the symmetry of the eightfold way diagrams can easily be explained by assuming that the baryons are made from combinations of three particles from a set of three particle/antiparticle groups. These are the up, down, and strange quarks in our last table. Mesons, in this model, are made from combinations of two particles from the set. Gell-Mann named the new particles *quarks*. Here is the meson graph but made from quarks.

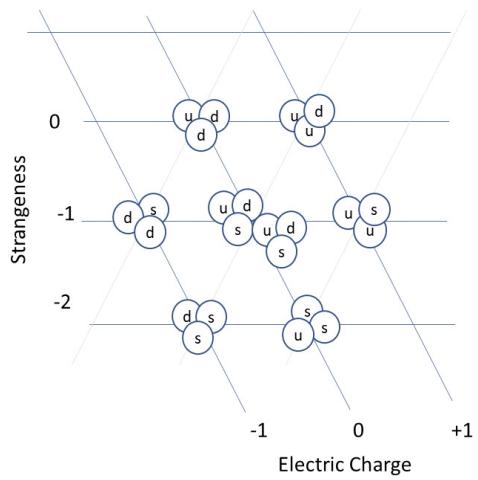
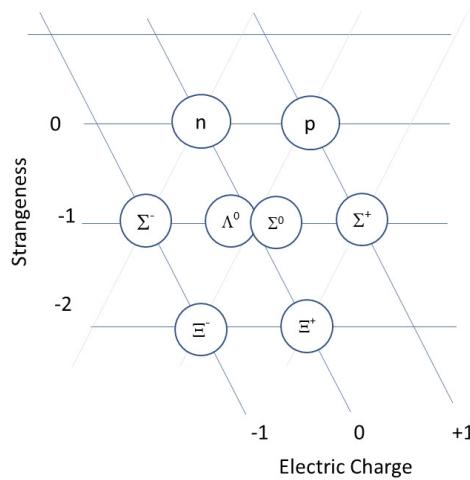


And by matching the mesons to the quark combination version of the graph, we can

identify the makeup of many mesons.



and many baryons.



Electron scattering experiments found that protons seem to have three charge centers within a single proton. This is consistent with the quark model for baryons. To make the electric charge of particles like the proton work in this model, the quarks must have fractional e amounts of charge. The same is true for baryon number. Here are more quark particle recipes.

Mesons		Baryons	
Particle	Recipe	Particle	Recipe
π^+	ud	p	uud
π^-	$\bar{u}s$	n	udd
K^+	$u\bar{s}$	Λ^0	uds
K^-	$u\bar{s}$	Σ^+	uus
K^0	$d\bar{s}$	Σ^0	uds
		Σ^-	dds
		Ξ^0	uss
		Ξ^-	dss
		Ω^-	sss

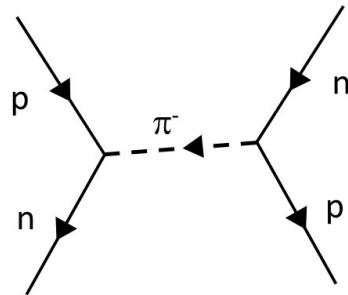
At the time the quark model was proposed, it explained the known particles. But particle physics didn't stop and soon many more particles were discovered. And some of the new particles seemed to break the quark model. By 1967 it was evident that the u, d, and s quarks couldn't be the only quarks. A new quark, symbolized with the letter c and called *charmed* was postulated. Charmedness, like strangeness, is conserved in strong and electromagnetic interactions, but not in weak interactions. The discovery of the J/ψ mesons lead to evidence for the c quark.

The discovery in 1975 of the τ lepton lead to two more quarks called top, t, and bottom, b. Of course adding in charm, bottomness, and topness to Gell-Mann's diagrams makes them four dimensional and hard to draw, but still useful in describing the quark makeup of particles and predicting new particles that are to be discovered.

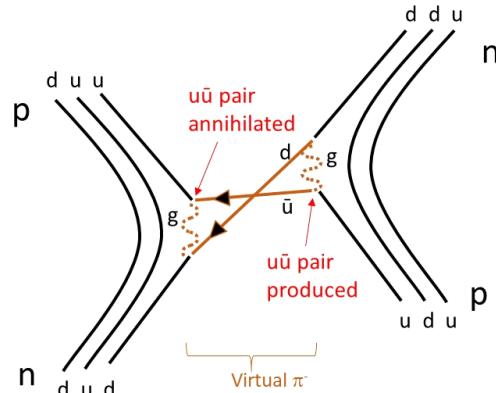
We should think of these properties, charm, topness, etc, as being a little like electrical charge. They represent wave forms that govern how the quarks will interact. So there are forces lurking behind these quantum numbers much like there is the electromagnetic force lurking behind our electric charge number in units of e . The calculations that show the force interactions are lengthy and beyond our course, but let's take a small look at how these interactions work.

Strong Force and QCD

Let's look again at strong force and exchange particles. Our early model for a strong force interaction might look like this proton-neutron interaction



where a π^- particle is the exchange particle. But now that we have quark theory, we can expand our understanding of strong force interactions. We know the proton and neutron are made of quarks. We could draw the interaction with a line for each quark.



In the interaction, a down quark in the neutron emits a virtual gluon, g , which then creates a $u\bar{u}$ pair through pair production. It must be a $u\bar{u}$ pair because pair production can only produce particle-antiparticle pairs. But what happens next is that the d joins with the \bar{u} part of the pair and they are transmitted to the proton. This is the same quark recipe as a π^- . The transmitted \bar{u} must combine with the proton's u to annihilate within a short time Δt to preserve conservation of energy by the Heisenberg uncertainty principle. This annihilation creates another virtual gluon. This leaves a d on the left where a u used to be. So the left hand side is now a neutron. And of course on the right hand side we left a u from the $u\bar{u}$ pair behind, so where the d was we now have a u . So the right hand side is now a proton.

The QCD interactions are a bit complicated, and for those who are lucky enough to go on to do graduate work there are whole courses in QCD (usually electives for those choosing to go into the field).

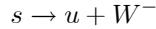
Quark changes by weak force

We really haven't dealt much with the weak force. It's time to figure out what it does.

The weak interaction can change quarks from one type into another. In each such case the reaction uses W^+ or W^- particles, the mediation particle for the weak force.

	Charge (e)	Baryon	Charm	Spin	Strangeness	Topness	Bottomness	Rest Energy (GeV)
W^+	1	0	0	1	0	0	0	80.4
W^-	-1	0	0	1	0	0	0	80.4

Here is an example



This is a little like changing a neutron into a proton plus an electron (and an antineutrino). Note that baryon number is conserved, and charge is conserved because W^- has a charge of -1 and u has a charge of $+2/3$. So the right hand side has a total charge of $-1/3$ which is the charge of the s quark. Strangeness changes by 1, but this is a weak force interaction, so that is fine.

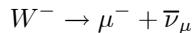
	s	\rightarrow	u	+	W^-
B	$\frac{1}{3}$	\rightarrow	$\frac{1}{3}$	+	0
Cg	$-\frac{1}{3}$	\rightarrow	$\frac{2}{3}$	+	-1
S	-1	\rightarrow	0	+	0

Other examples might be

	u	\rightarrow	d	+	W^+
B	$\frac{1}{3}$	\rightarrow	$\frac{1}{3}$	+	0
Cg	$\frac{2}{3}$	\rightarrow	$-\frac{1}{3}$	+	1
S	0	\rightarrow	0	+	0

	d	\rightarrow	u	+	W^-
B	$\frac{1}{3}$	\rightarrow	$\frac{1}{3}$	+	0
Cg	$-\frac{1}{3}$	\rightarrow	$\frac{2}{3}$	+	-1
S	0	\rightarrow	0	+	0

In each case the W^+ or W^- decay by the weak interaction into leptons. A typical decay might be



Quark changes only happen by the weak interaction.

Colored quarks

We have learned a lot about building particles from quarks. But you might object, if I put two up quarks and a down quark together to make a proton, don't I have three spin 1/2 particles, two of them identical, all in the same system. Won't that violate the Pauli exclusion principle? The predicted Ω^- particle is even worse with its recipe of sss.

The answer is that there must be some other property at play that splits the energy levels in nucleons so that we are not violating the Pauli exclusion principle. Back in 1965 Han Moo-Young and Nambu Yoichiro proposed a solution by postulating that there was an additional property of quarks. They called it *color charge*. Each of the quarks in a

baryon would need to have a different color charge to exist together. The three color charge values are given the name *red*, *green*, and *blue* even though this color charge has nothing at all to do with actual colors from optics. This model gives the name to this field of Quantum chromodynamics or QCD. This is one of the fields at the forefront of physics.

But if the energy levels separate when we bring in three s quarks to build a Ω^- particle, then there must be a force involved because different potentials require different forces. And that is exactly what QCD studies. The *color force* is the force between quarks. Like in electrostatics, two red charged quarks will repel but a red and a green will attract. Stable mesons and baryons must be *color neutral*, that is, the color force must essentially cancel out in the stable particles. But it doesn't quite cancel out, and the left over part is our model for what makes the strong force that holds nucleons together in the nucleus. This is a bit of a modification of our strong force model. There is still an exchange particle, but this time it is a massless gluon. So the color force is, in this sense, unified with the strong force.

Experimental evidence

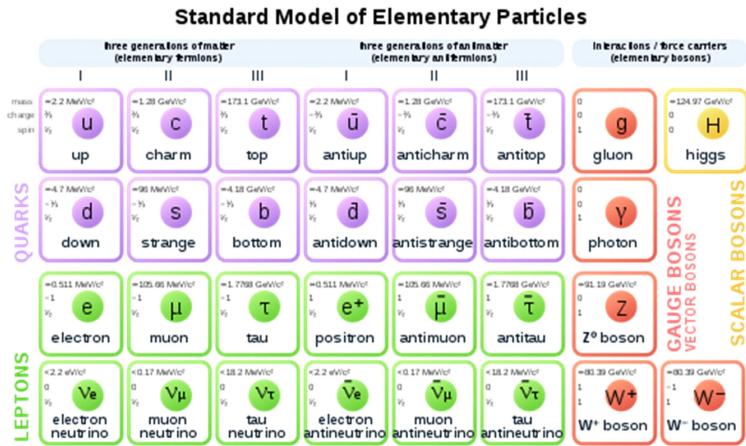
The QCD model has been very good at explaining what we see, but to date there has not been any direct measurements of single isolated quarks. This is not really surprising. It would be a little like looking at a major storm system, and expecting one funnel cloud to stay stable if we somehow take away the rest of the storm. The quark wavicles are only stable when they are inside the potential well of a hadron. Another example would be waves on strings. The standing waves are only stable so long as we don't free the string ends. But This means all of our evidence for quark theory is from groups of quarks and it is therefore secondary. Though the evidence is compelling, this should give us pause.

The Standard Model

Particle physicists have spent many years looking for the interaction particles associated with each fundamental force. The resulting model tells us that the world we can see and touch is made of four fermions and their antiparticles two of which are leptons and two quarks

$$\begin{aligned} & (e, \nu_e) \\ & (u, d) \end{aligned}$$

Protons and neutrons are made from u and d quarks. and atoms are made form protons, neutrons, and electrons. We need the neutrinos to include radioactive decay. With the possible addition of photons, this is our world. And we know that the world is held together by gluons, photons, Z bosons and W bosons. These are the force mediation particles. You can see these listed in the following table in column I (and the antiparticles in the next column I in the middle of the chart).



Standard Model of Elementary Particles (Image courtesy CUSH)

We should think about what this means. Let's start with what it doesn't mean. It does not mean that there are tones of electrons, neutrinos, up and down quarks sitting around in the universe and that we assemble matter from them. These are still wavicles. A better model is that these four wavicles, electrons, neutrinos, up and down quarks, are the stable wave functions that exist in the potential wells of atoms, baryons, and mesons.

High energy physics brings another four particles, two leptons and two quarks (column II)

$$\begin{aligned} &(\mu, \nu_\mu) \\ &(c, s) \end{aligned}$$

and even higher energy experiments bring four more

$$\begin{aligned} &(\tau, \nu_\tau) \\ &(t, b) \end{aligned}$$

This set of six leptons and six quarks (and their antiparticles) describe every one of the hundreds of particles we have discovered. At the time of this writing, there are no particles that don't fit this model. This model for particles is called the *standard model*, and it has been very successful. So are we done discovering the fundamental particles (still wavicles) that make up the universe?

There is current debate on this. Many physicists want to build a collider bigger than CERN's Large Hadron Collider in hopes to find more particles. Others point out that there isn't a theoretical bases for expecting new particles. Time (and funding) will tell.

The current research is in unifying the forces in the standard model. When we say "unifying" in this sense it is like when we found that magnetic forces and electrical forces really came from the same source.

The electromagnetic and the weak forces were unified (shown to be from the same origin) by Stephen Weinberg and Adus Salam in 1967. Their theory is known as the *electroweak theory*. Work is underway to unify the strong and electroweak forces. The candidates for this theory are given the name GUTS which stands for *grand unification theories*. The ultimate grand unification would be to combine gravity with the electroweakstrong force. But let's save that topic for a little later in our course.

If the standard model has been so successful, why are we still working? the answer is that the standard model doesn't explain why the particles we see have the masses we observe. A model of the masses has postulated another energy field that interacts with particles that have mass. The interaction slows the particles, giving them inertia. The field is called the Higgs field, and an evidence for the existence of this field is the Higgs Boson.

Another shortcoming of the standard model is that it assumes neutrinos have no mass. We gave them a mass upper bound in our tables, but the theory suggests that the neutrino mass should be exactly zero. But there are problems with solar neutrino measurements that have called this into question.

String hypothesis is a potential challenger to the standard model.¹⁹ String hypothesis postulates eleven dimensions of space time and that all particles are oscillations (standing waves) on one dimensional strings in the eleven dimensional space. The idea is intriguing because it would not only predict which particles exist, but could also predict their masses. But the theorists can't make exact predictions due to the tremendous mathematically complexity of solving partial differential equations in eleven dimensions. So, String Hypothesis has not been tested to date. This is why I am calling it *String Hypothesis* because it can't yet be *String Theory* until there has been experimental verification.

In our next few lectures, we will try to build an entire universe using the Standard Model.

¹⁹ See for example, *The Ellegant universe*, by Brian Greene.

35 Cosmology

15.1-15.3

Fundamental Concepts in the Lecture

- Expansion of the universe and Hubble's law
- Cosmic microwave background
- Dark Matter

Armed with the standard model and statistical physics and special relativity, it is almost time to build the universe! We will start with a description of the universe we experience. After all, if we want to design a universe we should know what a universe is like. And then we will find we need one more piece, General Relativity, to complete the picture. In this lecture we will start this process with a description of the universe.

We already know quite a lot about parts of the universe because we live in it. So we won't talk about the details of dirt and air and water, those we did in previous classes (and chemistry if you have been lucky enough to take it). But we will look at the universe on a universal scale.

Expansion of the universe and Hubble's law

The first thing we need to make sure we have in our model of the universe is the proper size and shape of the universe. In particular, the universe we live in seems to be expanding!

The evidence for this comes from atomic spectra and special relativity. Recall that earlier we found that the relativistic Doppler shift is given by

$$f' = \frac{f \sqrt{(1 - \frac{u}{c})}}{\sqrt{(1 + \frac{u}{c})}}$$

and we know that

$$c = \lambda f$$

so we can write this as

$$\frac{c}{\lambda'} = \frac{c}{\lambda} \frac{\sqrt{(1 - \frac{u}{c})}}{\sqrt{(1 + \frac{u}{c})}}$$

$$\lambda \frac{\sqrt{(1 + \frac{u}{c})}}{\sqrt{(1 - \frac{u}{c})}} = \lambda'$$

where λ is the frequency as measured if you are traveling with the moving object and λ' is the wavelength we would measure if we were in a reference frame watching the moving object rush by.

Suppose our moving object is a star, and we are watching the star move relative to us. If the star moves toward us we would find that u is negative so

$$\lambda' = \lambda \frac{\sqrt{(1 + \frac{-u}{c})}}{\sqrt{(1 - \frac{-u}{c})}}$$

and therefore the wavelength would be shorter (the numerator would be smaller than the denominator). Since blue light has shorter wavelength than red light we would say that the wavelength has been shifted in the blue direction. If our wavelength is part of an atomic or molecular spectrum then the spectral line has shifted to the blue side of the spectrum. This is called a *blue shift*. On the other hand if the star is moving away from us we would have $+u$ and then

$$\lambda' = \lambda \frac{\sqrt{(1 + \frac{+u}{c})}}{\sqrt{(1 - \frac{+u}{c})}}$$

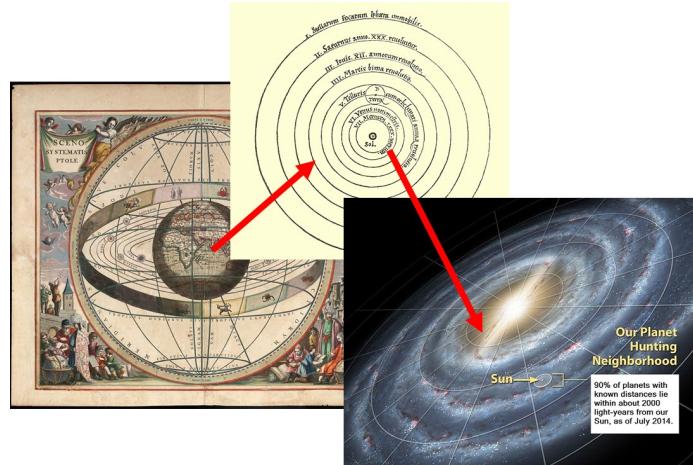
and the wavelength we measure on Earth would be larger than the one measured on the star. In the visible spectrum, red light has a long wavelength so we would say our wavelength has been shifted in the red direction or it has been *red shifted*.

Astronomers did the measurements for stars near us, and some were blue shifted and some red shifted. Relative speeds were on the order of $u = 30 \text{ km/s}$. This is a tiny fraction of the speed of light so the red and blue shifts were small.

Astronomers were fascinated with nebula. If you have seen a picture of the a nebula you can see why. Especially with the help of a telescope, nebula are large beautiful astronomical objects.

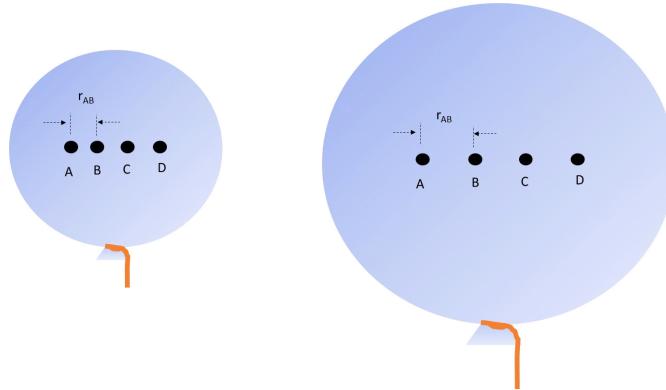


But once it was discovered that some of the fuzzy “nebula” were actually made of stars, the newly renamed *galaxies* could be recognized as being farther away. When these galaxies were studied astronomers had a shock. Nearly all the galaxies are red shifted, and the farther the galaxy is from us the more it is red shifted. Astronomy had already traveled the path that said the Earth was not the center of the universe, to realizing that the Sun was the center of the solar system and then that the sun and its solar system were a very small part of a galaxy.



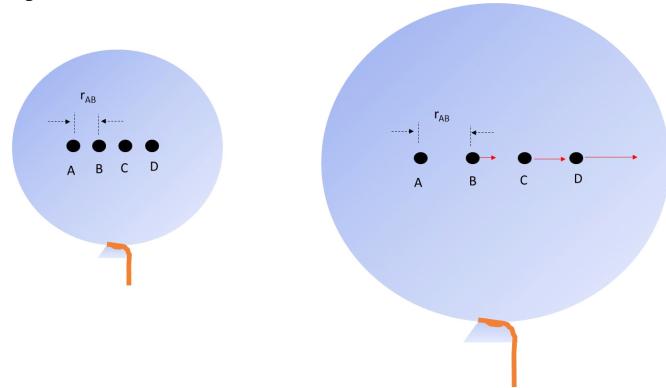
But now the red shift seemed to imply that outside of the local group of galaxies, everything is moving away from us. It is like we moved back to the theory that we are the center of everything again. This seemed unlikely to be true.

An alternate (and better) interpretation of the red shift data from galaxies is that the universe, itself, is expanding. Think of a balloon surface.



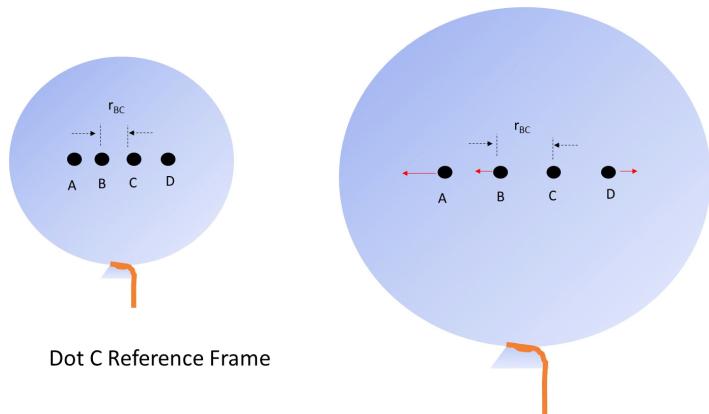
We can mark spots on the balloon surface as shown. As the balloon is inflated (an expansion) the dots move farther apart. This is not because the dots are traveling on the balloon surface. They are painted on, they can't move of their own accord. But rather, it is because the balloon material is stretching in between the dots, so the distance between the dots increases.

Let's take the point of view of dot *A*.



From *A*'s point of view *B* has a speed. That is why later it is farther away. And from *A*'s point of view *C* and *D* also have speeds. Notice that from *A*'s point of view, *C* and *D*'s distance from *A* increased more than the distance from *A* to *B* increased. And we moved everything in the same time Δt , so we would conclude that from *A*'s view, the farther from *A* the dot is, the faster it appears to go such that the dot can be in the final position we see it in later. This is just what we see when we look at galaxies!

But now look at the situation from the reference frame of point *C*.



A person viewing from dot C would see dot A move away from C and would also see dot B and dot D moving away. Notice once again the speed C associates with each dot depends on how far away from C the dot is. From C's point of view all the dots move away from C. So at A all the dots move away from A, but at C all the dots move away from C. This is the nature of an expansion, everything moves away from everything in an expanding system.

If we interpret our atomic spectral red shift data as coming from an expansion, then all parts of the universe would see the same thing, and we would not have the very special case of happening to be at the center of the universe. This seems more likely. Note also that C assigns different speeds (even different directions) to the points than A did. This is characteristic of an expansion.

Hubble discovered the recession of the galaxies and quantified his findings in an equation

$$v = H_o d$$

where v is the speed we assign to the galaxy and d is the relative distance of the galaxy from us. The quantity H_o is known as the *Hubble parameter*. Note that it would have units of inverse time ($1/\text{s}$).

$$H_o = 72 \frac{\text{km}}{\text{s} \cdot \text{Mpc}}$$

but the length units will cancel leaving just $1/\text{s}$.

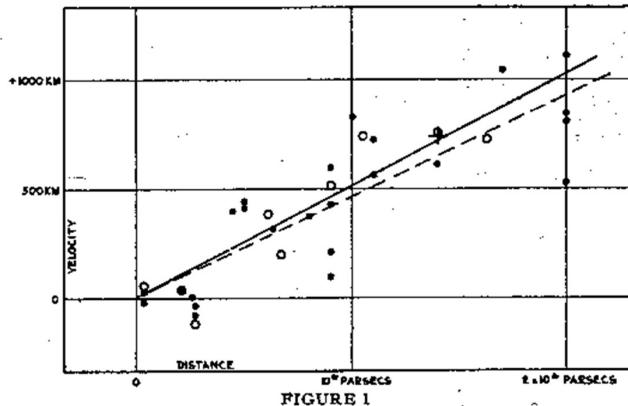


FIGURE 1

Hubble's original graph showing the linear Hubble law. (Hubble, E. 1929b, Proc. Nat. Acad. Sci., 15, 168., image courtesy NASA, https://apod.nasa.gov/debate/1996/hub_1929.html)

Let's try Hubble's law in a rectangular coordinate system. What Hubble says is that each dimension will increase in magnitude by some amount k . So the position $(x(t + \Delta t), y(t + \Delta t), z(t + \Delta t))$ depends linearly on the old position $(x(t), y(t), z(t))$

$$\begin{aligned} x(t + \Delta t) &= kx(t) \\ y(t + \Delta t) &= ky(t) \\ z(t + \Delta t) &= kz(t) \end{aligned}$$

Then the distance, d at time t

$$d(t) = \sqrt{x^2 + y^2 + z^2}$$

becomes

$$\begin{aligned} d(t + \Delta t) &= \sqrt{(kx)^2 + (ky)^2 + (kz)^2} \\ &= \sqrt{k^2x^2 + k^2y^2 + k^2z^2} \\ &= k\sqrt{x^2 + y^2 + z^2} \\ &= kd(t) \end{aligned}$$

and

$$\begin{aligned} v &= \frac{d(t + \Delta t) - d(t)}{\Delta t} \\ &= \frac{kd(t) - d(t)}{\Delta t} \\ &= d(t) \frac{k - 1}{\Delta t} \end{aligned}$$

and if we look at two galaxies to compare their speeds

$$\frac{v_1}{v_2} = \frac{d_1(t) \frac{k-1}{\Delta t}}{d_2(t) \frac{k-1}{\Delta t}} = \frac{d_1(t)}{d_2(t)} = \frac{d_1}{d_2}$$

then we could measure d_1 and d_2 to find that the speed of galaxy 1 as measured from

galaxy 2 would be

$$v_1 = \frac{v_2}{d_2} d_1$$

which is essentially Hubble's law. The model of a universal expansion seems to fit what we see very well.

If we borrow from our next lecture (you need a time machine to make this class work!) we can see from Einstein's theory of General Relativity that space-time can stretch, like our balloon surface (but space-time is three dimensional). This gives us a potential mechanism for our universal expansion red shift. While the photons from distant galaxy 1 are traveling to galaxy 2, the space they are in stretches, lengthening the wavelength because the wave is being stretched. We could say that

$$\frac{\lambda(t + \Delta t)}{\lambda(t)} = \frac{R(t + \Delta t)}{R(t)}$$

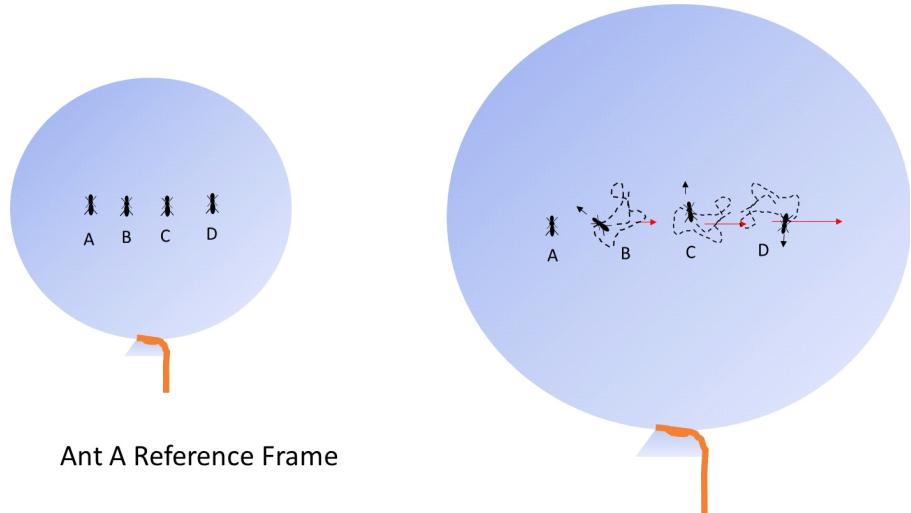
where R is a length scale for the universe that depends on time.

But now that we see that the universe is expanding, we should ask if that means the universe started from a much smaller space, or if as space expands somehow matter is created in the gaps so the universe seems to look somewhat the same over time.

The theory that says the universe was once more compact is called the *Big Bang Theory*. It tells us that the universe was born from a *singularity*, a point where all energy was located. In the original theory, this really was a point – no volume. The universe erupted from this point. That happened 15 – 20 billion years ago. One weakness of this theory is it has no explanation of what happened before this!

In the big bang theory, as the universe worked its way to the current system, the fundamental forces split from an ancestral unified force. It only took about a second after the initial “Big Bang” for protons to form. Shortly after, atoms could form. So for most of the age of the universe, the four forces seem to have been much as they are today.

Let's go back to our stars that were blue shifted. How come they weren't red shifted due to expansion as well? In our balloon analogy we had dots that didn't have any independent motion. All of their motion came from the universal expansion. But suppose we replace our dots with ants that can walk around on the balloon surface.



Now we see we have a set of ants wandering, with their little ant velocities shown as black arrows. Notice that near ant A the crawling velocity of the ant might be larger than the apparent velocity due to the expansion. So from ant A's view point ant B might be blue shifted. But the crawling velocity of ant C and ant D are not even close to the apparent velocity due to expansion. So ant C and D would be red shifted no matter which direction they crawled. Note also that if the ants were crawling away from ant A there would be two contributions to the red shift, a crawling contribution and an expansion contribution. It is the same way with stars. If they are close, their relative motions to us matter. If they are far, the relative motion of the star is not so important, the expansion motion dominates.

The Cosmic Microwave Background

By social media standards we are now “experts” in statistical mechanics. We know how to do calculations for a gas of photons. Suppose the universe did start as a hot dense high energy state, and has expanded and cooled. This would be modeled as a free expansion, and like anything, the universe would glow (emit photons) and the peak energy of those photons would change as it cooled and expanded. We should be able to predict what the peak energy of photons from the big bang would be now. This would be a test of our expanding universe model.

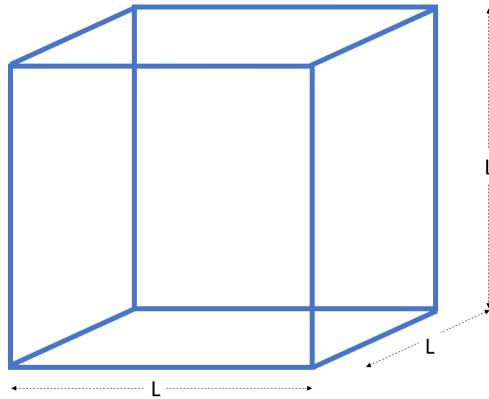
We want to use the number of photons per unit volume to find the temperature of the photons. If we know the energy density in a volume, and we know that energy is divided into N photons, we should be able to find the energy of the photons and this should depend on temperature. We are back to finding $N(E)dE$

The number of photons with energy between E and $E + dE$ would be found just in the same way we found this for any thermal radiator. We now just have a universe sized

thermal radiator.

$$N(E) dE = V g(E) f(E) dE$$

We could envision a cubical universe (to make the math easier). The universe would be a “box” confining the photons. That universal “box” would be an infinite potential well, because we don’t believe the photons can leave the universe. So we want the density of states function for a “gas” of photons confined in a reflective box with walls at temperature T .



Inside the box we have photons with energies from 0 to ∞ . And like before we take

$$f = \frac{E}{h}$$

so we get frequencies for 0 to ∞ . And we know

$$E = \frac{hc}{\lambda}$$

so we get

$$\lambda = \frac{hc}{E}$$

so our wavelengths go from ∞ to 0. And

$$E = pc$$

for photons so

$$E = c\sqrt{p_x^2 + p_y^2 + p_z^2}$$

We are going to get standing waves in all three directions in our box. The wave function must go to zero at the walls because they are reflective (photons can’t leave the universe, so let’s say the reflect at the end of the universe!). Then each photon would have quantized momentum

$$p_x = \frac{hn_x}{2L}$$

and similarly for p_y, p_z . So

$$\begin{aligned} E &= c\sqrt{\left(\frac{hn_x}{2L}\right)^2 + \left(\frac{hn_y}{2L}\right)^2 + \left(\frac{hn_z}{2L}\right)^2} \\ &= \frac{ch}{2L}\sqrt{n_x^2 + n_y^2 + n_z^2} \end{aligned}$$

and this seems very familiar. We could write this as

$$E = \frac{ch}{2L}n$$

where

$$n^2 = n_x^2 + n_y^2 + n_z^2$$

just like before.

Remember that photons have spin $s = 1$. So we expect a factor of $2s + 1 = 3$ but photons are polarized perpendicular to their direction of travel so one of the spin components must be zero. Let's set $s_z = 0$ so we only get 2 instead of 3 for our spin state multiplicity. So let's assemble our $g(E)$. Starting with

$$g(n) dn = \frac{1}{8} \frac{2s+1}{V} 4\pi n^2 dn$$

because of photon polarization

$$g(n) dn = \frac{1}{8} \frac{2}{V} 4\pi n^2 dn$$

and

$$E = \frac{ch}{2L}n$$

so

$$n = \frac{2L}{ch}E$$

and

$$dn = \frac{2L}{ch} dE$$

so

$$g(E) dE = \frac{1}{8} \frac{2}{L^3} 4\pi \left(\frac{2L}{ch} E \right)^2 \left(\frac{2L}{ch} dE \right)$$

$$g(E) dE = \pi \left(\frac{2}{ch} E \right)^2 \left(\frac{2}{ch} dE \right)$$

$$g(E) dE = \frac{8\pi}{h^3 c^3} E^2 dE$$

so our density of states function is

$$g(E) = \frac{8\pi}{h^3 c^3} E^2$$

And we know that photons are governed by Bose-Einstein statistics so

$$f_{BE}(E) = \frac{1}{A_{BE} e^{\frac{E}{k_B T}} - 1}$$

so, we expect²⁰

$$N(E) dE = V \left(\frac{8\pi}{h^3 c^3} E^2 \right) \left(\frac{1}{e^{\frac{E}{k_B T}} - 1} \right) dE$$

²⁰ Where we set $A_{BE} = 1$. This is an approximation. Because the number of photons is not constant we don't really know what A_{BE} should be (we can't be sure we can use normalization to find it). If we are content only knowing the shape of the $N(E) dE$ curve and rough estimates for our energy density, this is not too bad. We did this before for a photon gas.

To find the total number of photons per unit volume we need to integrate

$$\begin{aligned}\frac{N}{V} &= \frac{1}{V} \int_0^\infty N(E) dE \\ &= \frac{1}{V} \int_0^\infty V \left(\frac{8\pi}{h^3 c^3} E^2 \right) \left(\frac{1}{e^{\frac{E}{k_B T}} - 1} \right) dE \\ &= \frac{8\pi}{h^3 c^3} \int_0^\infty \left(\frac{E^2}{e^{\frac{E}{k_B T}} - 1} \right) dE\end{aligned}$$

and if we take

$$x = \frac{E}{k_B T}$$

then

$$dx = \frac{1}{k_B T} dE$$

so

$$dE = k_B T dx$$

and

$$E = k_B T x$$

then

$$\frac{N}{V} = \frac{8\pi}{h^3 c^3} \int_0^\infty \left(\frac{(k_B T x)^2}{e^x - 1} \right) k_B T dx$$

and the limits don't change

$$\frac{N}{V} = \frac{8\pi (k_B T)^3}{h^3 c^3} \int_0^\infty \left(\frac{x^2}{e^x - 1} \right) dx$$

and we are back to looking up integrals in tables or integrating numerically. Numerically I got

$$\int_0^\infty \left(\frac{x^2}{e^x - 1} \right) dx = 2.40411$$

so then

$$\frac{N}{V} = \frac{8\pi (k_B T)^3}{h^3 c^3} (2.40411)$$

Now we need the energy density in the volume. Fortunately we did this earlier as well! We once again start with

$$dN = N(E) dE = V g(E) f(E) dE = V \frac{8\pi}{(hc)^3} E^2 \frac{1}{e^{\frac{E}{k_B T}} - 1} dE$$

The photons will have different energies. We know that the number of photons with energy between E and $E + dE$ would be $\int EdN$ but for small dE we could approximate this by just $EdN = EN(E) dE$. So the energy density in the cavity (universe) would be.

$$\begin{aligned}u(E) dE &= \frac{EN(E) dE}{V} \\ &= \frac{8\pi}{(hc)^3} E^3 \frac{1}{e^{\frac{E}{k_B T}} - 1} dE\end{aligned}$$

and the total energy density is

$$\begin{aligned} U &= \int_0^\infty u(E) dE \\ &= \int_0^\infty \frac{8\pi}{(hc)^3} E^3 \frac{1}{e^{\frac{E}{k_B T}} - 1} dE \\ &= \frac{8\pi}{(hc)^3} \int_0^\infty \frac{E^3}{e^{\frac{E}{k_B T}} - 1} dE \end{aligned}$$

Another integral, this time from a table we find

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15}$$

and we can use this with

$$\begin{aligned} x &= \frac{E}{k_B T} \\ x k_B T &= E \end{aligned}$$

and

$$\begin{aligned} dx &= \frac{dE}{k_B T} \\ k_B T dx &= dE \end{aligned}$$

so that

$$\begin{aligned} U &= \frac{8\pi}{(hc)^3} \int_0^\infty \frac{(x k_B T)^3}{e^x - 1} k_B T dx \\ U &= \frac{8\pi (k_B T)^4}{(hc)^3} \int_0^\infty \frac{x^3}{e^x - 1} dx \\ U &= \frac{8\pi (k_B T)^4}{(hc)^3} \left(\frac{\pi^4}{15} \right) \\ U &= \left(\frac{8\pi^5 (k_B)^4}{15 (hc)^3} \right) T^4 \end{aligned} \tag{35.1}$$

Now we have the number of photons and the energy density, We just need to divide to get the average energy per photon per unit volume

$$\begin{aligned} E_{ave} &= \frac{U}{N/V} = \frac{\left(\frac{8\pi^5 (k_B)^4}{15 (hc)^3} \right) T^4}{\frac{8\pi (k_B T)^3}{h^3 c^3} (2.40411)} \\ &= 0.02773 \pi^4 T k_B \\ &= 0.02773 \pi^4 T \left(8.617385 \times 10^{-5} \frac{\text{eV K}^{-1}}{\text{photon}} \right) \\ &= 2.3277 \times 10^{-4} T \frac{\text{eV}}{\text{K (photon)}} \end{aligned}$$

So the average energy of the photons left from the big bang is temperature dependent.
²¹

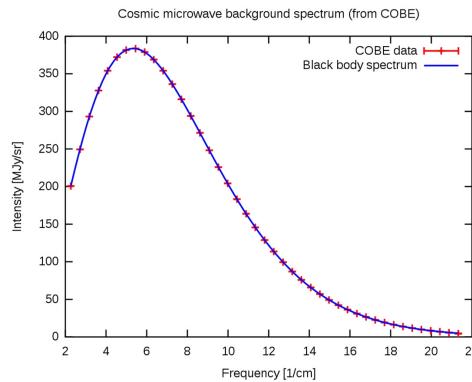
Measurements of this *cosmic background radiation* were done by accident when

²¹ We kind of ignored our A_{BE} constant. But it would have affected both the numerator and the denominator, so it are sort of OK in our approximation that $A_{BE} = 1$.

microwave communications dishes found an annoying hum on their communications. Isolating the hum proved difficult because the hum came from everywhere! And that is just what the big bang theory predicted. The universe should be a thermal system radiating but by now with a much cooler temperature. If we use the measured 2.7 K as the temperature we get

$$\begin{aligned} E_{ave} &= 2.3277 \times 10^{-4} (2.7 \text{ K}) \frac{\text{eV}}{\text{K (photon)}} \\ &= 6.2848 \times 10^{-4} \frac{\text{eV}}{\text{photon}} \end{aligned}$$

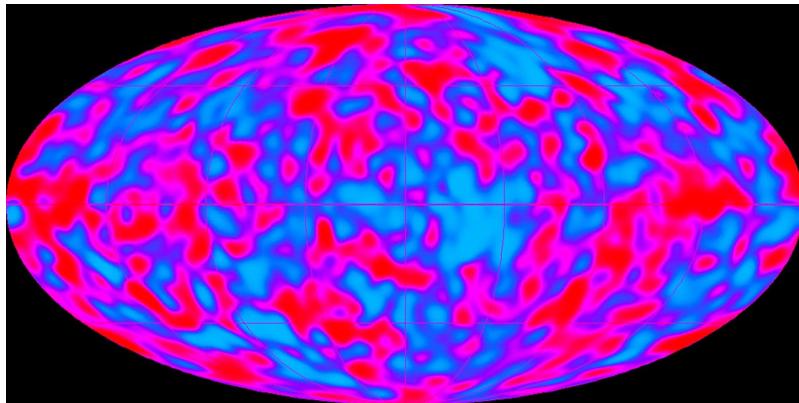
Plotting the predicted and measured intensity vs. frequency we see the cosmic background data fits the model of a thermal radiator very well.



Cosmic Microwave Background (CMB) spectrum plotted in waves per centimeter vs. intensity. The solid curve shows the expected intensity from a single temperature blackbody spectrum, as predicted by the hot Big Bang theory (Image courtesy NASA)

It was largely this measurement that solidified acceptance of the Big Bang Theory.

In 1989 NASA flew the Cosmic Background Explorer (COBE) mission that studied the cosmic microwave background radiation. The question to be solved was why is matter clumped into galaxies and stars and planets? Why would the big bang not be uniform. This is still a question in astrophysics, but if the clumpiness was fundamental to the big bang, itself, then there should be a clumpiness to the microwave background. COBE found that clumpiness. So cosmologists now work to explain this fundamental inhomogeneity in their models of the creation of the universe.



The extremely faint cosmic microwave background fluctuations are only one part in 100,000 compared to the 2.73 degree Kelvin average temperature of the radiation field. The cosmic microwave background radiation is a remnant of the Big Bang and the fluctuations are the imprint of density contrast in the early universe. The density ripples are believed to have given rise to the structures that populate the universe today: clusters of galaxies and vast regions devoid of galaxies. (https://lambda.gsfc.nasa.gov/product/cobe/dmr_image.cfm)

Dark Matter

We are familiar with pictures of galaxies full of stars. The stars are easy to see because they glow. But until 1992 we didn't know for sure that there were any non-glowing planets outside our solar system. That seems amazing when you consider audiences were treated to imaginary adventures of a young man and his droids on planet called Tatooine back in 1977! The matter that does not glow is called "dark." And this was the original version of "dark matter."

From way back in PH121 (or your introductory mechanics class) you learned Kepler's laws. Kepler's third law tells us that the period, T of an orbit is proportional to the distance from the center of the orbit r cubed. For a circular orbit this is

$$T^2 = \left(\frac{4\pi^2}{GM} \right) r^3$$

where G is the gravitational constant and M is the mass of the object being orbited. The orbiting object must have an orbital speed of

$$v = \frac{2\pi r}{T}$$

for a circular orbit. and we can find that orbital speed from Newtonian dynamics. The mass of the object being orbited would be the central mass of the galaxy plus all the galactic mass closer to the center than the star we are observing. Then the gravitational force on the star due to the galaxy would be

$$W_{SG} = G \frac{M_G m_S}{r_{GS}^2}$$

and from Newton's second law we know

$$F_{net_r} = m_s a_r$$

but we only have one force so

$$m_s a_r = G \frac{M_G m_s}{r_{GS}^2}$$

and W_{SG} is toward the center of the orbit so it is centripetal, thus

$$a_r = a_c = \frac{v_t^2}{r_{GS}}$$

or

$$m_s \frac{v_t^2}{r_{GS}} = G \frac{M_G m_s}{r_{GS}^2}$$

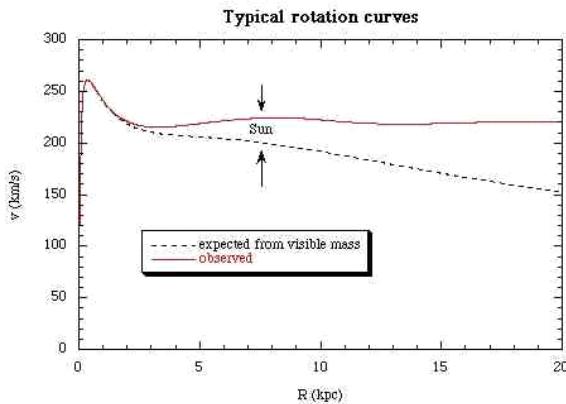
and some things cancel

$$v_t^2 = G \frac{M_G}{r_{GS}}$$

leaving the orbital speed

$$v_t = \sqrt{G \frac{M_G}{r_{GS}}}$$

This might be a little too simple, but it gives an estimate of the rotation speed of stars in a galaxy. The rotation rate should go down with r_{GS} for stars on the outer part of the galaxy. Near the center there is little mass closer to the center so for small r_{GS} the M_G grows with r_{GS} so for small (on a galactic scale) distances we expect the speed to increase with distance. Somewhere between the two extremes, there should be a peak velocity. Observations of stars do give the expected behavior near the center of the galaxy, and there is a maximum velocity. But past the maximum, the curve flattens out.



<https://www-sopof.gsfc.nasa.gov/stargaze/Sun4Adop3.htm>

One way to explain this departure from theory would be to assume that there is more mass in the flat galactic disk that we can account for by just counting the bright glowing stars. This dark matter would be distributed in a "halo" out away from the galactic center.

Now that we can find exoplanets, there do not seem to be enough of them (normal baryonic dark matter sources) to make up for the large discrepancy between prediction and reality. The dark matter content of a galaxy would need to be about 10 times the mass of the stellar objects. If our solar system (and the few exoplanet systems we know) are normal, then planets are not enough to fix the prediction.

Two solutions have been proposed. One called Massive Compact Halo Objects (MACHOs) postulate that we should find black holes, neutrons stars, white dwarf stars, and other massive but dark objects in the galactic halo.

The second proposed solution is that there are Weakly Interacting Massive Particles (WIMPs) that are not baryonic that could congregate in the galaxy. Neutrinos are the right kind of particle, except that neutrinos have very little mass. Particle physicists are looking for good WIMP candidates, but none have been found yet.

We now have a description of a universe, we need to patch our theory of relativity to start the job of building a universe to specification. We will do that in the next lecture.

36 General relativity

15.4-15.6

Fundamental Concepts in the Lecture

- General Relativity and gravity
- All reference frames are equivalent
- Gravitational fields are equivalent to accelerated reference frames
- Curved spacetime

So we described the universe, but we said that universe had an expansion of space. How can that be? Before we build a universe, we need to modify our understanding of space. In Newtonian physics, space is just a place where things happen. But for General Relativity, space is a lot more exciting.

General relativity

Einstein was not content with Special Relativity. It bothered him that his theory couldn't handle accelerating reference frames. And if you remember, it bothered us that Newtonian gravity didn't work in Special relativity. There was something missing. He spent the next decade after publishing Special Relativity working on the problem. His answer was the theory of General Relativity. General relativity has two fundamental ideas

1. All the laws of nature have the same form for observers no matter whether their frame of reference is accelerating or not.
2. The gravitational field is equivalent to an accelerated frame of reference without a gravitational field.

The first one lets us do problems like the twin paradox problem correctly accounting for the acceleration (it is a wonderful nasty problem, I did it on my comprehensive exams for my Ph.D.). It is an extension of Special Relativity where we said all inertial reference frames were equivalent. Now all reference frames, inertial or not, will be equivalent.

The second idea is more subtle. It says that if I put you in a box sitting on the Earth so you feel gravity, you could not tell the difference between this and if I am pulling the box, making it accelerate.

Let's look at these postulates in more detail.

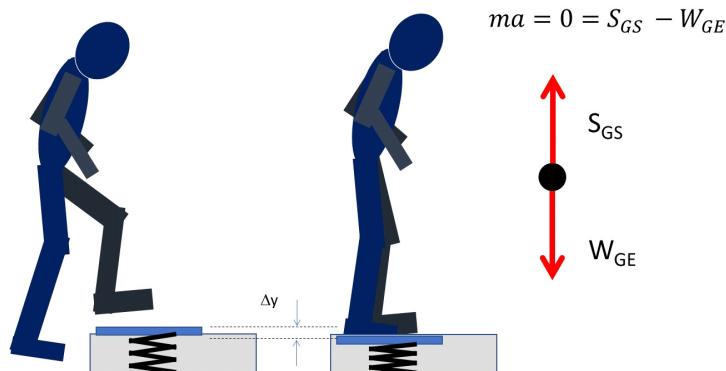
I said that Einstein didn't like the limitation built into special relativity that you could only use inertial reference frames. Einstein wanted any reference frame to be equivalent. That is, he wanted the laws of physics to be the same in any reference frame, even an accelerating reference frame. But how do these new postulates help?

In special relativity, to make c a constant in any inertial frame we had to give up the absoluteness of space and time. We got length contraction and time dilation. To make c constant in any reference frame, even accelerating reference frames, we will have to give up more classical ideas. But because we have watched enough science fiction, they are not strange ideas to us. A major thing we need to give up is viewing space and time as a container in which things happen. Space and time (combined into one, spacetime) are malleable, warpable, flexible and stretchable things. Spacetime can act on an object, and it is an active player in the universe. All this comes from the first postulate.

The second postulate directly addresses the missing law of gravity in special relativity. Special relativity didn't include an update to Newton's law of gravity, but it is immediately apparent that the Newtonian form when combined with special relativity did not give good predictions.

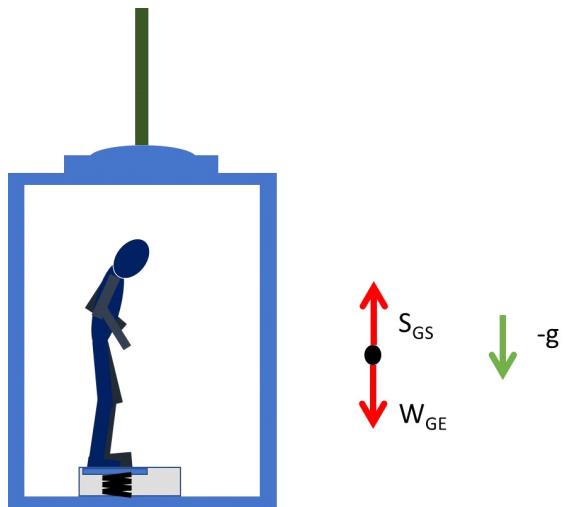
General relativity fills that void with a new theory of gravity that depends on both postulates. Let's see how it works.

Let's start by remembering Newton's second law. This is one of the laws of physics that we want to be true in any reference frame. Here is a person standing on a spring scale.



The person will come to equilibrium so the net force $F_{net} = ma = 0$ and we have two vertical forces, a spring force from the scale, and a weight force. Now suppose we put our person in a box that is sealed so the person can't see the environment around the

box.

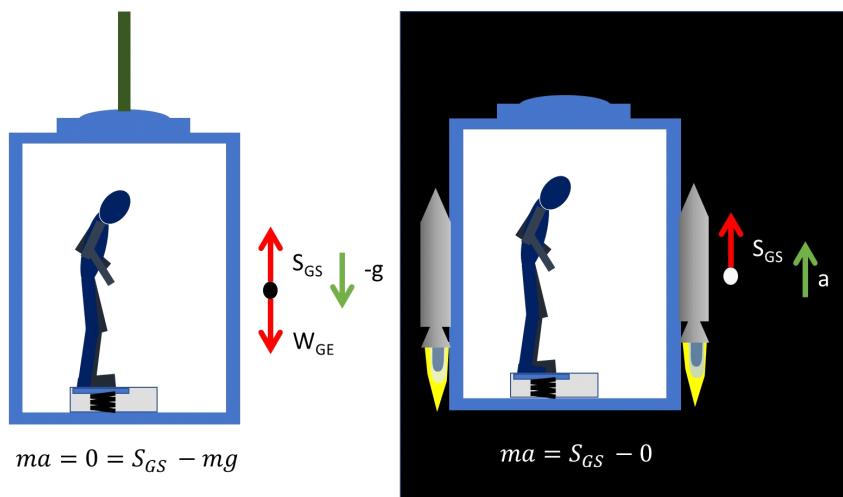


Here the box is suspended on a cable, and the person in the box would see the same weight on the scale as before. Classically we would say that this is because the box and its contents are in a gravitational field indicated by the gravitational acceleration $-g$. Newton's second law tells us that the scale would register a force

$$\begin{aligned} ma &= 0 = S_{GS} - W_{GE} \\ S_{GS} &= W_{GE} \\ &= mg \end{aligned}$$

and our person could write down this reading from the scale.

But suppose we move our box far away from everything else in the universe, and attach rocket boosters to our box, and accelerate the box through space.



The person in the box would see a reading on the scale

$$ma = S_{GS} - 0$$

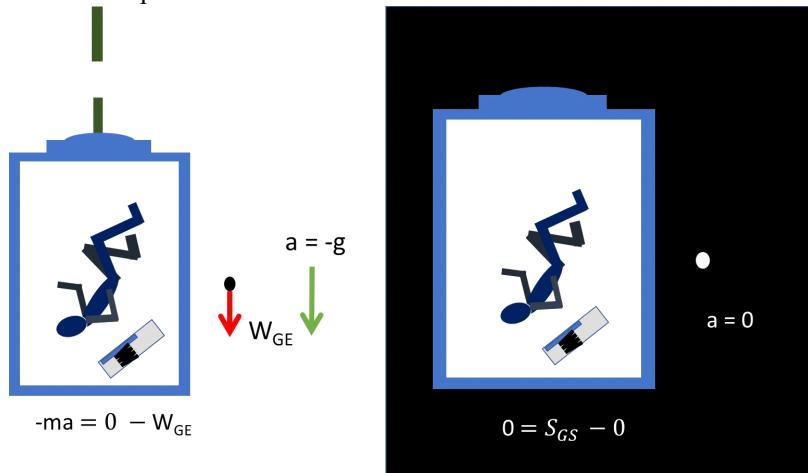
$$S_{GS} = ma$$

and if the box accelerates with $a = +g$ the person in the box would get exactly the same reading on the scale as the person saw when the box was hanging.

$$S_{GS} = mg$$

Einstein thought of these two experiments, and noted that the person in the box could not tell the difference between being in a gravitational field and being accelerated. And he asked the question, are the two situations different? Maybe a gravitational field *is* an acceleration.

Let's take another case, say we cut the cable for our Earth-bound box, and remove the rockets from our space box.



In the case of cutting the cable we would have a scale reading of

$$-ma = S_{GS} - W_{GE}$$

$$-mg = S_{GS} - mg$$

$$0 = S_{GS}$$

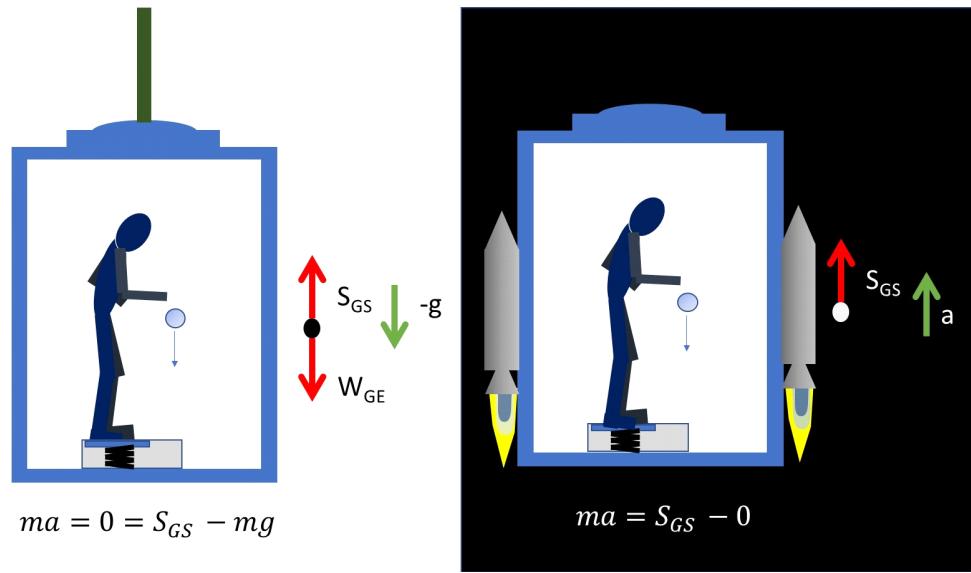
There would be no spring force. In the case of being in space with no rockets,

$$0 = S_{GS} - 0$$

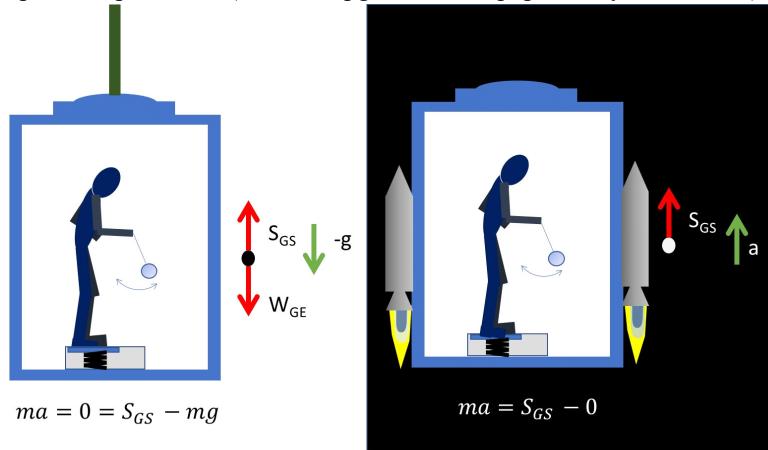
$$S_{GS} = 0$$

and once again the person could not tell the difference. Einstein decided that if we consider gravitation and acceleration equivalent, then all the laws of physics would already work in any coordinate system!

But you might object. If we gave the person some scientific equipment they could surely tell if they were accelerating or if they were in a gravitational field. So let's try that. A simple beginning physics lab experiment might include a ball.

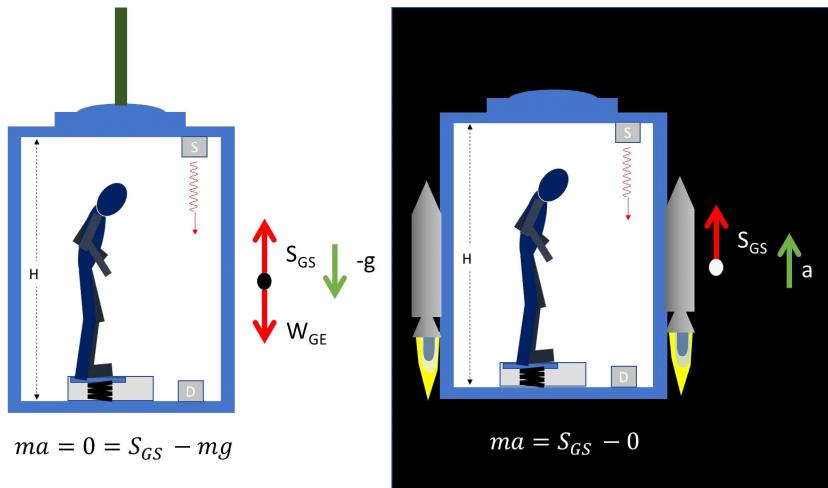


Would the ball bounce the same way in both cases? The answer is yes! Suppose we give the person a pendulum (we have upgraded our equipment by a semester!)



We would find that the pendula swing the same, but with one provision. The gravitational field of the Earth is not uniform. So if we make our box big enough we would see that the gravitational field lines are closer together at the bottom of the box than they are at the top. But if we have a sufficiently small box (or a real uniform gravitational field) the person could not tell the difference.

We could upgrade our equipment to an electrodynamics experiment.



If we have a light source at the top of the box and a detector at the bottom we could detect the light and measure its frequency. And here you might think there should be a difference. Shouldn't there be a Doppler shift for the case of the accelerating box? The light will leave the detector and the detector will move toward the light. That is exactly the situation that produces a Doppler shift.

$$f' = f \sqrt{\frac{1 + \frac{\Delta v}{c}}{1 - \frac{\Delta v}{c}}}$$

In our box, Δv would be small if the acceleration is equal to g , so $\Delta v/c$ will be very small so

$$\sqrt{1 - \frac{\Delta v}{c}} \approx 1$$

and so

$$f' = f \sqrt{\frac{1 + \frac{\Delta v}{c}}{1 - \frac{\Delta v}{c}}} \approx f \sqrt{1 + \frac{\Delta v}{c}}$$

and we can approximate the numerator as not too different than $1 + \frac{\Delta v}{c}$ without too much error if Δv is small. so

$$f' \approx f \left(1 + \frac{\Delta v}{c}\right)$$

$$f' \approx f + f \frac{\Delta v}{c}$$

then

$$\frac{f' - f}{f} \approx \frac{\Delta v}{c}$$

and if we start our box with $v_i = 0$ then $v_f = 0 + a\Delta t$

$$\frac{\Delta f}{f} \approx \frac{a\Delta t}{c}$$

and our light is traveling at

$$c = \frac{H}{\Delta t}$$

so

$$H = c\Delta t$$

so then

$$\frac{\Delta f}{f} \approx \frac{aH}{c^2}$$

But how about the box hanging in the gravitational field? If Einstein is correct, then we would predict that the person in the box sitting on the Earth should see a frequency shift as well! The experiment was done in 1959 at Harvard, and there was such a frequency shift!

This really destroys our theory of gravitational fields! But it really works. The GPS satellites direct a beam of light at the earth (radio waves) and as the beam travels downward the frequency shifts. The amount is tiny, but relevant to the operation of the GPS system! If we don't account for the frequency shift, your GPS in your phone would not work.

We could see this shift more easily if we were on a more dense planet or star.

Note that near the Earth we can use

$$U = -mgh$$

but for a larger h we have to use

$$U = -G \frac{mM}{r} = m \left(-G \frac{M}{r} \right)$$

In our box formula we recognize $aH = gh$ as the gravitational potential energy per unit mass. We could fix our formula by replacing gH with $(-G \frac{M}{r})$ for a distance comparable to the size of the star or planet. Then our frequency shift would be

$$\frac{\Delta f}{f} \approx G \frac{M}{c^2 r}$$

So our gravitational and accelerating cases are the same! And this new view point predicts a frequency shift that exists, but our old gravitational field model misses!

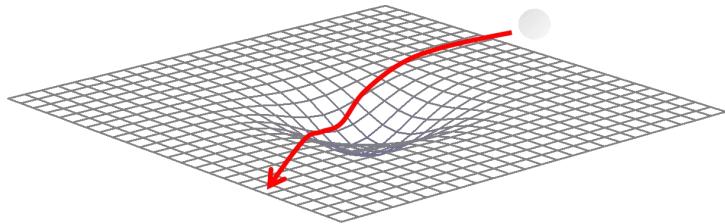
This is important. We are saying that there is no gravitational field.

The equating of gravitation and accelerated reference frames is called the *principle of equivalence* and really it was hiding in our Newtonian physics in a way. We have said for semesters now that the mass that makes things hard to accelerate

$$F_{net} = ma$$

is the same as the mass that causes gravitation. This is a statement that acceleration is equivalent to gravitation!

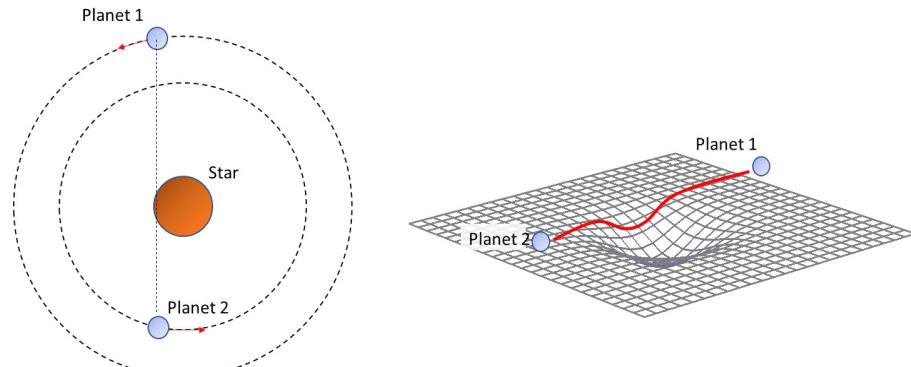
But we had to give up something classical to make all this work. What we gave up was a container spacetime. Einstein's spacetime warps and bends. A two dimensional analog of a warped three dimensional space is shown below.



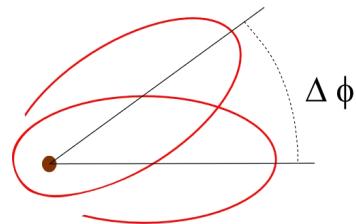
A particle traveling along this space will change direction because of the warped part. But what can warp space?

Einstein postulated that mass warps space. So the situation in the last figure could be a photon passing by a super heavy object like a star. This warping of space was the first experimental evidence for general relativity. The experiment took place during a total eclipse of the sun. Stars that were behind the sun became visible earlier than they should have because their light was bent in the warped space due to the mass of the Sun.

An artificial version of this experiment can be done by emitting a radio beam at another planet as it orbits the Sun

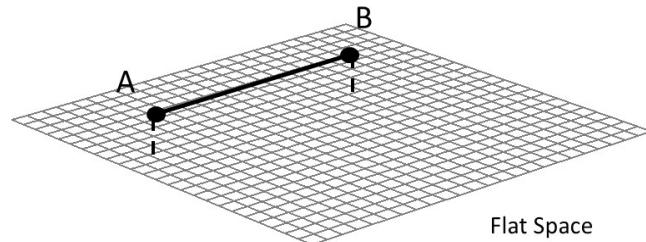


The light has to travel farther because it must traverse the warped region near the Sun, so there is a measurable delay in the signal. The closer to the mass you get, the more warped the space is. We would expect, then, to see an effect of this warping of space in the motion of planets that are close to the Sun. And we do! The orbit of Mercury is affected by the warping of space when it is nearest the Sun at its perihelion. In the next figure the eccentricity of an orbit is exaggerated. An orbit without space warping would be a perfect ellipse. But if the space is warped the planet must travel through more space in the warped part, and this delays the planet a little causing the ellipse to deform. The orbit *precesses* or wobbles into a new direction.

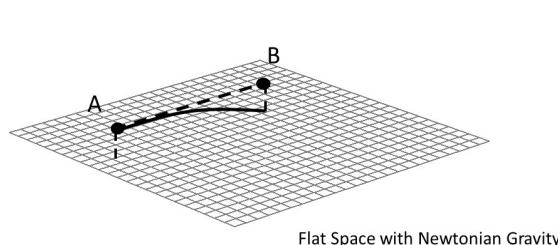


The change in direction is small, for Mercury it is only of 10^{-6} rad per orbit, but the effect is cumulative, making the orbit more off every time the planet goes around the Sun. Measuring this in Mercury's orbit is hard, because there are other reasons for a precession of Mercury's orbit due to the gravitational effect of the other planets. But in the orbit calculations there was always a missing piece, until Einstein.

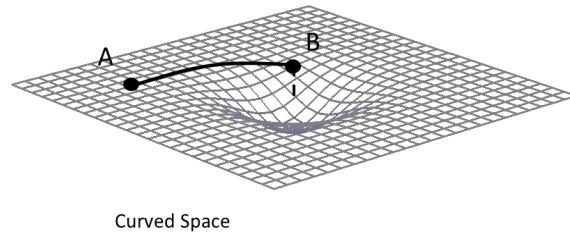
But there is more than just the thought that mass curves space. Look at the following figure. A particle moves in a flat space from *A* to *B*.



If the flat spacetime was in a classical gravitational field then the particle would “fall” as it traveled toward *B*.



But we get the same result if we allow a curved space.



Here the particle isn't attracted to the mass that causes the curve. The particle is just following along the curved space. But it looks just like the Newtonian gravity case. The implication is startling. There isn't a force of gravity! The apparent force of gravity is just movement of objects in a curved space time. The reason it looks like the objects are attracted to a mass is because the mass caused the curvature of the space, but not because the objects actually attract each other.

Indeed we did give up a cherished classical notion to make all reference frames equivalent. We gave up the force of gravity!

Curved spacetime

But what is a curved spacetime like? A small distance in space time is symbolized by ds . In general relativity we define

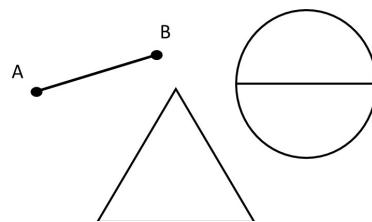
$$(ds)^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2$$

This quantity is invariant under the Lorentz transformation. So it is a good way to describe a change that does not depend on reference frame. Note that we have included time in our ds vectors (with a c to make distance units).

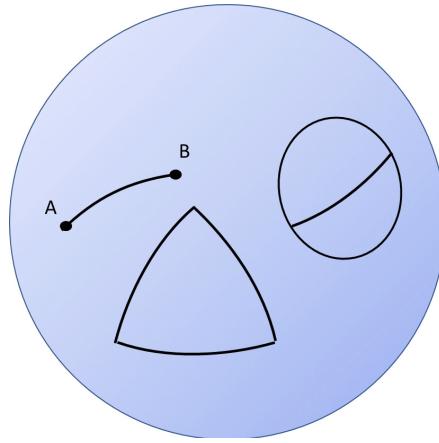
This definition of ds is for a flat spacetime. But in general relativity, spacetime is something that can stretch and bend. If our spacetime is curved we would have curvature coefficients.

$$(ds)^2 = g_0 (cdt)^2 - g_1 (dx)^2 - g_2 (dy)^2 - g_3 (dz)^2$$

The mathematics of general relativity is beyond what most of us have had. You need tensor calculus. But we can look at the results of thinking of the consequences of a curved spacetime. and we need to think about this to build our universe. If we draw shapes in a flat spacetime.



They will look different in a curved space time.

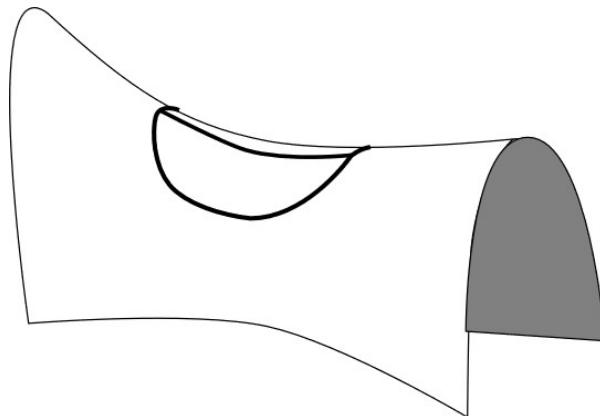


In this curved space time straight lines have become curved due to the curvature of the space. So triangles have interior angles that sum to greater than 180° and the diameter line of a circle is now longer so

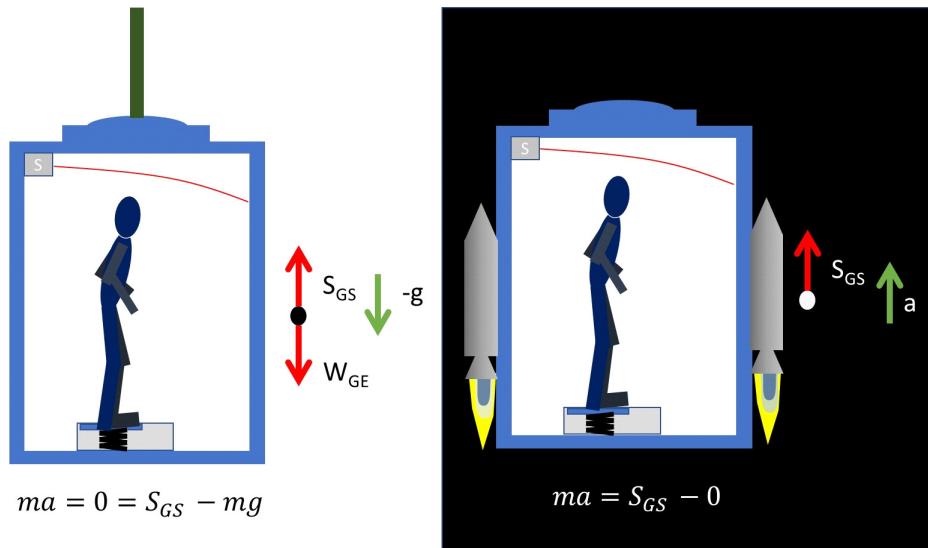
$$\frac{C}{D} < \pi$$

We could envision another way to curve spacetime where the circumference would grow so that

$$\frac{C}{D} > \pi$$



Let's look at this curvature in the context of our box experiment. If we have a light source that crosses the box, then the moving box would see the light hit lower on the opposite wall. In a warped space, the same would happen due to the curve of the space.



There must be a mass causing that curvature, but the light travels the curved path because the space it is in is curved. The light is not attracted to the mass. The light travels through space as it always does. But because of the mass, the space that is curved where the light is, so we see the light travel a curved path.

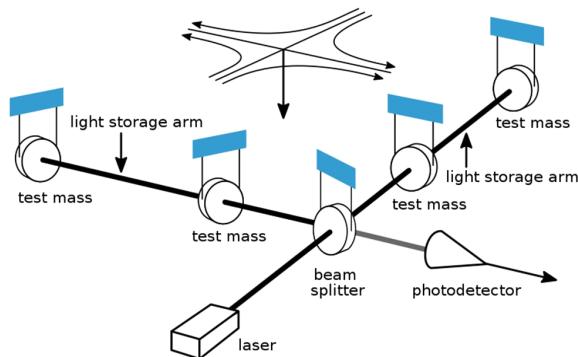
We can describe the curved spacetime with Einstein's equations. In the notation of this class we would write

$$\text{curvature of space} = \frac{8\pi G}{c^4} (\text{mass-energy density})$$

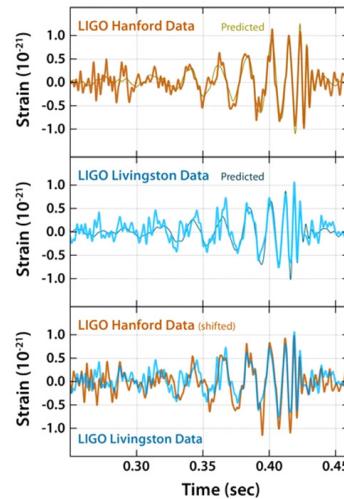
Why not use more symbols? The symbols use a notation we have not yet learned. If general relativity interests you, stay in physics and plan on graduate school. General relativity is usually a graduate level course.

Gravitational waves

In Einstein's general relativity a moving mass can make waves in the stretchy spacetime. The waves travel at the speed of light. But for normal masses (like stars and planets) the waves are very small. However, large events like the collision of two black holes or neutrons stars could make a measurable wave. The LIGO detector first found gravitational waves in 2018.



Concept of LIGO: A schematic diagram of a laser interferometer with light storage arms.

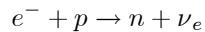


The top two plots show data received at Livingston and Hanford, along with the predicted shapes for the waveform. These predicted waveforms show what two merging black holes should look like according to the equations of Albert Einstein's general theory of relativity, along with the instrument's ever-present noise. Time is plotted on the X-axis and strain on the Y-axis. Strain represents the fractional amount by which distances are distorted.

Stellar Evolution

So now, with all of our studies, we can finish the story of the life of a star. We know from before that stars fuse elements starting with hydrogen. As the hydrogen is exhausted, helium is fused, and so on up until iron is produced. Then neutron capture can make heavier elements.. But at each stage, the star collapses inward and the kinetic

energy of the collapse provides the thermal energy needed to induce the fusion of heavier elements. But at some point the gravitational pull of the star's mass demands further collapse. If a star is about 1.4 times the mass of the sun, then the gravitational squashing of the star's matter (remember it is the contraction of space that is pulling the material together!) can overcome the Pauli force. The Pauli exclusion principle can hold a star from collapsing if the star's mass is less. But above 1.4 solar masses gravitation wins and the star further collapses. The 1.4 solar mass cutoff is called the *Chandrasekhar limit*. With a Fermi energy of about 0.3 MeV, the electrons on the higher energy end of the Fermi distribution will have enough kinetic energy to pass the threshold of the reaction

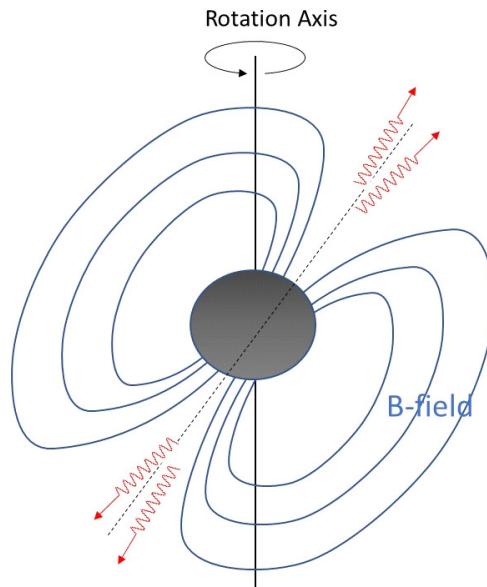


and then the protons in the atoms begin to turn into neutrons. This will destroy atoms. This causes a further collapse, and this fuels more proton to neutron conversion. Eventually the star is mostly made of neutrons. It is sort of one giant nucleus.

Neutrons stars

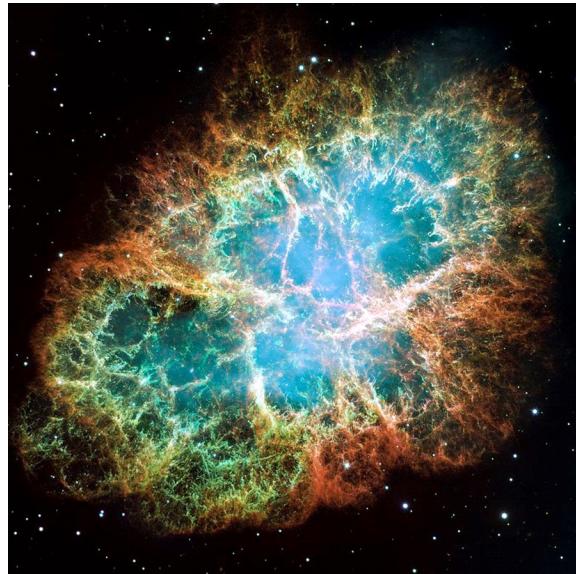
Neutrons are still fermions, and they also obey Pauli exclusion principle and the Pauli force can keep the star from further collapse for a while. This what we call a neutron star.

This seems very exotic, but neutrons stars are relatively easy to observe. This is because the neutron stars create strong magnetic fields. And the neutron star collapse increases the rotation rate of the star (like a solar sized ice skater pulling in arms and legs to increase the spin rate).



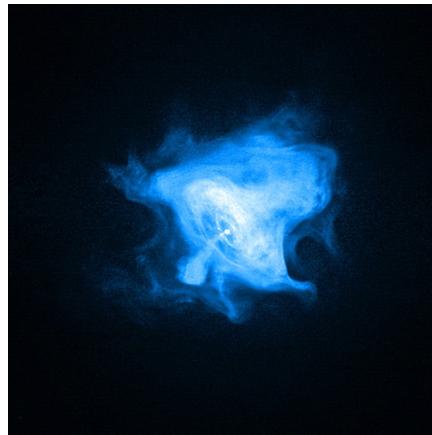
Charged particles will be propelled out along the magnetic field axis, and the star will spin on its spin axis. This makes a rotating beacon of radio frequency (and other frequency) waves due to the accelerated charged particle beam. As the star turns this beacon of radio frequency waves periodically hits Earth. Radio receivers hear a “pulse” repeated over and over. These pulsating neutron stars are called *pulsars*. Pulsars were discovered by Jocelyn Bell in 1976 when she was a graduate student. She found a regular series of radio frequency bursts that were found to be the rotating beacon of a neutron star.

The details of the collapse of a star into a neutron star seems to be a violent episode. Pulsars are often found in the middle of nebulae, which are the result of supernova. The Crab Nebula is an example.



NASA composite image of the Crab Nebula

In the center of the nebula is a pulsar. The Candra X-ray Observatory has imaged this pulsar.



The Crab Nebula's X-ray-emitting pulsar wind nebula. The pulsar is the white dot near the center of the image. The polar jets of charged material are visible, as well as a cloud of particles with the pulsar rotation in the middle which generates a particle "wind." (image courtesy

NASA/CXC/SAO/F.Seward et al, <https://chandra.harvard.edu/photo/2008/crab/>)

The images show the jet of charged particles emanating from the pulsar.

Black Holes

If a star is very massive, the gravitational squishing overcomes the neutron Pauli force, and the star further collapses. This time with little to stop the collapse. General relativistic equations are used to describe this collapse. Our spactime interval in spherical coordinates for a black hole would be

$$(ds)^2 = \left[1 - \frac{2GM}{c^2 r} \right] (cdt)^2 - \frac{1}{\left[1 - \frac{2GM}{c^2 r} \right]} (dr)^2 - r^2 (d\theta)^2 - r^2 \sin^2 \theta (d\phi)^2$$

where the terms in square brackets would reduce to 1 for low gravity. But black holes are *not* low gravity. Noted that in the radial term there is a problem. If

$$\left[1 = \frac{2GM}{c^2 r} \right]$$

then the equation has an infinity. We could rewrite this as

$$r = \frac{2GM}{c^2}$$

At this distance from the center of the black hole something must happen. But for a probe falling into a black hole it wouldn't notice anything different at this distance from the black hole. That is, nothing until it tried to get back out.

For observers outside the black hole, they would see the probe falling, but the fall would appear to get slower and slower. The probe would seem frozen in place and the gravitational warping of space would make the light from the probe ever more red shifted until we could no longer detect it. The probe would disappear.

Meanwhile the probe might have been programed to cross this distance and turn around

and fire it's rockets. But it would be to no avail. The escape velocity goes infinite at this special distance away from the center. Nothing can escape once it crosses this line, not even light. So we would have lost contact with the probe as well. The name of this distance away from the center of a black hole is the *Schwarzschild radius* but you may have heard of it referred to as the *event horizon*.

So what is the state of the matter inside a black hole? No one is really sure. At some point the neutrons collapse, but what form the matter makes is a mystery that is the subject of active research. The collapse of a black hole to a singular point would violate our quantum uncertainty principle. So most physicists believe that the collapse must end before this happens.

This is the part of physics called quantum gravity, where gravitation has shrunk size into the quantum realm. There are several quantum gravity hypothesis (string theory, loop quantum gravity, asymptotically safe gravity, causal dynamical triangulation, and emergent gravity and others).

Evaporation of black holes

in 1974 a young physicist found another way to combine general relativity and quantum behavior. He envisioned pair production right at the Schwarzschild radius. The tremendous potential energy just outside of the black hole can provide the energy for this to happen in the quantum sense that the pair can be produced and fall providing the energy needed to create the pair so long as this all happens within

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

(remember these are waves, so we have a large range of available energies if Δt is very small) so we can get a positron and an electron created just outside the Schwarzschild radius falling inward. If one of the pair (preferably the electron) has enough energy to escape, this leaves the positron to enter the black hole.

It took energy to do this, so effectively we will reduce the mass of the black hole. Eventually (like in the age of the universe eventually) the black hole will lose all of its mass. This is known as *black hole evaporation*.

Black holes are an active area of research. If they intrigue you, a graduate level degree in physics could lead to a career studying black holes. They may be a part of the solution to the dark matter problem. It is likely that there are supermassive black holes in the center of the Milky Way and other galaxies. Could there be black holes in the halo as well?

37 General Relativity and Cosmology

15.7 - 15.8

Fundamental Concepts in the Lecture

- Friedmann equation
- Mass dominated and photon dominated estimates for the age of the universe
- The history of the universe is governed by energy density because bonding of particles is governed by available energy to break bonds.

Size and Shape of the Universe

Now that we have some understanding of general relativity, Let's go back to describing our universe. Let's start with our general relativistic equation

$$\text{curvature of space} = \frac{8\pi G}{c^4} (\text{mass-energy density})$$

To understand how to make a universe, we should know how big the universe is, and how curved it is. The mass-energy term has to be all the mass energy of the whole universe, and on this scale we are not too worried about local density changes. Of course, if we have an expanding universe, the overall density of the universe must be decreasing. We won't go through the details, but Friedmann formed a solution to the general relativistic equation for the whole universe considering only large scale structure. He got

$$\left(\frac{dR}{dt} \right)^2 = \frac{8\pi}{3} G \rho R^2 - k c^2$$

where R is the distance scale factor (a characteristic size for the universe), G is our old friend, the universal gravitational constant, ρ is the mass-energy density of the universe, and k specifies the curvature structure of the universe.

If $k = 0$ the universe is flat. If $k = +1$ then the universe is closed and curved. If $k = -1$ the universe is open and curved. Because $R = R(t)$ if k is anything but exactly zero then the time dependence of $R(t)$ will increase the curvature over time. Note that

we are talking about the whole of the universe here. It is true no matter what k is that space is warped by matter.

Measurements tell us that if k is not zero, it is not very much different than 0, so let's take the case of a flat spacetime for a moment. For a rough estimate, we can take the amount of matter as constant (which is very close to being true) and assume the universe is homogenous, then

$$\rho_m \propto R^{-3}$$

since R is a length scale, and we assumed any part of the universe was pretty much like any other part. So exactly what length R is doesn't matter much. Then

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi}{3}G\rho R^2 - kc^2$$

becomes

$$\begin{aligned} \left(\frac{dR}{dt}\right)^2 &\propto \frac{8\pi}{3}G(R^{-3})R^2 - (0)c^2 \\ \left(\frac{dR}{dt}\right)^2 &\propto \frac{8\pi}{3R}G \\ \frac{dR}{dt} &\propto \sqrt{\frac{8\pi}{3R}G} \end{aligned}$$

assuming that the amount of mass/energy is not changing, we can fix the proportionality with a constant C which represents the total mass/energy of the universe.

$$\rho_m = \frac{C}{R^3}$$

so

$$\frac{dR}{dt} = \sqrt{\frac{8\pi C}{3R}G}$$

We can use this to estimate an age of the universe. Let's rearrange

$$\sqrt{R}dR = \sqrt{C\frac{8\pi}{3}G}dt$$

and we can integrate this

$$\begin{aligned} \int_0^R \sqrt{R}dR &= \sqrt{C\frac{8\pi}{3}G} \int_0^t dt \\ \frac{2}{3}R^{\frac{3}{2}} &= \sqrt{C\frac{8\pi}{3}G}t \end{aligned}$$

so that

$$\begin{aligned} R &= \left(\frac{3}{2}\sqrt{C\frac{8\pi}{3}G}t\right)^{\frac{2}{3}} \\ &= At^{\frac{2}{3}} \end{aligned}$$

where

$$A = \left(\frac{3}{2} \sqrt{C \frac{8\pi}{3} G} \right)^{\frac{2}{3}}$$

Then we can get an expression for the derivative of R in the Friedmann equation.

$$\frac{dR}{dt} = \frac{2}{3} A t^{-\frac{1}{3}}$$

And we could put this back into Friedmann's equation. Then we can eliminate R and find a time for the universe to have existed.

$$\left(\frac{dR}{dt} \right)^2 = \frac{8\pi}{3} G \rho R^2 - k c^2$$

becomes

$$\left(\frac{2}{3} A t^{-\frac{1}{3}} \right)^2 = \frac{8\pi}{3} G \rho \left(A t^{\frac{2}{3}} \right)^2 - (0) c^2$$

and we can solve this for t to get our estimate for the age of the universe

$$\begin{aligned} \frac{4}{9} A^2 t^{-\frac{2}{3}} &= \frac{8\pi}{3} G \rho A^2 \left(t^{\frac{4}{3}} \right) \\ \frac{1}{3} t^{-\frac{2}{3}} &= \frac{2\pi}{1} G \rho \left(t^{\frac{4}{3}} \right) \\ t^{-\frac{2}{3}} t^{-\frac{3}{4}} &= 6\pi G \rho \\ t^{-\frac{1}{2}} &= 6\pi G \rho \end{aligned}$$

$$t \approx \frac{1}{\sqrt{6\pi G \rho}} \quad (37.1)$$

for a mass-energy dominated universe. For the early universe this isn't so good because the early universe was photon dominated.

We can try to fix this using what we know about photon gasses. We know how to find the energy density of a gas of photons. We have done it several times. Let's use the result.

$$u(E) dE = \frac{8\pi hc}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} d\lambda$$

which says that our mass-energy density has a distance measure due to the wavelengths that is proportional to $d\lambda/\lambda^5$

$$\rho_r \propto \frac{d\lambda}{\lambda^5}$$

Because we assumed a homogeneous universe, all lengths scale like $R(t)$ so

$$\lambda \propto R(t)$$

$$\Delta\lambda \propto R(t)$$

which gives us a mass-energy density that is proportional to R^4

$$\rho_r = \frac{C}{R^4}$$

where once again we have C as a constant of proportionality that is the amount of

energy in the universe. We can put this into the Friedmann equation

$$\begin{aligned}\left(\frac{dR}{dt}\right)^2 &= \frac{8\pi}{3}G\left(\frac{C}{R^4}\right)R^2 - (0)c^2 \\ \left(\frac{dR}{dt}\right)^2 &= \frac{8\pi C}{3R^2}G \\ \frac{dR}{dt} &= \frac{1}{R}\sqrt{\frac{8\pi C}{3}}G\end{aligned}$$

so that

$$RdR = \sqrt{\frac{8\pi C}{3}}Gdt$$

and once again integrate

$$\int_0^R R dR = \int_0^t \sqrt{\frac{8\pi C}{3}}G dt$$

which once again gives us an expression for R

$$\begin{aligned}R &= \sqrt{2\sqrt{\frac{8\pi C}{3}}Gt^{\frac{1}{2}}} \\ &= Bt^{\frac{1}{2}}\end{aligned}$$

where

$$B = \sqrt{2\sqrt{\frac{8\pi C}{3}}G}$$

and we could find the derivative of R

$$\frac{dR}{dt} = \frac{1}{2}Bt^{-\frac{1}{2}}$$

and once again we can put this into the Friedmann equation

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi}{3}G\rho R^2 - kc^2$$

We find

$$\left(\frac{1}{2}Bt^{-\frac{1}{2}}\right)^2 = \frac{8\pi}{3}G\rho\left(Bt^{\frac{1}{2}}\right)^2 - kc^2$$

and solve for time to find an age of the universe estimate

$$\frac{1}{4t}B^2 = \frac{8\pi}{3}G\rho_r B^2 t$$

and the B terms cancel. We can solve for t

$$\frac{1}{4t} = \frac{8\pi}{3}G\rho_r t$$

$$\frac{1}{4\frac{8\pi}{3}G\rho_r} = t^2$$

$$t = \sqrt{\frac{3}{32\pi G\rho_r}} \quad (37.2)$$

for a photon dominated universe.

Of course our universe started as photon dominated and moved to mass dominated. So the actual time has to be somewhere in between our two age estimates. Note, that we assumed that the amount of mass/energy is not changing in the universe in our estimates for the age of the universe, If C was a function of time in either case, this would have been more difficult. We believe that energy can't leave the universe, so this is a reasonable assumption, but it is an assumption.

We won't show this, but we can write the Hubble parameter as

$$H = \frac{1}{R} \frac{dR}{dt}$$

and because $R(t)$ changes in time, so does H . So then, the current value we found, H_o , will change in time. If the expansion is at a constant rate, linear in t , then

$$R(t) \propto t$$

the age of the universe is

$$t_{age} = \frac{1}{H}$$

but we can see from our two estimates that $R(t) \propto t^{\frac{2}{3}}$ or $R(t) \propto t^2$. It turns out that in the first case (matter dominated) that we would get

$$t_{age} = \frac{2}{3} \frac{1}{H}$$

but that assumes that the universe was always matter dominated. The photon dominated case gives

$$t_{age} = \frac{1}{2} \frac{1}{H}$$

Either way, the age of the universe is about H^{-1} . The current best estimate is 13.7×10^9 y. And our two estimates are within about 20% of each other. Of course 20% of 13.7×10^9 y is a whopping 2.74×10^9 y. But in cosmology we can't let little things like three billion years upset us too much.

It remains as a problem yet to be solved to figure out what age is right. The fields of astrophysics and cosmology are closely related, and they are active fields of research. If these ideas intrigue you, these might be a good choices of fields of study for you.

A very brief history of time

Let's look at how the universe grew, according to our quantum and general relativistic models. Our expansion model tells us the universe started very small. Since all the energy that exists in the universe today was in the universe then, it must have also been hot. We found from our quantum theory that thermal energy will destroy atoms and particles, providing lots of dissociation energy to break all bonds, so this early universe was photon dominated. We can use our time equation for photon dominating (Equation 37.2) to find a relationship between time and temperature for this part of our history.

$$t \approx \sqrt{\frac{3}{32\pi G\rho_r}}$$

If we let

$$\rho_r = \frac{U}{c^2}$$

then we have

$$t = \sqrt{\frac{3c^2}{32\pi GU}}$$

and remember that we found the energy density U for a photon dominated universe when we described the cosmic background radiation (Equation 35.1).

$$U = \left(\frac{8\pi^5 (k_B)^4}{15 (hc)^3} \right) T^4$$

so our time is about

$$t \approx \sqrt{\frac{3c^2}{32\pi G \left(\frac{8\pi^5 (k_B)^4}{15 (hc)^3} \right) T^4}}$$

or, taking the temperature out of the square root

$$t \approx \sqrt{\frac{3c^2}{32\pi G \left(\frac{8\pi^5 (k_B)^4}{15 (hc)^3} \right)}} \frac{1}{T^2}$$

so that

$$T \approx \frac{D}{t^{\frac{1}{2}}}$$

where

$$D = \sqrt{\sqrt{\frac{3c^2}{32\pi G \left(\frac{8\pi^5 (k_B)^4}{15 (hc)^3} \right)}}}$$

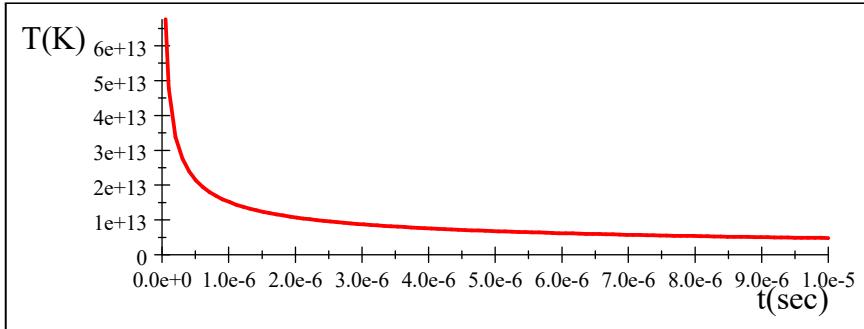
We can put in all the wonderful constants

$$\begin{aligned} D &= \sqrt{\sqrt{\frac{3 (2.99792458 \times 10^8 \text{ m s}^{-1})^2}{32\pi (6.67259 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}) \left(\frac{8\pi^5 (8.617385 \times 10^{-5} \text{ eV K}^{-1})^4}{15 (1240 \text{ eV nm})^3} \right)}}} \\ &= 1.5183 \times 10^{10} \sqrt{s} \text{ K} \end{aligned}$$

so then

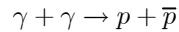
$$T \approx \frac{1.5183 \times 10^{10} \sqrt{s} \text{ K}}{t^{\frac{1}{2}}}$$

and we could plot this



And our conclusion is that as the universe starts out it is very hot, so hot that if photons participated in pair production of matter, that that matter would immediately be destroyed by other photons which could provide more than enough binding energy to tear apart any matter formed.

But the universe will cool as it expands. And with our temperature as a function of time equation we could find the time when the particles that our matter is built on would appear and not immediately be destroyed. The particles would appear as particle antiparticle pairs. So, say



The two photons must have at least the rest mass energy of both particles. Since $m_p = 938.27 \text{ MeV}/c^2$ we need each photon to have this amount of energy. We can approximate the energy as $E_\gamma \approx k_B T$ so

$$T = \frac{E_\gamma}{k_B} = \frac{m_p c^2}{k_B} = \frac{938.27 \text{ MeV}}{8.617385 \times 10^{-5} \text{ eV K}^{-1}} = 1.0888 \times 10^{13} \text{ K}$$

and we can put this in our equation for temperature as a function of time and solve for the time.

$$\begin{aligned} T &= \frac{1.5183 \times 10^{10} \sqrt{s} \text{ K}}{t^{\frac{1}{2}}} \\ t &= \left(\frac{1.5183 \times 10^{10} \sqrt{s} \text{ K}}{T} \right)^2 \\ t &= \left(\frac{1.5183 \times 10^{10} \sqrt{s} \text{ K}}{1.0888 \times 10^{13} \text{ K}} \right)^2 \\ &= 1.9445 \times 10^{-6} \text{ s} \end{aligned}$$

So we start making baryonic matter at around 2×10^{-6} s.

What happens before that? Using the same procedure we can trace back what could exist in the universe as a function of time. Let's do this for our Big Bang.

We can't go right back to $t = 0$ because of the unsolved quantum gravity problem for singularities. Before 10^{-43} seconds, all of what we have said breaks down. This is called the *Planck time*. We don't yet know how to solve the equations for times shorter than the Planck time, so we don't understand what this very early part of the universe's life was like. So let's start just after the Planck time.

After 10^{-43} s but before 10^{-6} s

As the universe cools we have the ability to create particles that are not immediately destroyed by high energy photons. So first the most massive particles are created, and then less massive particles are created. The temperature decreases as shown in our last figure.

After 10^{-6} s but before 10^{-2} s

There are about as many neutrons as protons and there are not enough high energy photons to destroy the nucleons. As time goes on the cooler universe can start to create small nuclei without those nuclei being immediately destroyed. Very few neutrons have decayed, because they just haven't existed long enough yet. We get particles like p , n , e^- , μ^- , π^0 , π^- and others plus their antiparticles. Increasingly matter begins to dominate, so our graph of temperature vs. time is not so accurate.

After 10^{-2} s but before 1 s

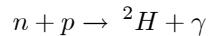
The universe is becoming increasingly matter-dominate. Our temperature vs. time graph is less and less accurate. We could switch to our equation for the matter dominated universe. But looking at our calculations we expect temperature will continue to fall. Photon energy falls with temperature, so now the photons can't make pions and muons, and because the universe is becoming old enough to exceed pion and muon lifetimes, the number of unstable particles drops off dramatically toward the end of this period. At this point there must have been an imbalance between antimatter and matter. All the particles were produced by particle-antiparticle production, but for some reason less antiparticles survive. The reason for the lack of visible antimatter in the universe is an active area of research with some experimental verification, but much work to do. If this imbalance didn't exist, then all the protons and antiprotons (and neutrons and antineutrons) would have annihilated, leaving no baryon matter. We are evidence that this didn't happen. In this time range we have p , n , e^- and their antiparticles. The ratio of neutrons to protons begins to fall.

After 1 s but before 6 s

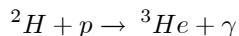
Temperature continues to fall. We get to the point when electron positron pair production is no longer routinely possible and we get the particle antiparticle culling for electrons that we had before for protons and neutrons. The ratio of neutrons to protons continues to fall due to neutron decay.

After 6 s

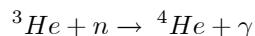
Nuclei and then atoms form. We start with reactions like



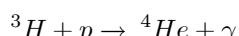
we at first we destroy as many nuclei as we create. But as the temperature drops there are fewer particles with high energy so as the nuclei are formed it is less likely that they are destroyed. Eventually we get reactions like



and then as time goes on

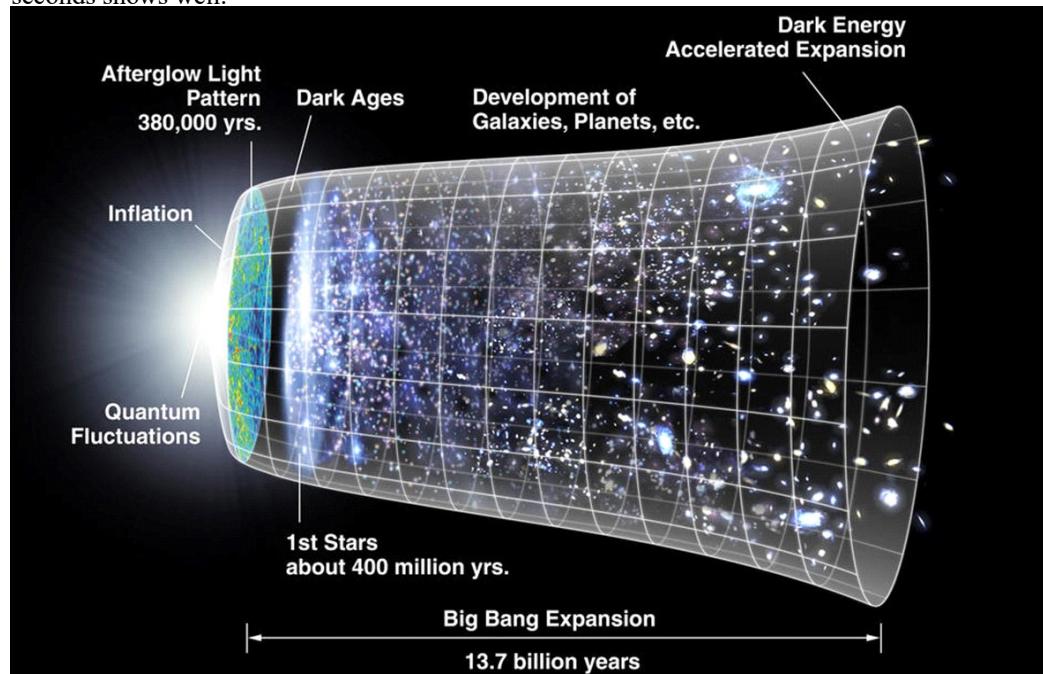


and



These are still nuclei, it takes a while before the universe cools enough that $k_B T < 13.6 \text{ eV}$ so electrons can be captured without being immediately freed again.

Now we need gravity to work. The inhomogeneities we saw in the cosmic background radiation tell us that the early universe wasn't uniform. The denser parts could at this point coalesce into stars, and we know that stars are element factories. NASA produced a nice figure that describes the history of the universe on a scale that the time after 6 seconds shows well.



History of the Universe (Image Courtesy NASA)

In the figure we see the big bang, then we see the cosmic radiation period (shown with the cosmic background radiation inhomogeniaty map superimposed on it) and then the time when photons are not dominant and we just have nuclei and atoms forming. Then we pick up the story of stars forming, then galaxies, and we get the universe we see today.

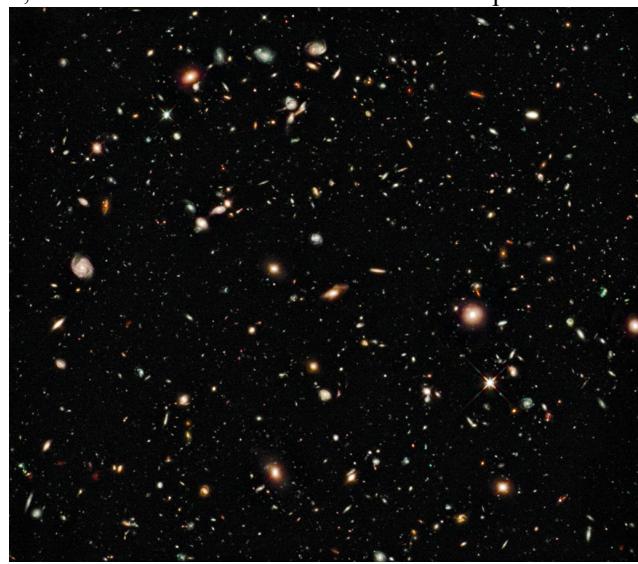
38 Current Cosmology

15.9 - 15.10

The universe today

We now have a universe that has matter in it and the matter is more prevalent than photons. We have stars formed, but the most prevalent form of matter is still hydrogen and helium.

We have an expanding universe. And there is some evidence that this expansion is accelerating. This brings up another area of research. If we look as far as we can see the universe is surprisingly uniform. Here is a picture taken by Hubble of a dark spot in the sky, at least, we saw it as dark with Earth-bound telescopes.

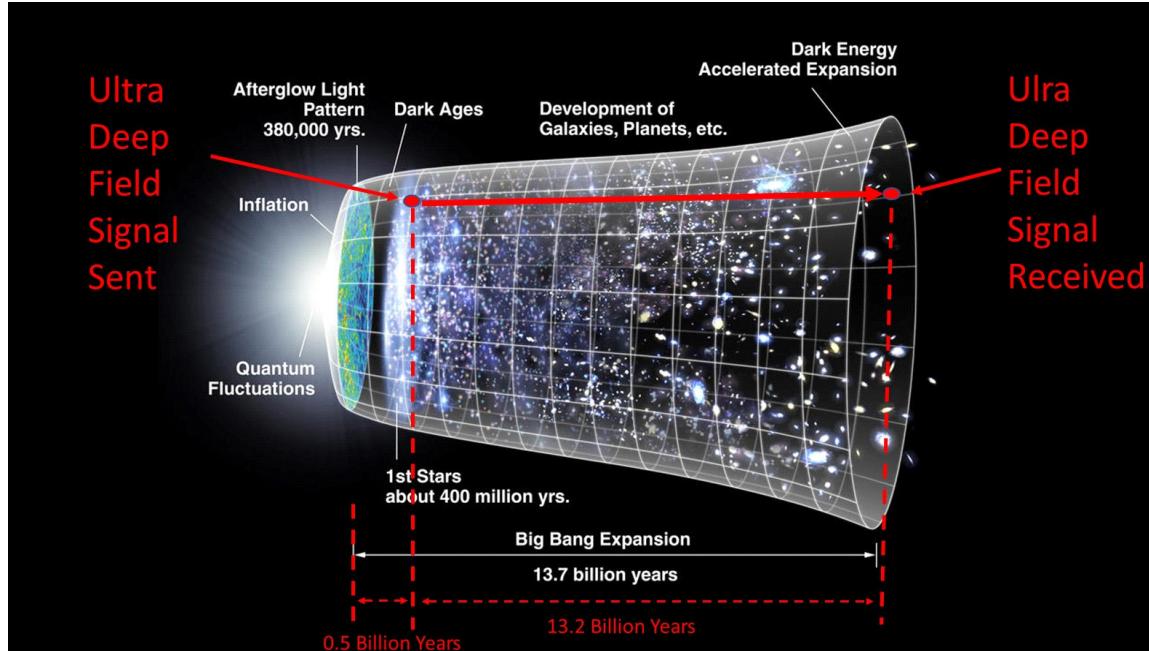


NASA Hubble Ultra Deep Space Image
and what do we see? Galaxies. Why is this surprising? Well if the universe is

13.7×10^9 y old then the farthest light could travel to us is

$$\begin{aligned}\Delta x &= c\Delta t \\ &= (2.99792458 \times 10^8 \text{ m s}^{-1}) (13.7 \times 10^9 \text{ y}) \\ &= 1.2961 \times 10^{26} \text{ m} \\ &= 4.1999 \times 10^9 \text{ pc}\end{aligned}$$

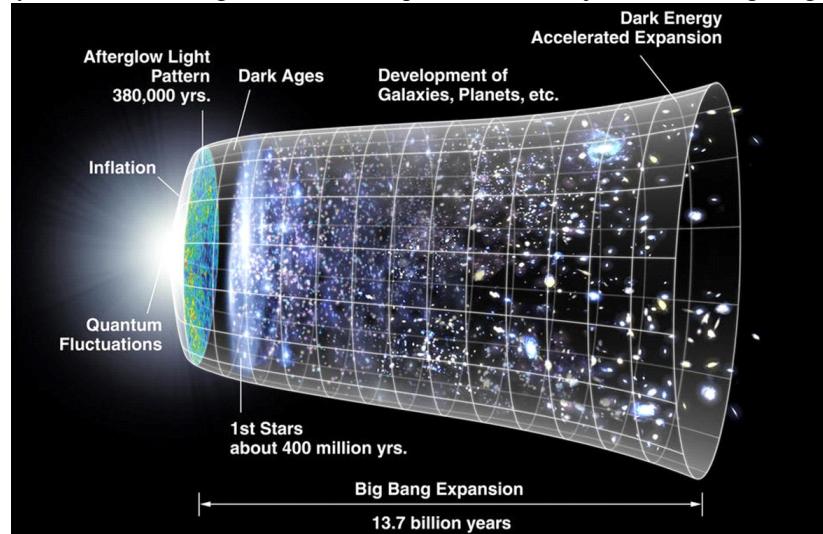
If there is something farther away from us, we would not have seen it yet. So this distance represents the edge of what we can know about the universe. The feeling is that whatever is along this edge would be seen by us as very young. After all, the light from it has traveled the entire time the universe has existed, so the image that this light gives us must be from the dawn of the universe. It would be like mailing a letter to yourself when you were two years old and just receiving it today. The letter would represent a much younger you.



Signal from the Ultra Deep Field Galaxies (Image Courtesy NASA)

The Ultra Deep Field galaxies are approximately 1.2488×10^{26} m away giving their light 13.2×10^9 y to get to us. Then the universe was only 0.5×10^9 y when the light left. and the picture looks very like the region of space around us. Worse yet, if we point our telescope in the other direction the deep space view is much the same. How can the universe be so uniform? It can't be that parts of the universe are cooperating. Think, if there were someone building the galaxies and groups of galaxies on the opposite ends of the universe, they would have to communicate at twice the speed of light to stay in sync.

Rather than believe this, it seems more likely that the uniformity came when the parts of the universe were close together. The *inflation* hypothesis explains this by assuming that the universal expansion was much bigger for a short time in the past. In our NASA history of the universe figure the inflation period is labeled just after the big bang.



I am not sure how you get spacetime to experience a rapid expansion. Einstein taught us that space is something and it is a stretchy something. But how do you make the whole universe of space stretch? The early universe had lots of energy, but what would the mechanism be for stretching space? But still, this would solve what has been called the *horizon problem*.

If you once again study the NASA history figure, you notice that the universe has a bell shape, with a flair on the end. That flair is the recently measured acceleration of the universal expansion. I don't know how you get spacetime to once again accelerate its expansion. But whatever the mechanism, it must need lots of energy to do so. And this is the origin of the *dark energy* hypothesis that seeks to solve the energy part of the expansion by assuming there are vast amounts of energy in the universe that we haven't yet found. This is another active area of research, one where I am anxious to see if it will prove true, or if we have some error in our theories.

There is a bit of frustration in this last thought. If there is an acceleration to the expansion, then some things will move farther away due to the expansion. Their light will head toward us, but because they are so far away, the expansion will win, and their light will never reach us. To be clear, they are not moving faster than light through space. That would be like our ant motion on our stretchy balloon. It is the cumulative effect of the universal expansion that is pushing them beyond our view. This is frustrating because we have to ask if the farthest thing we can see is really all there is. Maybe there are remnants of the big bang just beyond our view and we will never know. As the expansion accelerates, more and more will go beyond what we can know.

This is not a pleasant thought to astrophysicists and cosmologists.

Curvature of the Universe

One last question that we haven't answered (well, one that I am going to try to answer, there are others that we have left for other classes) would be, is the curvature of our universe such that it is closed, open, or is it open and flat?

If we go back to the Friedmann equation

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi}{3}G\rho R^2 - kc^2$$

and take our assertion (because we didn't show it in this class) that

$$H = \frac{1}{R} \frac{dR}{dt}$$

and combine these to find the density of the universe

$$\frac{dR}{dt} = HR$$

$$(HR)^2 = \frac{8\pi}{3}G\rho R^2 - kc^2$$

and for a flat universe $k = 0$

$$(HR)^2 = \frac{8\pi}{3}G\rho R^2$$

so that

$$\rho = \frac{3(HR)^2}{8\pi GR^2} = \frac{3(H)^2}{8\pi G}$$

If the universe has precisely this density, then it is flat. We call this the critical density. To find it we need a unit conversion with the Hubble constant

$$H_o = 72 \frac{\text{km}}{\text{s} \cdot \text{Mpc}} \frac{\text{Mpc}}{3.086 \times 10^{13} \text{ km} \times 1 \times 10^6} = 2.3331 \times 10^{-18} \frac{1}{\text{s}}$$

so our critical density is

$$\begin{aligned} \rho_{cr} &= \frac{3H^2}{8\pi G} = \frac{3(2.3331 \times 10^{-18} \frac{1}{\text{s}})^2}{8\pi (6.67259 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2})} \\ &= 9.7376 \times 10^{-27} \frac{\text{kg}}{\text{m}^3} \end{aligned}$$

If the universe has even slightly more density than this, the universe is curved and infinite and open. If it is slightly lower than this, the universe is curved and closed. It is customary to define

$$\Omega = \frac{\rho}{\rho_{cr}}$$

so that if $\Omega > 1$ the universe is curved and open, and if $\Omega < 1$ the universe is curved and closed, and $\Omega = 1$ would be an infinite flat universe. So what is the answer? We don't know. We can tell that Ω isn't too different than 1. But because we need Ω to be exactly 1 for a flat universe, the observations have to be very exact. And of course, accelerated expansion would imply that H might not be constant. So this might be

a confounding problem. Cosmologists are working on answering this question, but we don't have a definitive answer yet. So far the best we can say is that within our uncertainties, the universe looks pretty flat. But because of uncertainty, all we can say is that it is not very curved. Perhaps this is a field you would like to join and perhaps you will solve this problem.

Does it all work?

We talked about relativity, we have studied matter and it's motion from macroscopic classical ideas to particle physics and quantum mechanics. We should ask ourselves, is all this real?

Remember, in each age, scientists (or philosophers, or whatever the intellectuals of that era were called) were sure they had it right. Human kind has a poor track record of getting it right and a poorer track record of knowing when we are wrong. Aristotle's mostly wrong ideas lasted for thousands of years. Galileo was arrested for opposing Aristotelean ideas! So we should be careful and humble.

Our theories change in time. Might there be a breakthrough that radically changes our view point in the future. Undoubtedly there will! The race is on now in particle physics, quantum gravity, QCE and other fields for rewiring our theories to account for the current missing pieces and for new discoveries. Those new ideas will effect the rest of physics at some level.

Roger Penrose from the University of Oxford gave a rating for our theories back in 1989. For each theory he labeled them one of the following "SUPERB," "USEFUL," or "TENTATIVE." He thought of adding the classification of "MISGUIDED." But, In his words, "...then I thought better of it, since I do not want to lose half of my friends." Here is his classification including the theories we have studied.

Theory	Penrose Classification	Theory	Penrose Classification
Euclidian geometry	SUPERB	Gell-Mann-Zweig Quark Model	USEFUL
Newtonian Statics	SUPERB	Quantum Chromodynamics	USEFUL
Newtonian Mechanics	SUPERB	Standard Model	USEFUL
Maxwell's Electrodynamics	SUPERB	Big Bang	USEFUL
Special Relativity	SUPERB	Kaluza-Klein (supersymmetry)	TENTATIVE
General Relativity	SUPERB	GUT (including inflation)	TENTATIVE
Quantum Mechanics	SUPERB	Superstring	TENTATIVE
Quantum Electrodynamics (QED)	SUPERB (only on accuracy)		

Penrose says his category of "SUPERB" does not mean the theory should apply "without refutation to the phenomena of the world" But only that "the range and accuracy with which it applies should, in some appropriate sense, be *phenomenal*." To be in the TENTATIVE category, Penrose says you need a "lack of any significant experimental support."

This is interesting and brings even more humility. Yet even with this classification by one of the prominent mathematical physicists of our era, the categories do not tell us which theories are correct. They are all useful models, but we cannot really say they are right. Someone in the future may be looking at QED the way we look at Ether Theory!

Because of this, I believe the work of physics is far from over. I hope in your life time there will be many exciting changes. As physicists we have a chance to be part of the continual revolution.

This sounds a little like “Ever learning, and never able to come to the knowledge of the truth.”²² We should ask if this search for how the universe is made is of value. After all, the experiments are often expensive!

I believe it is. As we have understood the atom, we have learned to do chemistry which has lead to the development of new medical treatments and new materials. We have learned to take X-rays to form CAT scans, and we have built MRI and PET scanners based on our physics understanding. We have used what we learned to treat cancer in some forms. We have developed microwave ovens and we have electric lighting and LED flashlights. Even the internet was initially designed to allow the physics labs to share data and to allow the scientists to talk to each other²³. So the benefits to our curiosity are not just the answers, but the understanding we gain along the way that benefits our lives.

There are further classes, both at the undergraduate level and at the graduate level for most of the topics we covered (some are just at the graduate level). There is much to look forward to! If your curiosity was peaked by any of the topics we covered, consider going on to graduate school. But even if you don’t, we found that general relativity is used in engineering, quantum mechanics is used in electronics, and really all of what we have learned shows up in some form of our modern life. Knowing how it works is only a benefit. If nothing else, you can bask in the wonder of it all and share that wonder with the next generation.

²² 2 Tim 3:7

²³ No, Al Gore did not invent the internet, it was there years before he started his run for vice president.

A Fourier Series Review

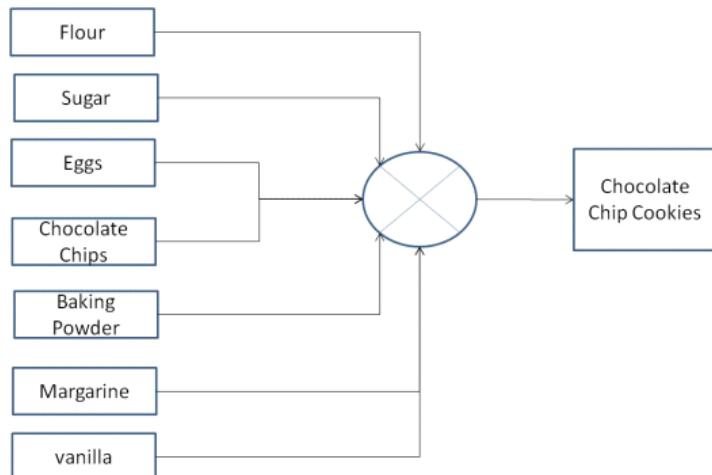
Back in section (3) we studied limiting waves in our study of the Heisenberg uncertainty principle. But Fourier series was more of a “review” for some of us than for others. In the real world, real waves are not perfect sinusoids. We need a way to represent non-sinusoidal waves (like our wave packets!).

Music and Non-sinusoidal waves

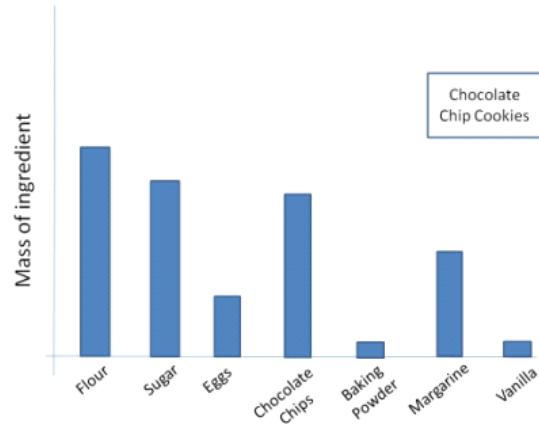
Let's take the example of music. From the example of standing waves on strings that we did back in PH123, we know that a string can support a series of standing waves with quantized frequencies—the harmonic series. We have also discussed that usually we excite more than one standing wave at a time. The fundamental mode tends to give us the pitch we hear, but what are the other standing waves for?

To understand, lets take an analogy. Making cookies and cakes.

Here is the beginning of a recipe for cookies.

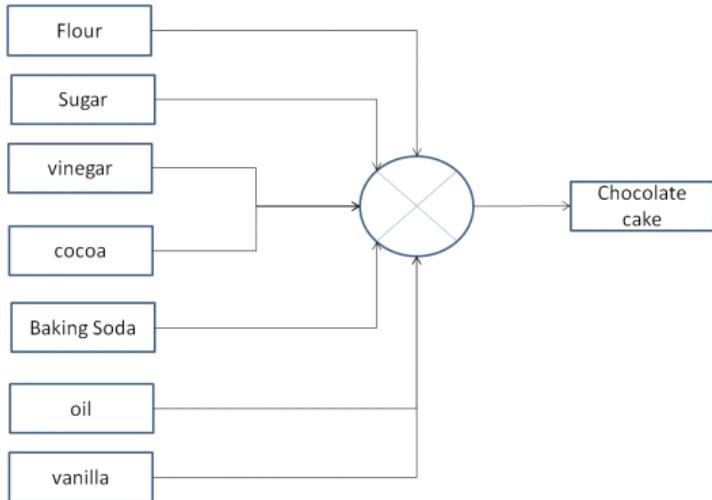


The recipe is a list of ingredients, and a symbolic instruction to mix and bake. The product is chocolate chip cookies. Of course we need more information. We need to know how much of each ingredient to use.



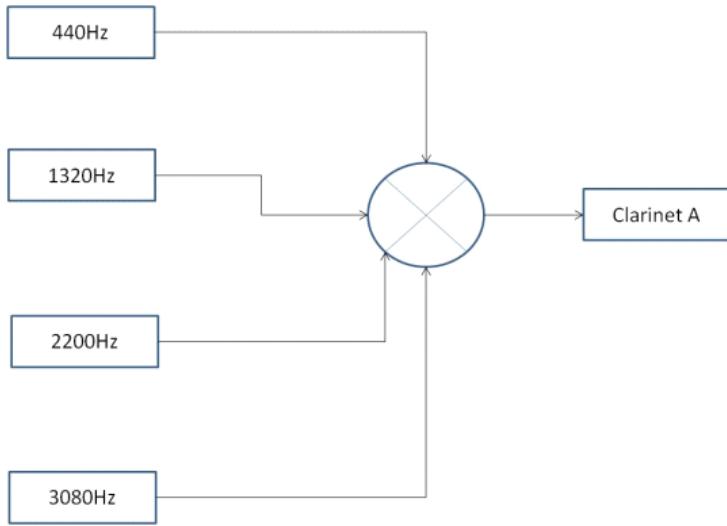
This graph gives us the amount of each ingredient by mass.

Now suppose we want chocolate cake.

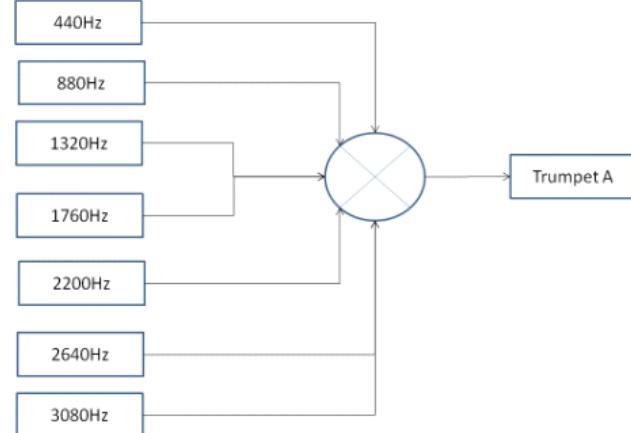


The predominant taste in each of these foods is chocolate. But chocolate cake and chocolate chip cookies don't taste exactly the same. We can easily see that the differences in the other ingredients make the difference between the "cookie" taste and the "cake" taste that goes along with the "chocolate" taste that predominates.

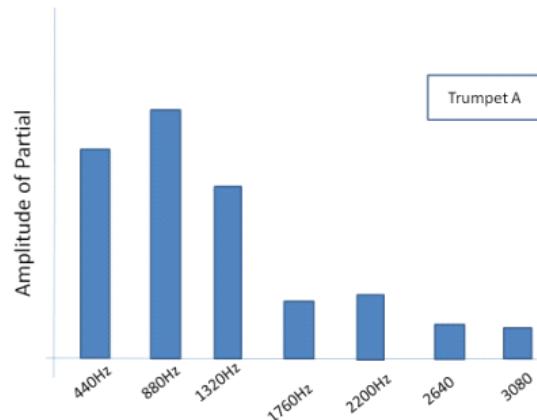
The sound waves produced by musical instruments work in a similar way. Here is a recipe for an "A" note from a clarinet.



and here is one for a trumpet playing the same “A” note.

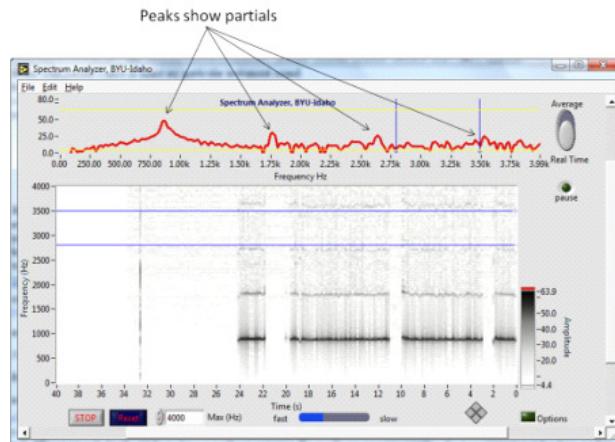


A trumpet sounds different than a clarinet, and now we see why. There are more harmonics involved with the trumpet sound than the clarinet sound. These extra standing waves make up the “brassiness” of the trumpet sound. As with our baking example, we need to know how much of each standing wave we have. Each will have a different amplitude. For our trumpet, we might get amplitudes as shown.



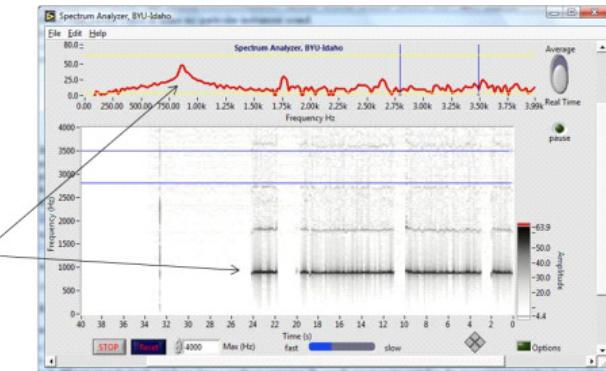
Note that the second harmonic has a larger amplitude, but we still hear the “A” as at 440 Hz. A fugal horn would still sound brassy, but would have a different mix of harmonics.

In music, the different harmonics are called *partials* because they make up part of the sound. A graph that shows which harmonics are involved is called a *spectrum*. The next figure is the spectrum of a six holed bamboo flute. Note that there are several harmonics involved.

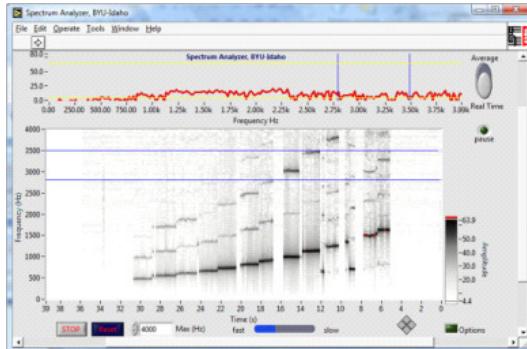


Note that our graph has two parts. One is the instantaneous spectrum, and one is the spectrum time history.

Notice that there
are non-harmonic
partials in this
instrument too!

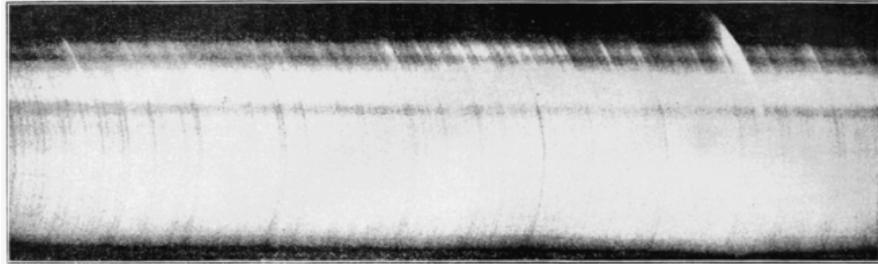


By observing the time history, we can see changes in the spectrum. We can also see that we don't have pure harmonics. The graph shows some response off the specific harmonic frequencies. This six holed flute is very "breathy" giving a lot of wind noise along with the notes, and we see this in the spectrum. In the next picture, I played a scale on the flute.

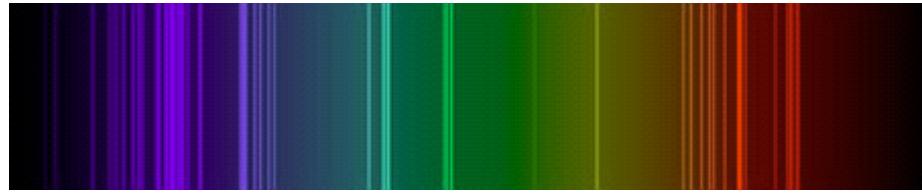


The instantaneous spectrum is not active in this figure (since it can't show more than one note at a time) but in the time history we see that as the fundamental frequency changes by shorting the length of the flute (uncovering holes), we see that every partial also goes up in frequency. The flute still has the characteristic spectrum of a flute, but shifted to new frequencies. We can use this fact to identify things by their vibration spectrum. In fact, that is how you recognize voices and musics within your auditory system!

The technique of taking apart a wave into its components is very powerful. With light waves, the spectrum is an indication of the chemical composition of the emitter. For example, the spectrum of the sun looks something like this



Solar coronal spectrum taken during a solar eclipse. The successive curved lines are each different wavelengths, and the dark lines are wavelengths that are absorbed. The pattern of absorbed wavelengths allows a chemical analysis of the corona. (Image in the Public Domain, originally published in Bailey, Solon, L, Popular Science Monthly, Vol 60, Nov. 1919, pp 244) The lines in this graph show the amplitude of each harmonic component of the light. Darker lines have larger amplitudes. The harmonics come from the excitation of electrons in their orbitals. Each orbital is a different energy state, and when the electrons jump from orbital to orbital, they produce specific wave frequencies. By observing the mix of dark lines in previous figure, and comparing to laboratory measurements from each element (see next figure) we can find the composition of the source. This figure shows the emission spectrum for Calcium. because it is an emission spectrum the lines are bright instead of dark. We can even see the color of each line!



Emission spectrum of Calcium (Image in the Public Domain, courtesy NASA)

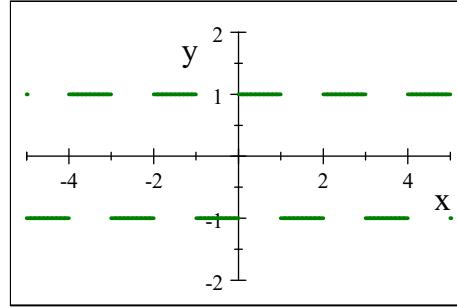
Fourier Series: Mathematics of Non-Sinusoidal Waves

We want to understand the mathematics of non-sinusoidal waves. The strategy we will take is to make a series expansion of basis functions that when we add up all the parts of the series are equal to the original wave. This is a lot like taking a Taylor series, but instead of the series being in powers of x we want a series that is like adding up many little waves to make a big complicated wave. So let's start with adding up waves with different amplitudes and different frequencies.

The math looks like this

$$y(t) = \sum_n (A_n \sin(2\pi f_n t) + B_n \cos(2\pi f_n t))$$

where A_n and B_n are a series of coefficients and f_n are the harmonic series of frequencies. Let's try a strange wave. Let's try a wave that has square parts. That is really really not a sinusoidal wave. Here is a picture.



Fourier Analysis

In our lecture we let Python do the heavy mathematics. But if you want to know what Python was doing, read this section. We could represent our square wave function $f(x)$ with the following series

$$f(x) = C_0 + C_1 \cos\left(\frac{2\pi}{\lambda}x + \varepsilon_1\right) \quad (1)$$

$$+ C_2 \cos\left(\frac{2\pi}{\lambda}x + \varepsilon_2\right) \quad (2)$$

$$+ C_3 \cos\left(\frac{2\pi}{\lambda}x + \varepsilon_3\right) \quad (3)$$

$$+ \dots \quad (4)$$

$$+ C_n \cos\left(\frac{2\pi}{\lambda}x + \varepsilon_n\right) \quad (5)$$

$$+ \dots \quad (6)$$

where we will let $\varepsilon_i = \omega_i t + \phi_i$

The C' s are just coefficients that tell us the amplitude of the individual cosine waves. The more terms in the series we take, the better the approximation we will have, with the series exactly matching $f(x)$ when the number of terms, $N \rightarrow \infty$.

Usually we rewrite the terms of the series as

$$C_m \cos(mkx + \varepsilon_m) = A_m \cos(mkx) + B_m \sin(mkx) \quad (7)$$

where k is the wavenumber

$$k = \frac{2\pi}{\lambda} \quad (8)$$

and λ is the wavelength of the complicated but still periodic function $f(x)$. Then we identify

$$A_m = C_m \cos(\varepsilon_m) \quad (9)$$

$$B_m = -C_m \sin(\varepsilon_m) \quad (10)$$

then

$$f(x) = \frac{A_o}{2} + \sum_{m=1}^{\infty} A_m \cos(mkx) + \sum_{m=1}^{\infty} B_m \sin(mkx) \quad (11)$$

where we separated out the $A_o/2$ term because it makes things nicer later.

The process of finding the coefficients of the series is called *Fourier analysis*. We won't do this much in our course, but here is how you could. We start by integrating equation ()

$$\int_0^\lambda f(x) dx = \int_0^\lambda \frac{A_o}{2} dx + \int_0^\lambda \sum_{m=1}^{\infty} A_m \cos(mkx) dx + \int_0^\lambda \sum_{m=1}^{\infty} B_m \sin(mkx) dx \quad (12)$$

We can see immediately that all the sine and cosine terms integrate to zero (we integrated over a wavelength) so

$$\int_0^\lambda f(x) dx = \int_0^\lambda \frac{A_o}{2} dx = \frac{A_o}{2} \lambda \quad (13)$$

We solve this for A_o

$$A_o = \frac{2}{\lambda} \int_0^\lambda f(x) dx \quad (14)$$

To find the rest of the coefficients we need to remind ourselves of the orthogonality of sinusoidal functions

$$\int_0^\lambda \sin(akx) \cos(bkx) dx = 0 \quad (15)$$

$$\int_0^\lambda \cos(akx) \cos(bkx) dx = \frac{\lambda}{2} \delta_{ab} \quad (16)$$

$$\int_0^\lambda \sin(akx) \sin(bkx) dx = \frac{\lambda}{2} \delta_{ab} \quad (17)$$

where δ_{ab} is the Kronecker delta.

To find the coefficients, then, we multiply both sides of equation () by $\cos(lkx)$ where l is a positive integer. Then we integrate over one wavelength.

$$\int_0^\lambda f(x) \cos(lkx) dx = \int_0^\lambda \frac{A_o}{2} \cos(lkx) dx \quad (18)$$

$$+ \int_0^\lambda \sum_{m=1}^{\infty} A_m \cos(mkx) \cos(lkx) dx \quad (19)$$

$$+ \int_0^\lambda \sum_{m=1}^{\infty} B_m \sin(mkx) \cos(lkx) dx \quad (20)$$

which gives

$$\int_0^\lambda f(x) \cos(mkx) dx = \int_0^\lambda A_m \cos(mkx) \cos(mkx) dx \quad (21)$$

that is, only the term with two cosine functions where $l = m$ will be non zero. So

$$\int_0^\lambda f(x) \cos(mkx) dx = \frac{\lambda}{2} A_m \quad (22)$$

solving for A_m we have

$$A_m = \frac{2}{\lambda} \int_0^\lambda f(x) \cos(mkx) dx \quad (23)$$

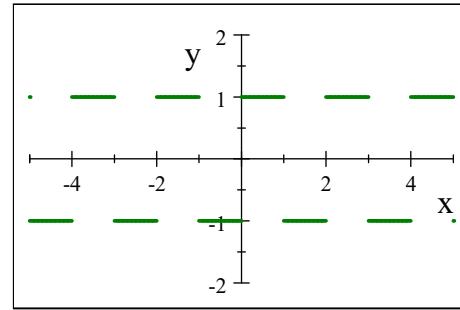
We can perform the same steps to find B_m only we use $\sin(lkx)$ as the multiplier. Then we find

$$B_m = \frac{2}{\lambda} \int_0^\lambda f(x) \sin(mkx) dx \quad (24)$$

Let's find the series for a square wave using our Fourier analysis technique. Let's take

$$\lambda = 2 \quad (25)$$

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < \frac{\lambda}{2} \\ -1 & \text{if } \frac{\lambda}{2} < x < \lambda \end{cases} \quad (26)$$



since $f(x)$ is odd, $A_m = 0$ for all m . We have

$$B_m = \frac{2}{\lambda} \int_0^{\frac{\lambda}{2}} (1) \sin(mkx) dx + \frac{2}{\lambda} \int_{\frac{\lambda}{2}}^\lambda (-1) \sin(mkx) dx \quad (27)$$

so

$$B_m = \frac{1}{m\pi} (-\cos(mkx)|_0^{\frac{\lambda}{2}} + \frac{1}{m\pi} (\cos(mkx)|_{\frac{\lambda}{2}}^\lambda) \quad (28)$$

Which is

$$B_m = \frac{1}{m\pi} \left(1 \cos\left(m \frac{2\pi}{\lambda} x\right)|_0^{\frac{\lambda}{2}} + \frac{1}{m\pi} \left(\cos\left(m \frac{2\pi}{\lambda} x\right)|_{\frac{\lambda}{2}}^\lambda \right) \right) \quad (29)$$

so

$$B_m = \frac{1}{m\pi} \left(\left(-\cos\left(m \frac{2\pi}{\lambda} \frac{\lambda}{2}\right) \right) + \cos\left(m \frac{2\pi}{\lambda} (0)\right) \right) \quad (30)$$

$$+ \frac{1}{m\pi} \left(\left(\cos\left(m \frac{2\pi}{\lambda} \lambda\right) - \cos\left(m \frac{2\pi}{\lambda} \frac{\lambda}{2}\right) \right) \right) \quad (31)$$

which is

$$B_m = \frac{2}{m\pi} (1 - \cos(m\pi)) \quad (32)$$

Back to our Square Wave

Whether we let Python find the coefficients or we do it ourselves, we now know that we can find the amplitude for each of the waves that we mix together to make our complicated non-sinusoidal waves. Our series is then just

$$f(x) = \sum_{m=1}^{\infty} \frac{2}{m\pi} (1 - \cos(m\pi)) \sin(mkx) \quad (33)$$

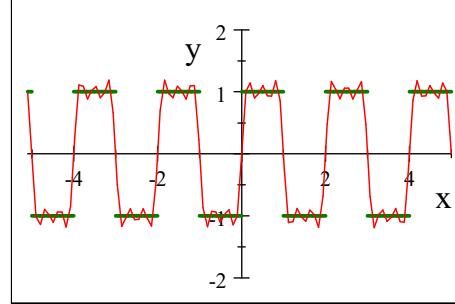
and we found the amplitudes for the square wave. It was these amplitudes that we had Python plot. We can write a few terms

Term	
1	$\frac{4}{\pi} \sin(kx)$
2	0
3	$\frac{4}{3\pi} \sin(3kx)$
4	0
5	$\frac{4}{5\pi} \sin(5kx)$

(34)

then the partial sum up to $m = 5$ looks like

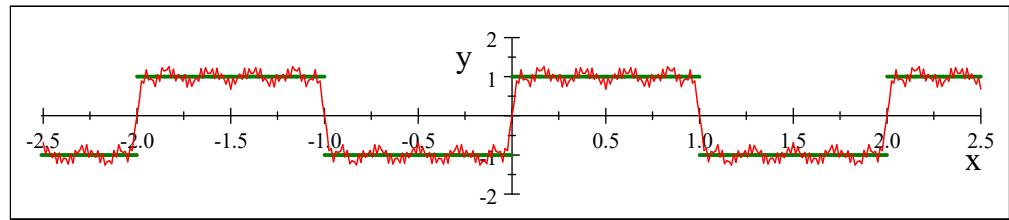
$$f(x) = \frac{4}{\pi} \sin(kx) + \frac{4}{3\pi} \sin(3kx) + \frac{4}{5\pi} \sin(5kx) \quad (35)$$



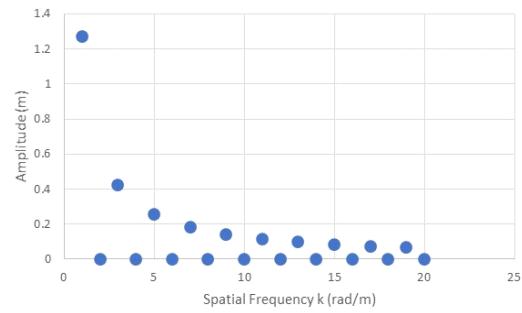
where the thin red line is our series (up to five terms). It is much less sinusoidal and much more square than a single sine wave. But it is still not perfect. If we take many terms,

$$\begin{aligned} f(x) &= \frac{4}{\pi} \sin(kx) + \frac{4}{3\pi} \sin(3kx) + \frac{4}{5\pi} \sin(5kx) + \frac{4}{7\pi} \sin(7kx) + \frac{4}{9\pi} \sin(95kx) \\ &\quad + \frac{4}{11\pi} \sin(11kx) + \frac{4}{13\pi} \sin(13kx) + \frac{4}{15\pi} \sin(15kx) + \frac{4}{17\pi} \sin(17kx) + \frac{4}{19\pi} \sin(19kx) \end{aligned} \quad (36)$$

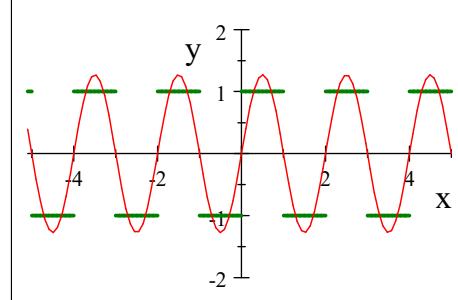
We see the function get closer and closer to a square wave. In the next figure I expanded our view so you can see the details of the red series wave. It is still not exactly equal to the square wave, but the red wave doesn't change much from the square wave.



In the limit of infinitely many waves, the match would be perfect. But we don't usually need an infinite number of terms. We can pick the part of the spectrum that best represents the phenomena we desire to observe. We can plot the amplitude of each of the 20 terms that we used. It looks like this

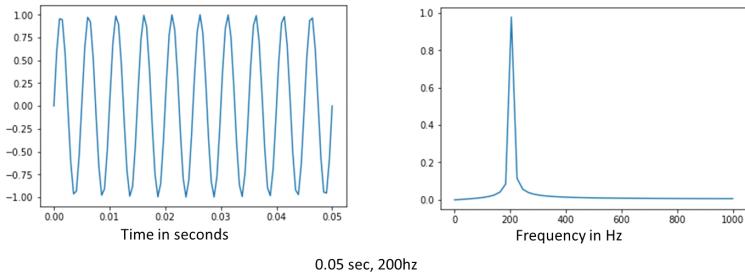


This is the spectrum of the wave amplitudes that make up this square wave. It says we have more of the low frequencies and less of the higher frequencies. It also shows that every other frequency has an amplitude of zero. If we mix waves of these frequencies and amplitudes we get our approximation to the square wave. We can see from this that the first wave is very important and the 19th wave is less important in some sense. If I plot just the first term we can see why



The first term in our series is giving the frequency of the square wave to our series. For sound, this is what our brains interpret as the musical note we hear. Our wave packets are built like this. For light, we expect a peak frequency in the spectrum graph that gives the color of the photon. We add together many partial waves to get the actual wave packet function that we want. Each of the partial waves will have a different amplitude and frequency. Usually these frequencies are a group of frequencies that are close together. So we will say they are in a "band" of frequencies that are Δf wide. The Δf says that we made our actual wave packet out of many partial waves and

that these partial waves give us many frequencies (and therefore many energies). The principal frequency will have a large amplitude and most others will have much smaller amplitudes. Here is one of our Python generated graphs.



The graph to the right tells us we have a large amplitude for 200 Hz but we have amplitudes that are not zero for other frequencies. These are all required to make a limited wave packet.

B Tables

Integral Table and Data Tables

Some useful integrals and trig identities

$$\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}$$

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}$$

$$\int_0^{\infty} x^{\frac{1}{2}} e^{-x} dx = \frac{1}{2} \sqrt{\pi}$$

$$\int_0^{\infty} x^{\frac{3}{2}} e^{-x} dx = \frac{3}{4} \sqrt{\pi}$$

$$\int_0^{\infty} x^{\frac{1}{2}} (e^{-nx}) dx = \frac{1}{2n} \sqrt{\frac{\pi}{n}}$$

$$\int_0^{\infty} x^4 e^{-ax^2} dx = -\frac{1}{8a^{\frac{5}{2}}} \left(6\sqrt{a}xe^{-ax^2} - 3\sqrt{\pi} \operatorname{erf}(\sqrt{ax}) + 4a^{\frac{3}{2}}x^3 e^{-ax^2} \right) \Big|_0^{\infty}$$

$$= -\frac{1}{8a^{\frac{5}{2}}} (-3\sqrt{\pi})$$

$$\int_0^{\infty} \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15}$$

$$\int \frac{x}{(1-x^2)^{\frac{3}{2}}} dx = \frac{1}{\sqrt{1-x^2}}$$

$$\int_{\infty}^0 \left(\frac{e^u}{(e^u + 1)^2} u^2 \right) du = \frac{\pi^2}{6}$$

$$\int_0^{\infty} e^{-a^2 x^2} \cos(bx) dx = \frac{\sqrt{\pi}}{2|a|} e^{\frac{-b^2}{4a^2}}$$

$$\int \sin(ax) dx = -\frac{1}{a} \cos ax$$

$$\int_0^{\infty} \left(\frac{x^2}{e^x - 1} \right) dx = 2.40411$$

$$\begin{aligned}
\int \sin^2(ax) dx &= -\frac{1}{4a} (\sin 2ax - 2ax) \\
\int x \sin^2(ax) dx &= -\frac{2ax(\sin(2ax) - ax) + \cos(2ax)}{8a^2} \\
\int_0^\infty x^{\frac{3}{2}} e^{-x} dx &= \frac{3}{4}\sqrt{\pi} \\
\int_0^\infty x^n e^{-cx} dx &= \frac{n!}{c^{n+1}} \\
\int x^m e^{ax} dx &= e^{ax} \sum_{r=0}^m (-1)^r \frac{m!x^{m-r}}{(m-r)!a^{r+1}} \\
\operatorname{erf}(\sqrt{a}\infty) &= \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt = 1 \\
\sin a + \sin b &= 2 \cos\left(\frac{a-b}{2}\right) \sin\left(\frac{a+b}{2}\right) \\
\cos a - \cos b &= -2 \sin\frac{a+b}{2} \sin\left(\frac{a-b}{2}\right) \\
\sin(a \pm b) &= \sin a \cos b \pm \cos a \sin b \\
\cos(a \pm b) &= \cos a \cos b \mp \sin a \sin b
\end{aligned}$$

$$\begin{aligned}
\sin(-x) &= -\sin x \\
\cos(-x) &= \cos x \\
\sin(kx) &= \frac{e^{ikx} - e^{-ikx}}{2i} \\
\cos(kx) &= \frac{e^{ikx} + e^{-ikx}}{2}
\end{aligned}$$

Other useful mathematical things

$$\begin{aligned}
(\sqrt{1-x})^{\frac{-1}{2}} &\approx 1 + \frac{1}{2}x + O(\text{higher}) \quad \text{for } x \ll 1 \\
\sum_{n=0}^{\infty} e^{nx} &= \frac{1}{(1-e^x)} \\
\sum_{n=0}^{\infty} n e^{nx} &= \frac{e^x}{(1-e^x)^2}
\end{aligned}$$

Some useful quantities

$$\begin{aligned}
m_{\text{electron}} &= 9.1093829 \times 10^{-31} \text{ kg} = 0.51099893 \text{ MeV}/c^2 \\
m_{\text{proton}} &= 1.67262178 \times 10^{-27} \text{ kg} = 938.27205 \text{ MeV}/c^2 \\
m_{\text{neutron}} &= 1.67492735 \times 10^{-27} \text{ kg} = 939.56538 \text{ MeV}/c^2 \\
1 \text{ u} &= 931.49406 \text{ MeV}/c^2 = 1.66053892 \times 10^{-27} \text{ kg} \\
h &= 6.6260755 \times 10^{-34} \text{ J s} \\
&= 4.1356692 \times 10^{-15} \text{ eV s} \\
\hbar &= 1.054571726 \times 10^{-34} \text{ J s} \\
&= 6.5821220 \times 10^{-16} \text{ eV s} \\
c &= 2.99792458 \times 10^8 \frac{\text{m}}{\text{s}} \\
hc &= 1240 \text{ eV nm} \\
\hbar c &= 197.326972 \text{ eV nm} \\
e &= 1.60217733 \times 10^{-19} \text{ C} \\
k_B &= 8.617385 \times 10^{-5} \text{ eV K}^{-1} \\
N_A &= 6.02 \times 10^{23} \frac{1}{\text{mol}}
\end{aligned}$$

$$\begin{aligned}
1 \text{ fm} &= 1.0 \times 10^{-15} \text{ m} \\
1 \text{ \AA} &= 1.0 \times 10^{-10} \text{ m} \\
1 \text{ nm} &= 1.0 \times 10^{-9} \text{ m}
\end{aligned}$$

$$R_\infty = \frac{m_e e^4}{64\pi^3 \epsilon_0^2 \hbar^3 c} = 1.0973731534 \times 10^7 \text{ m}^{-1}$$



Hydrogen Wave Functions

$$\psi(r, \theta, \phi) = R(r) \Theta(\theta) \Phi(\phi)$$

where the pieces look like

n	l	m_l	$R(r)$	$\Theta(\theta)$	$\Phi(\phi)$
1	0	0	$\frac{2}{(a_o)^{\frac{3}{2}}} e^{-\frac{r}{a_o}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2\pi}}$
2	0	0	$\frac{1}{(2a_o)^{\frac{3}{2}}} \left(2 - \frac{r}{a_o}\right) e^{-\frac{r}{2a_o}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2\pi}}$
2	1	0	$\frac{1}{\sqrt{3}(2a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o}\right) e^{-\frac{r}{2a_o}}$	$\sqrt{\frac{3}{2}} \cos \theta$	$\frac{1}{\sqrt{2\pi}}$
2	1	± 1	$\frac{1}{\sqrt{3}(2a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o}\right) e^{-\frac{r}{2a_o}}$	$\pm \frac{\sqrt{3}}{2} \sin \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$
3	0	0	$\frac{2}{(3a_o)^{\frac{3}{2}}} \left(1 - \frac{2r}{3a_o} + \frac{2r^2}{27a_o^2}\right) e^{-\frac{r}{3a_o}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2\pi}}$
3	1	0	$\frac{2}{9\sqrt{2}(3a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o} - \frac{r^2}{6a_o^2}\right) e^{-\frac{r}{3a_o}}$	$\sqrt{\frac{3}{2}} \cos \theta$	$\frac{1}{\sqrt{2\pi}}$
3	1	± 1	$\frac{2}{9\sqrt{2}(3a_o)^{\frac{3}{2}}} \left(\frac{r}{a_o} - \frac{r^2}{6a_o^2}\right) e^{-\frac{r}{3a_o}}$	$\pm \frac{\sqrt{3}}{2} \sin \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$
3	2	0	$\frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2}\right) e^{-\frac{r}{3a_o}}$	$\sqrt{\frac{5}{8}} (3 \cos^2 \theta - 1)$	$\frac{1}{\sqrt{2\pi}}$
3	2	± 1	$\frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2}\right) e^{-\frac{r}{3a_o}}$	$\pm \sqrt{\frac{15}{4}} \sin \theta \cos \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$
3	2	21	$\frac{2}{27\sqrt{10}(3a_o)^{\frac{3}{2}}} \left(\frac{r^2}{a_o^2}\right) e^{-\frac{r}{3a_o}}$	$\sqrt{\frac{15}{4}} \sin^2 \theta$	$\frac{1}{\sqrt{2\pi}} e^{\pm i\phi}$

Table of Leptons and their Quantum Numbers and Masses

Leptons	Symbol	Lepton Number	Charge (e)	Spin	Mean Life (s)	Anti-particle	Lepton Number	Rest Energy (MeV)	Decay Products
Electron	e^-	$L_e = +1$	-1	$\frac{1}{2}$	∞	e^+	$L_e = -1$	0.511	
Electron-Neutrino	ν_e	$L_e = +1$	0	$\frac{1}{2}$	∞	$\bar{\nu}_e$	$L_e = -1$	really small	
Muon	μ^-	$L_\mu = +1$	-1	$\frac{1}{2}$	2.2×10^{-6}	μ^+	$L_\mu = -1$	105.7	$e^- + \bar{\nu}_e + \nu_\mu$
Muon-Neutrino	ν_μ	$L_\mu = +1$	0	$\frac{1}{2}$	∞	$\bar{\nu}_\mu$	$L_\mu = -1$	really small	
Tau	τ^-	$L_\tau = +1$	-1	$\frac{1}{2}$	2.9×10^{-13}	τ^+	$L_\tau = -1$	1784	$\mu^- + \bar{\nu}_\mu + \nu_\tau$
Tau-Neutrino	ν_τ	$L_\tau = +1$	0	$\frac{1}{2}$	∞	$\bar{\nu}_\tau$	$L_\tau = -1$	really small	

Table of Baryons and their Quantum Numbers and Masses

Baryons	Symbol	Baryon Number	Particle Charge (e)	Spin	Strange-ness	Mean Life (s)	Anti-particle e	Baryon Number	Rest Energy (MeV)	Decay Products
Proton	p	+1	+1	$\frac{1}{2}$	0	∞	\bar{p}	-1	938.3	None
Neutron	n	+1	0	$\frac{1}{2}$	0	886	\bar{n}	-1	939.6	$-p + e^- + \bar{\nu}_e$
lambda	Λ^0	+1	0	$\frac{1}{2}$	-1	2.6×10^{-10}	$\bar{\Lambda}^0$	-1	1115.6	$p + \pi$
Sigma	Σ^+	+1	+1	$\frac{1}{2}$	-1	8.0×10^{-11}	$\bar{\Sigma}^+$	-1	1189.4	$p + \pi^0$
	Σ^0	+1	0	$\frac{1}{2}$	-1	7.4×10^{-20}	$\bar{\Sigma}^0$	-1	1192.5	$\Lambda^0 + \gamma$
	Σ^-	+1	-1	$\frac{1}{2}$	-1	1.5×10^{-10}	$\bar{\Sigma}^-$	-1	1197.3	$n + \pi^-$
Xi	Ξ^0		0	$\frac{1}{2}$	-2	2.9×10^{-10}	$\bar{\Xi}^0$	-1	1315	$\Lambda^0 + \pi^0$
	Ξ^-	+1	-1	$\frac{1}{2}$	-2	1.6×10^{-10}	$\bar{\Xi}^-$	-1	1321	$\Lambda^0 + \pi^-$
	Δ^*		+2, +1, 0, -1	$\frac{3}{2}$	0	5.6×10^{-24}	$\bar{\Delta}^*$		1232	$p + \pi$
	Σ^*		+1, 0, -1	$\frac{3}{2}$	-1	1.8×10^{-23}	$\bar{\Sigma}^*$		1385	$\Lambda^0 + \pi$
	Ξ^*		-1, 0	$\frac{3}{2}$	-2	7.2×10^{-23}	$\bar{\Xi}^*$		1533	$\Xi + \pi$
Omega	Ω^-	+1	-1	$\frac{3}{2}$	-3	8.2×10^{-11}	$\bar{\Omega}^-$	-1	1672	$\Lambda^0 + K^-$

Table of Mesons and their Quantum Numbers and Masses

Mesons	Symbol	Particle Charge (e)	Spin	Strange-ness	Mean Life (s)	Anti-particle	Rest Energy(MeV)	Decay Products
Pion	π^+	+1	0	0	2.6×10^{-8}	π^-	139.6	$\mu^+ + \nu_\mu$
	π^0	0	0	0	8.4×10^{-17}	π^0 (self)	135.0	$\gamma + \gamma$
Kaon	K^+	+1	0	+1	1.2×10^{-8}	K^-	493.7	$\mu^+ + \nu_\mu$
	K_S^0	0	0	+1	0.89×10^{-10}	\bar{K}_S^0	497.7	$\pi^+ + \pi^-$
	K_L^0	0	0	+1	5.2×10^{-8}	\bar{K}_L^0	497.7	$\pi^\pm + e^\mp + \bar{\nu}_e$
Eta	η	0	0	0	5.1×10^{-19}	η (self)	548.8	$\gamma + \gamma$
	η'	0	0	0	3.2×10^{-21}	η' (self)	958	$\eta + \pi^+ + \pi^-$
	ρ^+	+1	1	0	4.4×10^{-24}	ρ^-	775	$\pi^+ + \pi^0$
	D^+	+1	0	0	1.0×10^{-12}	D^-	1869	$K^- + \pi^+ + \pi^-$
	J/ψ	0	1	0	7.1×10^{-21}	J/ψ (self)	3097	$e^+ + e^-$
	B^+	+1	0	0	1.6×10^{-12}	B^-	5279	$D^- + \pi^+ + \pi^-$
	Υ	0	1	0	1.2×10^{-20}	Υ (self)	9460	$e^+ + e^-$

Some Hadrons and their Quantum Numbers and Masses

Table 1. Q is for charge, S is for strangeness, C is for charm, \bar{B} is for bottemness, B is the baryon number

	Q	S	C	B	\bar{B}	mass (MeV/ c^2)
π^+	+1	0	0	0	0	139.570
π^-	-1	0	0	0	0	139.570
π^0	0	0	0	0	0	134.977
η	0	0	0	0	0	547.853
ρ	0	0	0	0	0	775.49
ω^0	0	0	0	0	0	782.65
η'	0	0	0	0	0	957.66
ϕ	0	0	0	0	0	1019.46
K^+	+1	+1	0	0	0	493.667
K^-	-1	-1	0	0	0	493.667
K^0	0	+1	0	0	0	497.614
\bar{K}^0	0	-1	0	0	0	497.614
D^+	+1	0	+1	0	0	1869.12
D^-	-1	0	-1	0	0	1869.12
D^0	0	0	+1	0	0	1864.84
\bar{D}^0	0	0	-1	0	0	1864.84
D_s^+	+1	+1	+1	0	0	1968.49
D_s^-	-1	-1	-1	0	0	1968.49
B^+	+1	0	0	+1	0	5279.15
B^-	-1	0	0	-1	0	5279.15
B^0	0	0	0	+1	0	5279.53
\bar{B}^0	0	0	0	-1	0	5279.53
B_s^0	0	-1	0	+1	0	5366.3
\bar{B}_s^0	0	+1	0	-1	0	5366.3
B_c^+	+1	0	+1	+1	0	6276
B_c^-	-1	0	-1	-1	0	6276

Table 2. Q is for charge, S is for strangeness, C is for charm, \bar{B} is for bottemness, B is the baryon number

	Q	S	C	\bar{B}	B	mass (MeV/c^2)
η_c	0	0	0	0	0	2980.3
J/ψ	0	0	0	0	0	3096.92
η_b	0	0	0	0	0	9388.9
Y	0	0	0	0	0	9460.30
p	+1	0	0	0	1	938.272
n	0	0	0	0	1	939.565
Δ^{++}	+2	0	0	0	1	1232
Δ^+	+1	0	0	0	1	1232
Δ^0	0	0	0	0	1	1232
Δ^-	-1	0	0	0	1	1232
Λ	0	-1	0	0	1	1115.68
Σ^+	+1	-1	0	0	1	1189.37
Σ^0	0	-1	0	0	1	1192.64
Σ^-	-1	-1	0	0	1	1197.45
Ξ^0	0	-2	0	0	1	1314.86
Ξ	-1	-2	0	0	1	1321.71
Ω^-	-1	-3	0	0	1	1672.45
Λ_c^+	+1	0	+1	0	1	2286.46
Σ_c^{++}	+2	0	+1	0	1	2454.02
Σ_c^+	+1	0	+1	0	1	2452.9
Σ_c^0	0	0	+1	0	1	2453.76
Ξ_c^+	+1	-1	+1	0	1	2467.9
Ξ_c^0	0	-1	+1	0	1	2471.0
Ω_c^0	0	-2	+1	0	1	2697.5
Λ_b^0	0	0	0	-1	1	5620.2
Ξ_b^0	0	-1	0	-1	1	5792.4
Ξ_b	-1	-1	0	-1	1	5792.4