

Fundamentals of Physics II
Wave Motion, Thermal Physics, and Optics (under
construction)

R. Todd Lines
(©copyright, R. Todd Lines, All Rights Reserved)

6/13/2025

Contents

Preface	xiii
I Oscillation and Waves	1
1 Introduction to the Course, Simple Harmonic Motion 1.15.1	3
1.1 1 Oscillation and Simple Harmonic Motion	3
1.2 Velocity and Acceleration	7
1.3 The Idea of Phase	12
2 2 Energy and Dynamics of SHM 1.15.2 1.15.3	15
2.1 Initial Conditions	15
2.1.1 A second example	16
2.1.2 Example	19
2.2 Oscillators and Energy	22
2.3 Mathematical Representation of Simple Harmonic Motion	28
2.3.1 Hanging Springs	31
3 3 More Oscillators, Forces and Friction 1.15.4 1.15.5 1.15.6	33
3.1 Comparing Simple Harmonic Motion with Uniform Circular Motion	33
3.2 Our first problem type: Simple Harmonic Motion	38
3.3 The Simple Pendulum	39
3.3.1 Physical Pendulum	41
3.3.2 Torsional Pendulum	43
3.4 Damped Oscillations	44
4 4 Forces Oscillations and Waves 1.15.6	49
4.1 Forced Oscillations	49
4.2 What is a Wave?	53
4.2.1 Criteria for being a wave	53
4.2.2 Longitudinal vs. transverse	53
4.2.3 Examples of waves:	54
4.3 Wave speed	55
4.4 One dimensional waves	56

5	5 Sinusoidal Waves in one and Two Dimensions 1.16.2 1.16.3	61
5.1	Sinusoidal Waves	62
5.2	Parts of a wave	62
5.3	Wavenumber and wave speed	64
5.4	The Speed of Waves	70
5.4.1	Derivation of the string wave speed formula	70
5.4.2	Speed of the medium	72
5.5	The Linear Wave Equation	73
5.6	Waves in two and three dimensions	74
6	6 Power and Superposition 6 1.16.4 1.16.5	77
6.1	Energy and Power in waves	77
6.2	Power and Intensity	82
6.3	Reflection and Transmission	85
6.3.1	Case I: Fixed rope end.	85
6.3.2	Case III: Partially attached rope end	86
6.4	Superposition Principle	88
6.5	Consequences of superposition	90
6.6	Single Frequency Interference in one dimension	92
6.7	Single frequency interference in more than one dimension	95
7	7 Standing Waves and Sound Waves 1.16.6, 1.17.1	101
7.1	Mathematical Description of Superposition	101
7.1.1	Mathematical description of standing waves	104
7.1.2	Standing Waves in a String Fixed at Both Ends	106
7.1.3	Musical Strings	108
7.2	Sound waves	108
7.2.1	Periodic Sound Waves, Pressure	110
8	8 Sound Wave Speed and Intensity 1.17.2, 1.17.3	115
8.1	Speed of Sound Waves	115
8.1.1	Young's Modulus	116
8.1.2	Bulk Modulus	119
8.1.3	Finding the speed of sound in air @@@@ Don't do this here! @@@@	120
8.2	Intensity of Sound Waves @@ Maybe do this	126
8.3	Sound Levels in Decibels	129
9	9 Sound Standing Waves 1.17.4, 1.17.5	133
9.1	Sound Standing waves.	133
9.2	Pipes closed on one end	137
9.3	Standing Waves in Rods and Membranes	140

10 10 Beats and Doppler Shift 1.17.6, 1.17.7	143
10.1 Beats	143
10.2 Doppler Effect	146
10.2.1 Combined Doppler Equation	151
10.2.2 Doppler effect in light	151
11 11 Shock waves and Non-Sinusoidal Waves 1.17.8	155
11.1 Shock Waves	155
11.2 Non Sinusoidal Waves	157
11.2.1 Music and Non-sinusoidal waves	157
11.2.2 Vibrometry	163
11.2.3 Square wave	166
II Optics	169
12 12 Nature of Light and Reflection 3.1.1, 3.1.2	171
12.1 What is Light?	171
12.2 Measurements of the Speed of Light	173
12.2.1 Rømer's Measurement of the speed of light	173
12.2.2 Fizeau's Measurement of the speed of light	174
12.2.3 Faster than light	175
12.3 Reflection	177
12.4 Law of reflection	179
12.4.1 Retroreflection	180
12.5 Reflections, Objects, and seeing	182
13 13 Refraction 3.1.3, 3.1.4	185
13.1 Speed of light in a material	188
13.2 Change of wavelength	190
13.3 Index of refraction and Snell's Law	191
13.4 Total Internal Reflection	192
13.5 Fiber Optics	194
14 14 Dispersion and Huygen's Principle 3.1.5 3.1.6	197
14.1 Dispersion	197
14.1.1 Calculation of n using a prism	199
14.2 Huygen's Principle	201
14.3 Filters and other color phenomena	207
15 15 Polarization and Image formation 3.1.7, 3.2.1	209
15.1 Polarization of Light Waves	209
15.1.1 Polarization by removing all but one wave orientation . .	211
15.1.2 Polarization by reflection	212
15.1.3 Birefringence	216
15.1.4 Optical Stress Analysis	217

15.1.5	Polarization due to scattering	218
15.1.6	Optical Activity	218
15.1.7	Laser polarization	219
15.2	Imaging with Flat Mirrors	219
15.2.1	Virtual Images	221
15.2.2	Mirror reversal	222
16	16 Images formed by Curved Mirrors and by Refraction 3.2.2, 3.2.3	225
16.0.3	Concave Mirrors	226
16.0.4	Paraxial Approximation for Mirrors	227
16.1	Mirror Equation	229
16.1.1	Ray Diagrams for Mirrors	231
16.1.2	Principal rays for a convex mirror:	235
16.2	Aberrations	236
16.3	Semi-Infinite Lenses and the Semi-Infinite Lens Image Equation .	238
16.3.1	Flat Refracting surfaces	240
17	17 Thin Lenses 3.2.4, 3.2.5	243
17.1	Images from lenses	243
17.2	Virtual Images with Lenses	252
17.2.1	Diverging Lenses	253
17.3	Thin Lenses	254
17.3.1	Derivation of the lens equation	255
17.3.2	Lens Aberrations	259
17.3.3	Sign Convention	260
17.3.4	Magnification	261
17.4	The Eye	262
17.4.1	Nearsightedness	263
17.4.2	Farsightedness	264
17.4.3	Color Perception	264
18	18 Cameras and Magnifiers 3.2.6, 3.2.7	267
18.1	The Camera	267
18.2	Angular Magnification	271
19	19 Optical Systems and back to waves 3.2.8, 3.3.1	277
19.1	Combinations of Lenses	278
19.2	The Microscope	281
19.3	Telescopes	284
19.3.1	Refracting Telescopes	284
19.3.2	Reflecting Telescopes	285
19.4	Interference and Young's Experiment	286

20 20 Double Slits 3.3.2, 3.3.3	289
20.0.1 Constructive Interference	289
20.0.2 Destructive Interference	292
20.1 Multiple Slits	293
21 21 Thin Films and Light Interference 3.3.4, 3.3.5	297
21.1 Mathematical treatment of single frequency interference	297
21.1.1 Example of two wave interference: Stealth Fighter	300
21.1.2 Example of two wave interference: Soap Bubble	305
21.2 The Michelson Interferometer	308
22 22 Slit Intensity 3.4.1, 3.4.2	311
22.1 Single Slits	311
22.2 Narrow Slit Minima	314
22.3 Intensity of the single-slit pattern	316
22.3.1 Phasors	317
22.3.2 Using phasors to find the intensity pattern	321
23 23 Double Slit and Grating intensity patterns 3.4.3, 3.4.4	329
23.1 Double Slit Intensity Pattern	329
23.2 Diffraction Gratings	335
23.2.1 Resolving power of diffraction gratings	337
24 24 Resolution 3.4.5, 3.4.6, 3.4.7	341
24.1 Circular apertures	341
24.2 Resolution	344
24.3 Resolution	345
24.4 Holography	350
24.5 Diffraction of X-rays by Crystals	351
24.6 Ray Model vs. Wave Model	352
24.7 The Ray Approximation in Geometric Optics	354
24.8 The ray model and phase	356
24.9 Coherency	359
III Thermodynamics	361
25 25 Fluids and Pressure 1.14.1, 1.14.2	363
25.1 Fluids	363
25.1.1 What is a fluid?	365
25.2 Pressure	366
25.2.1 Working with the definition of pressure	367
25.2.2 Pressure Example:	368
25.3 Variation of Pressure with Depth	369
25.3.1 Pressure variation in liquids	369
25.3.2 Pressure variation in gasses	375

25.4 Pressure Measurements	377
25.4.1 The Barometer	378
25.4.2 Manometer	379
26 26 Pascal and Archimedes 1.14.3, 1.14.4	381
26.1 Hydraulic Press	381
26.2 Buoyant Forces, Archimedes' Principle	387
26.3 Buoyant Force: Two Cases	389
26.3.1 Totally submerged object	389
26.3.2 Partially submerged object	390
27 27 Conservation of Energy for Fluid Flow 1.14.5, 1.14.6	393
27.1 Moving Fluids	393
27.2 Equation of continuity	395
27.3 Bernoulli's Equation	397
27.3.1 Work done in Δt	397
27.3.2 Kinetic Energy	400
27.3.3 Potential Energy	401
27.3.4 Total Work done on the system	401
28 28 Turbulence and Temperature 1.14.7, 2.1.1	403
28.1 Viscosity	403
28.1.1 Turbulence	407
28.2 Temperature and the Zeroth Law of Thermodynamics	409
28.3 Numbers of Atoms and Number Densities	410
28.3.1 Atomic Mass	410
28.3.2 Moles	411
28.3.3 Molar mass	412
29 29 Measuring Temperature 2.1.2, 2.1.3	413
29.1 How to build a thermometer	414
29.1.1 Celsius temperature scale	414
29.1.2 The Constant-Volume Gas Thermometer and the Absolute Temperature Scale	416
29.1.3 Temperature scale conversions	417
29.2 Thermal Expansion	418
29.2.1 Expansion Coefficient	419
29.2.2 Volume expansion	420
29.2.3 Water is weird (and the Lord knows what He is doing!) .	422
30 30 Thermal Properties of Matter 2.1.4, 2.1.5	425
30.1 Heat	427
30.1.1 Units	427
30.1.2 Mechanical equivalent of heat	428
30.2 State Variables	429
30.3 Phase Changes and Phase Diagrams	434

30.4 Phase changes and heat of transformation	436
30.5 Calorimetry	439
31 31 Heat Transfer and Gas Laws 2.1.6, 2.2.1	443
31.1 Heat Transfer Mechanisms	443
31.1.1 Thermal Conduction	444
31.1.2 Convection	446
31.1.3 Radiation	448
31.1.4 Multiple transfer mechanisms	452
31.2 Gas Laws	452
31.2.1 What lead to the Ideal Gas Law	454
31.2.2 Alternate form of the Ideal Gas Law	456
31.3 PV diagrams	456
31.4 Van der Waals Gasses	457
32 32 Kinetic Theory of Gases 2.2.2, 2.2.3	461
32.1 Molecular Model of an Ideal Gas	462
32.2 Microscopic Definition of Pressure	462
32.3 Temperature	467
32.4 Mean Free Path	469
32.5 Molar Specific Heat	472
32.6 The Equipartition of Energy	473
32.7 Quantization	476
32.8 Molar specific heat of solids	477
32.9 Partial Pressure and Vapor Pressure	479
33 33 Speeds and Systems 2.2.4, 2.3.1	481
33.1 Speed of molecules	481
33.2 Distribution of Molecular Speeds	484
33.3 Thermodynamic systems	491
33.4 Reversible and Irreversible Processes	493
34 34 First Law and Work 2.3.2, 2.3.3	495
34.1 Conservation of Energy: First Law of Thermodynamics	495
34.1.1 Work and Internal energy	497
34.1.2 An Isolated System	498
34.1.3 A Cyclic Process	498
34.2 Work	499
35 35 Thermodynamic Processes 2.3.4, 2.3.5	511
35.1 Heat and systems	512
35.1.1 Constant Volume Process	514
35.1.2 Constant Pressure Process	515
35.2 Constant Temperature Process	519

36 36 Adiabatic Processes and 2nd Law 2.3.6, 2.4.1	521
36.1 Constant Q Processes	522
36.1.1 Path Dependence	527
36.2 Irreversibility	528
36.3 The Second Law of Thermodynamics	530
37 37 Producing Useful Work 2.4.2, 2.4.3	531
37.1 Useful work	531
37.2 Reservoirs	532
37.3 Heat Engines	534
37.3.1 Example: Steam engine	535
37.4 Heat Pumps and Refrigerators	537
37.4.1 Coefficient of Performance	538
38 38 Ideal Gas Heat Engines 2.4.4, 2.4.5	541
38.1 An example: The Otto Cycle	541
38.2 Carnot, the Limit of Efficiency for Heat Engines	556
38.2.1 The Carnot Cycle	557
38.3 Process $A \rightarrow B$ isothermal expansion	558
38.4 Process $B \rightarrow C$ adiabatic expansion	561
38.5 Process $C \rightarrow D$ isothermal compression	563
38.6 Process $D \rightarrow A$ adiabatic compression	565
39 39 Microscopic Energy Transfer Considerations 2.4.6, 2.4.7	571
39.1 Microscopic View of Energy Transfer	571
39.2 The arrow of time	574
39.3 Disorder	575
39.4 Extension to large spaces and many particles	576
39.5 A Measure of Disorder	577
39.6 Definition of entropy	578
39.6.1 Entropy and the Second Law	579
39.6.2 Macroscopic view	579
39.6.3 Heat death of the Universe	581
39.6.4 The First Law and LDS Thought ¹	582
39.6.5 The Second Law and LDS Thought	583
A Experiment 1	585
A.1 Introduction	585
A.2 Equipment	586
A.2.1 Assignment	586

¹Comparing physical theory to LDS thought has it's dangers, I am aware. But as scientists, we can't escape thinking about this. Just as a caution, John A. Widsoe wrote a book called *Joseph Smith as Scientist* in which Widsoe tried very hard to show that LDS thought is in harmony with Universal Ether Theory. But we now know Universal Ether Theory is not correct, so being in harmony with it is meaningless. So we should be cautious!

B Experiment 2	587
B.1 Introduction	587
B.2 Equipment	588
B.3 Assignment	588
C Experiment 3	589
C.1 Introduction	589
C.2 Equipment	590
C.3 Assignment	590
Afterword	595

Preface

This set of notes is intended to be an aid to the student. It is what I intend to present in class. There are likely errors and mistakes, so use these notes, but don't expect perfection. If there are things that are confusing, please talk to me or ask questions in class.

I consulted many authors and colleagues in creating these notes. Principle among the authors was Serway and Jewett and Haliday and Resnic. Colleagues of special note are Kevin Kelly, Brian Pyper, Steve Tercotte, and Ryan Nielson. I especially appreciate the PH123 students of the Fall of 2006. They were the first to test run this sets of notes, and I appreciate their feedback (and patience).

This work is © Copyright R. Todd Lines. Permission is given to BYU-Idaho for educational purposes. All other rights are reserved.

Part I

Oscillation and Waves

Chapter 1

Introduction to the Course, Simple Harmonic Motion

1.15.1

Fundamental Concepts

- Simple Harmonic Motion
- Frequency and Period
- Mathematical description of Simple Harmonic Motion

1.1 1 Oscillation and Simple Harmonic Motion

Last semester, in PH121 (or Dynamics) you studied how things move. We identified a moving object (I often refer to this object as the *mover object*) and other objects that exerted forces on the mover (I often refer to these other objects as the *environmental objects*). You learned about forces and torques which get mover objects moving. You should remember Newton's Second Law and Newton's Second Law for rotation.

You also learned and practiced a lot of math. We will continue to use the math you learned in PH121 this semester.

But we will go beyond what we learned in PH121 to study new types of motion, and new objects that move.

This semester, we will start with a very special type of motion. It is the motion that results from oscillation. We call this very special type of motion, *simple harmonic motion*.

Simple harmonic motion (SHM) means a motion that repeats in the most special, simple way. Some characteristics of this type of motion are as follows:

1. The motion repeats in a regular way, like a grandfather clock pendulum swings back and forth in a set amount of time. This set amount of time is called a period.

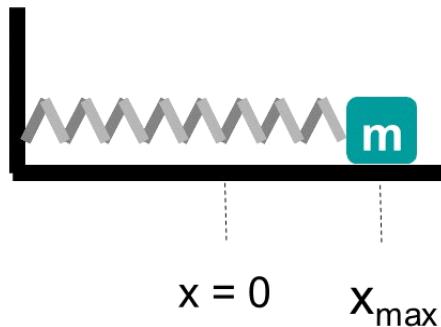
2. The mover object moves about a center position and is symmetric about that position. This center position is the bottom of the swing for our grandfather clock. The idea of symmetry just means that the pendulum reaches the same height on both sides as it swings

3. The mover object's motion can be traced out in a position vs. time graph. For SHM, the shape of the position vs. time graph is a sinusoid.

4. The velocity of the mover object constantly changes. It is zero at the extreme points (the points farthest from the center, or the largest positive and the largest negative displacements). That is where it turns around and goes back the other way. It stops there, but just for a split second. The velocity is largest at the equilibrium position (the center position). If we plotted a velocity vs. time graph, the velocity would also be a sinusoid or snake-like shape.

5. The acceleration also constantly changes. And it is also a sinusoid.

That is a lot of criteria for a motion that is supposed to be simple! Let's take an example system to see how simple harmonic motion works. I have drawn a mass (marked m) attached to a spring. And let's assume that the mass is on a frictionless surface and there is no air drag force. If we pull the mass so the string stretches, we would get simple harmonic motion.



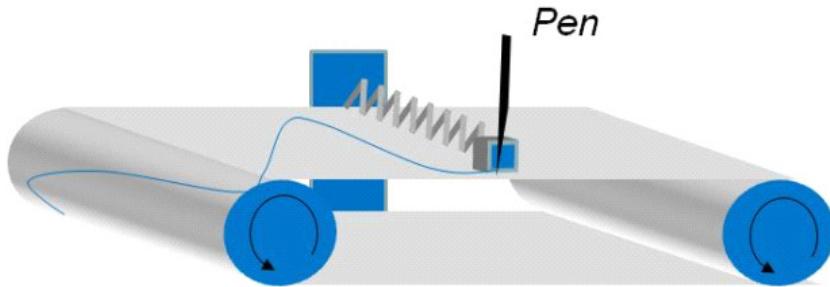
The center position for our mass is called the Equilibrium position, and for convenience we often define it as the origin of our coordinate system.

Equilibrium Position: The position of the mover mass (not the spring, which is an environmental object acting on the mover mass) when the spring is neither stretched nor compressed.

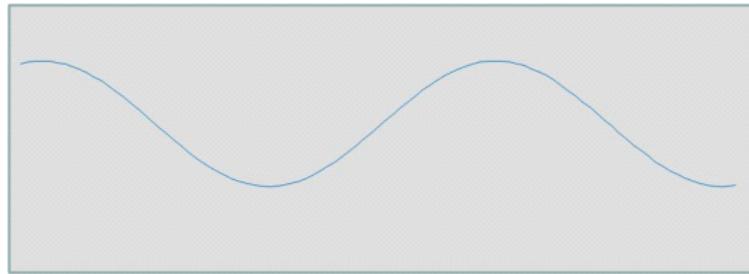
Let's draw a picture of simple harmonic motion. First we need a device to do this. Suppose you go to the supermarket¹. But instead of putting your ramen

¹Like Albertsons, where they still have checkers.

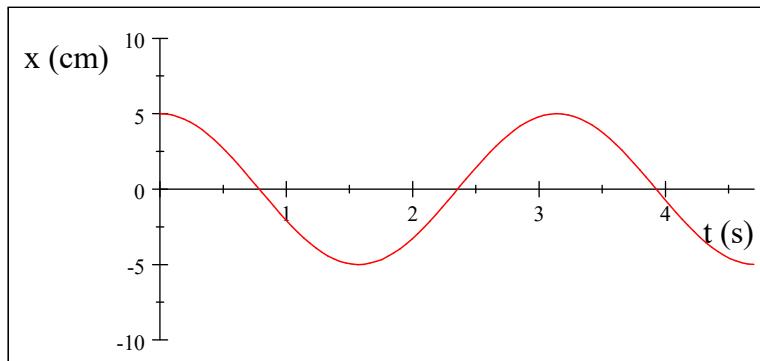
and granola bars on the belt at the checkout counter, you strap on a device like this²



You can see the mover object on the spring. But we have placed a pen on the object and that pen is tracing out a pattern as the belt moves. You will recognize this pattern as a trig function. The result might look something like this if you removed the belt.



Of course, what we have made is a position vs. time graph. We remember these from PH121! This gives us a record of the past motion of the object.



where in this graph, $x_{\max} = 5$ cm. Having the power of mathematics, we know we can write an equation that would describe this curve. From your Trigonometry

²Please don't really try this at the local supermarkets. They don't seem to have any sense scientific inqurey or even a sense of humor about such things at all.

experience, we can guess that an equation for our mover object motion might look something like

$$x(t) = x_{\max} \cos(\theta)$$

Notice that the instantaneous position, $x(t)$ has to be a function of time. Back in trigonometry we would have said a cosine function was a function of an angle, θ . But we know this is a position vs. time graph, so our angle must be different for different times. Let's write

$$\theta(t) = \omega t$$

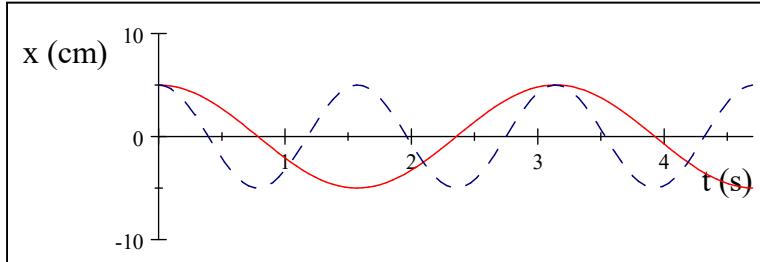
as strange sort of an angle. This “angle” is a function of time, and let's say how fast our “angle” is changing is given by

$$\frac{d\theta}{dt} = \omega$$

This is a sort of speed for how fast our angle is changing. So our equation must be

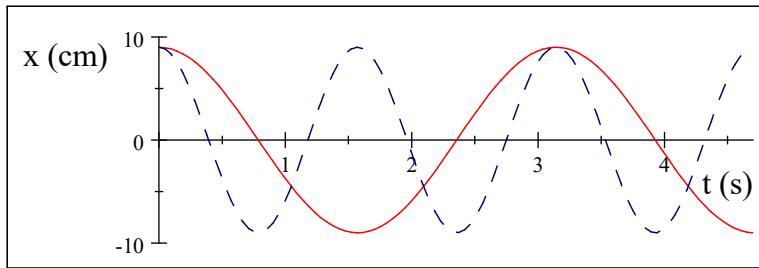
$$x(t) = x_{\max} \cos(\omega t)$$

In the next figure $x(t) = x_{\max} \cos(\omega t)$ is plotted with two different values for ω .



We can see what ω does for us. It stretches out or compresses our curve. In the blue (dashed) curve our angle changes quickly. In the red (solid) curve the angle changes slowly. Note that we are not plotting position vs. angle. Both plots would be the same if we did! We are plotting position vs. time. And for the blue curve our mass is oscillating much more quickly than it is for the red curve. The quantity ω tells us something about how fast our mover object is oscillating.

The quantity x_{\max} is the maximum displace of our mover object from the equilibrium position. This maximum displacement from equilibrium has a name. It is called the *amplitude*. In our equation I have used x_{\max} where most books use the letter A for amplitude to emphasize that the amplitude is the maximum displacement of the mover mass from the mover mass equilibrium position. In our coordinate system, the mass is going back and forth in the x direction. Since the amplitude means the maximum displacement, x_{\max} is a good way to write amplitude. Here is the same graph but with $x_{\max} = 9\text{ cm}$



The two graphs for the two different ω values are not more stretched out in time, but now they are taller along the position axis. This means that the mass is moving farther from the center as it oscillates.

You will remember from your trigonometry class that the *period*, T , tells us the time it takes for the oscillation to go through a complete cycle. A complete cycle is when the object, say, goes from x_{\max} to $-x_{\max}$ and then back to x_{\max} . You can probably guess that how long it takes to oscillate and how often it oscillates would be related. How often the oscillator completes a cycle is called the frequency. The longer the period, the lower the frequency.

$$f = \frac{1}{T}$$

Think of cars passing you on your way to class. Period is like how long you wait in between cars. Frequency is like how often cars pass. If you wait less time between cars, the cars pass more frequently. And that is just what our equation says! We can see from our graphs that our stretching quantity ω , must be related to the frequency of our oscillation. If the frequency is high, then ω must be large so that we reach different “angles” faster. But for the cosine function to work we need angle units. *We will choose radians* for our units and we will write our stretching quantity as

$$\omega = 2\pi f$$

This works! If ω is bigger then our oscillation happens more frequently. The 2π has units of radians. So ω has units of rad Hz or more commonly rad/s. That matches our derivative above. Let’s give a name to the quantity ω . Since it has radians in it we might guess that it has something to do with circular motion (more on this later) and it has frequency in it. So we will call ω the *angular frequency*.

1.2 Velocity and Acceleration

So far we have guessed the descriptive equation for SHM.

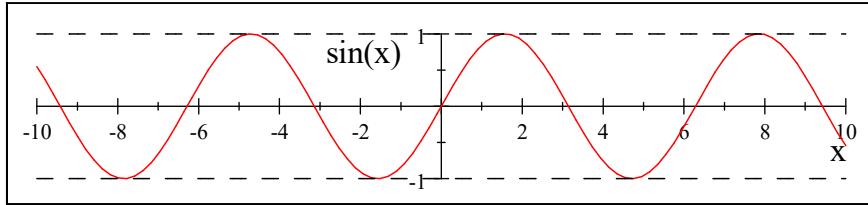
$$x(t) = x_{\max} \cos(\omega t) \quad (1.1)$$

This gives us the position of the particle at any given time. Since knowing the position as a function of time is a good description of the motion of our mover

mass, we can call this equation the *equation of motion* for our mass. We will see soon that this equation is correct. But let's pretend that we are earlier researchers and we just know the equation makes the right shape. Still, we can learn much from knowing this equation. Since we know how to take derivatives, and we know the derivative of position with respect to time is the velocity, we can see that the velocity of the mass at any given time is given by

$$v(t) = \frac{dx(t)}{dt} = -\omega x_{\max} \sin(\omega t) \quad (1.2)$$

From our fond memories of our trigonometry class we know the maximum of a sine function is always 1.



Notice that the ω and the x_{\max} aren't changing for our oscillation mover mass. They are constant. Then if we want the maximum speed we can simply set the $\sin(\omega t) = 1$. Then the maximum speed will be

$$v_{\max} = \omega x_{\max} (1) \quad (1.3)$$

Notice that this does not tell us when the speed is maximum. Just what the maximum speed is. We then have an equation for the speed of our object as a function of time

$$v(t) = -v_{\max} \sin(\omega t)$$

We will often use this trick of knowing the maximum of sine is one.

We can also find the acceleration. We just take another derivative.

$$\begin{aligned} a(t) &= \frac{dv}{dt} \\ &= \frac{d^2x(t)}{dt^2} \\ &= -\omega^2 x_{\max} \cos(\omega t) \end{aligned}$$

This is the acceleration of the object attached to the spring. It's also true that the maximum for a cosine function is 1, so the maximum acceleration would be

$$a_{\max} = \omega^2 x_{\max} (1) \quad (1.4)$$

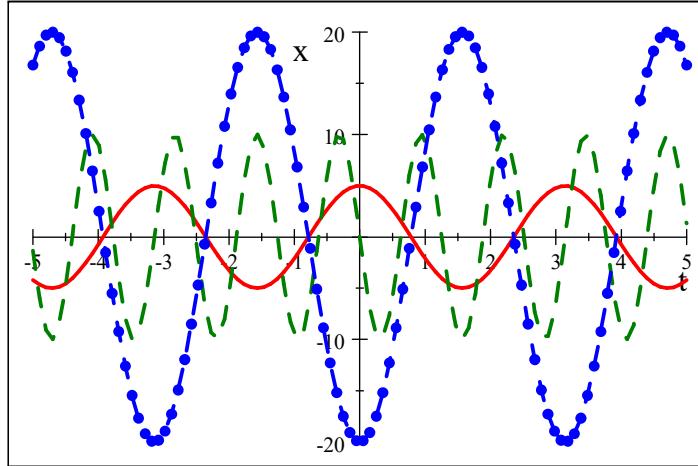
so we could write our instantaneous acceleration as

$$a(t) = -a_{\max} \cos(\omega t)$$

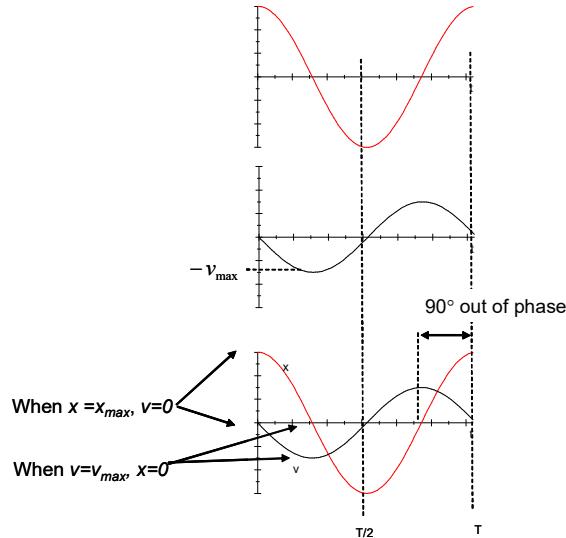
These are significant results, so let's summarize. For a simple harmonic oscillator, the instantaneous position, speed, and acceleration are given by

$$\begin{aligned} x(t) &= x_{\max} \cos(\omega t) \\ v(t) &= -\omega x_{\max} \sin(\omega t) \\ a(t) &= -\omega^2 x_{\max} \cos(\omega t) \end{aligned} \quad (1.5)$$

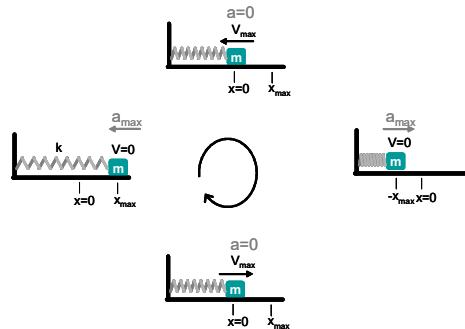
Let's plot $x(t)$, $v(t)$, and $a(t)$ for a specific case



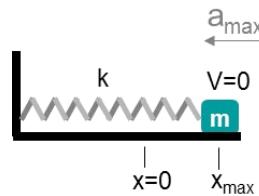
Red (solid) is the displacement, green (dashed) is the velocity, and blue (dot-dashed) is the acceleration. Note that each has a different maximum amplitude. Also note that they don't rise and fall at the same time. We will describe this as being *not in phase*.



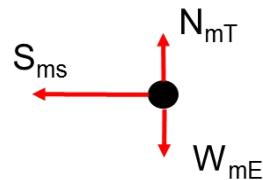
The acceleration is 90° *out of phase* from the velocity. Let's think about why this would be. Suppose we attach a mass to a spring and allow the mass to slide on a frictionless surface.



Let's start by stretching the spring by pulling the mass to the right and releasing. This is the situation in the left hand part of the last figure.



The spring is pulling strongly on the mass. We could draw a free body diagram for this situation. We will need the spring force



Back in PH121 we said that Hooke's Law is not something that is always true, but by "law" we mean a mathematical representation (and equation) that comes from our mental model of how the universe works. In this case, Hooke's law is an equation that comes from Hooke's model of how springs work. It is a good model for most springs as long as we don't stretch them too far. You remember Hooke's law from PH 121. It tells us that the spring force is proportional to how far we stretch or compress the spring (Δx_e) and how stiff the spring is (k).

$$\begin{aligned} F_s &= S = -k\Delta x_e & (1.6) \\ &= -k(x - x_e) \end{aligned}$$

where x_e is the *equilibrium position of the mass*. If we assume the equilibrium position is at $x = 0$ then we can write our spring force as

$$S = -kx \quad (1.7)$$

And if we write out Newton's second law we get

$$\begin{aligned} F_{net_x} &= -ma_x = -S_{ms} \\ F_{net_y} &= ma_y = N_{mT} - W_{mE} \end{aligned}$$

We can see that a_y should be zero because the mass won't lift off the table in the y -direction. but in the x -direction

$$-a_x = -\frac{S_{ms}}{m}$$

and knowing that

$$S_{ms} = -kx$$

from Hooke's law, we can see that

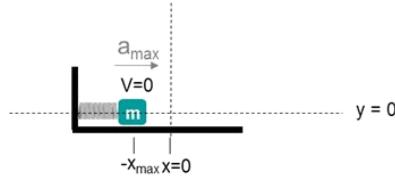
$$-a_x = +\frac{kx}{m}$$

or

$$a_x = -\frac{kx}{m}$$

Since at our starting point x is big our acceleration is big. But there is a minus sign. The minus sign tells us that a_x must be to the left.

But suppose the mass was to the left of $x = 0$



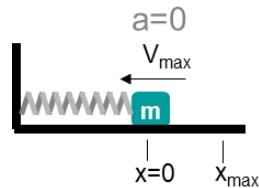
Now $x < 0$ so it is negative. That makes

$$a_x = -\frac{kx}{m}$$

positive. The acceleration always seems to point toward $x = 0$, the equilibrium position for the mass. This is an important part of the definition of simple harmonic motion, having an acceleration that always points toward the equilibrium position. We call a force that makes the acceleration point toward the equilibrium position *restoring force*.

Restoring force: A force that is always directed toward the equilibrium position

Now lets look at our object a short time later



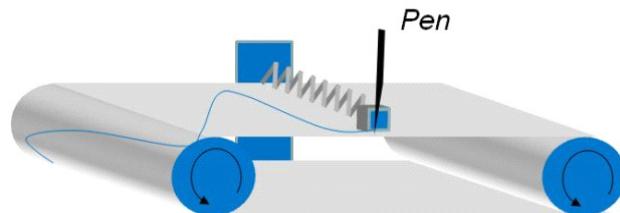
We can see that in this position, $x = 0$ so

$$a_x = \frac{k(0)}{m} = 0$$

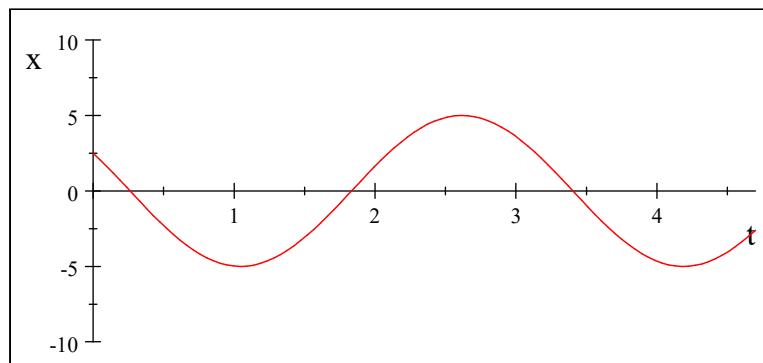
so at this position the acceleration is zero. That is because right at $x = 0$ the spring is neither pushing nor pulling. There is no net force so no acceleration when $x = 0$.

1.3 The Idea of Phase

We said before that $x(t)$, $v(t)$, and $a(t)$ are “out of phase.” Let’s look at the idea of “phase” more carefully. Suppose you return to the grocery store and start your SHM device.



But this time you work with a lab partner, and the lab partner tries to start a stopwatch when you let go of the mass. But, due to having a slow reaction time, your partner starts the clock too late. The resulting graph looks like this

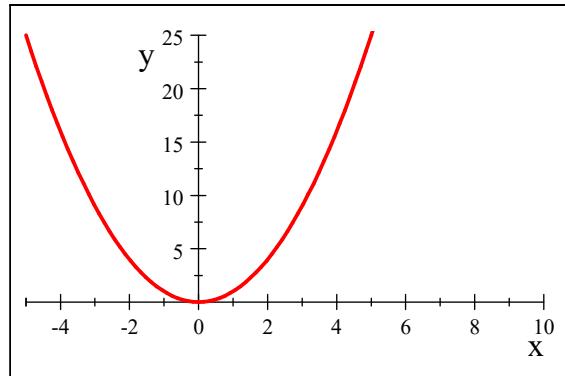


But we know that the only difference between this graph and the one we had before is that the lab partner was slow, so the graph is shifted on the time axes. We expect we can use the same mathematical model for SHM, but we must need to change something.

We know from our algebra classes what a shift looks like. Take the expression

$$y = x^2$$

We can plot this to get a parabola



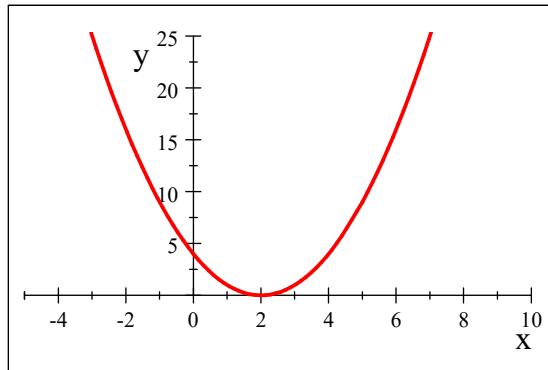
A shifted parabola would be expressed as

$$y = (x - \phi_o)^2$$

where ϕ_o is the amount of the shift. Suppose $\phi_o = 2$, then

$$y = (x - 2)^2$$

and our shifted parabola looks like this



Recall that a shift like

$$y = (x - \phi_o)^2$$

will move the parabola to the right while

$$y = (x + \phi_o)^2$$

will move the parabola to the left.

We can use this idea to write our SHM expression for our situation with the slow lab partner.

$$x(t) = x_{\max} \cos(\omega(t \pm \tau_o))$$

In our case, the lab partner was late by $\tau_o = 4.7124\text{ s}$. This is not usually how we express the shift, however. We usually distribute the ω so our equation looks like

$$\begin{aligned} x(t) &= x_{\max} \cos(\omega t \pm \omega \tau_o) \\ &= x_{\max} \cos(\omega t \pm \omega \tau_o) \end{aligned}$$

We usually use the symbol $\phi_o = \omega \tau_o$ so we will write our SHM expression for position as a function of time as

$$x(t) = x_{\max} \cos(\omega t \pm \phi_o)$$

We could call ϕ_o the *slow lab partner constant*, but that is long and not very kind. So let's call ϕ_o by the name *phase constant*. It is also customary to call the entire expression in parenthesis, $(\omega t \pm \phi_o)$, the phase of the cosine function. This is especially used in the fields of Optics and Electrodynamics.

From what we did before, we know that $x(t)$, $v(t)$, and $a(t)$ are out of phase, so there must be a phase constant involved somehow. Let's look for it in what follows.

Chapter 2

2 Energy and Dynamics of SHM 1.15.2 1.15,3

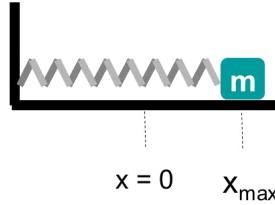
Back in PH121 we started with position, velocity, and acceleration to describe motion. We have done that again for a simple harmonic oscillator. In PH121, after basic motion, we found that we could use the idea of a force and Newton's second law to find the acceleration for our description of motion. Then we changed view points and used the idea of energy to find motion. The energy picture was easier in some ways. Let's try to develop an energy picture for simple harmonic oscillators. Of course the goal in physics is to describe the motion of the object, so we will say that we are looking for the *dynamics* of simple harmonic oscillators when we look for the position, velocity, and acceleration as a function of time.

Fundamental Concepts

- Initial Conditions
- Energy and SHM
- Equation of motion
- Vertical oscillations

2.1 Initial Conditions

Usually, we need to know how we start our oscillator to solve a problem. Let's see how this works.



Suppose we start the motion of a mass attached to a spring (a harmonic oscillator) by pulling the mass to $x = x_{\max}$ and releasing it at $t = 0$. Let's see if we can find the phase. Our initial conditions require

$$\begin{aligned} x(0) &= x_{\max} \\ v(0) &= 0 \end{aligned}$$

Using our formula for $x(t)$ and $v(t)$ we have

$$\begin{aligned} x(0) &= x_{\max} = x_{\max} \cos(0 + \phi_o) \\ v(0) &= 0 = -v_{\max} \sin(0 + \phi_o) \end{aligned}$$

If we choose $\phi_o = 0$, these conditions are met.

Notice that we needed to know the starting time and the position and the velocity at that time. These are what we call *initial conditions*. It is still true that $x(t)$ and $v(t)$ are out of phase. But we found $\phi_o = 0$. There is another phase term hiding in our expression for $v(t)$ and to find it we need a small trig identity.

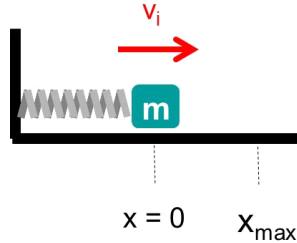
$$\begin{aligned} \sin(\alpha + \phi_o) &= \cos\left(\alpha - \left(\frac{\pi}{2} - \phi_o\right)\right) \\ &= \cos\left(\alpha - \frac{\pi}{2} + \phi_o\right) \end{aligned}$$

so we could write

$$\begin{aligned} v(t) &= -\omega x_{\max} \sin(\omega t + \phi_o) \\ &= -\omega x_{\max} \cos\left(\alpha - \frac{\pi}{2} + \phi_o\right) \end{aligned}$$

and we can see that there was a phase shift of $-\pi/2$ hiding in our sine function. So $v(t)$ really must be out of phase with $x(t)$.

2.1.1 A second example



Using the same equipment, let's start with

$$\begin{aligned}x(0) &= 0 \\v(0) &= v_i\end{aligned}$$

that is, the mover object is already moving when we start our experiment, and we start watching just as it passes the equilibrium point.

$$\begin{aligned}x(0) &= 0 = x_{\max} \cos(0 + \phi_o) \\v(0) &= v_i = -v_{\max} \sin(0 + \phi_o)\end{aligned}$$

from the first equation we have

$$0 = x_{\max} \cos(\phi_o)$$

and that gives us

$$\phi_o = \cos^{-1}(0)$$

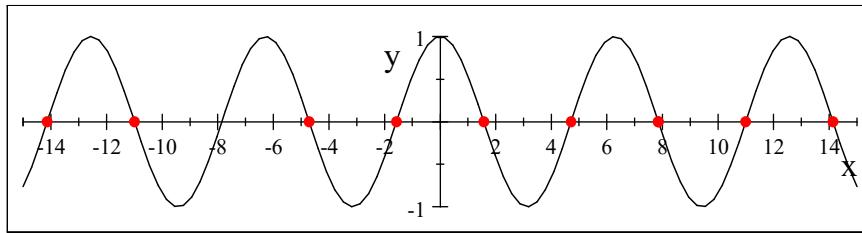
so

$$\phi_o = \pm \frac{\pi}{2}$$

really this is

$$\phi_o = \pm \frac{\pi}{2} \pm n\pi \quad n = 0, 1, 2, \dots$$

This gives the red dots in the next plot.



Notice that at each of these locations $\cos(\phi)$ is zero. but let's make an agreement that we will choose the smallest value for ϕ_o that makes $\cos(\phi_o) = 0$. That is our

$$\phi_o = \pm \frac{\pi}{2}$$

But positive and negative $\pi/2$ are the same “smallness.” We don’t know the sign. Using our initial velocity condition will let us determine the sign. Let’s try it

$$\begin{aligned} v_i &= -v_{\max} \sin\left(\frac{\pi}{2}\right) \\ v_i &= \pm v_{\max} \\ v_i &= \pm \omega x_{\max} \end{aligned}$$

From this we can see

$$x_{\max} = \pm \frac{v_i}{\omega}$$

We defined the initial velocity as positive, and we insist on having positive amplitudes, so then we choose

$$\phi_o = -\frac{\pi}{2}$$

so

$$\sin\left(-\frac{\pi}{2}\right) = -1$$

and then this minus sign will cancel the one in our initial velocity equation that to make our initial velocity positive.

$$\begin{aligned} v_i &= -v_{\max} \sin\left(\frac{\pi}{2}\right) \\ &= -v_{\max} (-1) \\ &= v_{\max} \end{aligned}$$

Our solutions are

$$\begin{aligned} x(t) &= \frac{v_i}{\omega} \cos\left(\omega t - \frac{\pi}{2}\right) \\ v(t) &= v_i \sin\left(\omega t - \frac{\pi}{2}\right) \end{aligned}$$

Note that our solution is a set of equations! It isn’t a set of numbers. We saw this sometimes in PH121. But we will see it more in our study of oscillations and waves. Sometimes the solutions are equations. But we did fill in our equations with the number for the phase constant. it would be even better if we had numbers for ω and v_i .

Generally to have a complete solution, you must find all the constants based on the initial conditions. This would mean we need x_{\max} , ω , and ϕ_o to have a complete solution.

Let’s try a complete example problem.

2.1.2 Example

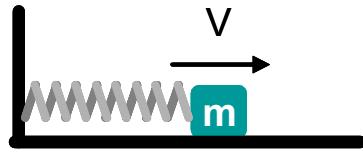
A particle moving along the x axis in simple harmonic motion starts from its equilibrium position, the origin, at $t = 0$ and moves to the right. The amplitude of its motion is 2.00 cm, and the frequency is 1.50 Hz.

a) show that the position of the particle is given by

$$x = (2.00 \text{ cm}) \sin(3.00\pi t)$$

determine

- b) the maximum speed and the earliest time ($t > 0$) at which the particle has this speed,
- c) the maximum acceleration and the earliest time ($t > 0$) at which the particle has this acceleration, and
- d) the total distance traveled between $t = 0$ and $t = 1.00 \text{ s}$



Basic Equations

$$\begin{aligned} x(t) &= x_{\max} \cos(\omega t + \phi_0) \\ v(t) &= -\omega x_{\max} \sin(\omega t + \phi_0) \\ a(t) &= -\omega^2 x_{\max} \cos(\omega t + \phi_0) \end{aligned}$$

$$\omega = 2\pi f$$

$$v_m = \omega x_m$$

$$a_m = \omega^2 x_m$$

$$T = \frac{1}{f}$$

Variables

t	time, initial time = 0	$t_i = 0$
x	Position, Initial position = 0	$x(0) = 0$
v		
a		
x_{\max}	x amplitude	$x_{\max} = 2.00 \text{ cm}$
v_m	v amplitude	
a_m	a amplitude	
ω	angular frequency	
ϕ_o	phase constant	
f	frequency	$f = 1.50 \text{ Hz}$

Symbolic Solution

Part (a)

We can start by recognizing that we know ω because we know the frequency.

$$\begin{aligned}\omega &= 2\pi f \\ &= 2\pi (1.50 \text{ Hz}) \\ &= 9.4248 \text{ rad Hz}\end{aligned}$$

We also know the amplitude $A = x_{\max}$ which is given.

$$A = 2.00 \text{ cm}$$

Knowing that at $t = 0$

$$x(0) = 0 = x_{\max} \cos(0 + \phi_o)$$

which we have seen before! We can guess that

$$\phi_o = \pm \frac{\pi}{2}$$

Using

$$v(0) = -\omega x_{\max} \sin\left(0 \pm \frac{\pi}{2}\right)$$

and demanding that amplitudes be positive values, and noting that at $t = 0$ the velocity is positive from the initial conditions:

$$\phi = -\frac{\pi}{2}$$

We also note from our trig identity that we used above

$$\cos\left(\theta - \frac{\pi}{2}\right) = \sin(\theta)$$

that we have

$$\begin{aligned}x(t) &= x_{\max} \cos\left(2\pi ft - \frac{\pi}{2}\right) \\ &= x_{\max} \sin(2\pi ft)\end{aligned}$$

which is in the form we want. All we have to do is put in numbers.

$$\begin{aligned}x(t) &= x_{\max} \sin(2\pi ft) \\&= (2.00 \text{ cm}) \sin(3.00\pi t)\end{aligned}$$

And once again our solution for a) is an equation.

Part (b)

We have a formula from our previous work for v_{\max}

$$\begin{aligned}v_{\max} &= \omega x_{\max} \\&= 2\pi f x_{\max}\end{aligned}$$

so let's use it. To find when v_{\max} happens, take a derivative of $x(t)$

$$v(t) = v_{\max} = -\omega x_{\max} \sin\left(2\pi ft - \frac{\pi}{2}\right)$$

and recognize that $\sin(\theta) = 1$ is at a maximum and this happens when $\theta = \pi/2$. So we take the stuff that is in the parenthesis for the sine function and set it equal to our angle that makes the sine function equal to 1.

$$\frac{\pi}{2} = 2\pi ft - \frac{\pi}{2}$$

We can simplify this and solve for t

$$\begin{aligned}\pi &= 2\pi ft \\ \frac{1}{2f} &= t \\ t &= \frac{1}{2(1.50 \text{ Hz})} = 0.33333 \text{ s}\end{aligned}$$

Part (c) Like with the velocity we must use the formula. But we know it is just taking the derivative of $v(t)$

$$a(t) = -\omega^2 x_{\max} \cos(\omega t + \phi_o)$$

but recognize that the maximum is achieved when $\cos(\omega t + \phi_o) = 1$ or when $\omega t + \phi_o = 0$. We can solve this for t .

$$\begin{aligned}t &= \frac{\phi_o}{\omega} \\&= \frac{-\frac{\pi}{2}}{2\pi f} \\&= \frac{-1}{4f} \\&= -0.16667 \text{ s}\end{aligned}$$

The formula for a_{\max} is also already in our set of equations

$$\begin{aligned} a_{\max} &= -\omega^2 x_{\max} \\ &= -(2\pi f)^2 x_m \end{aligned}$$

Putting in numbers gives

$$\begin{aligned} a_{\max} &= (2\pi 1.5 \text{ Hz})^2 (2.00 \text{ cm}) \\ &= 1.7765 \frac{\text{m}}{\text{s}^2} \end{aligned}$$

Part (d)

We know the period is

$$T = \frac{1}{f}$$

We should find the number of periods in $t = 1.00 \text{ s}$ and find the distance traveled in one period,

$$N = \frac{t}{T}$$

and multiply them together. In one period the distance traveled is

$$d = 4x_m$$

$$d_{\text{tot}} = d * N = d * \frac{t}{T} = 4fx_m t$$

Putting in numbers gives

$$\begin{aligned} d_{\text{tot}} &= 4fx_m t \\ &= 8.00 \text{ cm} * 1.50 \text{ Hz} * 1.00 \text{ s} \\ &= 0.12 \text{ m} \end{aligned}$$

We have come far in only one lecture! We have a set of new equations and a new problem type. In what we did in this lecture we used Newton's Second Law. But in PH121 we learned that often using the idea of energy made problems easier. Can we use energy with simple harmonic motion problems? That is what we will talk about in our next lecture.

2.2 Oscillators and Energy

You might have noticed that we are calling mover objects that experience simple harmonic motion (SHM) by the name *simple harmonic oscillators (SHO)*. Let's consider such a SHO. Because our SHO is moving, we know there must be energy associated with it. To understand the energy involved, let's start with kinetic energy. Recall from PH121 that

$$K = \frac{1}{2}mv^2 \tag{2.1}$$

and we recall that for a spring, we have the spring potential energy given by

$$U_s = \frac{1}{2}kx^2 \quad (2.2)$$

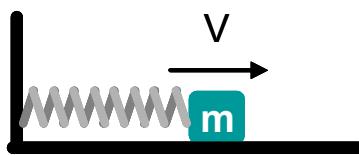
Let's try a problem.

Using energy techniques, find the maximum velocity of the mass in terms of the Amplitude and the angular frequency.

We want to use our problem solving steps:

Type of problem: This is an energy and a SHO problem:

Drawing:



Variables:

E	energy
K	Kinetic energy
x	Position
v	velocity or speed
m	mass of the object
k	spring stiffness constant

Basic Equations:

$$\begin{aligned} K &= \frac{1}{2}mv^2 \\ U_s &= \frac{1}{2}kx^2 \end{aligned}$$

Symbolic Answer

You might say, this is an easy problem, we know from last time that

$$v_{\max} = x_{\max}\omega$$

but let's find this again using the ideas of energy. In PH121 we often found that using energy made problems easier, so this might be worth a little more work now. The total mechanical energy is

$$E = K + U_s + U_g$$

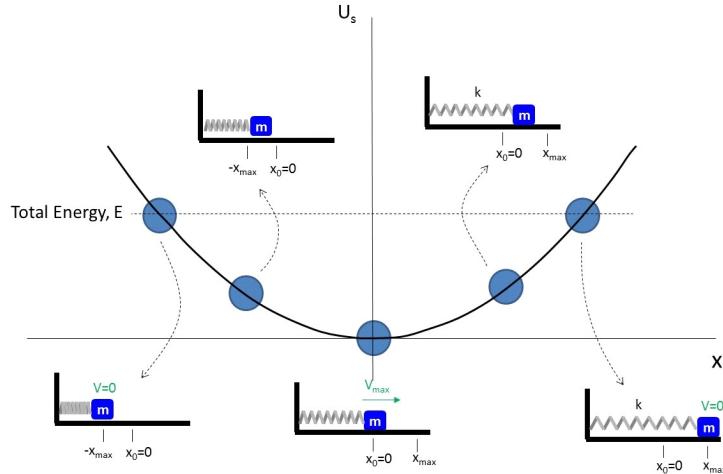
Let's say that our oscillator is moving horizontally just like last time, and that we define the center of mass of our object so that it's y -position is right at $y = 0$ so our object has zero gravitational potential energy.

$$E = K + U_s + 0$$

And let's say we have no friction, so that we can say that energy is conserved. We set the initial and final energies equal to each other.

$$\begin{aligned} E_i &= E_f \\ K_i + U_{si} &= K_f + U_{sf} \\ \frac{1}{2}mv_i^2 + \frac{1}{2}kx_i^2 &= \frac{1}{2}mv_f^2 + \frac{1}{2}kx_f^2 \end{aligned}$$

And let's start with our initial condition being where the spring is stretched to x_{\max} . Then at that moment, $v_i = 0$. And let's take our final position when the mass is moving as fast as possible (because that is what we are looking for, v_{\max}).



We know from our PH121 experience that this will be right when $U_{sf} = 0$ (so all the energy is kinetic). Then

$$0 + \frac{1}{2}kx_{\max}^2 = \frac{1}{2}mv_{\max}^2 + 0$$

or

$$\frac{1}{2}mv_{\max}^2 = \frac{1}{2}kx_{\max}^2$$

We can solve for v_{\max}

$$v_{\max}^2 = \frac{kx_{\max}^2}{m}$$

$$v_{\max} = x_{\max} \sqrt{\frac{k}{m}}$$

But is this what we wanted? We expected that

$$v_{\max} = x_{\max}\omega$$

And this is what we got so long as

$$\omega = \sqrt{\frac{k}{m}}$$

And this is always true for a mass on a spring. We don't need a numeric answer, This is reasonable (just what we expect) and the units check.

Let's look at kinetic and potential energy as a function of time. For our Simple Harmonic Oscillator (SHO) we know the velocity as a function of time,

$$v(t) = -\omega x_{\max} \sin(\omega t + \phi)$$

so the kinetic energy as a function of time must be

$$\begin{aligned} K &= \frac{1}{2}m(-\omega x_{\max} \sin(\omega t + \phi))^2 \\ &= \frac{1}{2}m\omega^2 x_{\max}^2 \sin^2(\omega t + \phi) \end{aligned}$$

and now we know that $\omega = \sqrt{k/m}$, so we can write the kinetic energy as

$$\begin{aligned} K &= \\ &= \frac{1}{2}m\left(\sqrt{\frac{k}{m}}\right)^2 x_{\max}^2 \sin^2(\omega t + \phi) \\ &= \frac{1}{2}m\frac{k}{m}x_{\max}^2 \sin^2(\omega t + \phi) \\ &= \frac{1}{2}kx_{\max}^2 \sin^2(\omega t + \phi) \end{aligned}$$

As the spring is stretched or compressed we store energy as spring potential energy. The potential energy due to a spring acting on our SHO (mover mass) is given by (from your PH121 class)

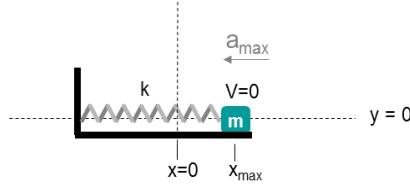
$$U_s = \frac{1}{2}kx^2 \quad (2.3)$$

For our SHO we also know the position as a function of time

$$x(t) = x_{\max} \cos(\omega t + \phi_0)$$

So, the potential energy as a function of time must be

$$U_s = \frac{1}{2}kx_{\max}^2 \cos^2(\omega t + \phi)$$



Let's say, again, that our oscillator is moving horizontally, and that we define the center of mass of our object so that it's y -position is right at $y = 0$ so our object has zero gravitational potential energy

$$U_g = mgy = 0$$

so the total mechanical energy is given by

$$E = K + U_s$$

which we can write out as

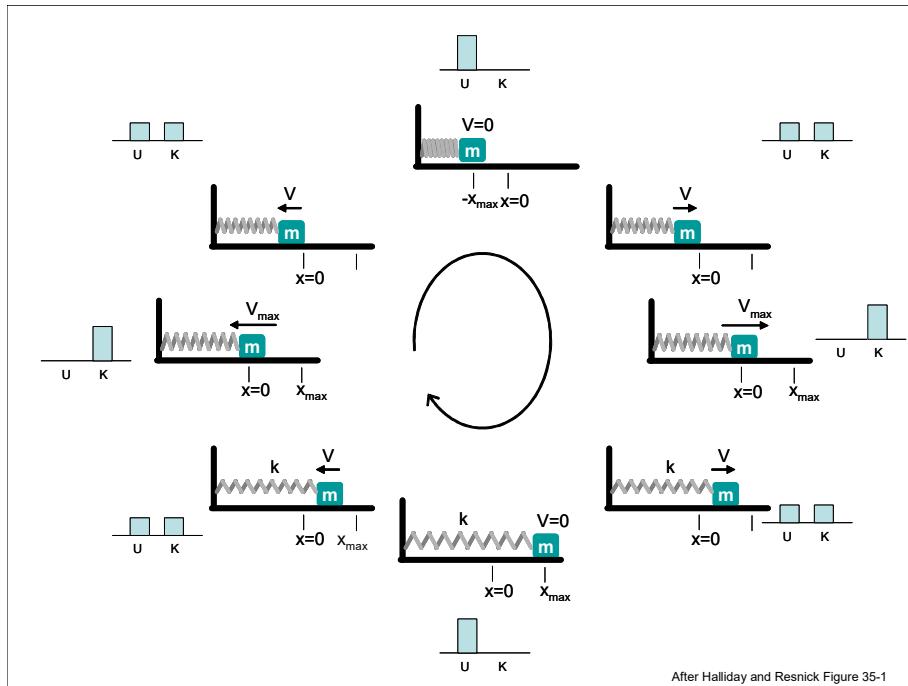
$$\begin{aligned} E &= K + U_s \\ &= \frac{1}{2}kx_{\max}^2 \sin^2(\omega t + \phi) + \frac{1}{2}kx_{\max}^2 \cos^2(\omega t + \phi) \\ &= \frac{1}{2}kx_{\max}^2 (\sin^2(\omega t + \phi) + \cos^2(\omega t + \phi)) \\ &= \frac{1}{2}kx_{\max}^2 \end{aligned}$$

Where we have used the trig identity that $\sin^2(\theta) + \cos^2(\theta) = 1$.

It tells us that if there are no loss mechanisms (e.g. no friction) then the energy in a harmonic oscillator never changes. And we remember that we would say that energy is conserved for such a situation, which is not surprising because we have already used conservation of energy for springs in PH121 and in the previous example. But let's give a new name to a quantity that does not change. Let's call it a *constant of motion*. So we can make a statement about the total energy for our SHO.

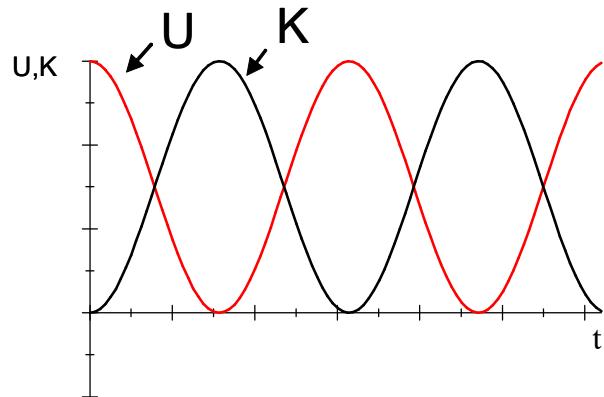
The total mechanical energy of an ideal SHO is a constant of motion

If we plot the amount of kinetic and potential energy for an oscillator we might find something like this:



After Halliday and Resnick Figure 35-1

Note that the kinetic and potential energy are out of phase with each other. If we plot them on the same scale (for the case $\phi = 0$) we have



Let's try another problem using energy of a SHO.

2.3 Mathematical Representation of Simple Harmonic Motion

We have a mathematical representation of simple harmonic motion from looking at our graph of position vs. time. But as a problem, let's use math and what we know from PH121 to show that our equation

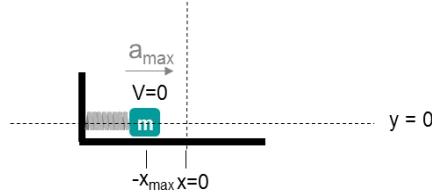
$$x(t) = x_{\max} \cos(\omega t)$$

must be right.

Recall from PH 121

$$a = \frac{dv}{dt} = \frac{d^2x}{dt^2} \quad (2.4)$$

and let's assume we have SHO moving only in the x -direction.



Further assume the surface the object rests on is frictionless. Also let's say that the equilibrium position $x_{em} = 0$. Then we can write Newton's second law as

$$\begin{aligned} F_{net_x} &= ma_x = -kx & (2.5) \\ m \frac{d^2x}{dt^2} &= -kx \end{aligned}$$

We have a new kind of equation. If you are taking this freshman class as a... well... freshman, you may not have seen this kind of equation before. It is called a differential equation. The solution of this equation is a function or functions that will describe the motion of our mass-spring system as a function of time. It says that the way the object moves is an equation where the second derivative is almost the same as the original function. The only difference is some constants that are multiplied.

It is this function that we want, so let's see how we can find it.

Start by getting all the constants on the same side of the equation.

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x$$

It would be tidier if we defined a quantity ω as

$$\omega^2 = \frac{k}{m} \quad (2.6)$$

why define ω^2 ? Think of our previous example. We found that $\omega = \sqrt{k/m}$. This is just the square of what we found in our example. Then we can write our differential equation as

$$\frac{d^2x}{dt^2} = -\omega^2 x \quad (2.7)$$

We need a function who's second derivative is the negative of itself with just a constant out front. From Math 112 we know a few of these

$$\begin{aligned} x(t) &= A \cos(\omega t + \phi_o) \\ x(t) &= A \sin(\omega t + \phi_o) \end{aligned}$$

where A , ω , and ϕ_o are constants that we must find. Let's choose the cosine function and explicitly take it's derivatives to see if this function does solve our differential equation

$$\begin{aligned} x(t) &= A \cos(\omega t + \phi_o) & (2.8) \\ \frac{dx(t)}{dt} &= -\omega A \sin(\omega t + \phi_o) \\ \frac{d^2x(t)}{dt^2} &= -\omega^2 A \cos(\omega t + \phi_o) \end{aligned}$$

Let's substitute these expressions into our differential equation for the motion

$$\begin{aligned} \frac{d^2x}{dt^2} &= -\omega^2 x \\ -\omega^2 A \cos(\omega t + \phi_o) &= -\omega^2 A \cos(\omega t + \phi_o) \end{aligned}$$

As long as the constant ω^2 is our $\omega^2 = k/m$ we have a solution. We could have found that

$$\omega^2 = \frac{k}{m}$$

by solving this differential equation, but it might not have been as meaningful that way. We can identify ω as the angular frequency.

$$\omega = 2\pi f$$

Thus

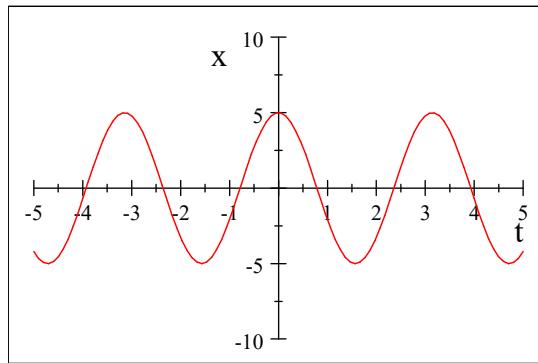
$$\omega = \sqrt{\frac{k}{m}} = 2\pi f \quad (2.9)$$

This says that if the spring is stiffer, we get a higher frequency, or if the mass is larger, we get a lower frequency.

We still don't have a complete solution, because we don't know A and ϕ_o . We recognize ϕ_o as the initial phase angle. We will have to find this by knowing the initial conditions of the motion. A is the amplitude. That must be the maximum displacement x_{\max} . Let's look at a specific case

$x_{\max} = 5$	
$\phi_o = 0$	
$\omega = 2$	

(2.10)



We can easily see that the amplitude A corresponds to the maximum displacement x_{\max} . (how would you prove this?). We know from trigonometry that a cosine function has a period T .

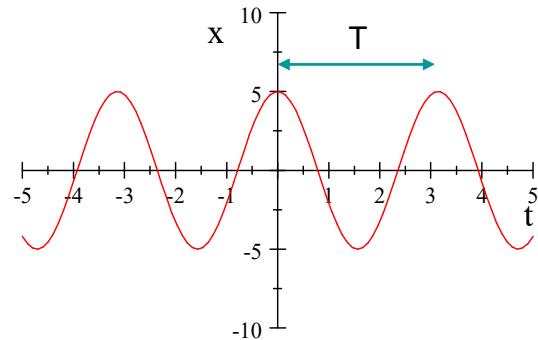


Figure 2.1:

The period is related to the frequency

$$T = \frac{1}{f} = \frac{2\pi}{\omega} \quad (2.11)$$

We can write the period and frequency in terms of our mass and spring constant

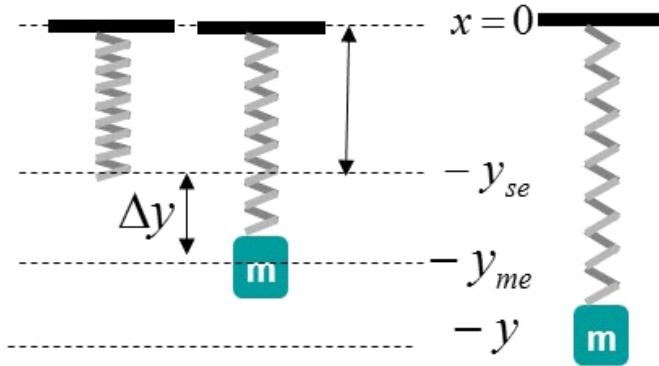
$$T = 2\pi\sqrt{\frac{m}{k}} \quad (2.12)$$

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \quad (2.13)$$

From our graph we were able to complete our problem. But more importantly we have two new equations for T and f that will help us solve more problems.

2.3.1 Hanging Springs

Let's do a problem. We now know we have forces involved with our SHOs. So let's do a force problem. In class so far, I have always hung our masses and springs, but we have used horizontal systems for calculations. Let's find the equation of motion for a mass hanging from a spring.



From observation, we would guess that we can just choose the new equilibrium point to be $y = 0$ and use the same equation

$$y(t) = y_{\max} \cos(\omega t - \phi_0)$$

let's see if that is true. Start with a free body diagram for the mass (the hanging mass is our mover, the spring is part of the environment). There are two forces acting on the mass. The force due to gravity, and the force due to the spring.

$$\Sigma F_y = ma_y = S_{ms} - W_{mE}$$

We know the form of these forces

$$\begin{aligned} S_{ms} &= -k\Delta y \\ W_{mE} &= -mg \end{aligned}$$

but we need to carefully choose our origin. Let's try the top of the spring where it attaches to the stand. If the mass just hangs there we would expect the spring to stretch to an equilibrium length y_e and any other motion would either shorten or lengthen the spring. We can write our force equation as

$$\begin{aligned} ma_y &= S_{ms} - W_{mE} \\ &= -k(y - y_e) - mg \end{aligned}$$

where y can be positive or negative. If $y = 0$ and the mass is just sitting there, not oscillating, there is no acceleration. Then the stretched length is just y_e . In

this case

$$\begin{aligned} m(0) &= k(0 - y_e) - mg \\ ky_{em} &= mg \end{aligned}$$

But now let's let our mass move again. We can substitute our stationary mass answer this into the previous equation for a moving mass

$$\begin{aligned} ma &= -k(y - y_e) - mg \\ &= kye - ky - mg \\ &= mg - ky - mg \\ &= -ky \end{aligned}$$

This gives a net force for the system of

$$F_{net} = -ky$$

It is as though the system were horizontal with no gravitational force and only a spring force. We can see that we are justified in claiming that we could simply choose the origin at the distance y_{em} from the top of the spring, and we can use the equation

$$y(t) = y_{\max} \cos(\omega t - \phi_0)$$

as our equation of motion.

Chapter 3

3 More Oscillators, Forces and Friction 1.15.4 1.15.5 1.15.6

You have probably wondered if anyone actually uses mass-spring systems. And we do. They were more common in the past where springs were used to store energy (U_s) to run clocks and toys and machines. But simple harmonic motion can describe other systems as well. You have probably seen an old fashioned clock with a pendulum. A pendulum almost experiences simple harmonic motion. Let's see how this works.

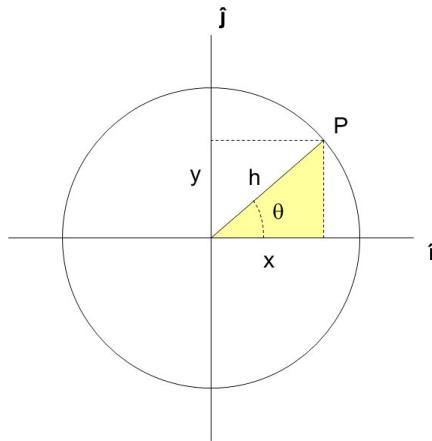
Fundamental Concepts

- Relationship between Simple Harmonic Motion and circular motion.
- Pendula
- Damping
- Driving
- Resonance

3.1 Comparing Simple Harmonic Motion with Uniform Circular Motion

You might have objected to our use of ω as angular frequency. Didn't ω mean angular speed in PH121?

That circular motion and SHM are related should not be a surprise once we found the solutions to the equations of motion were trig functions. Recall that the trig functions are defined on a unit circle

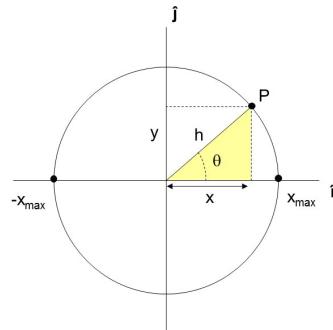


$$\tan \theta = \frac{x}{y} \quad (3.1)$$

$$\cos \theta = \frac{x}{h} \quad (3.2)$$

$$\sin \theta = \frac{y}{h} \quad (3.3)$$

Let's relate this to our SHM equations of motion.



Look at the projection x of the point P on the x axis. This is just the x -component of the position! Lets follow this projection as P travels around the circle. We find the projection ranges from $-x_{\max}$ to x_{\max} . If we watch closely we find the projection's velocity is zero at the extreme points and is a maximum in the middle. This projection is given as the cosine of the vector from the origin to P . It is just taking the x -component! This projection, indeed fits our SHO solution.

3.1. COMPARING SIMPLE HARMONIC MOTION WITH UNIFORM CIRCULAR MOTION 35

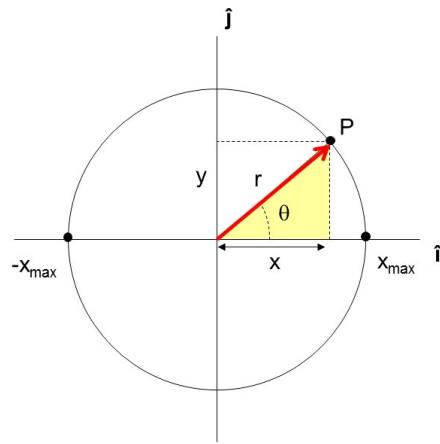
Now let's define a projection of P onto the y axis. Again we have SHM, but this time the projection is a sine function. Because

$$\cos\left(\theta - \frac{\pi}{2}\right) = \sin(\theta) \quad (3.4)$$

we can see that this is just a SHO that is $90^\circ = \frac{\pi}{2}$ rad out of phase. It is probably worth recalling that a projection of one vector on another can be expressed by a dot product. We could express our length x as

$$x = \vec{r} \cdot \hat{i} = r \cos \theta$$

where r is the radius of the circle.



We can see that when $\theta = 0$ we have $x = r$ and this will be the largest x value, so $r = x_{\max}$.

So by projecting circular motion onto the x -axis

$$x(t) = x_{\max} \cos(\theta)$$

But θ changes in time. We can recall from our PH121 or Dynamics experience that the angular speed

$$\omega = \frac{\Delta\theta}{\Delta t}$$

or, if we agree to start from $\theta_i = 0$ and $t_i = 0$,

$$\omega = \frac{\theta}{t}$$

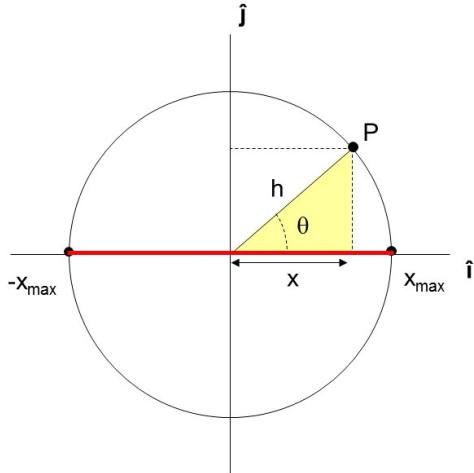
so

$$\theta = \omega t$$

then we have

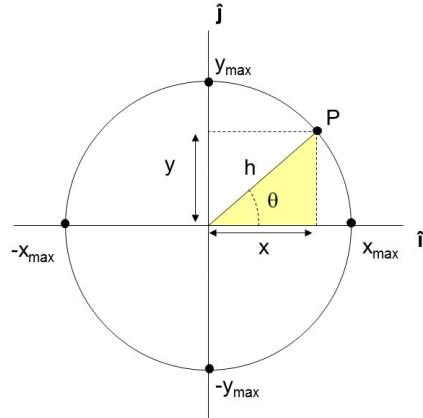
$$x(t) = x_{\max} \cos(\omega t)$$

Which is just our equation for SHM. Now we can see why we used “ ω ” for both angular speed and angular frequency. Really they are very related. Both tell us something about how fast a cyclic event happens. For motion around a circle, one is how fast the point P goes around the circle, and the other is how often the projection goes back and forth. It makes sense that these have to be the same rate.



The projection of circular motion onto the x -axis gives simple harmonic motion.

Let's go back to our projection on the y -axis.



We found that we can describe this projection as

$$y(t) = y_{\max} \sin(\omega t)$$

We will choose the cosine function, but from our trig experience it should be clear that these projections are equivalent, just 90° out of phase

$$\cos\left(\theta \pm \frac{\pi}{2}\right) = \mp \sin \theta$$

So

$$x(t) = x_{\max} \cos(\omega t + \phi_o)$$

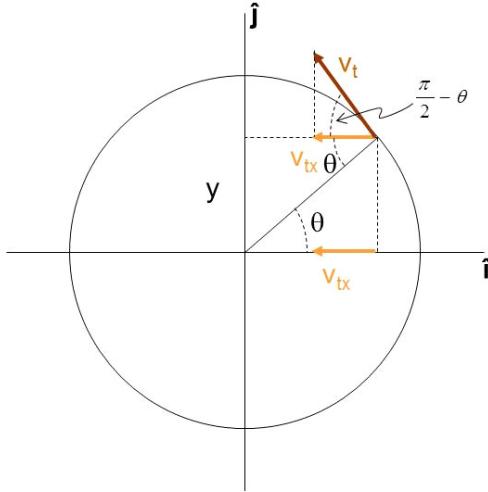
could be a sine function if $\phi_o = \pm\pi/2$. In this way we have incorporated both possibilities into one function.

Note that this is just the function we guessed from our observation!

We see that uniform circular motion can be thought of as the combination of two SHOs, with a phase difference of $\pi/2$ rad.

The angular velocity is given by

$$\omega = \frac{v}{r} \quad (3.5)$$



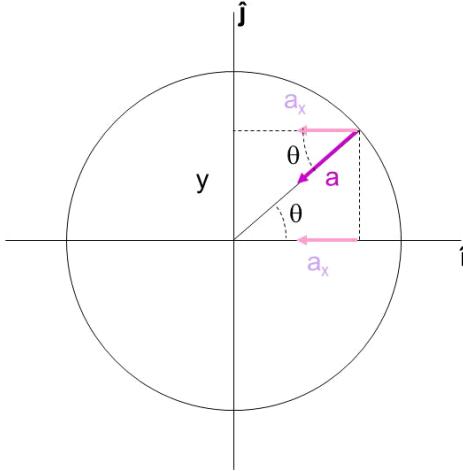
A particle traveling on the x -axis in SHM will travel from x_{\max} to $-x_{\max}$ and from $-x_{\max}$ to x_{\max} (one complete period, T) while the particle traveling with P makes one complete revolution. Thus, the angular frequency ω of the SHO and the angular velocity of the particle at P are the same. The magnitude of the tangential velocity is then

$$v_t = \omega r = \omega x_{\max} \quad (3.6)$$

and the projection of this velocity onto the x -axis is

$$v_{tx} = -\omega x_{\max} \sin(\omega t + \phi_o) \quad (3.7)$$

Which is just what we expected from our earlier observation!



The centripetal acceleration of a particle at P is given by

$$a_c = \frac{v_t^2}{r} = \frac{v_t^2}{x_{\max}} = \frac{\omega^2 x_{\max}^2}{x_{\max}} = \omega^2 x_{\max} \quad (3.8)$$

The direction of the acceleration is inward toward the origin. Of course, we just want the x -component of this, so again we make a projection. If we project this onto the x -axis we have

$$a_x = -\omega^2 x_{\max} \cos(\omega t + \phi) \quad (3.9)$$

Again this is just what we expected from our observation.

So now we have shown that our set of equations

$$\begin{aligned} x(t) &= x_{\max} \cos(\omega t) \\ v(t) &= -\omega x_{\max} \sin(\omega t) \\ a(t) &= -\omega^2 x_{\max} \cos(\omega t) \end{aligned} \quad (3.10)$$

is correct for our harmonic oscillator.

3.2 Our first problem type: Simple Harmonic Motion

You have probably thought by now that we have a new problem type. We can call it the *simple harmonic motion* problem type or SHM. The equations we have so far for this problem type are

$$\begin{aligned} x(t) &= x_{\max} \cos(\omega t + \phi_o) \\ v(t) &= -\omega x_{\max} \sin(\omega t + \phi_o) \\ a(t) &= -\omega^2 x_{\max} \cos(\omega t + \phi_o) \end{aligned} \quad (3.11)$$

$$\omega = 2\pi f$$

$$v_m = \omega x_m$$

$$a_m = \omega^2 x_m$$

$$T = \frac{1}{f}$$

$$U_s = \frac{1}{2} k x_{\max}^2 \cos^2(\omega t + \phi)$$

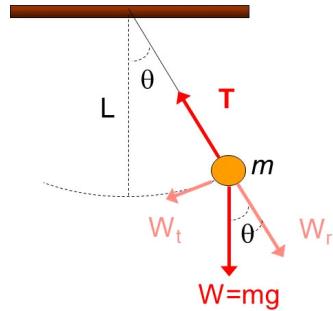
$$K = \frac{1}{2} m \omega^2 x_{\max}^2 \sin^2(\omega t + \phi)$$

$$T = 2\pi \sqrt{\frac{m}{k}} \quad (3.12)$$

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (3.13)$$

We have used energy to describe simple harmonic motion, and we have found our equation of motion using a differential equation for mass-spring systems. And we have simple harmonic motion in the y -direction including the weight force due to gravity for mass-spring systems. In our next lecture, we will study simple harmonic motion for different systems, pendula, and other things. It turns out that SHM is a good model for many different systems.

3.3 The Simple Pendulum



A simple pendulum is a mass on a string. The mass is called a “bob.” Usually we study the motion of the pendulum bob, so let’s consider the pendulum the mover object. A simple pendulum bob exhibits periodic motion, but not exactly simple harmonic motion. The forces on the bob are \vec{W} , and \vec{T} the tension on

the string. The tangential component of W is always directed toward $\theta = 0$. This is a restoring force!

Let's call the path the bob takes "s." Then from Jr. High geometry we recall¹

$$s = L\theta \quad (3.14)$$

We will use a the cylindrical or rtz coordinate system. The radial axis is directed along the string. The tangential direction is along the circular path the bob takes and is always tangent to the path. In this coordinate system, we can solve for the part of the force directed along the path. This is the restoring part of the net force. Remember from Newton's second law the tangential and radial components of the force are

$$\begin{aligned} F_t &= ma_t \\ F_r &= ma_r \end{aligned}$$

and

$$a_t = \frac{d^2 s}{dt^2}$$

Let's write out Newton's second with the sum of the forces part.

$$\begin{aligned} ma_t &= -W \cos(90 - \theta) \\ -ma_r &= -T + W \sin(90 - \theta) \end{aligned}$$

then, using a trig identity (but only a small one)

$$\begin{aligned} a_t &= -\frac{W}{m} \sin(\theta) \\ &= -g \sin \theta \end{aligned}$$

We have two expressions for a_t . We can set them equal

$$\frac{d^2 s}{dt^2} = -g \sin \theta \quad (3.15)$$

Remember that $s = L\theta$ so we could write the left hand side as

$$\frac{d^2 s}{dt^2} = \frac{d^2}{dt^2}(L\theta) = L \frac{d^2}{dt^2}(\theta)$$

then equation (3.15) becomes

$$L \frac{d^2}{dt^2}(\theta) = -g \sin \theta$$

or

$$\frac{d^2 \theta}{dt^2} = -\frac{g}{L} \sin \theta$$

¹Really I did not recall this, I have to look it up every time, but s is called the *arclength*.

This is a differential equation much like our differential equation for a harmonic oscillator,

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

except it has a sine function in it. But, if we take θ as a very small angle, then

$$\sin(\theta) \approx \theta \quad (3.16)$$

This approximation has a name, it is called the “small angle approximation.”

In this approximation

$$\frac{d^2\theta}{dt^2} = -\frac{g}{L}\theta$$

and we have a differential equation we recognize! If we compare to

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

we see that it is a match if

$$\omega^2 = \frac{g}{L} \quad (3.17)$$

we have all the same solutions for θ that we found last time for x . Since ω changed, the frequency and period will now be in terms of g and L .

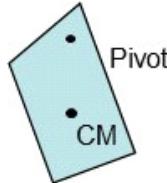
$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{L}{g}} \quad (3.18)$$

For a pendulum that oscillates only over small angles, the period and frequency depend only on L and g !

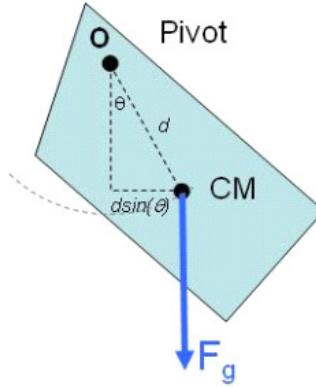
The analysis we just did for the pendulum we can do for other simple harmonic oscillators (or near simple harmonic oscillators). Let's try a few.

3.3.1 Physical Pendulum

Usually when we build a pendulum, we assume the string is so small compared to the bob, that we can ignore its mass. What if this is not true? Suppose we build a pendulum by making a large solid object swing from one point. Can we describe its motion?



Let's pull it to the right



then we consider that because of F_g we will have a torque about an axis through O . Last pendulum we used Newton's Second Law. This time let's use Newton's second law for rotation.

$$\tau = \mathbf{r} \times \mathbf{F}$$

In our case this is

$$\begin{aligned}\tau &= \mathbf{d} \times \mathbf{F}_g \\ &= -F_g d \sin \theta \\ &= -mgd \sin \theta\end{aligned}$$

Remember that an extended body has a moment of inertia, \mathbb{I} . Remember also that angular acceleration is given by

$$\alpha = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}$$

From the rotational form of Newton's Second Law for rotation.

$$\Sigma \tau = \mathbb{I} \alpha$$

we can write

$$-mgd \sin \theta = \mathbb{I} \frac{d^2\theta}{dt^2}$$

Getting all the constants together gives

$$\frac{d^2\theta}{dt^2} = -\frac{mgd}{\mathbb{I}} \sin(\theta)$$

and again if we let θ be small so that $\sin(\theta) \approx \theta$

$$\frac{d^2\theta}{dt^2} = -\frac{mgd}{\mathbb{I}} \theta$$

which we can compare to

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

and we can see that we have the same differential equation if

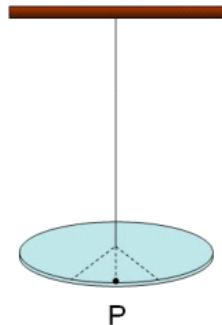
$$\omega^2 = \frac{mgd}{I} \quad (3.19)$$

In this case

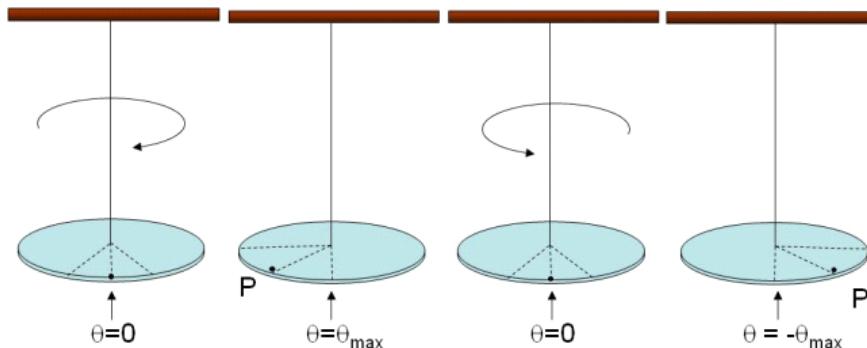
$$T = 2\pi \sqrt{\frac{I}{mgd}} \quad (3.20)$$

3.3.2 Torsional Pendulum

Physics majors and those taking PH220 in the future, you will see a torsional pendulum. Mechanical engineering majors might use a torsional pendulum. A torsional pendulum is made by suspending a rigid object from a wire.



In this case, a rigid disk. The object can rotate. But wires don't like to twist. The spring-like molecular bonds resist the twisting.



The twisted wire exerts a restoring torque on the body that is proportional to the angular position (sound familiar).

$$\tau = -\kappa\theta$$

This looks like a greek version of $F = -kx$! Again, let's use Newton's Second Law for rotation.

$$\begin{aligned}\Sigma\tau &= \mathbb{I}\alpha \\ &= \mathbb{I}\frac{d^2\theta}{dt^2}\end{aligned}$$

so

$$\begin{aligned}\mathbb{I}\frac{d^2\theta}{dt^2} &= -\kappa\theta \\ \frac{d^2\theta}{dt^2} &= -\frac{\kappa}{\mathbb{I}}\theta\end{aligned}$$

Once again we have our favorite differential equation so long as

$$\omega^2 = \frac{\kappa}{\mathbb{I}} \quad (3.21)$$

which makes the period of the oscillation

$$T = 2\pi\sqrt{\frac{\mathbb{I}}{\kappa}} \quad (3.22)$$

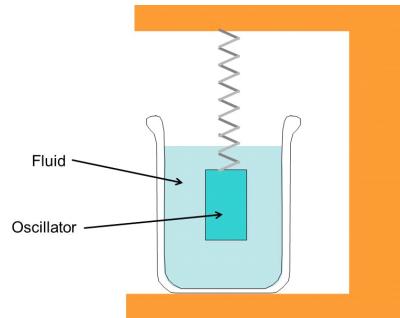
So far we have used only idea springs or wires or frictionless pendula with no air drag. This made the equations easier. But what if there is friction of some sort? Let's deal with this next.

3.4 Damped Oscillations

You remember friction from PH121. So far we have only allowed frictionless oscillators to make the math easy. But what if we do have friction? To investigate this, suppose we add in another force

$$\mathbf{D} = -b\mathbf{v} \quad (3.23)$$

This force is proportional to the velocity. This a dissipative (friction-like) force typical of what we find when we moves objects through viscous fluids. This is a drag force, but a more extreme drag force than we used in PH121 where we only had air to make the drag force.



We call b the damping coefficient and it depends on how much friction the fluid can supply. Now, from Newton's second law,

$$\Sigma F = -kx - bv_x = ma$$

We can write the acceleration and velocity as derivatives of the position just like we have done before

$$-kx - b\frac{dx}{dt} = m\frac{d^2x}{dt^2}$$

This is another differential equation! But it is harder to guess its solution, and finding that solution is a subject for a differential equations class like M316 or PH332, so we won't learn how to find the solution here, but we can use the results from our trusted colleagues in the math department². Here is the solution:

$$x(t) = x_{\max} e^{-\frac{b}{2m}t} \cos(\omega t + \phi_o) \quad (3.24)$$

which looks simple enough, but now we have the added complication that ω is more complex

$$\omega = \sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2} \quad (3.25)$$

so we get quite a mess if we write equation (3.24) with this new ω .

$$x(t) = x_{\max} e^{-\frac{b}{2m}t} \cos \left(\left(\sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2} \right) t + \phi_o \right) \quad (3.26)$$

To see what this solution means, we should study three cases:

1. the damping force is small: ($bv_{\max} < kx_{\max}$) The system oscillates, but the amplitude is smaller as time goes on. We call this "underdamped."
2. the damping force is large: ($bv_{\max} > kx_{\max}$) The system does not oscillate. we call this "overdamped." We can also say that $\frac{b}{2m} > \omega_o$ (after we define ω_o below)
3. The system is "critically damped" (see below).

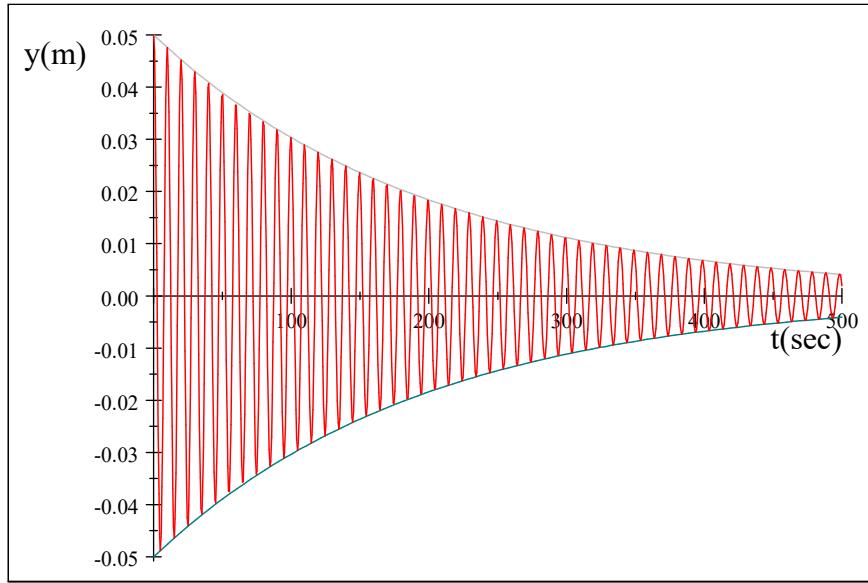
Let's look at an example. Suppose we have an oscillator with the following characteristics:

1.

$x_{\max} = 5 \text{ cm}$
$b = 0.005 \frac{\text{kg}}{\text{s}}$
$k = 0.2 \frac{\text{N}}{\text{m}}$
$m = .5 \text{ kg}$
$\phi_o = 0$

²That is, until you finish M316, then you will know how to do this kind of problem yourself!

Graphing the equation of motion $x(t)$, we get a graph that looks like this



The gray lines are given by

$$\pm x_{\max} e^{-\frac{b}{2m}t}$$

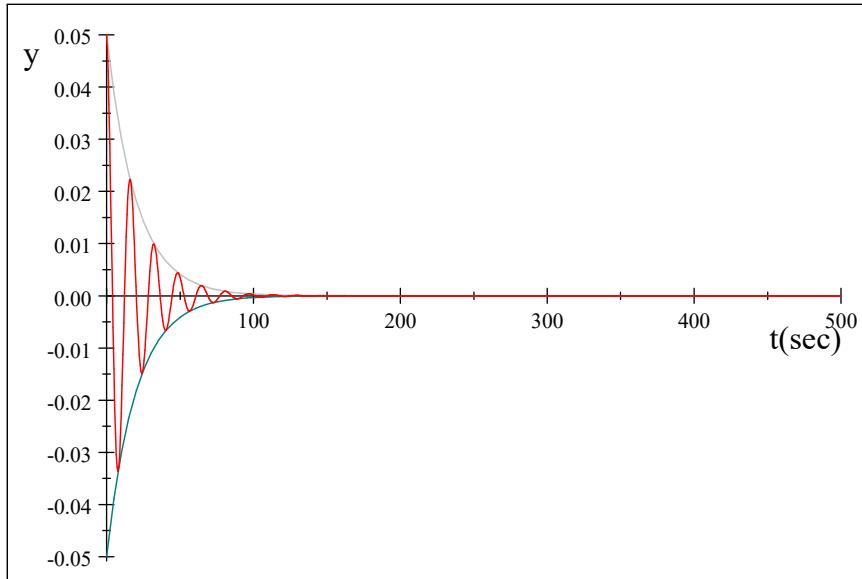
Notice that this quantity is as a collection of terms multiplies the cosine part. It is marked in curly braces below

$$x(t) = \left\{ x_{\max} e^{-\frac{b}{2m}t} \right\} \cos \left(\left(\sqrt{\frac{k}{m}} - \left(\frac{b}{2m} \right)^2 \right) t + \phi_o \right)$$

We know that the part of the equation that multiplies the cosine part is the amplitude. But now the amplitude is more than just x_{\max} . And notice that the amplitude $\left\{ x_{\max} e^{-\frac{b}{2m}t} \right\}$ changes with time. The gray lines in the figure show how the amplitude changes. We call this the *envelope* of the curve. The oscillation fits within the gray lines (like an old fashioned letter fits inside a paper envelope).

Now let's change b to a larger value

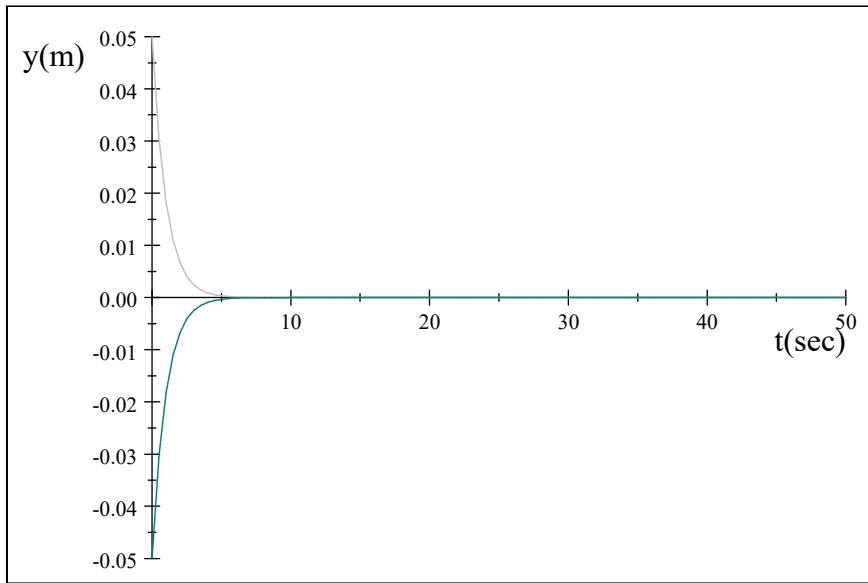
$x_{\max} = 5 \text{ cm}$
$b = 0.05 \frac{\text{kg}}{\text{s}}$
$k = 0.2 \frac{\text{N}}{\text{m}}$
$m = .5 \text{ kg}$
$\phi_o = 0$



we see we have less oscillation. The envelope has become more restrictive, making the oscillation die out more quickly. This is a little bit like going over a bump in your car. The car may go up and down a few times, but not many. Now let's increase b even more.

$x_{\max} = 5 \text{ cm}$
$b = 0.5 \frac{\text{kg}}{\text{s}}$
$k = 0.2 \frac{\text{N}}{\text{m}}$
$m = .5 \text{ kg}$
$\phi_o = 0$

(3.27)



What happened?

When the damping force gets bigger, the oscillation eventually stops. Only the exponential decay is observed. This happens when

$$\frac{b}{2m} = \sqrt{\frac{k}{m}}$$

When that is true,

$$\omega = \sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2} = 0$$

We call this situation *critically damped*. We are just on the edge of oscillation. We define

$$\omega_o = \sqrt{\frac{k}{m}}$$

as the *natural frequency* of the system. Then the value of b that gives us critically damped behavior is

$$b_c = 2m\omega_o$$

When $\frac{b}{2\pi} \geq \omega_o$ the solution in equation (3.24) is not valid! If you are a physicist or a mechanical engineer you will find out more about this situation in your advanced mechanics classes.

Chapter 4

4 Forces Oscillations and Waves 1.15.6

We studied oscillation with mass-spring systems and pendula. In today's lecture we will add in forces to make the oscillators start to move. Oscillation is a motion of a mass, but with another object (a spring or string) participating in the motion. The air in our classroom is an example of this type of motion in a way, but instead of one object on a spring, there are millions of objects, the air molecules. The molecules move randomly, but with a specific distribution of speeds. But what would happen if all the molecules moved together in the same direction (more or less) and at the same speed (more or less)? This is what we call wind! Of course we can have bulk motion of millions of objects, like wind. We are going to study an even more specific motion of millions of object that is not random like thermal motion. This specific motion of the objects we will call a *wave*.

Fundamental Concepts

1. Forced Oscillations
2. A wave requires a disturbance, and a medium that can transfer energy
3. Waves are categorized as longitudinal or transverse (or a combination of the two).

4.1 Forced Oscillations

We found in the last section that if we added a force like

$$\mathbf{F}_d = -b\mathbf{v}$$

our oscillation died out. An example would be a small child on a swing. You give them a push, but eventually they stop swinging.

Suppose we want to keep the child going? You know what to do, you push! But you have to push at just the right time.

Suppose we make a machine that can push the child on the swing. We could apply a periodic force like

$$F(t) = F_o \sin(\omega_f t)$$

where ω_f is the angular frequency of this new driving force and where F_o is a constant.

$$\Sigma F = F_o \sin(\omega_f t) - kx - bv_x = ma$$

When this system starts out, the solution is very messy. It is so messy that we will not give it in this class! (So maybe this isn't really a good way to drive a child on a swing). But after a while, a steady-state is reached. In this state, the energy added by our driving force $F_o \sin(\omega_f t)$ is equal to the energy lost by the drag force, and we have

$$x(t) = A \cos(\omega_f t + \phi)$$

our old friend! BUT NOW the amplitude is given by

$$A = \frac{\frac{F_o}{m}}{\sqrt{\left(\omega_f^2 - \omega_o^2\right)^2 + \left(\frac{b\omega_f}{m}\right)^2}}$$

where

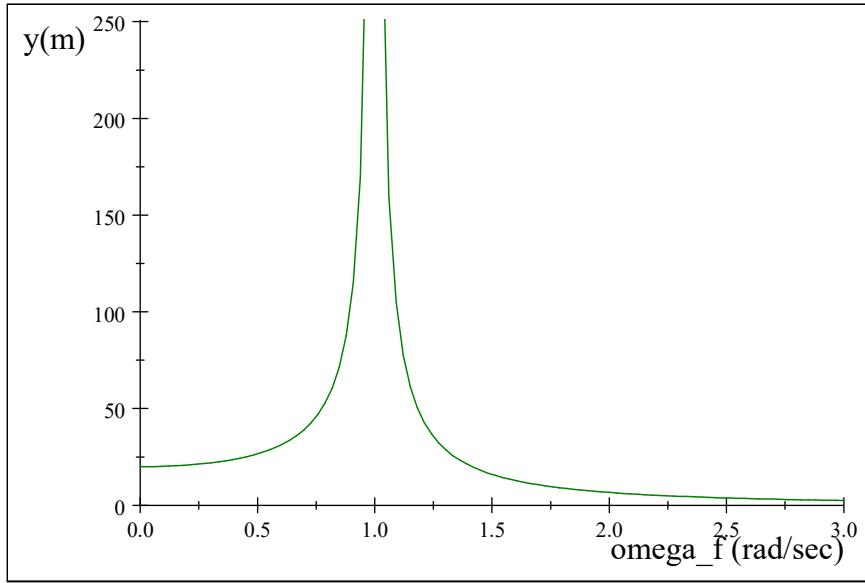
$$\omega_o = \sqrt{\frac{k}{m}} \quad (4.1)$$

is the natural frequency. This is the frequency we had before when we weren't pushing the oscillator. And recall that ω_f is the frequency of our force. So now our solution looks more like our original SHM solution except for the wild formula for A . In fact, it operates very like our SHM solution. But what does the new version of A mean?

Lets look at A for some values of ω_f . I will pick some nice numbers for the other values.

$F_o = 2 \text{ N}$
$b = 0.5 \frac{\text{kg}}{\text{s}}$
$k = 0.5 \frac{\text{N}}{\text{m}}$
$m = 0.5 \text{ kg}$
$\phi_o = 0$

The graph looks like this:



Now let's calculate ω_o

$$\begin{aligned}\omega_o &= \sqrt{\frac{0.5 \text{ N}}{0.5 \text{ kg}}} \\ &= \frac{1.0}{\text{s}}\end{aligned}$$

Notice that right at $\omega_f = \omega_o$ our solution gets very big. This is called *resonance*. To see why this happens, think of the velocity for a simple harmonic oscillator

$$\frac{dx(t)}{dt} = -\omega A \sin(\omega t + \phi_o)$$

Our driving force is still

$$F(t) = F_o \sin(\omega t)$$

And remember that work is given by

$$w = \int \vec{F} \cdot d\vec{x}$$

or just

$$w = \vec{F} \cdot \Delta \vec{x}$$

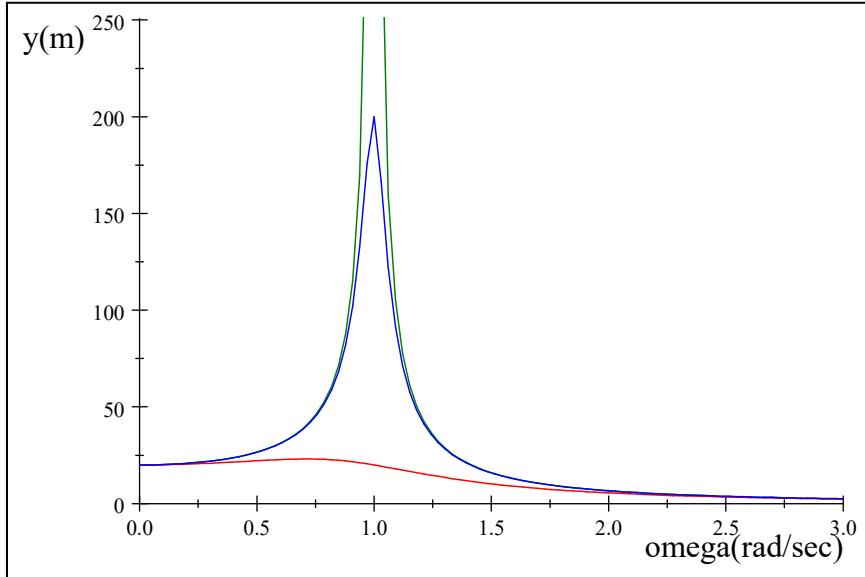
if our force is constant. The rate at which work is done (power) is

$$\mathcal{P} = \frac{\vec{F} \cdot \Delta \vec{x}}{\Delta t} = \vec{F} \cdot \vec{v} \quad (4.2)$$

$$= -F_o \omega A \sin(\omega t) \sin(\omega t + \phi_o) \quad (4.3)$$

if F and v are in phase ($\phi_o = 0$), the power will be at a maximum! Think of pushing the child on the swing. You push in phase with the oscillation of the child. When you do this the amplitude of the swing gets bigger.

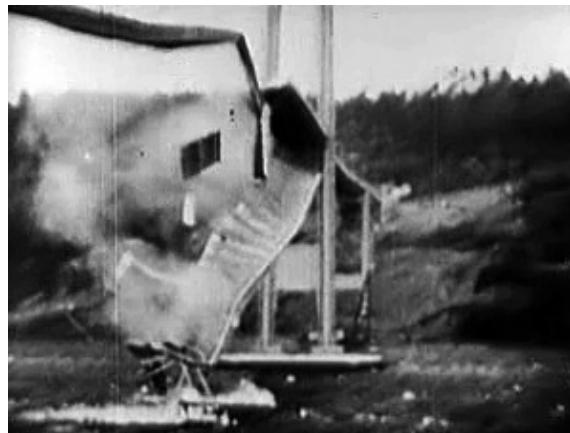
We can plot A for several values of b , our damping (friction) coefficient.



Green: $b=0.005\text{kg/s}$; Blue: $b=0.05\text{kg/s}$; Red $b=0.01\text{ kg/s}$

As $b \rightarrow 0$ we see that our resonance peak gets larger. In real systems b can never be zero, but sometimes it can get small. As $b \rightarrow$ large, the resonance dies down and our A gets small.

Resonance can be great, it can make a musical instrument sound louder (more on this later). But it can also be bad. Here is a picture of the Tacoma Narrows Bridge.



Fall of the Tacoma Narrows Bridge (Image in the Public Domain)

The wind gusts formed a periodic driving force that allowed a driving harmonic oscillation to form. Since the bridge was resonant with the gust frequency, the amplitude grew until the bridge materials broke. Resonance can be a bad thing for structures.

4.2 What is a Wave?

Waves are organized motions in a group of objects. We will give a name to the group of objects, we will call a *medium*.

4.2.1 Criteria for being a wave

Waves involve energy transfer, but in the case of waves the energy is transferred through space without transfer of matter. Winds transfers both energy and matter. So waves are different than wind. In a wave the objects don't move far from where they start. Think of an oscillation. The mass does move, but never gets too far away from its equilibrium position. Waves are very like this. This is a very specific kind of motion. To be a wave, the motion must have the following characteristics"

1. some source of disturbance
2. a medium (group of objects) that can be disturbed
3. some physical mechanism by which the objects of the medium can influence each other

A wave made by a rock thrown into a pond will go out in all directions away from the place where the rock started the wave. This is a normal way that waves are formed. A disturbance starts the wave (the rock disturbs the water) and the energy from the disturbance moves away from the disturbance as a wave. But if we have a wave in a rope or string, the wave can't go in all directions because the string does not go in all directions.

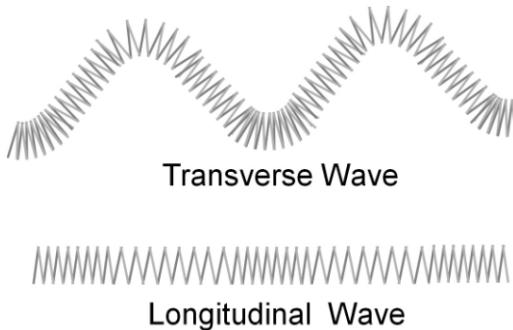
Let's take on this one-dimensional case of a wave on a rope or string first. In the limit that the string mass is negligible we represent a one-dimensional wave mathematically as a function of two variables, position and time, $y(x, t)$.

4.2.2 Longitudinal vs. transverse

We divide the various kinds of waves that occur into two basic types:

Transverse wave: a traveling wave or pulse that causes the elements of the disturbed medium to move perpendicular to the direction of propagation

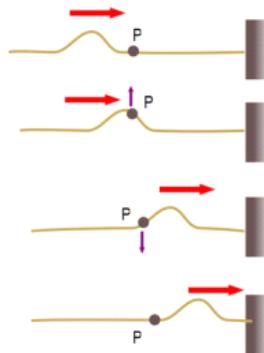
Longitudinal wave: a traveling wave or pulse that causes the elements of the medium to move parallel to the direction of propagation



4.2.3 Examples of waves:

Let's look at some specific cases of wave motion.

A pulse on a rope:

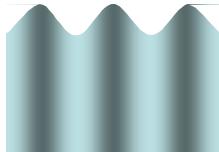


In the picture above, you see wave that has just one peak traveling to the right. We call such a wave a *pulse*. Notice how the piece of the rope marked *P* moves up and down, but the wave is moving to the right. This pulse is a transverse wave because the parts of the medium (observe point *P*) move perpendicular to the direction the wave is moving.

An ocean wave:

Of course, some waves are a combination of these two basic types. You may have noticed that in Physics we tend to define basic types of things, and then use these basic types to define more complex objects. Water waves, for example,

are transverse at the surface of the water, but are longitudinal throughout the water.



Earthquake waves:

Earthquakes produce both transverse and longitudinal waves. The two types of waves even travel at different speeds! P waves are longitudinal and travel faster, S waves are transverse and slower.

4.3 Wave speed

We can perform an experiment with a rope or a long spring. Make a wave on the rope or spring. Then pull the rope or spring tighter and make another wave. We see that the wave on the tighter spring travels faster.

It is harder to do, but we can also experiment with two different ropes, one light and one heavy. We would find that the heavier the rope, the slower the wave. We can express this as

$$v = \sqrt{\frac{T_s}{\mu}}$$

where T_s is the tension in the rope, and μ is the linear mass density

$$\mu = \frac{m}{L}$$

where m is the mass of the rope, and L is the length.

The term μ might need an analogy to make it seem helpful. So suppose I have an iron bar that has a mass of 200 kg and is 2 m long. Further suppose I want to know how much mass there would be in a 20 cm section cut off the end of the rod. How would I find out?

This is not very hard, We could say that there are 200 kg spread out over 2 m, so each meter of rod has 100 kg of mass, that is, there is 100 kg/m of mass per unit length. Then to find how much mass there is in a 0.20 m section of the rod I take

$$m = 100 \frac{\text{kg}}{\text{m}} \times 0.20 \text{ m} = 20.0 \text{ kg}$$

The 100 kg/m is μ . It is how much mass there is in a unit length segment of something. In this example, it is a unit length of iron bar, but for waves on string, we want the mass per unit length of string.

If you are buying stock steel bar, you might be able to buy it with a mass per unit length. If the mass per unit length is higher then the bar is more massive. The same is true with string. The larger μ , the more massive equal string segments will be.

We should note that in forming this relationship, we have used our standard introductory physics assumption that the mass of the rope can be neglected. Let's consider what would happen if this were not true. Say we make a wave in a heavy cable that is suspended. The mass at the lower end of the cable pulls down on the upper part of the cable. The tension will actually change along the length of the cable, and so will the wave speed. Such a situation can't be represented by a single wave speed. But for our class, we will assume that any such changes are small enough to be ignored.

4.4 One dimensional waves

To mathematically describe a wave we will define a function of both time and position.

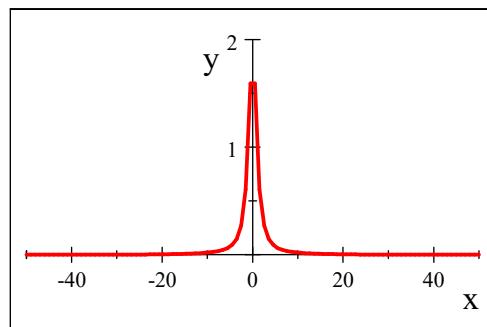
$$y(x, t) \quad (4.4)$$

Note, that this is new in our physics experience. Before we usually were concerned about only one variable at a time. For oscillation we had just $y(t)$ for example. But now we will be concerned about two variables, x , and t .

let's take a specific example¹

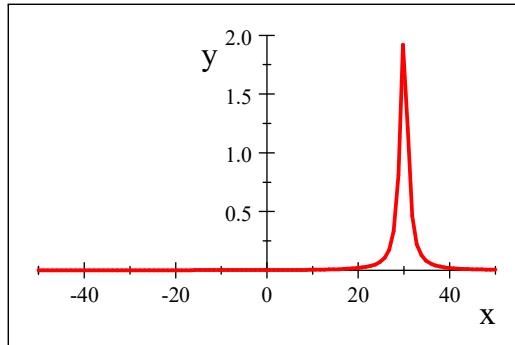
$$y = \frac{2}{(x - 3.0t)^2 + 1} \quad (4.5)$$

Let's plot this for $t = 0$



what will this look like for $t = 10$?

¹This is not an important wave function, just one I picked because it makes a nice graphic example.



The pulse travels along the x -axis as a function of time. Note that there is a value for y for every x position and that these y values change for different times. That is what we mean by saying we have a function of both x and t .

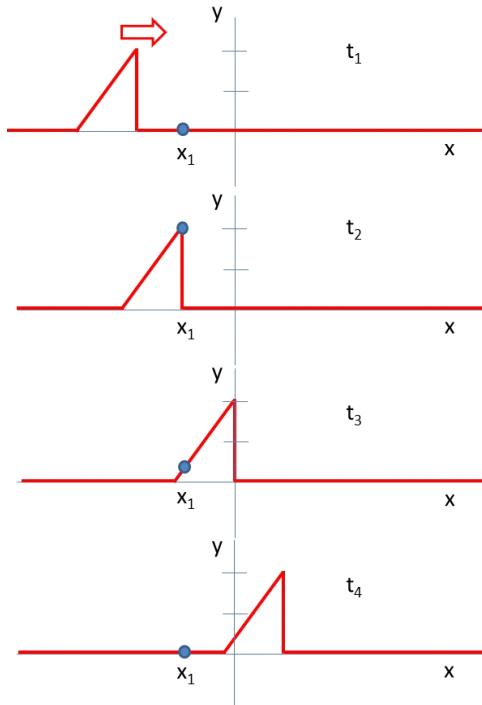
We denote the speed of the pulse as v , then we can define a function

$$y(x, t) = y(x - vt, 0) \quad (4.6)$$

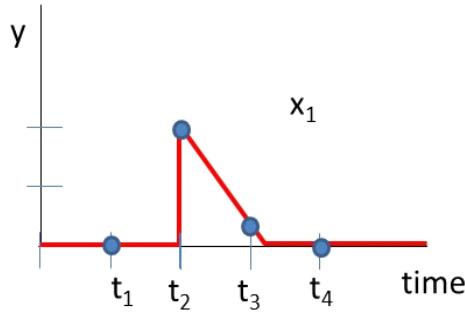
that describes a pulse as it travels. An element of the medium (rope, string, etc.) at position x at some time t , will have the displacement that an element had earlier at $x - vt$ when $t = 0$.

We will give $y(x - vt, 0)$ a special name, the *wave function*. It represents the y position, the transverse position in our example, of any element located at a position x at any time t .

Notice that wave functions depend on two variables, x , and t . It is hard to draw a wave so that this dual dependence is clear. Often we draw two different graphs of the same wave so we can see independently the position and time dependence. So far we have used one of these graphs. A graph of our wave at a specific time, t_o . This gives $y(x, t_o)$. This representation of a wave is very like a photograph of the wave taken with a digital camera. It gives a picture of the entire wave, but only for one time, the time at which the photograph was taken. Of course we could take a series of photographs, but still each would be a picture of the wave at just one time. Here is such a series of graphs at times t_1 through t_4 .



The second representation is to observe the wave at just one point in the medium, but for many times. This is very like taking a video camera and using it to record the displacement of just one part of the medium for many times. You could envision marking just one part of a rope, and then using the video recorder to make a movie of the motion of that single part of the rope. We could then go frame by frame through the video, and plot the displacement of our marked part of the rope as a function of time. Such a graph is sometimes called a history graph of the wave. Here is such a graph for the position x_1 .



To go from the time graphs to the history graph you observe what happens at the location x_1 for each of the times. Then plot those y positions on the y vs.

t graph. Then you must connect the points. This takes some thinking to make sure you connect them right (or a whole lot of points). But with practice, this is not hard and both view points are valuable ways to look at a wave.

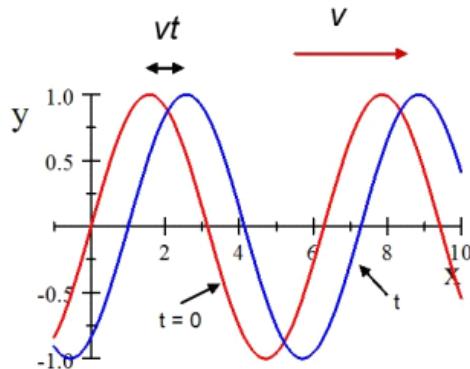
Chapter 5

5 Sinusoidal Waves in one and Two Dimensions 1.16.2 1.16.3

Fundamental Concepts

- It takes two variables to describe a wave, position and time
- Wave graphs have many named parts, amplitude, period, wavelength crest, trough, etc.
- There is a spatial analog to temporal frequency called spacial frequency and it is represented by the factor k in our equations
- Waves from point sources are spherical

5.1 Sinusoidal Waves



A sinusoidal graph should be familiar from our study of oscillation. Simple harmonic oscillators are described by the function

$$y(t) = y_{\max} \cos(\omega t + \phi) \quad \text{SHM} \quad (5.1)$$

but this only gave us a vertical displacement. Now our sinusoidal function must also be a function of position along the wave.

$$y(x, t) = y_{\max} \sin(ax - \omega t + \phi) \quad \text{waves} \quad (5.2)$$

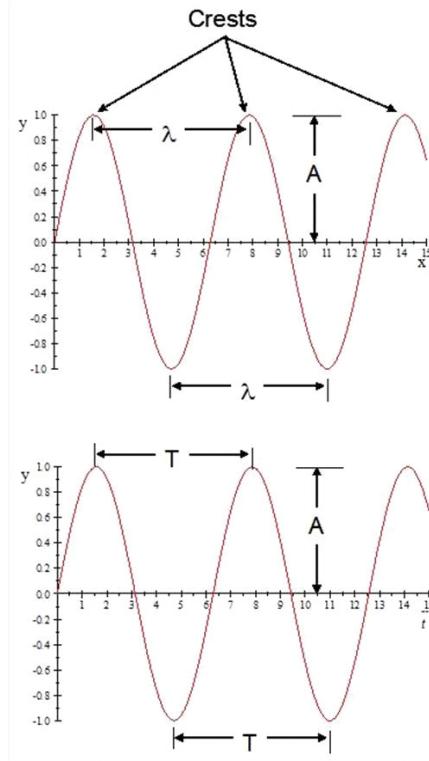
but before we study the nature of this function, let's see what we can learn from the graph of a sinusoidal wave. We will need both our two views, the camera snapshot and the video of a point. Look at figure ???. This is two camera snapshots superimposed. The red curve shows the wave (y position for each value of x) at $t = 0$. At some later time t , the wave pattern has moved to the right as shown by the blue curve. The shift is by an amount $\Delta x = v\Delta t$.

5.2 Parts of a wave

The peak of a wave is called the crest. For a sine wave we have a series of crests. We define the wavelength as the distance between any two identical points (e.g. crests) on adjacent waves in a snapshot view.

Notice that this is very similar to the definition of the period, T , when we graphed SHM on a y, t set of axes. In fact, this similarity is even more apparent if we plot a sinusoidal wave using our two wave pictures. In the next figure, the snapshot comes first. We can see that there will be crests. The distance between the crests is given the name *wavelength*. This is not the entire length of the whole wave. But it is a characteristic length of part of the wave that is easy to identify. The next figure shows all this using our snapshot and history

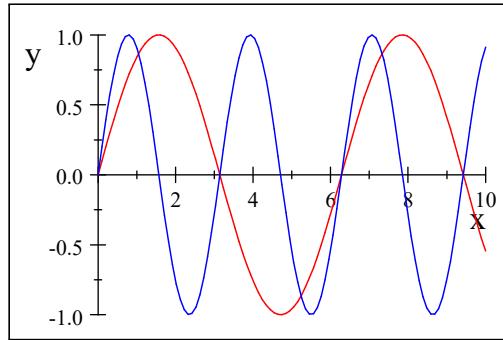
graph for a sine wave.



Note that there are crests in the history graph view as well. That is because one marked part of the medium is being displaced as a function of time. Think of a marked piece of the rope going up and down, or think of floating in the ocean at one point, you travel up and down (and a little bit back and forth) as the waves go by. But now the horizontal axis is time. There will be a characteristic time between crests. That time is called the *period*. This is just like oscillatory motion—because it is oscillatory motion! Like the wavelength is not the length of the whole wave, the period is not the time the whole wave exists. It is just the time it takes the part of the medium we are watching to go through one complete cycle. Notice that this video picture is exactly the same as a plot of the motion of a simple harmonic oscillator! For a sinusoidal wave, each part of the medium experiences simple harmonic motion.

We remember frequency from simple harmonic motion. But now we have a wave, and the wave is moving. We can extend our view of frequency by defining it as follows:

The frequency of a periodic wave is the number of crests (or any other point of the wave) that pass a given point in a unit time interval.



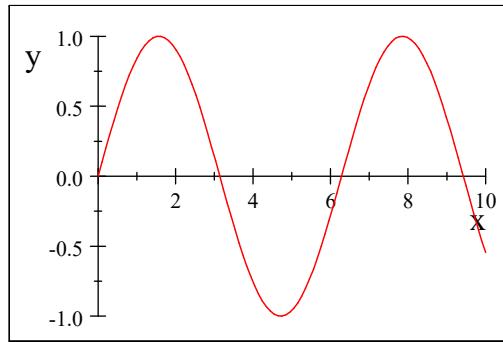
In the previous figure, the blue curve has twice the frequency as the red curve. Notice how it has two crests for every red crest. The maximum displacement of the wave is called the *amplitude* just as it was for simple harmonic oscillators.

5.3 Wavenumber and wave speed

Consider again a sinusoidal wave.

$$y(x, t) = y_{\max} \sin(kx - \omega t + \phi_0) \quad (5.3)$$

We have drawn the wave in the snapshot picture mode



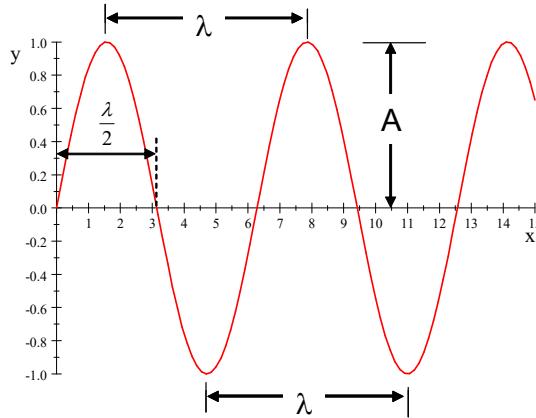
To make this graph, we set $t = 0$ and plotted the resulting function

$$y(x, 0) = y_{\max} \sin(kx + 0) \quad (5.4)$$

y_{\max} is the amplitude. I want to investigate the meaning of the constant k . Lets find k like we did for SHM when we found ω . Consider the point $x = 0$. At this point

$$y(0, 0) = y_{\max} \sin(k(0)) = 0 \quad (5.5)$$

The next time $y = 0$ is when $x = \frac{\lambda}{2}$



then

$$y\left(\frac{\lambda}{2}, 0\right) = y_{\max} \sin\left(k \frac{\lambda}{2}\right) = 0 \quad (5.6)$$

From our trigonometry experience, we know that this is true when

$$k \frac{\lambda}{2} = \pi \quad (5.7)$$

solving for k gives

$$k = \frac{2\pi}{\lambda} \quad (5.8)$$

Then we now have a feeling for what k means. It is 2π over the spacing between the crests. The 2π must have units of radians attached. Then

$$y(x, 0) = A \sin\left(\frac{2\pi}{\lambda} x + 0\right) \quad (5.9)$$

We have a special name for the quantity k . It is called the *wave number*.

$$k \equiv \frac{2\pi}{\lambda} \quad (5.10)$$

Both the name and the symbol are somewhat unfortunate. Neither gives much insight into the meaning of this quantity. But from what we have done in studying oscillation, we can understand this new quantity. For a harmonic oscillator, we know that

$$y(t) = y_{\max} \sin(\omega t)$$

where

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

T is how far, *in time*, the crests are apart, and the inverse of this, $\frac{1}{T}$ is the frequency. The frequency tells us how often we encounter a crest as we march along in time. So $\frac{1}{T}$ must be how many crests we have in a unit amount of time.

Now think of the relationship between the snapshot and the video representation for a sinusoidal wave. We have a new quantity

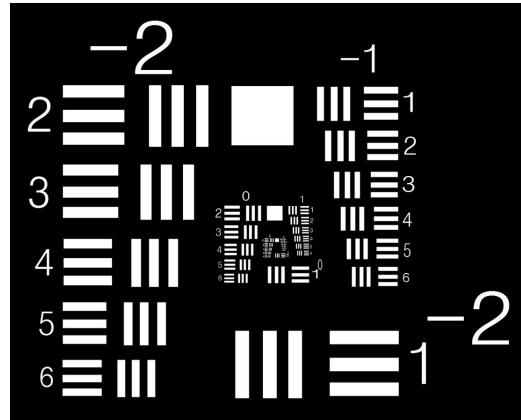
$$k = \frac{2\pi}{\lambda}$$

where λ is how far, *in distance*, the crests are apart. This implies that $\frac{1}{\lambda}$ plays the same role in the snap shot graph that f plays in the video graph. It must tell us how many crests we have, but this time it is how many crests in a given amount of distance. We found above that k told us something about how often the zeros (well, every other zero) will occur. But the crests must occur at the same rate. So k tells us how often we encounter a crest in our snapshot graph. This is not too strange. It would be meaningful, for example, to say that a farmer had plowed his or her field to have two furrows per meter.



The frequency in the video graph is how often we encounter a crest, $\frac{1}{\lambda}$ is how often we encounter a crest in the snap shot graph. Thus $\frac{1}{\lambda}$ is playing the same role for a snap shot graph as frequency plays for a history graph. We could call $\frac{1}{\lambda}$ a *spatial frequency*. It is how often we encounter a crest as we march along in position, or how many crests we have in a unit amount of distance. And λ could be called a *spatial period*. Both $1/T$ and $1/\lambda$ answer the question “how often something happens in a unit of something” but one asks the question in time and the other in position along the wave.

My mental image for this is the set of groves on the edge of a highway. There is a distance between them, like a wavelength, and how often I encounter one as I move a distance along the road is the spatial frequency. You could say that there are, maybe, $3/m$. That is a spatial frequency! It is how many of something happens in a unit distance. We use this concept in optics to test how well an optical system resolves details in a photograph. The next figure is a test image. A good camera will resolve all spatial frequencies equally well. Notice the test image has sets of bars with different spatial frequencies. By forming an image of this pattern, you can see which spacial frequencies are faithfully represented by the optical system.



Resolution test target based on the USAF 1951 Resolution Test Pattern (not drawn to exact specifications).

In class you will see that our projector does not represent all spatial frequencies equally well! You can also see this now in the copy you are reading. If you are reading on-line or an electronic copy, your screen resolution will limit the representation of some spatial frequencies. Look for the smallest set of three bars where you can still tell for sure that there are three bars without zooming. A printed version that has been printed on a laser printer will usually allow you to see even smaller sets of three bars clearly.

Let's place k in the full equation for the sine wave for any time, t .

$$y(x, t) = y_{\max} \sin(kx - \omega t + \phi_o) \quad (5.11)$$

Sometimes you will see this equation written in a slightly different way. We could change our variable from x to $x - vt$ so that

$$y(x, t) = y(x - vt, 0)$$

With a little algebra we can do this

$$\begin{aligned} y(t) &= y_{\max} \sin(kx - \omega t + \phi_o) \\ &= y_{\max} \sin\left(\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t + \phi_o\right) \\ &= y_{\max} \sin\left(\frac{2\pi}{\lambda}\left(x - \frac{\lambda}{T}t\right) + \phi_o\right) \end{aligned}$$

This is in the form of a wave function so long as

$$v = \frac{\lambda}{T} \quad (5.12)$$

which is just how far the wave travels from crest to crest divided by how long it takes to travel that distance. The wave travels one wavelength in one period.

Indeed, this is the wave speed! Then

$$y(x, t) = y_{\max} \sin \left(\frac{2\pi}{\lambda} (x - vt) + \phi_o \right) \quad (5.13)$$

or, using k

$$y(x, t) = y_{\max} \sin (k(x - vt) + \phi_o) \quad (5.14)$$

Wave speed forms

We just found

$$v = \frac{\lambda}{T} \quad (5.15)$$

but it is easy to see that

$$v = \frac{2\pi\lambda}{2\pi T} = \frac{\omega}{k} \quad (5.16)$$

and

$$v = \lambda f \quad (5.17)$$

This last formula is, perhaps, the most common form encountered in our study of light.

Phase

You may be wondering about the phase constant we learned about in our study of SHM. We have ignored it up to now. But of course we can shift our sine just like we did for our plots of position vs. time for oscillation. Only now with a wave we have two graphs, a history and snapshot graphs, so we could shift along the x in a snapshot graph or along the t axes in a history graph. So the sine wave has the form.

$$y(x, t) = A \sin (kx - \omega t + \phi_o) \quad (5.18)$$

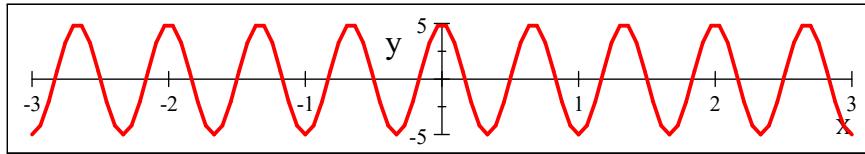
were ϕ_o will need to be determined by initial conditions just like in SHM problems and those initial conditions will include initial positions as well as initial times.

Let's consider that we have two views of a wave, the snapshot and history view. Each of these looks like sinusoids for a sinusoidal wave. Let's consider a specific wave,

$$y(x, t) = 5 \sin \left(3\pi x - \frac{\pi}{5}t + \frac{\pi}{2} \right)$$

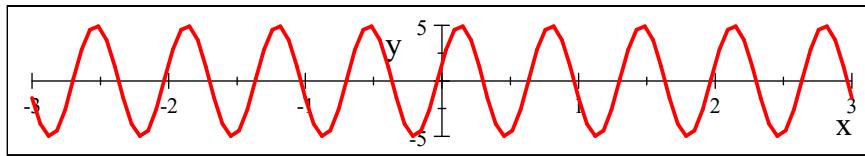
we look at a snapshot graph at $t = 0$

$$y(x, 0) = 5 \sin \left(3\pi x - \frac{\pi}{5}(0) + \frac{\pi}{2} \right)$$



and another at $t = 2\text{ s}$

$$y(x, 2\text{ s}) = 5 \sin \left(3\pi x - \frac{\pi}{5} (2\text{ s}) + \frac{\pi}{2} \right)$$



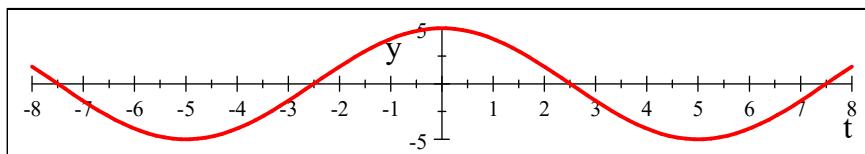
Comparing the two, we could view the latter as having a different phase constant that is the sum of what we have called the phase constant, ϕ_o plus $\omega\Delta t$, which tells us how different the times are between the two graphs. This time difference is what is shifting the graph

$$\phi_{total} = \omega\Delta t + \phi_o = -\frac{\pi}{5} (2\text{ s}) + \frac{\pi}{2}$$

that is, within the snapshot view, the time dependent part of the argument of the sign acts like an additional phase constant.

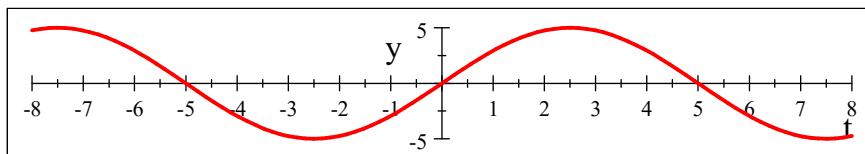
Likewise, in the history view, we can plot our wave at $x = 0$

$$y(0, t) = 5 \sin \left(3\pi (0) - \frac{\pi}{5} t + \frac{\pi}{2} \right)$$



and at $x = 1.5\text{ m}$

$$y(1.5\text{ m}, t) = 5 \sin \left(3\pi (1.5\text{ m}) - \frac{\pi}{5} t + \frac{\pi}{2} \right)$$



Within the history view, the kx part of the argument acts like a phase constant.

$$\phi_{total} = k\Delta x + \phi_o = 3\pi (1.5 \text{ m}) + \frac{\pi}{2}$$

Of course neither kx nor ωt are constant, But within individual views of the wave we have set them constant to form our snapshot and history representations. We can see that any part of the argument of the sine, $kx - \omega t + \phi_o$ could contribute to a phase shift, depending on the view we are taking.

Because of this, it is customary to call the entire argument of the sine function, $\phi = kx - \omega t + \phi_o$ the *phase of the wave*. Where ϕ_o is the phase constant, ϕ is the phase. Of course then, ϕ must be a function of x and t , so we have a different value for $\phi(x, t)$ for every point on the wave for every time. This can be a little confusing, ϕ and ϕ_o look a lot the same, but they are different.

5.4 The Speed of Waves

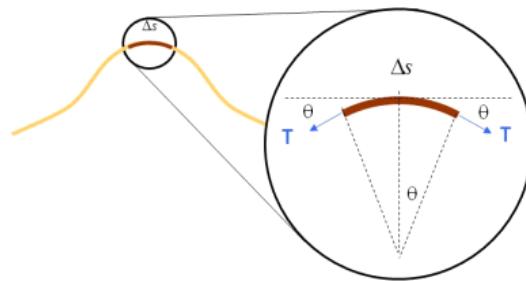
We have already considered the speed of waves on strings

$$v = \sqrt{\frac{T}{\mu}}$$

we should see where this comes from so we know its limitations. We would expect that there may be things that change the speed of sound waves in air as well (like a tension or a massiveness of the medium). Let's start with a formal derivation of our string speed formula, then take on sound waves in air.

5.4.1 Derivation of the string wave speed formula

Let's work a problem together. Let's find an expression for the speed of the wave as it travels along a string.



What are the forces acting on an element of string?

1. Tension on the RHS of the element from the rest of the string on the right, T_r

2. Tension on the LHS of the element from the rest of the string on the left,
 T_l
3. F_g

Lets assume that the element of string, Δs , at the crest is approximately an arc of a circle with radius R .

There is a force pulling left on the left end of the element that is tangent to the arc, there is a force pulling right at the right end of the element which is also tangent to the arc. These forces produce centripetal accelerations

$$a = \frac{v^2}{R} \quad (5.19)$$

The horizontal components of the forces cancel ($T \cos \theta$). The vertical component, ($T \sin(\theta)$) is directed toward the center of the arc. If the rope is not moving in the x direction, then $T_l = T_r$.

Then, the radial force F_r will have matching components from each side of the element that together are $2T \sin(\theta)$. Notice that to make sure the components are the same we must assume that the string is uniform and that it is not too massive (almost massless string approximation!). Since the element is small,

$$F_r = 2T \sin(\theta) \approx 2T\theta \quad (5.20)$$

The element has a mass m .

$$m = \mu \Delta s \quad (5.21)$$

where μ is the mas per unit length. Using the arc length formula

$$\Delta s = R(2\theta) \quad (5.22)$$

so

$$m = \mu \Delta s = 2\mu R\theta \quad (5.23)$$

and finally we use the formula for the radial acceleration

$$F_r = ma = (2\mu R\theta) \frac{v^2}{R} \quad (5.24)$$

Combining these two expressions for F_r

$$2T\theta = (2\mu R\theta) \frac{v^2}{R} \quad (5.25)$$

$$T = (\mu R) \frac{v^2}{R} \quad (5.26)$$

$$\frac{T}{\mu} = v^2 \quad (5.27)$$

and we find that

$$v = \sqrt{\frac{T}{\mu}} \quad (5.28)$$

which is just what we stated before, only now we see just how much approximation there was in building this formula. We might guess that it would not work so well for large cables supporting bridges or with cables that change size along their length.

5.4.2 Speed of the medium

Let's review for a moment. Suppose we take a jump rope, and shake one end up and down while a lab partner keeps his or her end stationary. You can make a sine wave in the rope. But being good physics students, let's make our "shake" very symmetric. Let's start our wave using simple harmonic motion.

Let's call an element of the rope Δx . Here the " Δ " is being used to mean "a small amount of." We are taking a small amount of the rope and calling its length Δx . And it is the motion of this small amount of rope that I want to think about now.

Each element of the rope (Δx_i) will also oscillate with SHM (think of a driven SHO). Note that the elements of the rope oscillate in the y direction, but the wave travels in x . This is a transverse wave.

Let's describe the motion of an element of the string at point P .

At $t = 0$,

$$y = A \sin(kx - \omega t + \phi_o) \quad (5.29)$$

(where I have chosen $\phi_o = 0$ for this example). The element does not move in the x direction. So we define the *transverse speed*, v_y , and the *transverse acceleration*, a_y , as the velocity and acceleration of the element of rope in the y direction. These are not the velocity and acceleration of the wave, just the velocity and acceleration of the element Δx at a point P .

Because we are doing this at one specific x location we need partial derivatives to find the velocity

$$v_y = \left. \frac{dy}{dt} \right|_{x=\text{constant}} = \frac{\partial y}{\partial t} \quad (5.30)$$

That is, we take the derivative of y with respect to t , but we pretend that x is not a variable because we just want one x position. Then

$$v_y = \frac{\partial y}{\partial t} = -\omega A \cos(kx - \omega t + \phi_o) \quad (5.31)$$

and

$$a_y = \frac{\partial v_y}{\partial t} = -\omega^2 A \sin(kx - \omega t + \phi_o) \quad (5.32)$$

These solutions should look very familiar! We expect them to be the same as a harmonic oscillator except that we now have to specify which oscillator—which part of the rope—we are looking at. That is what the kx part is doing.

5.5 The Linear Wave Equation

We added in some new math in our last two lectures. Differential equations! We identified simple harmonic motion using a differential equation

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

Perhaps we could find an equation like this for waves and then identify waves using that equation. We can start the same way we did for oscillation. Let's write the acceleration for a part of the medium.

$$a_y = \frac{d^2y}{dt^2} = -\omega^2 A \sin(kx - \omega t + \phi_o) \quad (5.33)$$

But for waves we have two variables. Let's guess that we might need to take derivatives with respect to x as well as for t .

$$\frac{dy}{dx} = kA \cos(kx - \omega t + \phi_o)$$

This isn't all that useful on its own, but it does tell us how the slope of the curve changes in a snapshot graph. Let's take another derivative

$$\frac{d^2y}{dx^2} = -k^2 A \sin(kx - \omega t + \phi_o)$$

We are back to a sine function. That is not a surprise. But it looks a lot like our time second derivative. We could take the ratio of the two equations to see just how different they are

$$\frac{\frac{d^2y}{dt^2}}{\frac{d^2y}{dx^2}} = \frac{-\omega^2 A \sin(kx - \omega t + \phi_o)}{-k^2 A \sin(kx - \omega t + \phi_o)}$$

Of course the sine parts cancel, and so do the A 's

$$\frac{\frac{d^2y}{dt^2}}{\frac{d^2y}{dx^2}} = \frac{-\omega^2}{-k^2}$$

and we know that

$$\frac{\omega}{k} = v$$

the wave speed (not the speed of a piece of the medium, but the speed of the wave)

$$\frac{\frac{d^2y}{dt^2}}{\frac{d^2y}{dx^2}} = v^2$$

If a function satisfies this equation, we can identify it as a wave. But this form is awkward, let's rewrite it as

$$\frac{d^2y}{dt^2} = v^2 \frac{d^2y}{dx^2}$$

This is the *linear wave equation*. There are waves that don't follow this equation. They are called non-linear waves. But that is a subject for an upper division course.

5.6 Waves in two and three dimensions

So far we have written expressions for waves, but our experience tells us that waves don't usually come as one dimensional phenomena. In the next figure, we see the disturbance (a drop) creating a water wave.

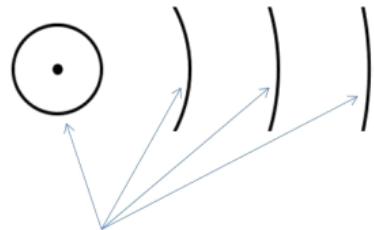


Picture of a water drop (Jon Paul Johnson, used by permission)

You can see the wave forming, but the wave is clearly not one dimensional. It appears nearly circular. In fact, it is closer to hemispherical, and this limit is only true because the disturbance is at the air-water boundary. Most waves in a uniform medium will be roughly spherical. As such a wave travels away from the source, the energy traveling gets more spread out. This causes the amplitude to decrease. Think of a sound wave, it gets quieter the farther you are from the source. We change our equation to account for this by making the amplitude a function of the distance, r , from the source

$$y(r, t) = A(r) \sin(kr - \omega t + \phi_0) \quad (5.34)$$

Of course, if we look at a very large wave, but we only look at part of the wave, we see that our part looks flatter as the wave expands.



Portion of a Spherical Wave: Wave becomes more flat as it expands

Very far from the source, our wave is flat enough that we can ignore the curvature across its wave fronts. We call such a wave a *plane wave*. There are no true plane waves in nature, but this idealization makes our mathematical solutions simpler and many waves come close to this approximation.

We have said that sound is a longitudinal wave with a medium of air. Really any solid, liquid, or gas will work as a medium for sound. For our study, we will take sound to be a longitudinal wave and treat liquids and gasses. Solids have additional forces involved due to the tight bonding of the atoms, and therefore are more complicated. Technically in a solid sound can be a transverse wave as well a longitudinal wave, but we usually call transverse waves of this nature *shear waves*.

Chapter 6

6 Power and Superposition

6 1.16.4 1.16.5

Fundamental Concepts

- Because waves are three dimensional, describing the power or energy delivered per time of the wave is not enough, We describe how spatially spread out that power is. We call this spread power *intensity*
- Waves invert on reflection from “fixed ends” and don’t invert from “free ends”
- If we make more than one wave in a medium, the waves “add up” or *superimpose*.
- When two waves interfere we get a new wave with a more complicated amplitude

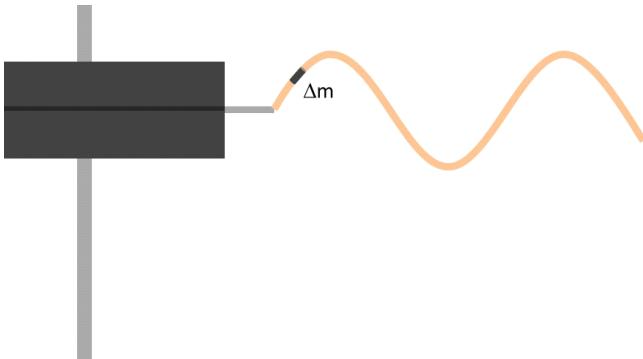
6.1 Energy and Power in waves

We have said that waves don’t transport mass, they transport energy. We used the example of sound waves. If the air molecules move to another location , we call it wind. But if the air molecules just vibrate around an equilibrium position, we called that sound waves. The molecules don’t end up somewhere else. But something does move.

Maybe a better example is a water wave in a swimming pool. Suppose someone jumps into a calm pool. The person is a disturbance and the water is a medium. We will get waves. But the water doesn’t end up bunched together on the other side of the pool. The water is still everywhere in the pool, even around the person that jumped in. So what did move away from the person? It is energy.

For sinusoidal waves, the amount of energy in the wave is related to both it's amplitude and it's frequency. And the specifics of that relationship depend on the type of wave. We will study both sound and light waves later in our course and we will find that relationship between energy, amplitude, and frequency are different for the two different kinds of waves.

Let's look at a specific case of a rope attached to a mechanical oscillator. You can see the situation in the next figure. The oscillator (black thing) has a piece that goes up and down (silver thing sticking out to the right). The rope is attached to this silver oscillating piece. The oscillatory acts as a disturbance. A wave is formed in the rope.



Once again let's take a small part of our medium. Before we took a pice that was Δx long, but this time let's describe our piece of rope as a small bit of mass Δm . We already know about linear mass densities.

$$\mu = \frac{M}{L}$$

and since our rope is uniform, the linear mass density is the same for each part of the rope. So for our marked part

$$\mu = \frac{\Delta m}{\Delta x}$$

Then for this part of the rope medium there will be a kinetic energy

$$\begin{aligned} K &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}\Delta m v_y^2 \end{aligned}$$

using our linear mass density we can write

$$\Delta m = \mu \Delta x$$

so the kinetic energy is

$$K = \frac{1}{2}(\mu \Delta x) v_y^2$$

and from before we know that $v_y = -\omega A \cos(kx - \omega t + \phi_o)$ so our kinetic energy is

$$K = \frac{1}{2} (\mu \Delta x) (-\omega A \cos(kx - \omega t + \phi_o))^2$$

we can clean this up a bit

$$K = \frac{1}{2} \mu \Delta x \omega^2 A^2 \cos^2(kx - \omega t + \phi_o)$$

Now let's simplify our situation by letting $\phi_o = 0$ and let's take a snap shot view with $t = 0$.

$$K = \frac{1}{2} \mu \Delta x \omega^2 A^2 \cos^2(kx)$$

Then, if we take only part of the energy associated with one very tiny part of our rope medium, our Δx becomes just dx . We don't have all the energy of the wave, just a small part of it. So we can write this as

$$dK = \frac{1}{2} \mu \omega^2 A^2 \cos^2(kx) dx$$

and we can integrate this up to find the energy of just one wavelength of the wave.

$$\int_0^{K_\lambda} dK = \int_0^\lambda \frac{1}{2} \mu \omega^2 A^2 \cos^2(kx) dx$$

A lot of the parts of our equation are constants, let's take them out front.

$$\int_0^{K_\lambda} dK = \frac{1}{2} \mu \omega^2 A^2 \int_0^\lambda \cos^2(kx) dx$$

The left hand side of this equation is just K_λ

$$K_\lambda = \frac{1}{2} \mu \omega^2 A^2 \int_0^\lambda \cos^2(kx) dx$$

but if you are like me, the right hand side is an integral that you have conveniently forgotten. But we can look up this integral in an integral table. My table gave the form

$$\int (\cos^2(ax)) dx = \frac{1}{2}x + \frac{1}{4a} \sin(2ax)$$

We have to match this to our equation. We see $a = k$ so

$$\begin{aligned}
 K_\lambda &= \frac{1}{2}\mu\omega^2 A^2 \left[\frac{1}{2}x + \frac{1}{4k} \sin(2kx) \right]_0^\lambda \\
 &= \frac{1}{2}\mu\omega^2 A^2 \left[\left\{ \frac{1}{2}\lambda + \frac{1}{4k} \sin(2k\lambda) \right\} - \left\{ \frac{1}{2}(0) + \frac{1}{4k} \sin(2k(0)) \right\} \right] \\
 &= \frac{1}{2}\mu\omega^2 A^2 \left[\left\{ \frac{1}{2}\lambda + \frac{1}{4\frac{2\pi}{\lambda}} \sin\left(2\frac{2\pi}{\lambda}\lambda\right) \right\} - \frac{1}{4k} \sin(2k(0)) \right] \\
 &= \frac{1}{2}\mu\omega^2 A^2 \left[\left\{ \frac{1}{2}\lambda + \frac{\lambda}{8\pi} \sin(4\pi) \right\} \right] \\
 &= \frac{1}{4}\mu\omega^2 A^2 \lambda
 \end{aligned}$$

This wavelength sized piece of rope will also have some potential energy because the fibers that make the rope are stretched to make the wave. They aren't really quite like little springs, but it's not too far wrong to model them as though they were. So let's take the spring potential energy

$$U = \frac{1}{2}k_s(y - y_e)^2$$

and let $y_e = 0$ for our piece of rope. We know that our piece of rope will go up and down in SHM. The natural frequency of our rope piece is

$$\omega = \sqrt{\frac{k_s}{m}}$$

We can solve this for the spring constant

$$k_s = m\omega^2$$

or for our little piece of the rope

$$k_s = \Delta m \omega^2$$

Let's put this into our potential energy equation

$$U = \frac{1}{2}\Delta m \omega^2 (y)^2$$

and substitute $\Delta m = \mu \Delta x$ again

$$U = \frac{1}{2}\mu \Delta x \omega^2 (y)^2$$

and of course, let our Δx get small so we have a small bit of potential energy from a small bit of rope.

$$dU = \frac{1}{2}\mu\omega^2 (y)^2 dx$$

and integrate this to find the potential energy in the stretchy bonds of one wavelength worth of rope.

$$\int_0^{U_\lambda} dU = \int_0^\lambda \frac{1}{2} \mu \omega^2 (y)^2 dx$$

We need to put in our equation for the position for our wave.

$$y(x, t) = A \sin(kx - \omega t + \phi_o)$$

but once again let's choose $\phi_o = 0$ and $t = 0$ so that

$$y(x, 0) = A \sin(kx)$$

then

$$\int_0^{U_\lambda} dU = \int_0^\lambda \frac{1}{2} \mu \omega^2 (A \sin(kx))^2 dx$$

The left hand side is easy. And there are, once again, constants

$$U_\lambda = \frac{1}{2} \mu \omega^2 A^2 \int_0^\lambda (\sin(kx))^2 dx$$

and we have another integral I don't remember how to do. But I still have my table of integrals and my table tells me

$$\int (\sin^2(ax)) dx = \frac{1}{2}x - \frac{1}{4a} \sin(2ax)$$

Which feels very familiar. Lets let $a = k$ again. Then

$$\begin{aligned} U_\lambda &= \frac{1}{2} \mu \omega^2 A^2 \left[\frac{1}{2}x - \frac{1}{4k} \sin(2kx) \right]_0^\lambda \\ &= \frac{1}{2} \mu \omega^2 A^2 \left[\left\{ \frac{1}{2}\lambda - \frac{1}{4k} \sin(2k\lambda) \right\} - \left\{ \frac{1}{2}0 - \frac{1}{4k} \sin(2k(0)) \right\} \right] \\ &= \frac{1}{4} \mu \omega^2 A^2 \lambda \end{aligned}$$

The total energy in the λ -sized rope piece would be

$$\begin{aligned} E_\lambda &= K_\lambda + U_\lambda \\ &= \frac{1}{4} \mu \omega^2 A^2 \lambda + \frac{1}{4} \mu \omega^2 A^2 \lambda \\ &= \frac{1}{2} \mu \omega^2 A^2 \lambda \end{aligned}$$

We know that energy is being transferred by the wave, whether it is a light or sound wave or any other mechanical wave. We should wonder, how fast is energy transferring. This can mean the difference between a warm ray of sun on a cool

spring day and being burned by a laser beam. We will start by considering the rate of energy transfer, *power*. The concept of power should be familiar to us from PH121. We can find the power as the rate at which energy is transferred.

$$\mathcal{P} = \frac{\Delta E}{\Delta t}$$

Since we picked the amount of energy in one wavelength, and we know the time it takes for one wavelength of the wave to pass by is T . Then the power is

$$P = \frac{\frac{1}{2}\mu\omega^2 A^2\lambda}{T}$$

but remember

$$v = \frac{\lambda}{T}$$

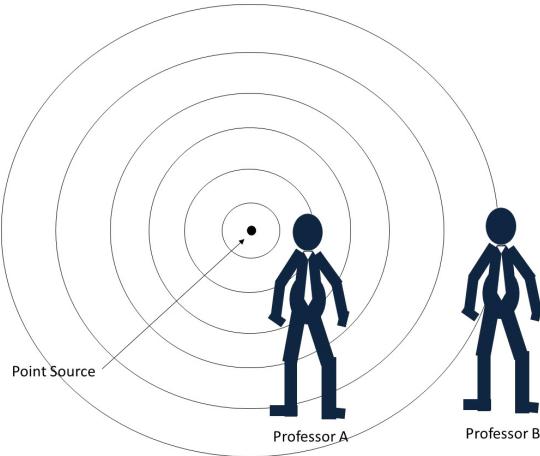
so we can write our power as

$$P = \frac{1}{2}\mu\omega^2 A^2 v$$

Power is important. Most detectors (like our ears and eyes) really detect the power delivered by a wave. And we can see that the power is proportional to the amplitude squared and the frequency squared for the wave. More on this as we study light and sound.

6.2 Power and Intensity

What we have done works fine for linear waves. But if we consider our spherical wave from a point source, we can see that this description isn't good enough. In the next figure we have two professors.



Thinking from our experience we would say that the sound will seem louder for

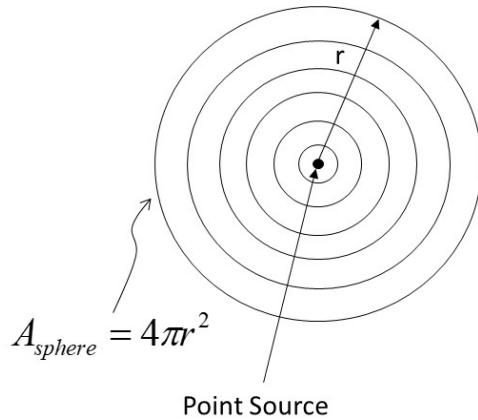
professor A. This is because the energy in the sound wave is being spread over the surface of the wave, and that surface is getting bigger as the wave moves outward. The energy is more spread out by the time it gets to professor B.

To be able to describe how much energy we get from our wave we need define something new.

$$\mathcal{I} \equiv \frac{\mathcal{P}}{A} \quad (6.1)$$

that is, the power divided by an area. But what does it mean?

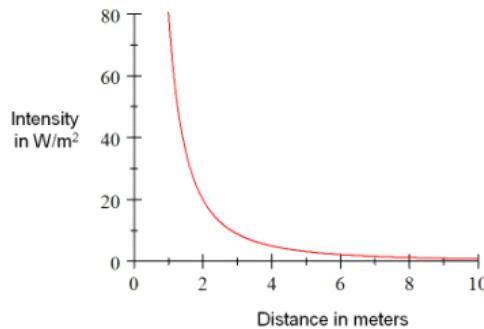
Consider a point source. This could be a loud speaker, or a buzzer, or a baby crying, etc.



it sends out waves in all directions. The wave crests will define a sphere around the points source (the figure shows a cross section but remember it is a wave from a point source, so we are really drawing concentric spheres like balloons.). Then form our point source

$$\mathcal{I} = \frac{\mathcal{P}}{4\pi r^2} \quad (6.2)$$

As the wave travels, its the power per unit area decreases with the square of the distance (think gravity) because the area is getting larger.



This quantity that tells us how spread out our power has become is called the *intensity* of the wave. Professor A would agree with us that the wave he heard was more intense than the wave heard by Professor B. That is because the wave was less spread out for Professor A.

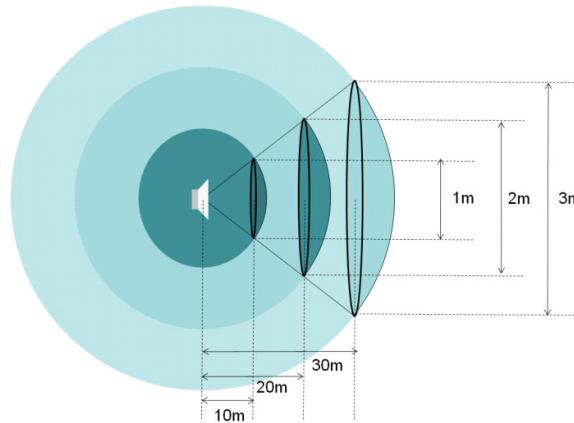
Suppose we cup our hand to our ear. What are we doing? We are increasing the area of our ear. Our ears work by transferring the energy of the sound wave to a mechanical-electro-chemical device that creates a nerve signal.¹ The more energy, the stronger the signal. If we are a distance r away from the source of the sound then the intensity is

$$\mathcal{I} = \frac{\mathcal{P}_{source}}{A_{wave}}$$

But we are collecting the sound wave with another area, the area of our hand. The power received is

$$\begin{aligned}\mathcal{P}_{received} &= \mathcal{I}A_{hand} \\ &= \frac{A_{hand}}{A_{wave}}\mathcal{P}_{source}\end{aligned}$$

and we can see that, indeed, the larger the hand, the more power, and therefore more energy we collect. This is the idea behind a dish antenna for communications and the idea behind the acoustic dish microphones we see at sporting events. In next figure, we can see that it would take an increasingly larger dish to maintain the same power gathering capability as we get farther from the source.

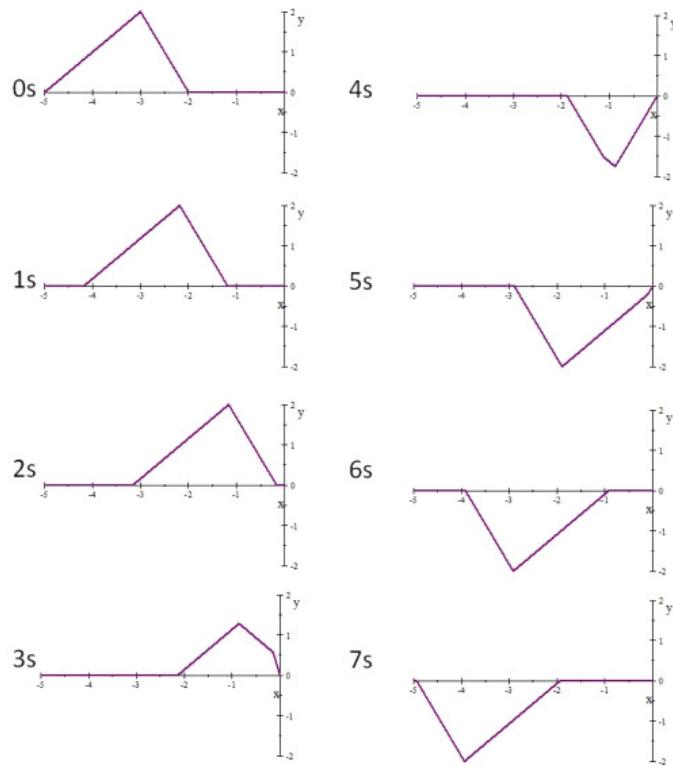


¹The inner hair cells in the organ of Corti in the cochlea.

6.3 Reflection and Transmission

In our examples so far, we have not explained how we got two waves into a medium. One way is to simply reflect one wave back on top of itself.

In class we will make pulses on a spring with one end of the rope fixed (held by a class member). What happened when the pulse reached the end of the rope?

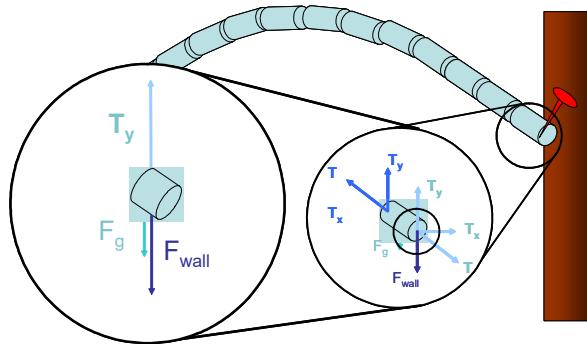


6.3.1 Case I: Fixed rope end.

There is a big change in the medium at the end of the rope. The rope ends. There is a person or (as in the next figure) some thing holding the rope in place. This change in medium causes a reflection.

In the fixed end case, the pulse is inverted. Why?

In the next figure I have envisioned a rope made of small rope segments.



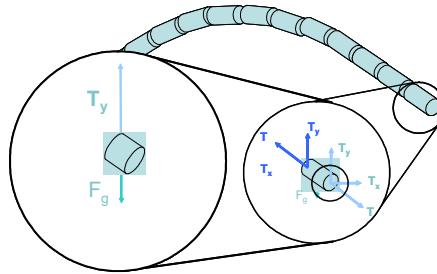
The end of the rope pushes up on the support (person in class, or nail in the figure). By Newton's third law there must be a normal force downward on the rope end. Compressing the molecules in the support stores potential energy in those compressed atoms. They will release that potential energy and create kinetic energy in the rope end. That kinetic energy will have a rope end velocity

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

and that velocity will be downward. This will pull the rope down, inverting the wave.

But what happens if the rope end is not fixed?

The rope end rises, and therefore there is no force exerted. The pulse (or at least part of the pulse energy) is still reflected, but there is no inversion because there was not downward force or no stored potential energy in the support!

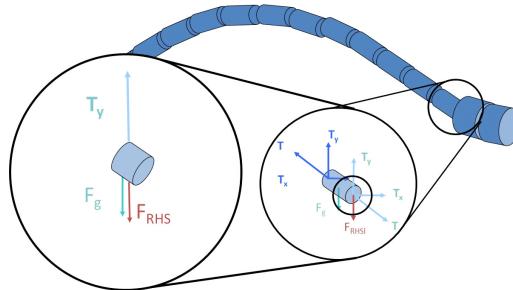


6.3.2 Case III: Partially attached rope end

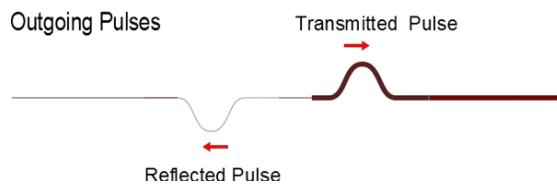
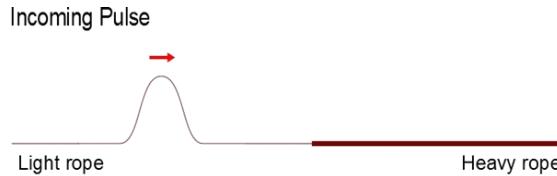
Now lets tie the rope to another rope that is larger, more dense, than the rope we have been using, what will happen?

The light end of the rope exerts a force on the heavy beginning of the new rope. The atoms in the heavy rope will be pulled and compressed. The heavy rope will move, but the heaviness of the rope will prevent them from going very far. The fibers on the end of the heavy rope will build up potential energy

because their bonds will be compressed. They will push downward on the light end of the rope. Once again the potential energy from the compressed heavy rope atoms will transfer to kinetic energy in the light rope end. Once again that light rope end will move downward.



We expect part of the energy to reflect back along the light rope, and this pulse will be inverted. Notice that the heavy rope did move upward a little, so there will also be a pulse on the heavy rope. Since this pulse formed from the light rope pulling up the heavy rope, and the light rope atoms were not compressed, this pulse will not be inverted.



This is all great for pulses. But suppose a sinusoidal wave approaches the boundary. We can envision the crests like pulses, and we expect the first crest to slow down when it reaches the boundary, letting the other crests catch up. Once the wave passes the boundary, the crests will be closer together. The wavelength changes as we move to the slower medium.

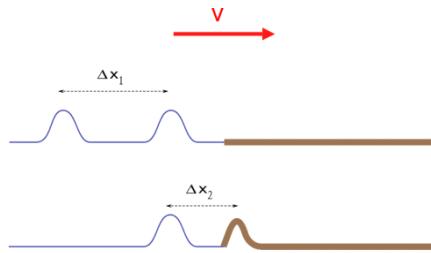
But does the frequency change? We know that

$$v = \lambda f$$

so

$$f = \frac{v}{\lambda}$$

both the speed and the wavelength have changed, but did they change proportionately so f is constant? This must be so. Think that the change in wavelength is due to the relative speed of the wave in the two media. If Δv is small the change in λ will be small because the crests are not delayed too long. If Δv is large, the crests are delayed by a large amount and so the change in λ is large. We won't derive the fact that f is constant, but we can see that is is very believable that it is true.



This is true for all waves, even light. When a wave crosses a boundary from a fast to a slow or a slow to a fast medium, λ will change and f will remain constant.

6.4 Superposition Principle

Wave
Demo Machine

What happens if we have more than one wave propagating in a medium? If you remember being a little child in a bath tub, you will probably remember making waves in the water. If you made a wave with each hand, the two waves seemed to “pile up” in the middle and make a big splash. We should expect something like this for any kind of wave. We call the “piling up” of waves *superposition*. The word literally means putting one wave on top of another.

Superposition: If two or more traveling waves are moving through a medium, the resultant wave formed at any point is the algebraic sum of the values of the individual wave forms.

So if we have

$$y_1(x, t) \quad (6.3)$$

and

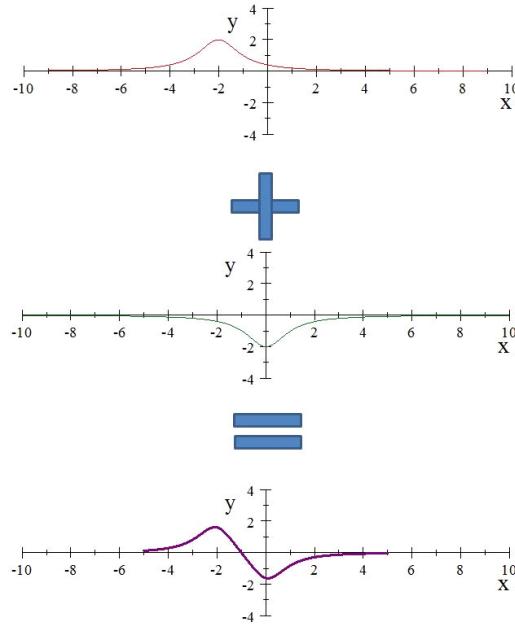
$$y_2(x, t) \quad (6.4)$$

both propagating on a string, then we would see

$$y_r(x, t) = y_1(x, t) + y_2(x, t) \quad (6.5)$$

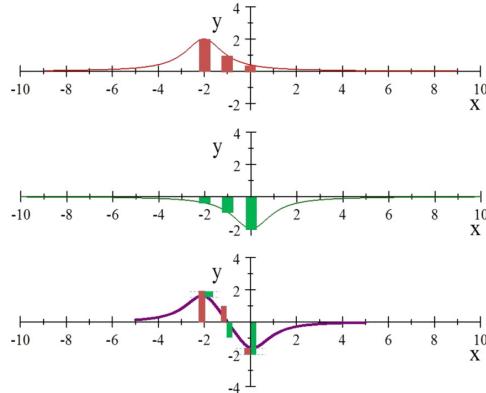
This is a fantastically simple way for the universe to act!

Let's look at an example. let's add the top wave (red) to the middle wave(green). We get the bottom wave (purple)



Of course we are adding these in the snapshot view. So this is all done for just one instant of time.

Let's see how to do this.



Start at $x = -2$. In the figure, I drew a red bar to show the y value at $x = -2$ for the red curve. Likewise, I have a green bar showing the value of y at $x = -2$ for the green wave. Note that this is negative. On the bottom graph, the bars have been repeated, and we can see that the red bar minus the green bar brings

us to the value for the resulting wave at the point $x = -2$. We need to do this at every point along all the waves for this instant of time.

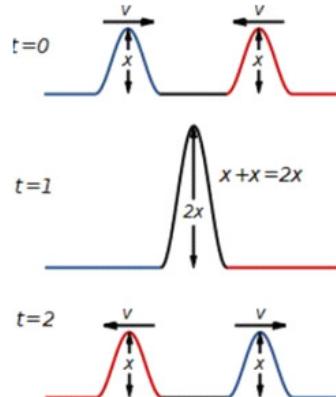
This is tedious by hand, so we won't generally do this calculation by hand. But a computer can do it easily.

Note that this is really only true for *linear* systems. Let's take the example of a slinky. If we form two waves in the slinky, they behave according to the superposition principle most of the time. But suppose we make the amplitude of the individual waves large. They may travel individually OK, but when the amplitudes add we may overstretch the slinky. Then it would never return to its original shape. The wave form would have to change. Such a wave is not linear. There is a good rule of thumb for when waves are linear.

A wave is generally linear when its amplitude is much smaller than its wavelength.

6.5 Consequences of superposition

Linear waves traveling in media can pass through each other without being destroyed or altered!

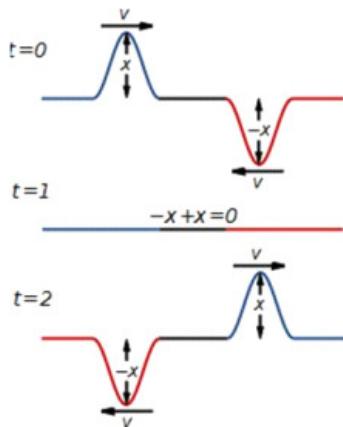


Constructive Interference (Public Domain image by Inductiveload,
http://commons.wikimedia.org/wiki/File:Constructive_interference.svg)

Our wave on the string makes the string segments move in the y direction. Both waves do this. So putting the two waves together just makes the string segments move more! There is a special name for what we observe

1. *interference*: The combination of separate waves in the same region of space to produce a resultant wave.

What happens if one of the pulses is inverted?

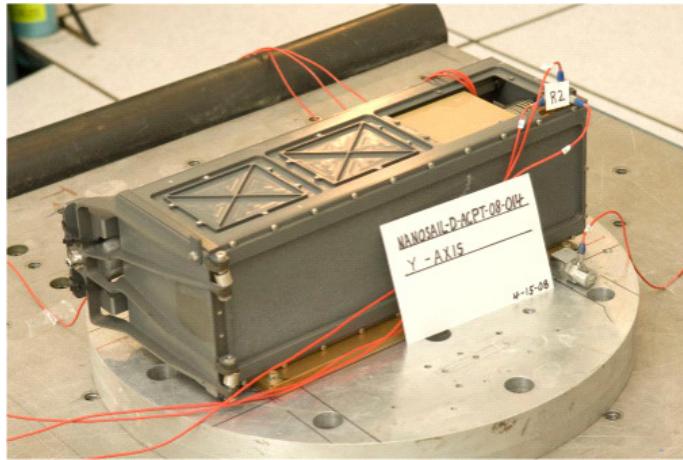


Destructive Interference (Public Domain image by Inductiveload,
http://commons.wikimedia.org/wiki/File:Destructive_interference1.svg)

When the two pulses meet, they “cancel each other out.” But do they go away? No! the energy is still there, the string segment motions have just summed vectorially to zero, the energy carried by each wave is still there. We have a few more definitions. The type of interference we have just seen is the first

1. *Destructive Interference*: Interference between waves when the displacements caused by the two waves are opposite in direction
2. *Constructive Interference*: interference between waves when the displacements caused by the two waves are in the same direction

The combination of waves is important for both scientists and engineers. In engineering this is the heart of vibrometry.



Marshall and Cal Poly technicians wired the NanoSail-D spacecraft to accelerometers, instruments which measure vibration response during simulated launch conditions. Image courtesy NASA, image in the Public Domain.

Mechanical systems have moving parts. These moving parts can be the disturbance that creates a wave. If more than one wave crest arrives at a location in the device, the amplitude at that location could become large. The oscillation of this part of the device could rattle apart welds or bolts, destroying the device. Later, as we study spectroscopy, we will see how to diagnose such a problem and hint at how to correct it.

6.6 Single Frequency Interference in one dimension

We have seen interference from superimposing two waves in the same medium. Whether we get constructive or destructive interference really depends on the total phase difference. Let's review this, but let's make our two waves as general as possible.

$$\begin{aligned}y_1 &= y_{\max} \sin(k_1 x_1 - \omega_1 t_1 + \phi_1) \\y_2 &= y_{\max} \sin(k_2 x_2 - \omega_2 t_2 + \phi_2)\end{aligned}$$

where we have allowed the wavelengths to be different (so then the wave numbers, k_1 and k_2 could be different), the distances traveled by the waves could be different (x_2 might not be equal to x_1), the frequencies might be different, the times observed might be different, and the phase constants might be different. We have not made the amplitudes different (but it is likely to happen in an upcoming homework problem).

Then the resulting wave would be

$$\begin{aligned} y_r &= y_{\max} \sin(k_2 x_2 - \omega_2 t_2 + \phi_2) + A \sin(k_1 x_1 - \omega_1 t_1 + \phi_1) \\ &= 2y_{\max} \cos\left(\frac{(k_2 x_2 - \omega_2 t_2 + \phi_2) - (k_1 x_1 - \omega_1 t_1 + \phi_1)}{2}\right) \\ &\quad \times \sin\left(\frac{(k_2 x_2 - \omega_2 t_2 + \phi_2) + (k_1 x_1 - \omega_1 t_1 + \phi_1)}{2}\right) \end{aligned}$$

just as before. We can rewrite this as

$$\begin{aligned} y_r &= 2y_{\max} \cos\left(\frac{1}{2} [(k_2 x_2 - \omega_2 t_2 + \phi_2) - (k_1 x_1 - \omega_1 t_1 + \phi_1)]\right) \\ &\quad \times \sin\left(\frac{(k_2 x_2 - \omega_2 t_2 + \phi_2) + (k_1 x_1 - \omega_1 t_1 + \phi_1)}{2}\right) \end{aligned}$$

and recognize that the sine part is still a function of position and time (just a complicated one) so it must be a wave. The cosine part must be part of the amplitude. Let's write out this amplitude

$$A = 2y_{\max} \cos\left(\frac{1}{2} [(k_2 x_2 - \omega_2 t_2 + \phi_2) - (k_1 x_1 - \omega_1 t_1 + \phi_1)]\right)$$

The part in square brackets is what we have been calling the phase difference

$$\Delta\phi = (k_2 x_2 - \omega_2 t_2 + \phi_2) - (k_1 x_1 - \omega_1 t_1 + \phi_1)$$

Notice that to make constructive interference we want the amplitude to be as big as possible. This happens when the cosine is ± 1 . We can find when this happens by recalling that cosine is ± 1 when the argument of the cosine is a multiple of π . Notice that the amplitude has the cosine of half the phase difference

$$A = 2y_{\max} \cos\left(\frac{1}{2} \Delta\phi\right)$$

so

$$\frac{1}{2} \Delta\phi = \pi m \quad m = 0, \pm 1, \pm 2, \pm 3, \dots$$

or

$$\Delta\phi = 2\pi m \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \quad \text{Constructive}$$

For total destructive interference we want the amplitude to be zero. To achieve this the cosine must be zero and this happens at odd multiples of $\pi/2$.

$$\frac{1}{2} \Delta\phi = \frac{\pi}{2} m \quad m = \pm 1, \pm 3, \pm 5, \dots$$

Then we could say that our phase difference

$$\Delta\phi = \pi m \quad m = \pm 1, \pm 3, \pm 5, \dots$$

another way to say this is to allow m to be any integer but to write the condition as

$$\Delta\phi = \left(m + \frac{1}{2}\right) 2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \quad \text{Destructive}$$

Let's write these out in detail

$$\begin{aligned} [(k_2x_2 - \omega_2t_2 + \phi_2) - (k_1x_1 - \omega_1t_1 + \phi_1)] &= 2\pi m \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \\ [(k_2x_2 - \omega_2t_2 + \phi_2) - (k_1x_1 - \omega_1t_1 + \phi_1)] &= \left(m + \frac{1}{2}\right) 2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \end{aligned}$$

We can see that there are at least four sources of phase difference here. We can change the phase difference by changing wavelength ($k_2 \neq k_1$), changing frequency ($\omega_2 \neq \omega_1$), having the waves travel different distances ($x_2 \neq x_1$), having different phase constants ($\phi_2 \neq \phi_1$) or even consider somehow mixing the waves at different times ($t_2 \neq t_1$). This last one is not as interesting, we want our waves mixed together at the same time so let's set

$$t_2 = t_1 = t$$

but let's let any of the other variables be changeable. This gives us many ways to produce constructive or destructive interference. In our next lecture we will use this analysis to solve specific problems. The pattern will be to use our two criteria

$$\begin{aligned} \Delta\phi &= 2\pi m \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \quad \text{Constructive} \\ \Delta\phi &= \left(m + \frac{1}{2}\right) 2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \quad \text{Destructive} \end{aligned}$$

to determine if we have constructive, destructive, or partially constructive or destructive interference.

If both waves have the same frequency, we can simplify this.

$$\begin{aligned} \Delta\phi &= [(k_2x_2 - \omega t_2 + \phi_2) - (k_1x_1 - \omega t_1 + \phi_1)] = 2\pi m \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \\ \Delta\phi &= [(k_2x_2 - \omega t_2 + \phi_2) - (k_1x_1 - \omega t_1 + \phi_1)] = \left(m + \frac{1}{2}\right) 2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \end{aligned}$$

becomes

$$\begin{aligned} \Delta\phi &= (k_2x_2 - k_1x_1) + (\phi_2 - \phi_1) = 2\pi m \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \\ \Delta\phi &= (k_2x_2 - k_1x_1) + (\phi_2 - \phi_1) = \left(m + \frac{1}{2}\right) 2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \end{aligned}$$

and our wave equation becomes

$$\begin{aligned} y_r &= 2y_{\max} \cos \left(\frac{1}{2} [(k_2x_2 - k_1x_1) + (\phi_2 - \phi_1)] \right) \\ &\quad \times \sin \left(\frac{k_2x_2 + k_1x_1}{2} - \frac{(\omega t_2 + \omega_1 t_1)}{2} + \frac{(\phi_2 + \phi_1)}{2} \right) \end{aligned}$$

6.7 Single frequency interference in more than one dimension

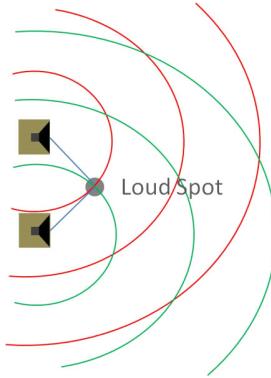
But what happens if our waves don't travel along the same line? Suppose you are at a dance, and there are two speakers. Further suppose that you are testing the system with a constant tone so $\omega_2 = \omega_1$ (either that or you have somewhat boring music with long, sustained tones). Suppose the two speakers make waves in phase ($\phi_1 = \phi_2$). There is no slower medium, so we expect $k_2 = k_1$. If you are equal distance from the two speakers, you would expect constructive interference because both $\Delta\phi_o = 0$ and $\Delta x = 0$ for this case.

$$\begin{aligned}\Delta\phi &= (k_2 x_2 - \omega_2 t_2 + \phi_2) - (k_1 x_1 - \omega_1 t_1 + \phi_1) \\ &= (kx_2 - \omega t + \phi) - (kx_1 - \omega t + \phi) \\ &= k(\Delta x)\end{aligned}$$

If we pick a spot where $x_1 = x_2$ we would expect

$$\Delta\phi = 0$$

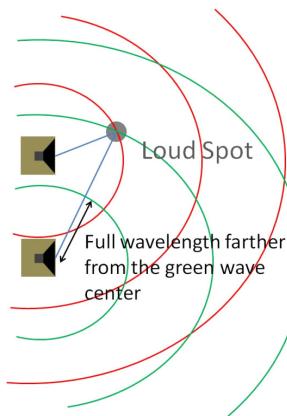
and since 0 is a multiple of 2π in our criteria, then we expect this will give us constructive interference.



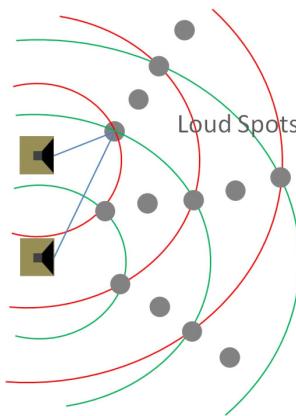
But there are more places where we expect constructive interference, because we know the sound wave is really spherical. Any time the path difference, $\Delta x = n\lambda$, then

$$\Delta\phi = \frac{2\pi}{\lambda} (n\lambda) = n2\pi$$

and we will have constructive interference. The next figure shows an example where the path difference is one wavelength.



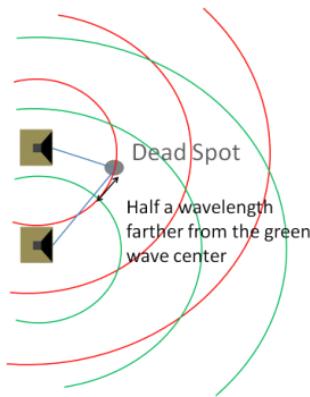
But any of the spots in the next figure will experience constructive interference. Note the loud spots are where there are two crests or two troughs together.



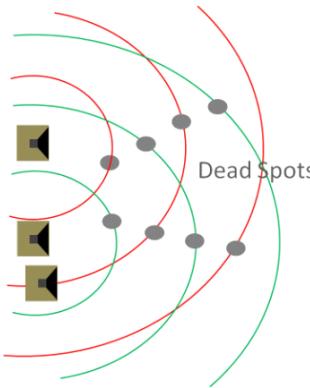
We also expect to see destructive interference. This should occur where path differences are multiples of $\lambda/2$ so that

$$\begin{aligned}\Delta\phi &= \frac{2\pi}{\lambda} \Delta x \\ &= \frac{2\pi}{\lambda} \left(n \frac{\lambda}{2} \right) \\ &= n\pi\end{aligned}$$

The next figure shows a case where $\Delta x = \lambda/2$



and the next figure shows many places where there will be destructive interference because the two waves are out of phase by half a wavelength.



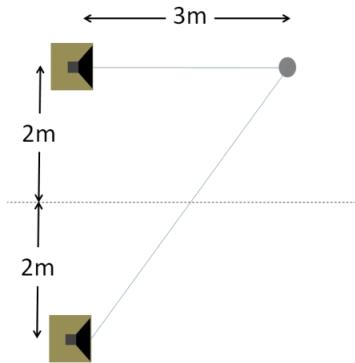
When you moved from one dimension to two dimensions in PH121, we changed from the variables x and y to the variable r where

$$r = \sqrt{x^2 + y^2}$$

Thus for this case our phase becomes

$$\Delta\phi = \left(2\pi \frac{\Delta r}{\lambda} + \Delta\phi_o \right)$$

In our dance example, suppose we have speakers that are 4 m apart and we are standing 3 m directly in front of one of the speakers. Further suppose that we play an *A* just above middle *C* which has a frequency of 440 Hz. The speed of sound is 343 m/s. Our speakers are connected to the same stereo with equal length wires. What is the phase difference at this spot?



From the geometry we can tell that the path from the second speaker must be 5 m. So

$$\begin{aligned}\Delta x &= 5 \text{ m} - 3 \text{ m} \\ &= 2 \text{ m}\end{aligned}$$

We can tell that the wavelength is

$$\begin{aligned}\lambda &= \frac{v}{f} \\ &= \frac{343 \text{ m/s}}{440 \text{ Hz}} \\ &= 0.77955 \text{ m}\end{aligned}$$

Since the speakers are connected to the same stereo with equal length wires, $\Delta\phi_o = 0$. Then

$$\begin{aligned}\Delta\phi &= \frac{2\pi}{\lambda} \Delta x + \Delta\phi_o \\ &= \frac{2\pi}{0.77955 \text{ m}} (2 \text{ m}) + 0 \\ &= 5.1312\pi \\ &= 2\pi + 3.1312\pi\end{aligned}$$

We should ask, is this constructive or destructive interference? Well, it is neither purely constructive interference nor total destructive interference. Our amplitude would be

$$2A \cos\left(\frac{1}{2} \left(\frac{2\pi}{\lambda} \Delta x + \Delta\phi_o \right)\right)$$

so in this case we get

$$2A \cos\left(\frac{1}{2} (2\pi + 3.1312\pi)\right) = -0.40927A$$

which is smaller (in magnitude) than A , so it is partial destructive interference. It would be quieter at this spot than if we just had one speaker playing.

*6.7. SINGLE FREQUENCY INTERFERENCE IN MORE THAN ONE DIMENSION*99

You might guess that this sort of analysis plays a large part in design of concert halls. It also is important in mechanical designs. But you should have seen a deficit in what we have learned so far. Up to this point, we have only mixed waves that have the same frequency. Can we mix waves that have different frequencies?

Chapter 7

7 Standing Waves and Sound Waves 1.16.6, 1.17.1

Fundamental Concepts

- Waves traveling different directions can create static patterns of constructive and destructive interference.
- These patterns are called “standing waves”
- The frequencies that make the standing wave patterns are quantized
- Sound is a wave
- We can express sound waves in terms displacement of air molecules or as pressure changes

7.1 Mathematical Description of Superposition

We know what superposition is, but we don’t really want to add values for millions of points in a medium to find out what a combination of waves will look like. At the very least, we want to make a computer do that (and programs like OpenFoam do something very akin to this!). But where we can, we would like to combine wave functions algebraically. Let’s see how this can work.

Lets define two wave functions

$$y_1 = y_{\max} \sin(kx - \omega t)$$

and

$$y_2 = y_{\max} \sin(kx - \omega t + \phi_o)$$

These are two waves with the same frequency and wave number traveling the same direction in the medium, but they at $t = 0$ the y values are not the same because of the phase shift. The graph of y_2 is shifted by an amount ϕ_o .

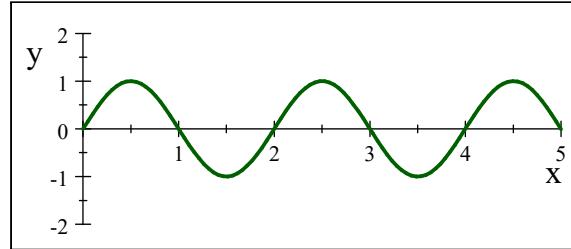
I will pick some values for the constants

$$\begin{aligned}\lambda &= 2 \\ k &= \frac{2\pi}{\lambda} \\ \omega &= 1 \\ \phi_o &= \frac{\pi}{6} \\ t &= 0 \\ A &= 1\end{aligned}$$

then for y_1 we have

$$\begin{aligned}y_1 &= (1) \sin\left(\frac{2\pi}{\lambda}x - (1)t\right) \\ &= \sin\left(\frac{2\pi}{2}x - (1)t\right) \\ &= \sin(\pi x - t)\end{aligned}$$

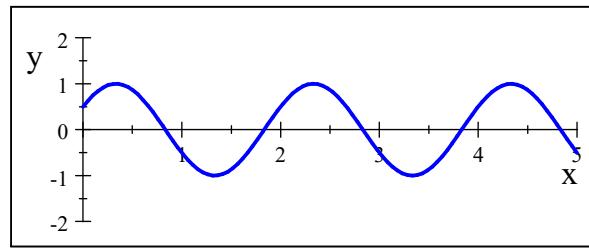
here is a plot of the wave function, y_1



Now let's consider y_2 . Using the values we chose, y_2 can be written as

$$\begin{aligned}y_2 &= y_{\max} \sin(kx - \omega t + \phi_o) \\ &= \sin\left(\pi x - t + \frac{\pi}{6}\right)\end{aligned}$$

which looks like this



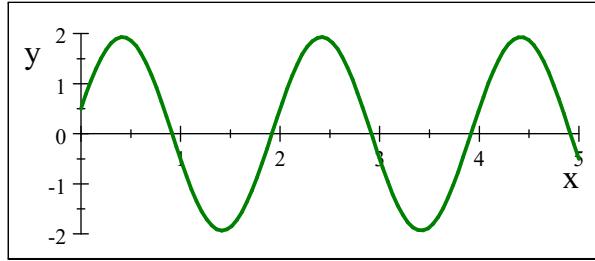
What does it look like if we add these waves using superposition? Symbolically we have

$$y_r = y_{\max} \sin(kx - \omega t) + y_{\max} \sin(kx - \omega t + \phi_o) \quad (7.1)$$

and putting in the numbers gives

$$y_r = \sin(\pi x - t) + \sin\left(\pi x - t + \frac{\pi}{6}\right)$$

which is shown in the next graph.



Notice that the wave form is taller (larger amplitude). Noticed it is shifted along the x axis.

We can find out by how much by rewriting y_r . We need a trig identity

$$\sin a + \sin b = 2 \cos\left(\frac{a-b}{2}\right) \sin\left(\frac{a+b}{2}\right)$$

Then let $a = kx - \omega t$ and $b = kx - \omega t + \phi_o$

$$\begin{aligned} y_r &= y_{\max} \sin(kx - \omega t) + A \sin(kx - \omega t + \phi_o) \\ &= 2y_{\max} \cos\left(\frac{(kx - \omega t) - (kx - \omega t + \phi_o)}{2}\right) \sin\left(\frac{(kx - \omega t) + (kx - \omega t + \phi_o)}{2}\right) \\ &= 2y_{\max} \cos\left(\frac{-\phi_o}{2}\right) \sin\left(\frac{2kx - 2\omega t + \phi_o}{2}\right) \\ &= 2y_{\max} \cos\left(\frac{-\phi_o}{2}\right) \sin\left(kx - \omega t + \frac{\phi_o}{2}\right) \\ &= 2y_{\max} \cos\left(\frac{\phi_o}{2}\right) \sin\left(kx - \omega t + \frac{\phi_o}{2}\right) \end{aligned}$$

where we use the fact that $\cos(-\theta) = \cos(\theta)$.

Let's look at the parts of this expression. First take the sine part.

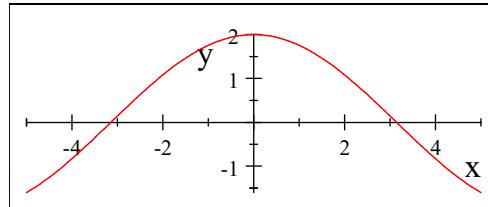
$$\sin\left(kx - \omega t + \frac{\phi_o}{2}\right) \quad (7.2)$$

This part is a traveling wave with the same k and ω as our original waves, but it has a phase of $\phi_o/2$. So our combined wave is shifted by $\phi_o/2$ or half the phase shift of y_2 .

Now let's look at other factor

$$2y_{\max} \cos\left(\frac{\phi_o}{2}\right) \quad (7.3)$$

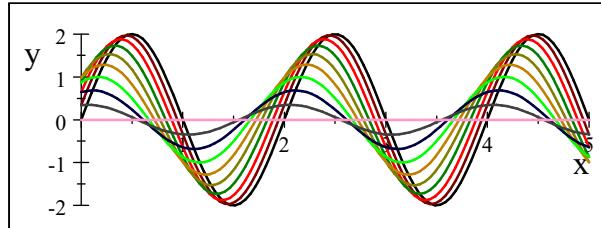
The sine part of our wave equation is multiplied by all of this factor. So all of this part is the new amplitude. It has a maximum value when $\phi_o = 0$



When $\phi_o = \pi$, then

$$2y_{\max} \cos\left(\frac{\pi}{2}\right) = 0$$

so when $\phi_o = 0$ we have a new maximum amplitude of $2y_{\max}$ and when $\phi_o = \pi$ we have a zero amplitude. Here is our wave for several choices of ϕ_o .



7.1.1 Mathematical description of standing waves

Now that we have a way to make two waves to superimpose, we can study the special case of standing waves.

A standing wave pattern is the result of the superposition of two traveling waves with the same frequency going in opposite directions

$$\begin{aligned} y_1 &= y_{\max} \sin(kx - \omega t) \\ y_2 &= y_{\max} \sin(kx + \omega t) \end{aligned}$$

The sum is

$$y_r = y_1 + y_2 = y_{\max} \sin(kx - \omega t) + A \sin(kx + \omega t)$$

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

$$\begin{aligned}
 y_r &= y_{\max} \sin(kx - \omega t) + y_{\max} \sin(kx + \omega t) \\
 &= y_{\max} \sin(kx) \cos(\omega t) - y_{\max} \cos(kx) \sin(\omega t) + y_{\max} \sin(kx) \cos(\omega t) + y_{\max} \cos(kx) \sin(\omega t) \\
 &= 2y_{\max} \sin(kx) \cos(\omega t) \\
 &= (2y_{\max} \sin(kx)) \cos(\omega t)
 \end{aligned}$$

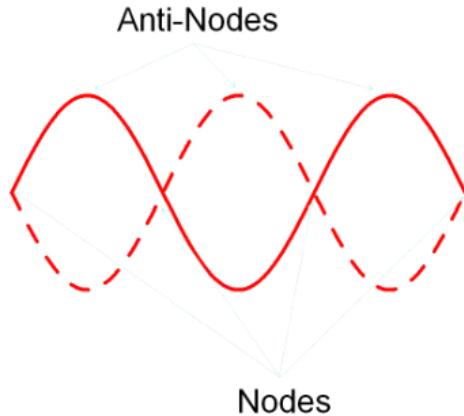
This looks like the harmonic oscillator equation

$$y = y_{\max} \cos(\omega t + \phi_o)$$

with $\phi_o = 0$ and with another complicated amplitude

$$2y_{\max} \sin(kx)$$

That is, we have a set of harmonic oscillators who's amplitude is different for each value of x .



We can identify spots along the x axis where the amplitude is always zero! we will call these spots *nodes*. These happen when $\sin(kx) = 0$ or when

$$kx = n\pi$$

By using

$$k = \frac{2\pi}{\lambda}$$

we have

$$\frac{2\pi}{\lambda}x = n\pi$$

We can cancel things, and solve for x

$$\begin{aligned}
 \frac{2}{\lambda}x &= n \\
 x &= n\frac{\lambda}{2}
 \end{aligned}$$

We can also find the places along x where the amplitude will be largest. this occurs when $\sin(kx) = 1$ or when

$$kx = n\frac{\pi}{2}$$

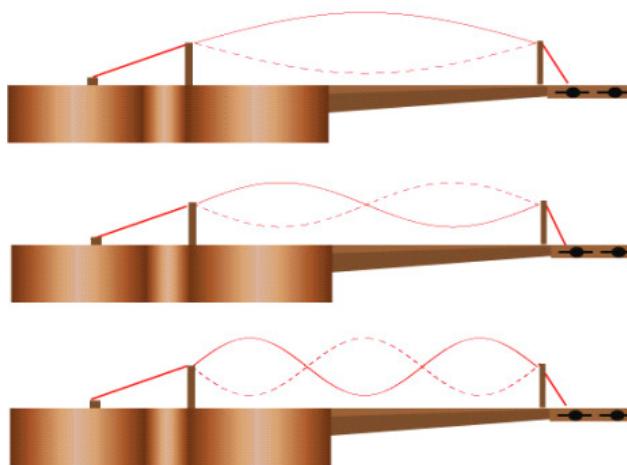
or

$$\begin{aligned}\frac{2\pi}{\lambda}x &= n\frac{\pi}{2} \\ x &= n\frac{\lambda}{4}\end{aligned}$$

these are called *antinodes*.

We can also create standing waves with sound or even light waves!

7.1.2 Standing Waves in a String Fixed at Both Ends



If we attach a string to something on both ends, we find something interesting in the standing wave pattern. Not all possible standing waves can be realized. Some frequencies are preferred, and some never show up. The standing wave pattern is *quantized*. The patterns that are allowed are called *normal modes*. We will see this any time a wave confined by boundary conditions (light in a resonant cavity, radio waves in a wave guide, electrons in an atom, etc.).

The figure shows some normal modes for a string.

We find which modes are allowed by first imposing the boundary condition that each end must be a node. We start with

$$y = 2y_{\max} \sin(kx) \cos(\omega t)$$

and recognize that we have one condition met because

$$y = 0$$

when

$$x = 0$$

we need $y = 0$ when $x = L$. That happens when

$$kL = n\pi$$

I will write this as

$$k_n L = n\pi$$

to indicate there are many values of k that can work. Solving this for λ_n gives

$$\begin{aligned} \frac{2\pi}{\lambda_n} L &= n\pi \\ \frac{2L}{n} &= \lambda_n \end{aligned}$$

Which just says there are many wavelengths that will work. Let's see how to use this, the first mode will have

$$\lambda_1 = 2L$$

where L is the length of the string. Looking at the figure, this works.

The second mode has three nodes (one on each end and one in the middle). This gives

$$\lambda_2 = L$$

We can keep going, the third mode will have five nodes

$$\lambda_3 = \frac{3L}{2}$$

and so forth to give

$$\lambda_n = \frac{2L}{n}$$

We use our old friend, the equation for wave speed,

$$v = f\lambda$$

to find the frequencies of the modes

$$f = \frac{v}{\lambda}$$

$$f_1 = \frac{v}{\lambda_1} = \frac{v}{2L}$$

or, in general

$$\begin{aligned} f_n &= \frac{v}{\lambda_n} = n \frac{v}{2L} \\ &= \frac{n}{2L} v \\ &= \frac{n}{2L} \sqrt{\frac{T}{\mu}} \end{aligned}$$

The lowest frequency has a special name, the *fundamental frequency*. The higher frequencies are integer multiples of the fundamental. When this happens we say that the frequencies form a *harmonic series*, and the modes are called *harmonics*. Since only certain frequencies work, we say that the frequencies of waves on the string that make standing waves are *quantized!* Quantization means that some values are allowed. This idea is the basis behind Quantum mechanics (which views light and even matter as waves).

7.1.3 Musical Strings

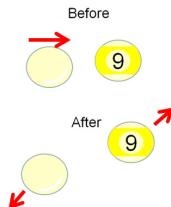
So how do we get different notes on a guitar or Piano?

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\mu}} \quad (7.4)$$

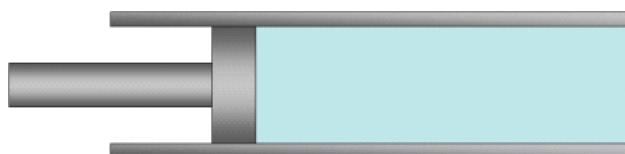
A guitar uses tension to change the frequency or pitch (tuning) and length of string (your fingers pressing on the strings) to change notes. A Piano uses both tension and length of string (and mass per unit length as well!). What do you expect an organ will do?

7.2 Sound waves

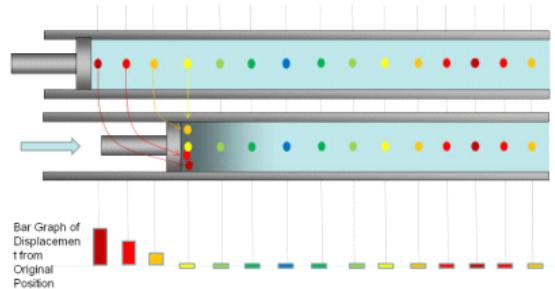
Sound is a wave. The medium is air particles. The transfer of energy is done by collision.



The wave will be a longitudinal wave. Let's see how it forms. We can take a tube with a piston in it.

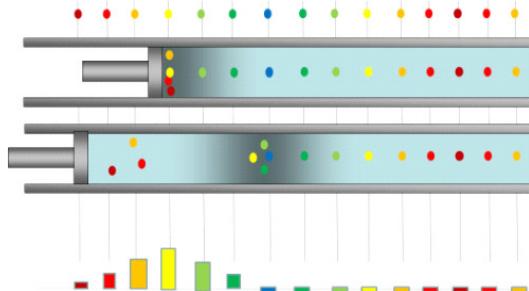


As we exert a force on the piston, the air molecules are compressed into a group. In the next figure, each dot represents a group of air molecules. In the top picture, the air molecules are not displaced. But when the piston moves, the air molecules receive energy by collision. They bunch up. We see this in the second picture.

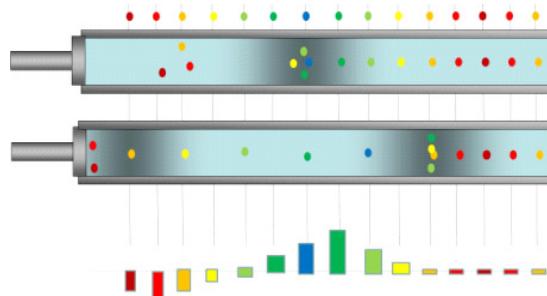


The graph below the two pictures shows how much displacement each molecule group experiences.

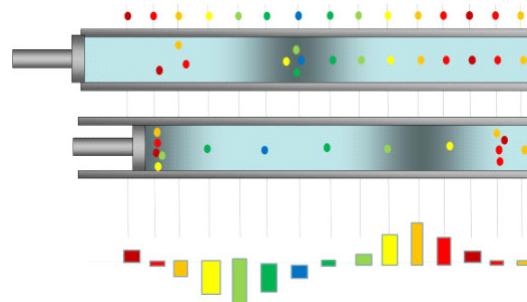
Suppose we now pull the piston back. This would allow the molecules to bounce back to the left, but the molecules that they have collided with will receive some energy and go to the right. This is shown in the next figure. Color coded dots are displayed above the before and after picture so you can see where the molecule groups started.



If we pull the piston back further, the molecules can pass their original positions.



Then we can push inward again and compress the gas.



This may seem like a senseless thing to do, but it is really what a speaker does to produce sound. In particular, a speaker is a harmonic oscillator. The simple harmonic motion of the speaker is the disturbance that makes the sound wave.



7.2.1 Periodic Sound Waves, Pressure

Jollyball Demo

Suppose I have a ball and I ask six people from the class to come up and press on the ball from all directions. This is a new force situation, unlike most we dealt with in Dynamics or PH121. Each person exerts a force on the ball. The person uses the area of their hand to exert the force. The motion of the ball, and even its shape depend on both the force (magnitude and direction) and the area involved in each push.

From our demo, it seems that the force and area of the ball could be related to better describe the situation. Let's look at the ratio

$$\frac{F}{A}$$

What does it represent? This ratio tells us how spread out an applied force may be. The area is important. Think of the sides of a the eraser of a pencil.

It is convenient to give this concept a name. We will call it *pressure*.

Pressure is a scalar value that describes how a force acts over an area

$$P = \frac{F}{A}$$

But in a fluid, what is the force? Lets consider a ball hitting a wall. Is there a force? During the collision, there is a force of the ball on the wall, and a Newton's third law force of the wall molecules pushing back on the ball. The ball will bounce back.



This force is small and only lasts during the collision. But now suppose we have many balls, and all the balls impact the wall. Further suppose that every time a ball bounces back it ends up headed back to the wall and bounces again. If the balls keep coming, there will be a force on the wall quite a bit of the time. At least, on average there is a force, anyway. This is the force that causes air pressure. The air molecules in this room are like the small balls. we will find that they have an amount of kinetic energy. They impact the walls (and us) all over our surface area. The result is air pressure.

The water pressure in a swimming pool is caused by moving water particles. You should convince yourself that the reason the water stays in the pool is partly because the air molecules bounce against the water surface exerting a pressure on the water!

Let's go back to making sounds. Suppose we push our piston as we did before in figure ???. When we push in the piston, it creates a region of higher pressure next to it.

Compression: A local region of higher pressure in a fluid

Rarefaction: A local region of lower pressure in a fluid

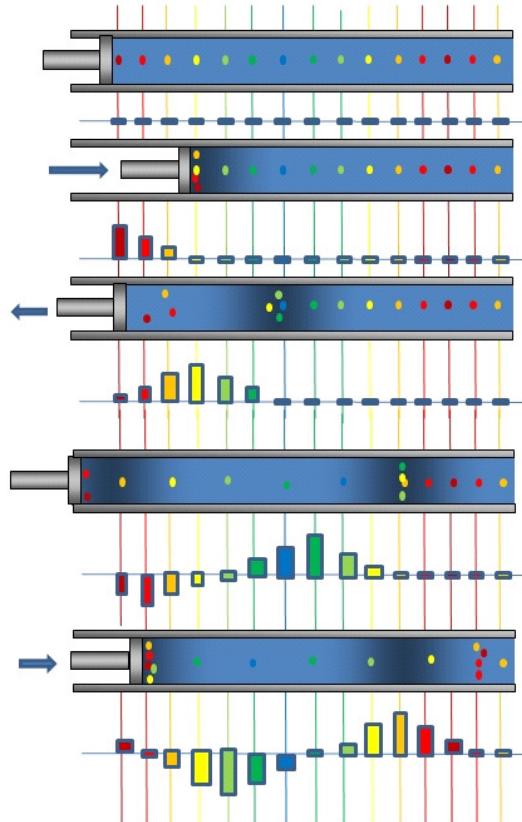
When we pull back the piston the fluid expands to fill the void. We create a rarefaction next to the piston.

Suppose we drive the piston sinusoidally. Can we describe the motion of the particles and of the wave? We can identify the distance between two compressions as λ . We define $s(x, t)$ (like we defined a wave function, $y(x, t)$) as the displacement a particle of fluid relative to its equilibrium position.

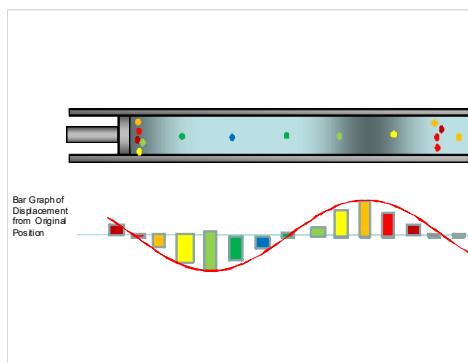
$$s(x, t) = s_{\max} \cos(kx - \omega t) \quad (7.5)$$

but what is s_{\max} ?

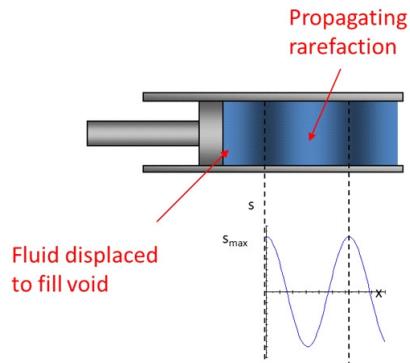
We remember that s_{\max} is the maximum displacement of a particle of fluid from its equilibrium position. We plotted this using a bar graph to show displacement from the equilibrium position for our molecules. As we push the piston in and out we will get something like this.



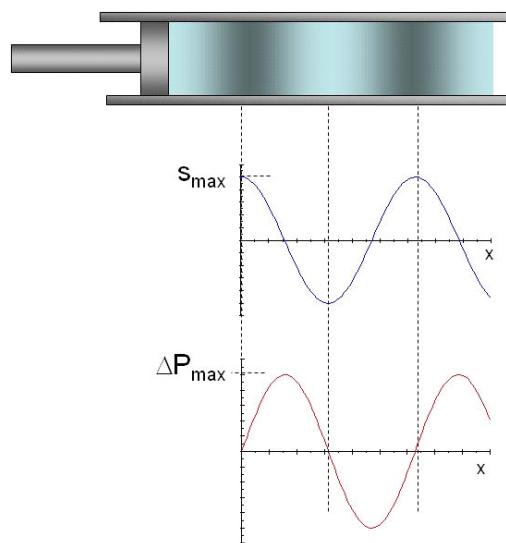
found before that we get something that looks like a sine wave, but remember what the bars represent. They represent the displacement from original position.



We don't usually draw bar graphs, we usually just draw the sine wave.



The variation of the gas pressure ΔP measured from its equilibrium is also periodic



which is why we often refer to a sound wave as a pressure wave. Think of when the wave gets to your ear. the wave consists of a group of particles all headed for your ear drum. When they hit, they exert a force. Pressure is a force spread over an area,

$$P = \frac{F}{A}$$

so in a sense, we hear changes in air pressure!

Chapter 8

8 Sound Wave Speed and Intensity 1.17.2, 1.17.3

Fundamental Concepts

- We don't hear all frequencies equally well
- Waves from point sources are spherical
- Light waves are waves in the electromagnetic field

8.1 Speed of Sound Waves

We found that waves on ropes and strings had a speed that depended on the stretchy bonds in the medium. Our equation for this is

$$v = \sqrt{\frac{T}{\mu}}$$

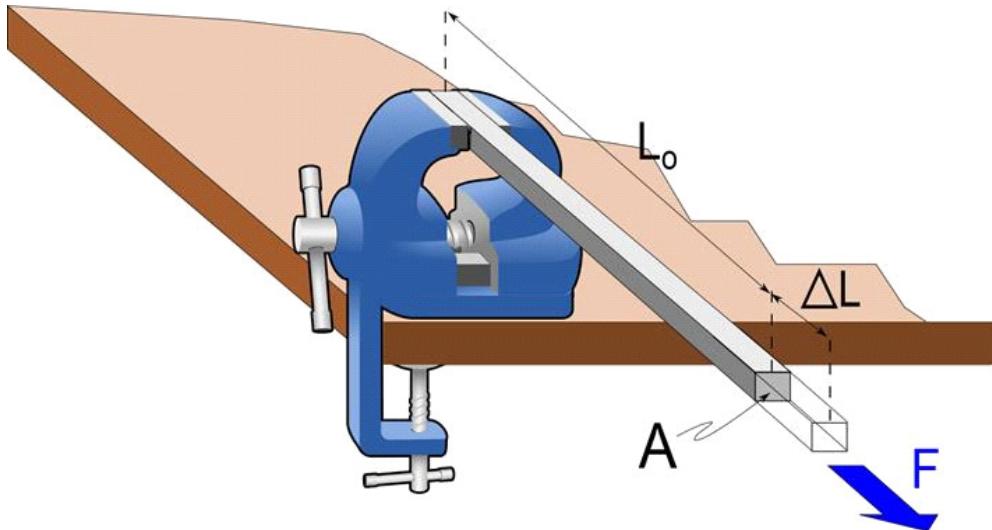
Let's think about this equation. The tension is an elastic property of the string or rope medium. And μ has the mass in it. It is an inertial property of the string or rope. It tells us how hard the string or rope is to move.

We need to have something like this for sound waves. And we know that sound waves have to do with pressure and the compression of the molecules that make up the medium. Long ago in PH121 you should have studied elastic properties of materials. But let's review them here.

8.1.1 Young's Modulus

Remember that a modulus is a constant that describes how a material can be deformed. It tells us how elastic that material is. Think of it like the spring constant k that tells us how hard it is to stretch or compress a spring.

Suppose we pull on a rod with a force \vec{F} .



The rod will be stressed. We call this *tensile stress* which comes from a pull. This pull will stretch the solid's molecular bonds (like a tension force).

We write the stress as

$$\frac{F}{A} \quad (8.1)$$

which looks like a pressure, but this time we mean that we are pulling on the beam so that the pull is spread over the cross sectional area (marked A in the figure above). That is not much like a pressure!

By pulling on the beam, the beam will stretch. That stretching is a strain. We write the strain as the percent change in shape. In this case, the percent change in length

$$\frac{\Delta L}{L_0} \quad (8.2)$$

If the stress is not too great, then we can write a linear equation that relates the stress to the strain.

$$\frac{F}{A} = Y \frac{\Delta L}{L_0} \quad (8.3)$$

this linear equation only works in the elastic region of a stress vs. strain curve for the particular object. The constant of proportionality, Y , is called *Young's*

Modulus. If we write this as

$$F = \frac{YA}{L_o} \Delta L \quad (8.4)$$

$$F = k \Delta L \quad (8.5)$$

it looks very much like Hook's law. Our restoring force or our "pull" is the rod pulling back with a force like $F = -k(x - x_o)$.

Young's modulus depends on the microscopic properties of the solid (which we will not study in detail, but these are our stretchy/squishy forces between atoms that make normal and tension forces). We will look up Young's modulus for each material in tables like the following:

Material ^{1,2}	Young's Modulus (N/m ²)
Aluminum*	7×10^{10}
Titanium*	11×10^{10}
Steel*	20×10^{10}
Carbon Steel*	21×10^{10}
Lead†	1.6×10^{10}
Brass†	10×10^{10}
Concrete†	2.0×10^{10}
Nylon†	0.5×10^{10}
Bone† (arm or leg)	1.5×10^{10}
Pine Wood† (parallel to grain) (perpendicular to grain)	1.0×10^{10} 0.1×10^{10}

Let's take an example. Suppose we have a steel piano wire. One end of the piano wire is fixed in place, and the other is connected to a peg that can be rotated so the piano wire is tightened. This is what you do when you tune a piano, you turn the peg and tighten the wire. Suppose we tighten the wire so that there will be 980 N of tension on the wire. The wire is 1.6 m long and has a diameter of 2 mm. How much will the wire stretch as we tighten the peg? We will need to know that the Young's modulus for steel is 20×10^{10} N/m³.

Here is a summary of what we know

$$\begin{aligned} T &= 980 \text{ N} \\ L_o &= 1.6 \text{ m} \\ d &= 2 \text{ mm} \\ Y &= 200 \times 10^9 \text{ N/m}^2 \end{aligned}$$

Let's start with our stress/strain basic equation

$$\frac{F}{A} = Y \frac{\Delta L}{L_o}$$

¹*Ledbetter, H. M., Physical Properties Data Compilations Relevant to Energy Storage, US National Bureaus of Standards, 1982. Different alloys have different properties, so for any real work see the original tables in the original publication.

²†Average values from numbers given in various text books. These numbers should be taken as example values and more exact numbers found for any real work.

and solve for ΔL

$$\frac{F}{A} \frac{L_o}{Y} = \Delta L$$

the force is our tension, and the cross sectional area will be

$$A = \pi \frac{d^2}{4}$$

$$\frac{T}{\pi \frac{d^2}{4}} \frac{L_o}{Y} = \Delta L$$

$$\frac{4T}{\pi d^2} \frac{L_o}{Y} = \Delta L$$

then

$$\begin{aligned} \Delta L &= \frac{4(980 \text{ N})}{\pi (2 \text{ mm})^2} \frac{(1.6 \text{ m})}{(200 \times 10^9 \text{ N/m}^2)} = \\ &= 2.4955 \times 10^{-3} \text{ m} \end{aligned}$$

so the wire stretched about a quarter of a centimeter.

Now that we remember what a modulus is, let's consider the speed of sound waves in our bar of metal. This is going to be different than our speed of sound in air. In air the molecules have to travel and crash into each other to make a sound wave, but in our bar the molecules are bound together. They can't move. We might guess that this will speed up the sound because there is no travel time for the medium pieces. And we might guess that the springy bonding forces would be what transfers energy in the medium. But what is our speed of sound?

Say, we hit the bar with a hammer. Sound waves would be created in the bar. But suppose we make our bar thin and stretch it across a guitar. Now as we hit the string we expect a wave. We need a measure of the elasticity of the metal and a measure of the inertial of the metal material. The tension in the (now) wire would be our force, but we now know how to express this force in terms of the elasticity of the wire material.

$$\begin{aligned} v &= \sqrt{\frac{T}{\mu}} = \sqrt{\frac{Y \frac{\Delta L}{L_o} A}{\frac{m}{L_o}}} \\ &= \sqrt{\frac{Y \Delta L A}{m}} \\ &= \sqrt{\frac{Y V}{m}} \\ &= \sqrt{\frac{Y}{\rho}} \end{aligned}$$

And we have the wave speed in terms of the an elastic quantity (Young's modulus) and an inertial property (the metal density). This gives a wave speed of

$$v = \sqrt{\frac{Y}{\rho}}$$

But this won't do for sound waves in air. We need a different modulus for gasses.

8.1.2 Bulk Modulus

This one is harder. It tells us how much a solid can be squashed or, compressed. Let's define

$$\Delta P = \frac{\Delta F}{A} \quad (8.6)$$

where now A is the outer surface area of the solid. If we were dealing with a fluid, this would be pressure. Often fluids are very hard to compress (we assumed they were not at all compressible). Many solids are easy to compress. Think of Styrofoam or any foam rubber. For solids $\Delta F/A$ where the force is all around the solid and the area is the surface area is a stress. The strain is the percentage change in volume (change amount over the original volume³)

$$\frac{\Delta V}{V} \quad (8.7)$$

and the relationship is

$$\Delta P = -B \frac{\Delta V}{V} \quad (8.8)$$

B is the *bulk modulus*, which again, we look up in tables.

Material ⁴	Bulk Modulus (N/m ²)
Steel [‡]	140×10^9
Cast Iron [‡]	90×10^9
Brass [‡]	80×10^9
Aluminum [‡]	70×10^9
Water [‡]	2×10^9
Ethyl Alcohol [‡]	1×10^9
Mercury [‡]	2.5×10^9
Air [‡]	1.01×10^5

The minus sign may be surprising. But in our definition of bulk modulus, we are thinking of compression. So usually ΔV is negative. The minus sign means that for compression this formula gives positive values. Also notice that there is a ΔP or ΔF in our formula. Usually we start compressing a solid while it

³I am using a V with a line through it for volume so it doesn't look like velocity, v , or voltage, V .

⁴‡Average values from numbers given in various text books. These numbers should be taken as example values and more exact numbers found for any real work.

is experiencing air pressure. So we don't start from zero stress. We change the force from the force due to air pressure to some larger compressive force.

Once again we have a quantity that shows how elastic our medium is, B . And we can still use the density for the inertial property of the gas, ρ . So our speed of sound in a gas is

$$v = \sqrt{\frac{B}{\rho}}$$

But how do we find B ? That is something we will do much later in this course. So for now let's borrow a result from our future. For an ideal gas

$$v = \sqrt{\frac{\gamma RT_K}{M}}$$

where γ is the adiabatic index (related to the type of gas we have) and R is the universal gas constant, and T_K is the temperature in kelvins and M is the gas molar mass. For now we can just use this approximation.

This might not be exactly clear at this point. But what is important is that the speed changes with temperature. The hotter the gas, the faster the waves. If we assume the air is like an ideal gas, this reduces to just

$$v = v_o \sqrt{1 + \frac{T_c}{T_o}} \quad (8.9)$$

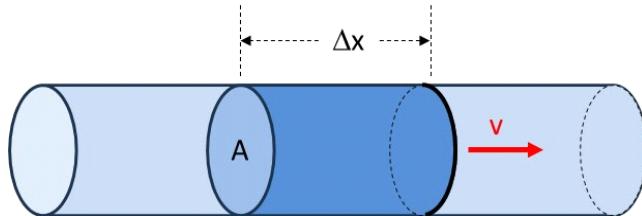
where $v_o = 331 \frac{\text{m}}{\text{s}}$ and $T_o = 273 \text{ K}$ (0°C).⁵

The speed of sound in air at normal room temperature is around 340 m/s.

8.1.3 Finding the speed of sound in air @@@@ Don't do this here! @@@@

Now that we know about adiabatic processes, let's go back to the beginning of our course and think about the speed of sound in air. The disturbance that makes the wave will change the pressure and compress the gas a bit. So work will be done on the gas. But the disturbance happens quickly and we wouldn't expect energy transfer by heat. This is an adiabatic process!

We used pipes when we studied sound waves. So, let's consider our air in a pipe.



⁵ $v = v_o \sqrt{1 + \frac{T_C}{T_o}} = v_o \sqrt{\frac{T_o}{T_o} + \frac{T_C}{T_o}} = v_o \sqrt{\frac{T_o + T_C}{T_o}} = v_o \sqrt{\frac{T_K}{T_o}}$

The pipe has a cross sectional area, A . And let's look at a small volume of air in the pipe. The darker region is our small volume of air and let's say it has a mass Δm where here the delta means "a small amount of" and not a difference. The density of the air is given by

$$\rho = \frac{\Delta m}{\Delta V}$$

so we can write

$$\Delta m = \rho \Delta V$$

But look at our little mass' volume

$$V = A \Delta x$$

so we can write Δm as

$$\Delta m = \rho A \Delta x$$

Now let's consider the amount of mass that passes a point in our pipe. That is our mass flow rate and we know that is

$$\frac{dm}{dt}$$

If we make our small amount of mass very small, we would have

$$dm = \rho A dx$$

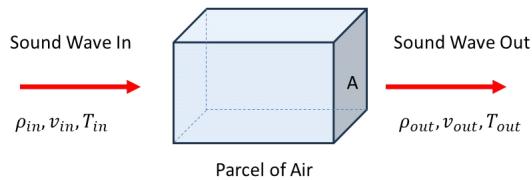
and our mass flow rate would be

$$\frac{dm}{dt} = \frac{\rho A dx}{dt} = \rho A v$$

This looks familiar. It is one side of our equation of continuity

$$\rho_{in} A_{in} v_{in} = \rho_{out} A_{out} v_{out}$$

Now consider a "parcel of air." This is just a part of the air that we will pay attention to. It is surrounded by more air and is only different because we are thinking about it. But concert such a parcel of air and let's make it sort of box shaped like in the next figure.



We could write $\rho_{in} = \rho$ and $v_{in} = v$ and $T_{in} = T$ and we could write $\rho_{out} = \rho + d\rho$, $v_{out} = v + dv$, and $T_{out} = T + dT$. Let's put these in our continuity equation.

$$\rho Av = (\rho + d\rho) A (v + dv)$$

The area, A , didn't change. So it cancels, and we can do some algebra

$$\begin{aligned}\rho v &= (\rho + d\rho)(v + dv) \\ \rho v &= \rho v + \rho dv + vd\rho + d\rho dv\end{aligned}$$

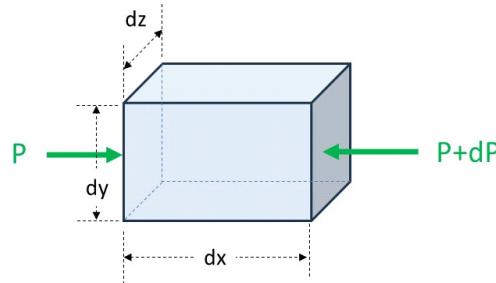
We can assume that $d\rho$ will be small if our parcel of air isn't too big. And likewise dv will be small. So $d\rho dv$ should be very small. Let's drop this term.

$$\rho v = \rho v + \rho dv + vd\rho$$

and the ρv terms cancel

$$\begin{aligned}0 &= \rho dv + vd\rho \\ \rho dv &= -vd\rho\end{aligned}\tag{8.10}$$

This doesn't seem to have helped, but we will use it later. Let's go back to our parcel of air and this time think of the forces on the input and output sides.



We will have an input pressure and an output pressure with $P_{in} = P$ and $P_{out} = P + dP$

We remember that $F = PA$ so we will have a force on the input side

$$F_{in} = P_{in}A = PA = Pdydz$$

and an output force

$$F_{out} = -P_{out}A = -(P + dP) dydz$$

The net force would be

$$\begin{aligned}F_{net} &= Pdydz - Pdydz - dPdydz \\ &= -dPdydz\end{aligned}$$

Then

$$ma = -dPdydz$$

We can divide by the mass to get the acceleration

$$a = -\frac{dPdydz}{m}$$

and now think,

$$\rho = \frac{m}{V} = \frac{m}{dxdydz}$$

so

$$m = \rho dxdydz$$

so our acceleration is

$$\begin{aligned} a &= -\frac{dPdydz}{\rho dxdydz} \\ &= -\frac{dP}{\rho dx} \end{aligned}$$

and $a = dv/dt$

$$\begin{aligned} \frac{dv}{dt} &= -\frac{dP}{\rho dx} \\ dv &= -\frac{dPdt}{\rho dx} \end{aligned}$$

and recall that $v = dx/dt$

$$dv = -\frac{dP}{\rho v}$$

or

$$dv\rho v = -dP$$

Finely we can use our result from our continuity equation (8.10) $\rho dv = -vd\rho$

$$\begin{aligned} v\rho dv &= -dP \\ v(-vd\rho) &= -dP \\ v(-vd\rho) &= -dP \\ v^2 &= \frac{dP}{d\rho} \end{aligned}$$

and we are close. the wave speed, v , is the square root of the pressure change over the density change.

$$v = \sqrt{\frac{dP}{d\rho}} \quad (8.11)$$

This is a little like $\sqrt{B/\rho}$. So we are close.

Since we are modeling this as an adiabatic process we know

$$PV^\gamma = \text{constant}$$

Let's look at that volume, V .

$$V = \frac{m}{\rho}$$

and we can write the mass, m , in terms of the molar mass

$$M = \frac{m}{n}$$

or

$$m = nM$$

so our volume is

$$V = \frac{nM}{\rho}$$

and let's put this into our adiabatic process equation

$$PV^\gamma = P \left(\frac{nM}{\rho} \right)^\gamma = \text{constant}$$

we could write this as

$$P \frac{n^\gamma M^\gamma}{\rho^\gamma} = \text{constant}$$

and rearrange it to be

$$P \frac{1}{\rho^\gamma} = \frac{\text{constant}}{n^\gamma M^\gamma}$$

and since n and M are not changing

$$P \frac{1}{\rho^\gamma} = \text{new constant}$$

Let's take the logarithm of both sides

$$\begin{aligned} \ln \left[P \frac{1}{\rho^\gamma} \right] &= \ln [\text{new constant}] \\ \ln P - \gamma \ln \rho &= \ln [\text{new constant}] \\ \ln P - \gamma \ln \rho &= \text{another new constant} \end{aligned}$$

Now let's take the derivative of both sides

$$\frac{d}{d\rho} (\ln P - \gamma \ln \rho) = \frac{d}{d\rho} (\text{another new constant})$$

The right hand side is just zero (which is why we didn't spend time figuring out the number for the constant!).

$$\frac{d}{d\rho} (\ln P - \gamma \ln \rho) = 0$$

But the right hand side gives

$$\frac{1}{P} \frac{dP}{d\rho} - \frac{\gamma}{\rho} \frac{d\rho}{d\rho} = 0$$

$$\begin{aligned}\frac{1}{P} \frac{dP}{d\rho} &= \frac{\gamma}{\rho} \\ \frac{dP}{d\rho} &= \frac{\gamma}{\rho} P\end{aligned}\quad (8.12)$$

Now let's pretend that the air is an ideal gas. Then we know that

$$\begin{aligned}PV &= nRT \\ &= \frac{m}{M} RT\end{aligned}$$

and we can solve for P and put it into equation (8.12)

$$\begin{aligned}P &= \frac{m}{VM} RT \\ &= \frac{\rho}{M} RT\end{aligned}$$

and then equation (8.12) becomes

$$\frac{dP}{d\rho} = \frac{\gamma}{\rho} \frac{\rho}{M} RT = \frac{\gamma}{M} RT \quad (8.13)$$

and we can put this directly into equation (8.11)

$$v = \sqrt{\frac{dP}{d\rho}} = \sqrt{\frac{\gamma}{M} RT}$$

We can get our approximation that we used earlier for temperature dependence for sound speed in air from putting in some numbers. First, let's define $T_o = 273\text{ K}$. Then

$$\begin{aligned}v &= \sqrt{\frac{\gamma}{M} RT \frac{T_o}{T_o}} \\ &= \sqrt{\frac{\gamma RT_o}{M}} \sqrt{\frac{T}{T_o}}\end{aligned}$$

It turns out that for air at normal air pressure and at $T = 10^\circ\text{C}$, $\gamma = 1.4$ and $M = 0.02897 \frac{\text{kg}}{\text{mol}}$, and of course $R = 8.31 \frac{\text{J}}{\text{mol K}}$ then our speed is

$$\begin{aligned}v &= \sqrt{\frac{(1.4)(8.31 \frac{\text{J}}{\text{mol K}})(273\text{ K})}{0.02897 \frac{\text{kg}}{\text{mol}}}} \sqrt{\frac{T}{T_o}} \\ &= 331.1 \frac{\text{m}}{\text{s}} \sqrt{\frac{T}{T_o}} \\ &= 331.1 \frac{\text{m}}{\text{s}} \sqrt{\frac{T}{273\text{ K}}}\end{aligned}$$

where T is in kelvins. We can change this to having T in Celsius if we

$$\begin{aligned} v &= 331.1 \frac{\text{m}}{\text{s}} \sqrt{\frac{T_o + T_c}{T_o}} \\ &= 331.1 \frac{\text{m}}{\text{s}} \sqrt{1 + \frac{T_c}{T_o}} \end{aligned}$$

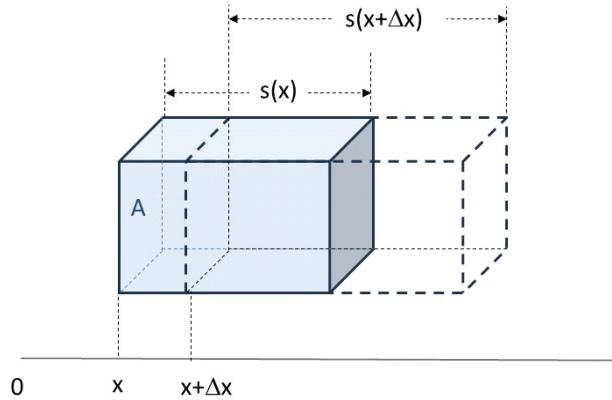
which is what we used before.

8.2 Intensity of Sound Waves @@ Maybe do this

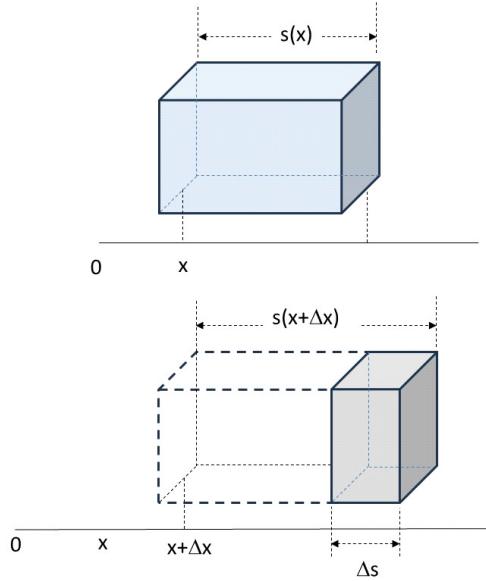
We have already learned that knowing the power of a sound source isn't enough to tell us how loud that sound source will appear to us. We know the sound gets spread out over a big spherical area. And we use the quantity called intensity to describe how spread out the power is at a particular distance.

$$\mathcal{I} = \frac{\mathcal{P}}{4\pi r^2} \quad (8.14)$$

Let's consider a parcel of air. The air might both move and stretch as the sound wave passes. So the displacement from equilibrium $s(x)$ might be different as the air moves to location $x + \Delta x$. The next figure tries to show this.



Let's look at the stretch of the parcel of air for a moment. Let's take the before size of the parcel of air and compare it to the stretched parcel size.



We can see that the size difference for the displacement from equilibrium is

$$\Delta s = s(x + \Delta x) - s(x)$$

The change in volume would be

$$\Delta V = A\Delta s$$

The fractional change in volume is defined as

$$\frac{\Delta V}{V}$$

and we can see that this will be

$$\frac{\Delta V}{V} = \frac{A\Delta s}{A\Delta x} = \frac{\Delta s}{\Delta x}$$

If we take the limit as \$\Delta x \rightarrow 0\$ this is just

$$\frac{dV}{V} = \frac{ds}{dx}$$

Now let's go back to our bulk modulus, \$B\$. A form for the bulk modulus is

$$B = -\frac{\Delta P}{dV/V}$$

We can solve this for \$\Delta P\$

$$\Delta P = -B \frac{dV}{V} = -B \frac{ds}{dx}$$

For a sinusoidal wave,

$$s(x, t) = s_{\max} \sin(kx - \omega t + \phi_o) \quad (8.15)$$

(where ϕ_o could be $\pi/2$ to make this really a cosine). and we can take a derivative of this

$$\frac{ds}{dx} = -ks_{\max} \cos(kx - \omega t + \phi_o)$$

then

$$\Delta P = -B \frac{ds}{dx} = Bks_{\max} \cos(kx - \omega t + \phi_o)$$

We want the intensity and the intensity is

$$\mathcal{I} = \frac{\mathcal{P}}{A} \quad (8.16)$$

We could write the power as the force multiplied by the velocity

$$\mathcal{I} = \frac{Fv}{A} = Pv \quad (8.17)$$

We know something about the pressure, we need the speed. But this is the speed of parts of the medium. And we know how to find that for waves.

$$v(x, t) = \frac{d}{dt} s(x, t) = -\omega s_{\max} \cos(kx - \omega t + \phi_o)$$

We can put this all together to get the intensity

$$\mathcal{I} = \Delta Pv \quad (8.18)$$

$$= (Bks_{\max} \cos(kx - \omega t + \phi_o)) (-\omega s_{\max} \cos(kx - \omega t + \phi_o)) \quad (8.19)$$

$$= -Bk\omega s_{\max}^2 \cos^2(kx - \omega t + \phi_o) \quad (8.20)$$

The intensity is proportional to the amplitude of the displacement, s_{\max} squared.

We often can only measure a time average intensity. To average our intensity function over one period we would integrate

$$\begin{aligned} \langle \mathcal{I} \rangle &= \int_0^T -Bk\omega s_{\max}^2 \cos^2(kx - \omega t + \phi_o) dt \\ &= -Bk\omega s_{\max}^2 \int_0^T \cos^2(kx - \omega t + \phi_o) dt \\ &= -Bk\omega s_{\max}^2 \frac{1}{2} \end{aligned}$$

With a little more math we can write this in terms of $\Delta P_{\max} = Bks_{\max}$

$$\begin{aligned} \langle \mathcal{I} \rangle &= -\frac{Bk\omega s_{\max}^2}{2} = -\frac{\omega}{2} (Bks_{\max}) (s_{\max}) \\ &= -\frac{\omega}{2} (Bks_{\max}) \left(\frac{Bks_{\max}}{Bk} \right) \\ &= -\frac{\omega}{2} (\Delta P_{\max}) \left(\frac{\Delta P_{\max}}{Bk} \right) \\ &= -\frac{\omega}{2Bk} (\Delta P_{\max})^2 \end{aligned}$$

and we know for sound the wave speed is given by

$$v = \frac{\omega}{k} = \sqrt{\frac{B}{\rho}}$$

so

$$B = \frac{\omega^2}{k^2} \rho$$

$$\langle I \rangle = -\frac{\omega}{2 \left(\frac{\omega^2}{k^2} \rho \right) k} (\Delta P_{\max})^2$$

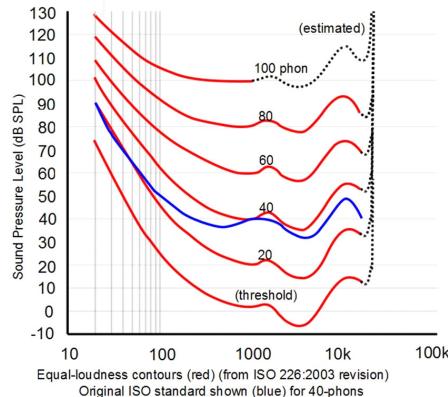
$$\langle I \rangle = -\frac{1}{2(v\rho)} (\Delta P_{\max})^2$$

This gives us the average intensity in the pressure picture for sound waves and once again the intensity is proportional to the amplitude squared.

8.3 Sound Levels in Decibels

The intensity of a sound wave gives us how much power per unit area is being delivered by the wave. This is a great way to measure the power delivered by a sound wave because most instruments have a detection area. So power per unit area is a natural measurement for instruments.

Frequency Range
Demo



Robinson-Dadson equal loudness curves (Image in the Public Domain courtesy Lindosland)

but our ears are not designed to be quantitative scientific instruments (though they are truly amazing in their range and ability). Sounds with the same intensity at different frequencies do not appear to us to have the same loudness.

The frequency response graph above shows how this relationship works for test subjects⁶.

Our Design Engineer made an interesting choice in building us. We need to hear very faint sounds, and very loud sounds too. In order to make us able to hear the soft sounds without causing extreme discomfort when we hear the loud, He gave up linearity. That is, we don't hear twice the sound intensity as twice as loud.

The mathematical expression that matches our perception of loudness to the intensity is

$$\beta = 10 \log_{10} \left(\frac{I}{I_o} \right) \quad (8.21)$$

Sound
Demo

Meter where the quantity I_o is a reference intensity.

We call β the *sound level*. The quantity I_o we choose to be the *threshold of hearing*,

$$I_o = 1 \times 10^{-12} \frac{\text{W}}{\text{m}^2}$$

the intensity that is just barely audible for humans (it will be different for electronic devices or other animals). Measured this way, we say that intensity is in units of decibels (dB). The decibel, is useful because it can describe a non-linear response in a linear way that is easy to match to our human experience. But it is tricky because, like radians, it is dimensionless because we are comparing two intensities (I/I_o). We expect the same strange unit behavior with dB that we see with radians.

Suppose we double the intensity by a factor of 2.

$$\begin{aligned} \beta &= 10 \log_{10} \left(\frac{2I_o}{I_o} \right) \\ &= 10 \log_{10} 2 \\ &= 3.0103 \text{dB} \end{aligned}$$

The sound intensity level is not twice as large, but only 3dB larger. It is a tiny increase. This is what we hear. A good rule to remember is that 3dB corresponds to a doubling of the intensity.

The tables that follow give some common sounds in units of dB and W/m^2 . Just for reference, I have measured a Guns n Roses concert at 120 dB outside the stadium.

⁶Ones that have not gone to Guns n Roses concerts

Sound	Sound Level (dB)
Jet Airplane at 30m	140
Rock Concert	120
Siren at 30m	100
Car interior when Traveling 60mi/hr	90
Street Traffic	70
Talk at 30cm	65
Whisper	20
Rustle of Leaves	10
Quietest thing we can hear (Io)	0

Chapter 9

9 Sound Standing Waves

1.17.4, 1.17.5

Fundamental Concepts

- Sound Standing waves (music)

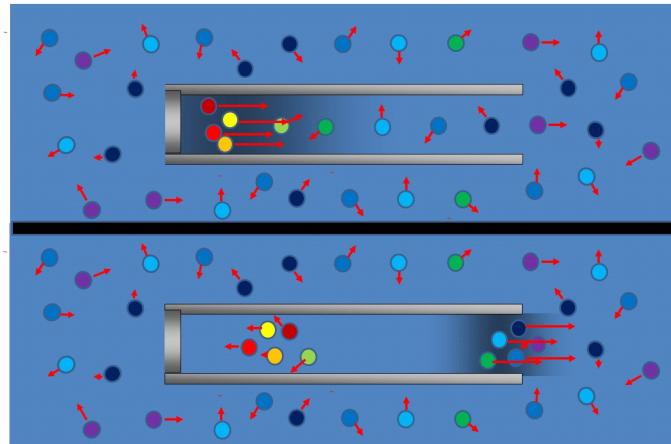
9.1 Sound Standing waves.

We have spent some time studying standing waves on strings. But we also have studied sound waves, could we make sound standing waves?

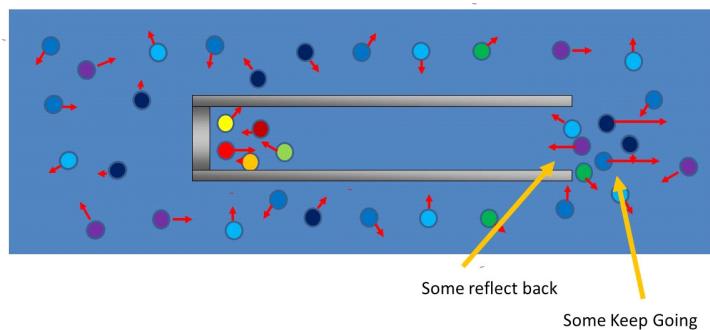
For our attempt, let's take a pipe as shown in the next figures. The pipe will control the direction of the sound wave. The sound wave will (mostly) go down the pipe. But will there be a reflection at the end of the pipe like there was for the string?

It turns out that there will be a reflection. Let's consider our sound wave as a change in pressure to see why. For the string wave, we had a reflected pulse because the person (or guitar bridge, or whatever is holding the string end) exerted a force on the string. If a sound wave travels down our pipe, the pressure will change as the wave goes down the pipe. Remember that pressure is a force spread over an area. When the sound wave gets to the end of the pipe, the rest of the room's air is waiting. And that air mass pushes back on the air molecules that are trying to leave the pipe. The air mass of the room won't change its pressure much. That resistance to pressure change (the force due to the rest of the room air molecules colliding with our wave molecules, pushing them back) will send our tube molecules back down the tube. And that starts a reflection. Think about this for a moment. We need to picture more than we have before. The room is full of air, and it turns out that the air molecules aren't sitting still. So when we make our wave start but hitting molecules with

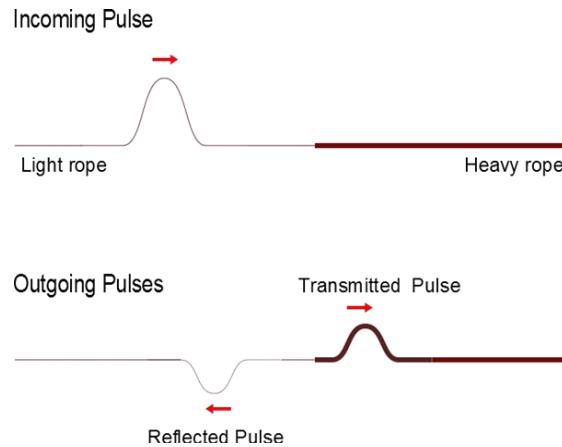
a piston, the molecules at the end of the tube will hit room air molecules that are already moving.



Think of pool balls. If the cue ball and another ball, say, the 6 ball, collide, usually the 6 ball is stationary at the beginning. But if it is not, that will change the outcome of our conservation of momentum problem. if the 6 ball is initially moving toward the cue ball, the cue ball will bounce back the way it came. That will happen with some of the molecules at the end of the pipe. But some will keep going.

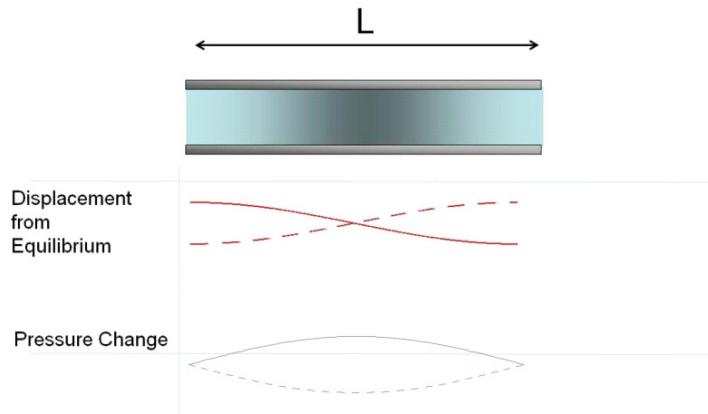


We will split the energy from the wave into two waves, much like we did with waves on strings when we came to an interface.



Now that we have a reflected wave in the tube, we can end up with two waves so long as the piston keeps working at making new waves. And with two waves in the same medium, we have the possibility of having standing waves!

If we have a pipe open at both ends, we can see that air molecules are free to move in and out of the ends of the pipes. If the air molecules can move, the ends must not be nodes. In our string case, the part of the string that experienced destructive interference and did not move was called a node and we always had a node on the end of the string. But the pipe is different than the string case! We can see that air molecules will move out and then bounce back in at the end of the pipe. Still, we expect that there must be a node somewhere. We can reasonably guess that there will be a node in the middle of the pipe. Of course, the pressure on both ends must be atmospheric pressure. So, remembering that pressure and displacement are 90° out of phase for sound waves, we can guess that there are pressure nodes on both ends. But there is a displacement anti-node at the ends of the pipe.



Then for the first harmonic we can draw a displacement node in the middle and

we see that

$$\lambda_1 = 2L \quad (9.1)$$

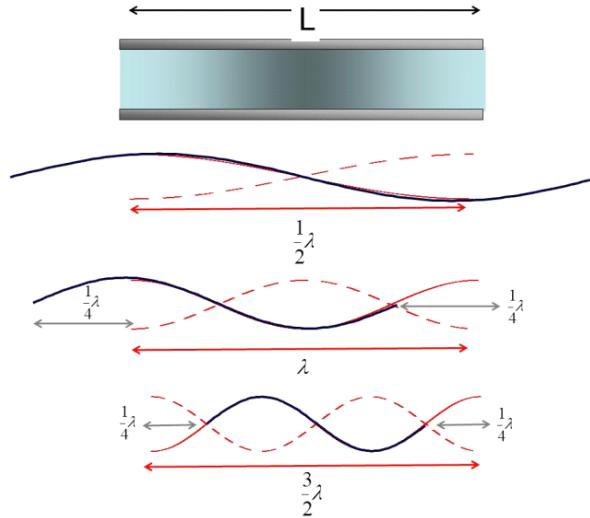
and

$$f_1 = \frac{v}{2L} \quad (9.2)$$

The next mode fits a whole wavelength

$$\lambda_2 = L \quad (9.3)$$

$$f_2 = \frac{v}{L} \quad (9.4)$$



but the next mode fits a wavelength and a half

$$\lambda_3 = \frac{3}{2}L \quad (9.5)$$

$$f_3 = \frac{2v}{3L} \quad (9.6)$$

If we keep going

$$\lambda_n = \frac{2}{n}L \quad (9.7)$$

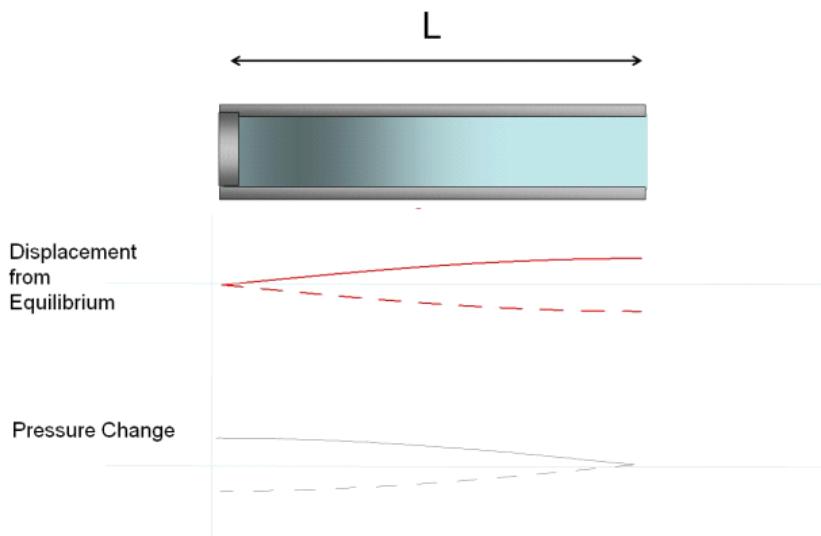
$$f_n = n \frac{v}{2L} \quad n = 1, 2, 3, 4 \dots \quad (9.8)$$

This is the same mathematical form that we achieved for a standing wave on a string!

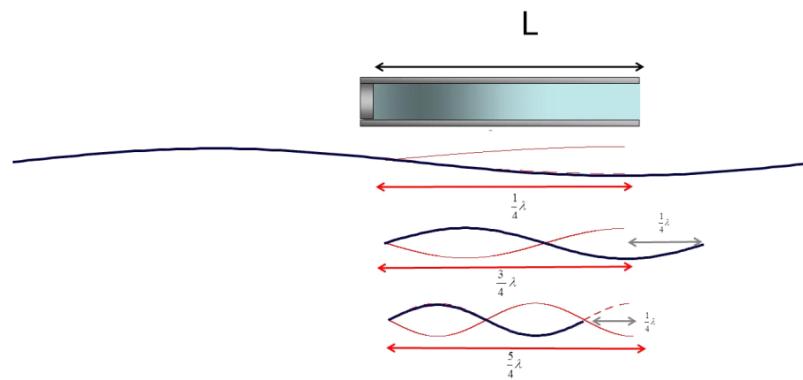
9.2 Pipes closed on one end

But what happens if we put a cap on one end of the pipe? The air molecules cannot move longitudinally once they hit the cap. And where the molecules don't move, that is a node. So our capped end must be a displacement node. Question 123.24.2

The open end is still a displacement antinode. Our standing wave will have a node on the capped side, and an anti-node on the open side. It might look like this



In the next figure we draw the first few harmonics for the “closed on one end” case.



The first harmonic for the closed pipe are found by using

$$v = \lambda f \quad (9.9)$$

$$f = \frac{v}{\lambda} \quad (9.10)$$

We know the speed of sound, so we have v . Knowing that the first harmonic has a node at one end and an anti node at the other end gives us the wavelength. If the pipe is L in length, then L must be

$$L = \frac{1}{4}\lambda_1 \quad (9.11)$$

or

$$\lambda_1 = 4L \quad (9.12)$$

then

$$f_1 = \frac{v}{\lambda_1} = \frac{v}{4L} \quad (9.13)$$

The next configuration that will have a node on one end and an antinode on the other will have

$$L = \frac{3}{4}\lambda_2 \quad (9.14)$$

which gives

$$\lambda_2 = \frac{4}{3}L \quad (9.15)$$

and

$$f_2 = \frac{v}{\lambda_2} = \frac{3v}{4L} \quad (9.16)$$

If we continued, we would find

$$\lambda_n = \frac{4}{n}L \quad (9.17)$$

and

$$f_n = n \frac{v}{4L} \quad n = 1, 3, 5 \dots \quad (9.18)$$

Example: organ pipe



Organ Pipe Demo

The organ pipe shown is closed at one end so we expect

$$f_n = n \frac{v}{4L} \quad n = 1, 3, 5 \dots \quad (9.19)$$

Measuring the pipe, and assuming about 20°C for the room temperature we have

$$\begin{aligned} L &= 0.41 \text{ m} \\ R &= 0.06 \text{ m} \\ v &= 343 \frac{\text{m}}{\text{s}} \end{aligned} \quad (9.20)$$

There is a detail we have ignored in our analysis, the width of the pipe matters a little. I will include a fudge factor to account for this. With the fudge factor, the wavelength is

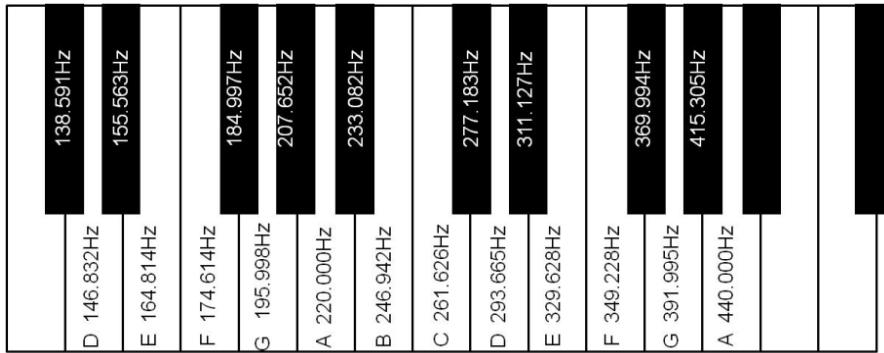
$$\begin{aligned} \lambda_1 &= 4(L + 0.6R) \\ &= 178.4 \text{ cm} \end{aligned}$$

then our fundamental frequency is

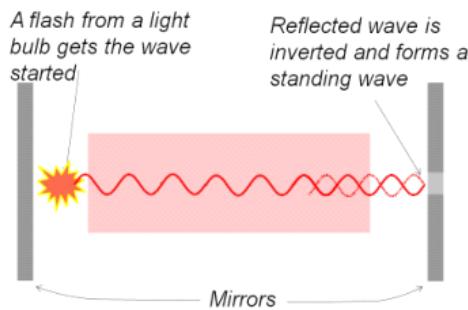
$$f_1 = \frac{v}{\lambda_1} \quad (9.21)$$

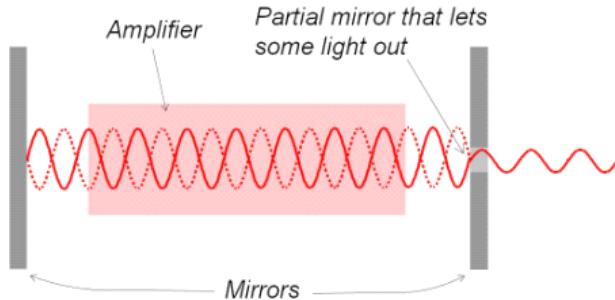
$$= 192.26 \quad (9.22)$$

We can identify this note, and compare to a standard, like a tuning fork or a piano to verify our prediction.



Lasers and standing waves



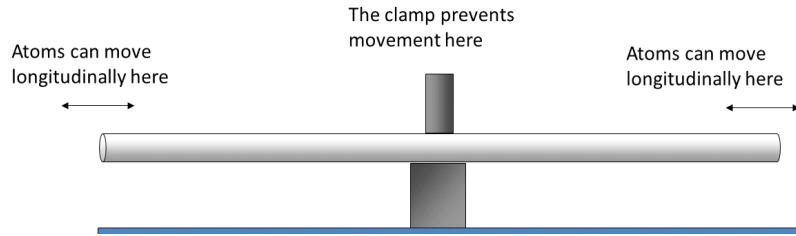


Question 123.24.3

Question 123.24.4

9.3 Standing Waves in Rods and Membranes

We have hinted all chapter that the analysis techniques we were building apply to structures. We need more math and computational tools to analyze complex structures like bridges and buildings, but we can tackle a simple structure like a rod that is clamped. The atoms in the rod can vibrate longitudinally

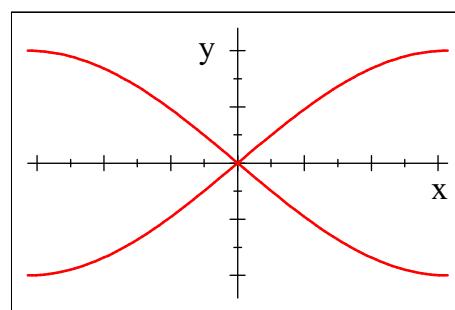


Since we have motion possible on both ends and not in the middle, we surmise that this system will have similar solutions as did the open ended pipe.

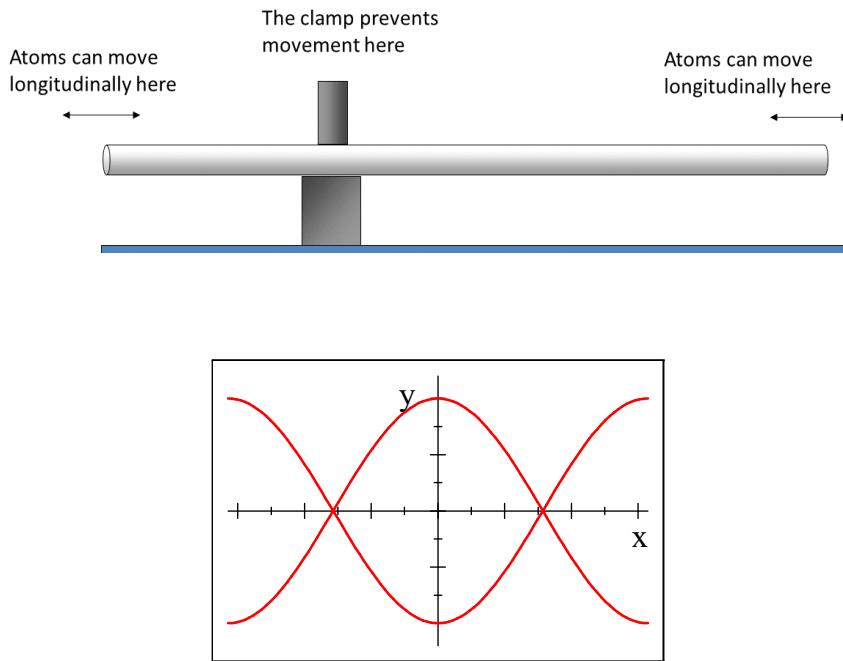
$$f_n = n \frac{v}{2L}$$

The fundamental looks like

$$\sin(x)$$



But suppose we move the clamp. The clamp forces a node where it is placed. If we place the clamp at $L/4$



We can perform a similar analysis for a drum head, but it is much more complicated. The modes are not points, but lines or curves, and the frequencies of oscillation are not integer multiples of each other. See for example <http://physics.usask.ca/~hirose/ep225/animation/drum/anim-drum.htm>.

Of course structures can also waggle on the ends. the ends can rotate counter to each other, etc. These are more complex modes than the longitudinal modes we have considered.

Chapter 10

10 Beats and Doppler Shift 1.17.6, 1.17.7

Fundamental Concepts

- If we mix waves of different frequencies we get a periodic change in loudness
- If we move the wave source or detector the relative motion causes a change in detected frequency

10.1 Beats

Beat Demo

Up till now we only superposed waves that had the same frequency. But what happens if we take waves with different frequencies? Let's take the case where $\phi_2 = \phi_1 = \phi_o$, $k_2 = k_1 = k$ and let's let the waves mix at the same location $x_2 = x_1$ so

$$y_1 = y_{\max} \sin(kx - \omega_1 t + \phi_o)$$

$$y_2 = y_{\max} \sin(kx - \omega_2 t + \phi_o)$$

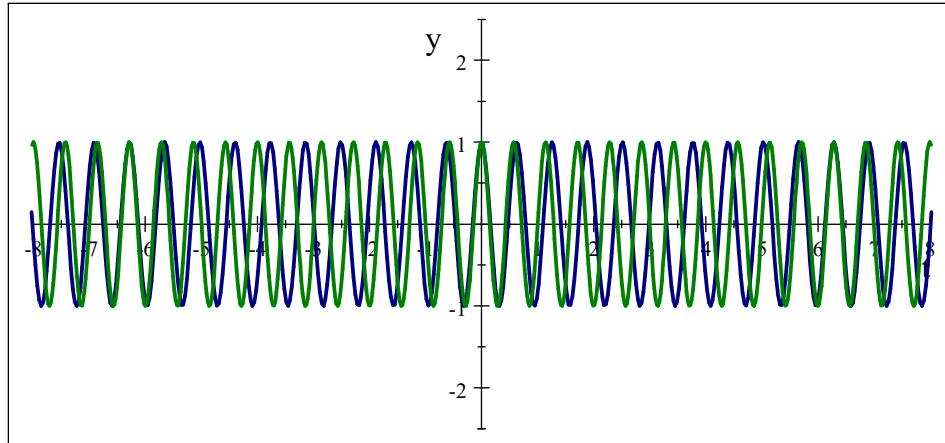
$$\begin{aligned}\Delta\phi &= (kx - \omega_1 t + \phi_o) - (kx - \omega_2 t + \phi_o) \\ &= (\omega_1 - \omega_2) t\end{aligned}$$

We can use our criteria for constructive and destructive interference, but before going on let's put this back into the total amplitude function for two wave mixing.

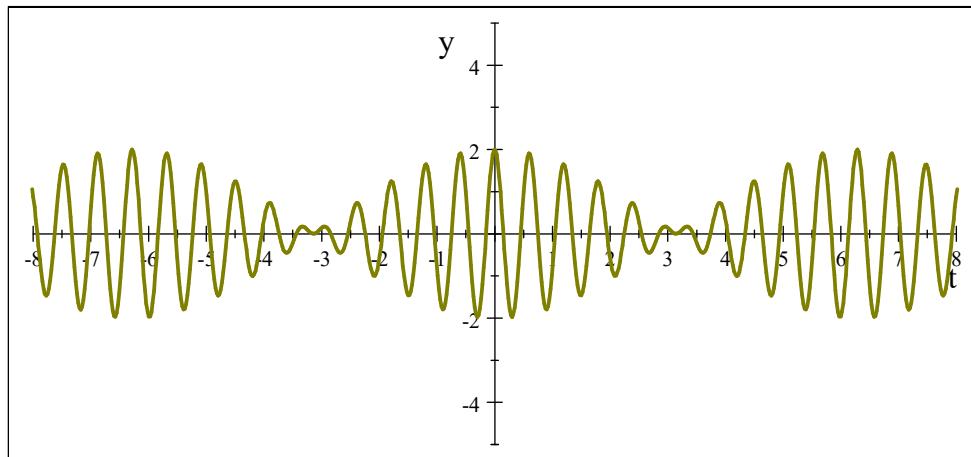
$$\begin{aligned}A &= 2y_{\max} \cos\left(\frac{1}{2}\Delta\phi\right) \\ &= 2y_{\max} \cos\left(\frac{1}{2}(\omega_1 - \omega_2)t\right)\end{aligned}$$

and we can see that our amplitude will change in time. Sometimes it will be constructive interference and sometimes destructive interference and often in between.

We can plot both waves on the same graph and see that this will happen.



Notice that there are places where the waves are in phase, and places where they are not. The superposition looks like this



where there is constructive interference, the resulting wave amplitude is large, where there is destructive interference, the resulting amplitude is zero. We get a traveling wave who's amplitude varies. We can find the amplitude function algebraically.

We can write out the entire resultant wave in our usual way. Our two waves are

$$y_1 = y_{\max} \sin(kx - \omega_1 t + \phi_o)$$

$$y_2 = y_{\max} \sin(kx - \omega_2 t + \phi_o)$$

and the resultant

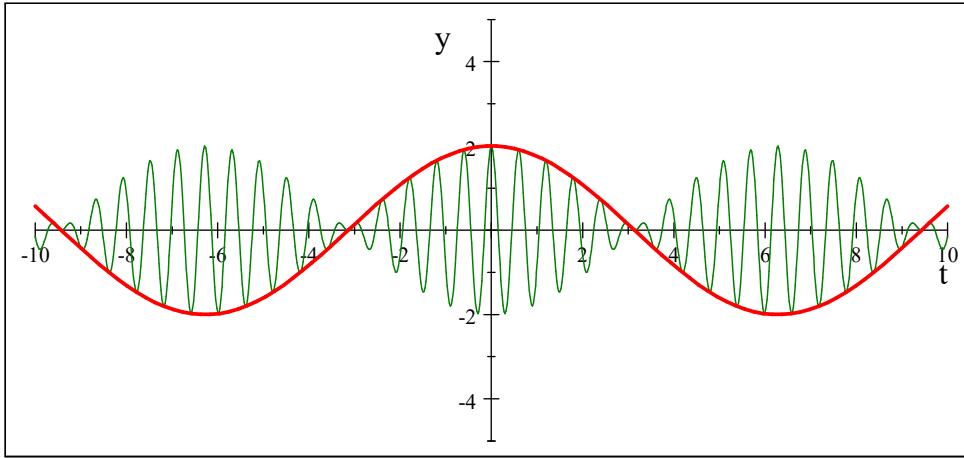
$$\begin{aligned}
 y_r &= 2y_{\max} \cos \left(\frac{(kx - \omega_2 t + \phi_o) - (kx - \omega_1 t + \phi_o)}{2} \right) \sin \left(\frac{kx - \omega_2 t + \phi_o + kx - \omega_1 t + \phi_o}{2} \right) \\
 &= 2y_{\max} \cos \left(\frac{kx - 2\pi f_2 t - (kx - 2\pi f_1 t)}{2} \right) \sin \left(\frac{kx - 2\pi f_2 t + kx - 2\pi f_1 t + 2\phi_o}{2} \right) \\
 &= 2y_{\max} \cos \left(2\pi \frac{f_1 - f_2}{2} t \right) \sin \left(kx - 2\pi \frac{f_1 + f_2}{2} t + \phi_o \right) \\
 &= \left[2y_{\max} \cos \left(2\pi \frac{f_1 - f_2}{2} t \right) \right] \sin \left(kx - 2\pi \frac{f_1 + f_2}{2} t + \phi_o \right)
 \end{aligned}$$

The sine part is a wave, it is a function of position and time. We see that it has a frequency that is the average of f_1 and f_2 . This is the frequency we hear. But we have another complicated amplitude term, and this time it is a function of time just as we suspected. The amplitude has its own frequency that is half the difference of f_1 and f_2 .

$$A_{\text{resultant}} = 2A \cos \left(2\pi \frac{f_1 - f_2}{2} t \right)$$

So the sound amplitude will vary in time for a given position in the medium.

The situation is odder still. We have a cosine function, but it is really an envelope for the higher frequency motion of the air particles. The air molecules move back and forth for both the crest and the trough of the envelope function.



So we will hear two maxima for every period! This frequency with which we hear the sound get loud at a given location as the wave goes by is called the *beat frequency*. The red envelope (solid heavy line in the last figure) has a frequency of

$$f_A = \frac{f_1 - f_2}{2}$$

but it is just the envelope. We can see that the green (thin line) wave will push and pull air molecules, and therefore our ear drums, with maximum loudness at twice this frequency. So our beat frequency is

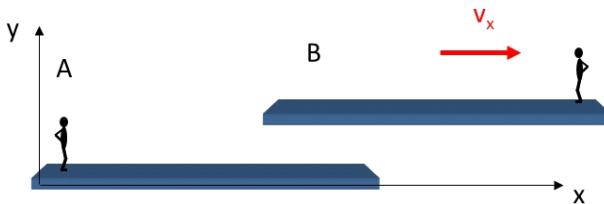
$$f_{beat} = |f_1 - f_2|$$

Any time we mix waves of different frequencies we get beating. Often the beat frequency is very fast, and our hearing system can't track the beats, so we don't hear them. And if we mixed more than two waves the beats might not come at perfectly regular intervals. The mixing of waves can become quite complicated. Yet even a barbershop quartet is a mixing of at least four waves. So complicated superpositions are common. In the next lecture we will try to see how we could take on these complicated combined waves.

10.2 Doppler Effect

Let's start by considering an inertial reference frame (remember this from PH121?)

Suppose we pick two reference frames, one traveling with a velocity v_r with respect to the other. Let's also place them far away from any other object.

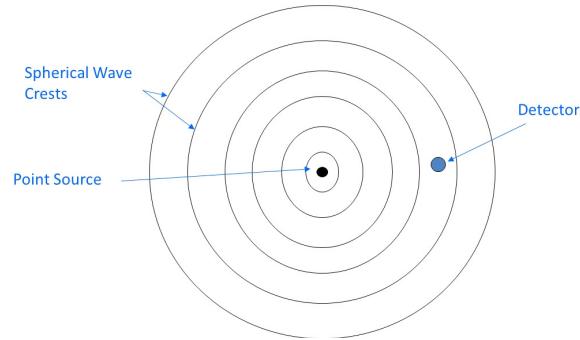


Person *A* sees himself as stationary and sees person *B* traveling with velocity v_x . Person *B* sees himself as stationary, and person *A* traveling with velocity $-v_x$.

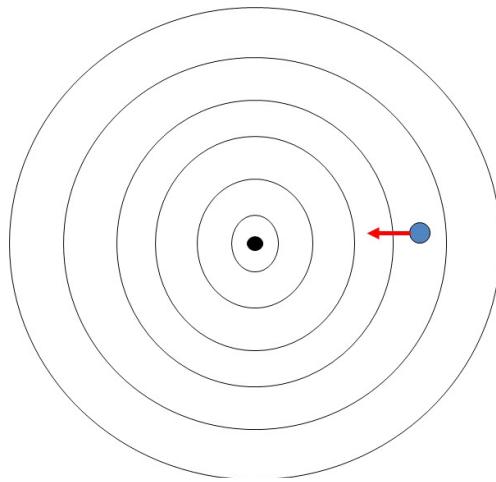
In looking at this situation it is the *relative speed* v_x that we must consider. We recall that we could write the speed of guy *b* as seen from platform *A* as

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

where v_{bB} is guy *b*'s speed seen from platform *B* and v_{BA} is the speed of platform *B* as seen from platform *A*. In this case $v_{BA} = v_x$. Now suppose we have a wave generator (a point source) creating spherical waves. Let the point source be at rest. We will call this point source an emitter and use a subscript *e* for it.

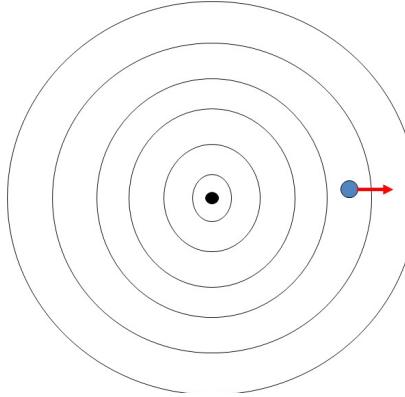


Let's also assume a detector. If the detector is stationary with respect to the emitter so the detector is in the emitter's reference frame, it sees a frequency of the wave in the emitter frame, f_{we} . But let's have the detector move relative to the emitter.



Remember, that the frequency is the number of crests that pass by a given point in a unit time. Does the moving detector see the same number of crests per unit time as when it was stationary?

No, the frequency appears to be higher! How about if we let the detector move the other way?



Again the frequency seen by the detector is different, but this time lower.

We can quantify this change. Take our usual variables f_{we} , λ_{we} , and the velocity of sound $v_{sound} = v_{we}$ for the wave speed of the sound wave because the air itself is not moving with respect to the emitter and the wave travels through the air. When the detector moves toward the source, it sees the wave velocity as

$$v_{wd} = v_{we} + v_{de} \quad (10.1)$$

where v_{wd} is the velocity of the wave in the detector frame, and v_{de} is the velocity of the detector in the emitter frame. And to be clear, for this case the detector is moving and we are in the emitter reference frame so the emitter is not moving.

The wavelength will not be changed, so

$$v_{we} = \lambda_{we} f_{we}$$

can just be written as

$$v_{we} = \lambda f_{we}$$

which tells us the frequency must change in the detector frame.

$$f_{wd} = \frac{v_{wd}}{\lambda} = \frac{v_{we} + v_{de}}{\lambda}$$

We can eliminate λ from this expression for the change in f by using $v_{we} = \lambda f_{we}$ again, this time solving for λ we get

$$\lambda = \frac{v_{we}}{f_{we}}$$

and substitute this into our f_{wd} equation

$$f_{wd} = \frac{v_{we} + v_{de}}{\frac{v_{we}}{f_{we}}}$$

or, after rearranging

$$f_{wd} = \frac{v_{we} + v_{de}}{v_{we}} f_{we} \quad \begin{array}{l} \text{detector moving toward the emitter} \\ \text{emitter not moving} \end{array} \quad (10.2)$$

or recognizing that v_{we} is the speed of sound in the stationary frame of the emitter

$$f_{wd} = \frac{v_{sound} + v_{de}}{v_{sound}} f_{we} \quad \begin{array}{l} \text{detector moving toward the emitter} \\ \text{emitter not moving} \end{array} \quad (10.3)$$

Now if the detector is going the other way

$$v_{wd} = v_{we} - v_{de}$$

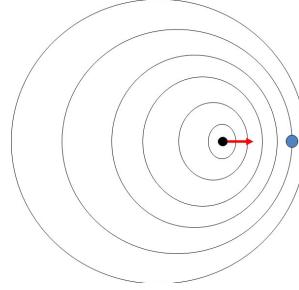
and the same reasoning gives

$$f_{wd} = \frac{v_{we} - v_{de}}{v_{we}} f_{we} \quad \begin{array}{l} \text{detector moving toward the emitter} \\ \text{emitter not moving} \end{array} \quad (10.4)$$

or

$$f_{wd} = \frac{v_{sound} - v_{de}}{v_{sound}} f_{we} \quad \begin{array}{l} \text{detector moving toward the emitter} \\ \text{emitter not moving} \end{array} \quad (10.5)$$

From our thinking about the motion of two inertial reference frames, we should expect a similar situation if the detector is stationary and the source moves. In this case the detector will see a different wavelength.



In fact, if we measure the distance between the crests we must account for the fact that the source moved by an amount

$$\Delta x_{ed} = v_{ed} T_{we} = \frac{v_{ed}}{f_{we}}$$

during one period of oscillation of the emitter before making the next crest. We can see that in this case, $v_{wd} = v_{sound}$ because this time the air is not moving with respect to the detector, so they are in the same reference frame. Then the wavelength will be shorter by this amount! That is, if we take the wavelength that we would get if the emitter were at rest, and subtract Δx_{ed} we should have our new wavelength at the detector. Let's start by finding the wavelength we expect if the emitter were at rest.

$$\lambda_{we,rest} = \frac{v_{we}}{f_{we}} = \frac{v_{sound}}{f_{we}}$$

This is because if the emitter were at rest, then $v_{we} = v_{sound}$. But in our case the emitter is moving, so we must subtract Δx_{ed} from the rest wavelength.

$$\lambda_{wd} = \lambda_{we,rest} - \Delta x_{ed}$$

Then the wavelength we would see at the detector would be

$$\lambda_{wd} = \lambda_{we,rest} - \frac{v_{ed}}{f_{we}}$$

Using

$$\lambda_{wd} = \frac{v_{wd}}{f_{wd}}$$

once more, we can write the frequency in the detector frame as

$$\begin{aligned} f_{wd} &= \frac{v_{wd}}{\lambda_{wd}} = \frac{v_{wd}}{\lambda_{we,rest} - \frac{v_{ed}}{f_{we}}} \\ &= \frac{v_{wd}}{\frac{v_{sound}}{f_{we}} - \frac{v_{ed}}{f_{we}}} \\ &= \frac{v_{wd}}{v_{sound} - v_{ed}} f_{we} \end{aligned}$$

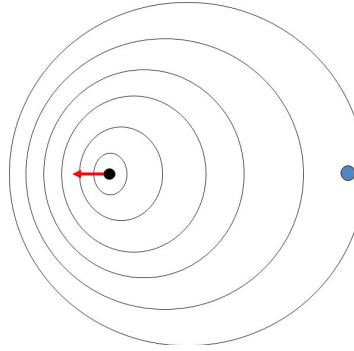
so then we have

$$f_{wd} = \frac{v_{wd}}{v_{sound} - v_{ed}} f_{we} \quad \begin{array}{l} \text{emitter moving toward detector} \\ \text{detector not moving} \end{array} \quad (10.6)$$

Now recall that it is the detector that is stationary with respect to the air this time so $v_{sound} = v_{wd}$

$$f_{wd} = \frac{v_{sound}}{v_{sound} - v_{ed}} f_{we} \quad \begin{array}{l} \text{emitter moving toward detector} \\ \text{detector not moving} \end{array} \quad (10.7)$$

When the source is moving away from the detector,



we expect the wavelength to be larger. The same reasoning gives

$$f' = \frac{v_{wd}}{v_{wd} + v_{ed}} f_{we} \quad \begin{array}{l} \text{emitter moving away from detector} \\ \text{detector not moving} \end{array} \quad (10.8)$$

or

$$f' = \frac{v_{sound}}{v_{sound} + v_{ed}} f_{we} \quad \begin{array}{l} \text{emitter moving away from detector} \\ \text{detector not moving} \end{array} \quad (10.9)$$

10.2.1 Combined Doppler Equation

We can combine these formulae to make one expression, but to do so we need to remember what v_{de} and v_{ed} mean. The first was the speed of the detector when the emitter was not moving. The second was the speed of the emitter when the detector was not moving. But we are experienced with relative motion. We should ask, “not moving with respect to what?” Let’s envision a reference frame that is not tied to either the emitter or detector. In this reference frame v_{dR} is the speed of the detector, and v_{eR} is the speed of the emitter. In this R reference frame our first Doppler equation for a moving detector with a stationary emitter might be written as

$$f_{wd} = \frac{v_{sound} \pm v_{dR}}{v_{sound}} f_{we} \quad \begin{array}{l} \text{detector moving emitter stationary in R frame} \\ \text{emitter stationary in R frame} \end{array} \quad (10.10)$$

and our second equation for a moving emitter with a stationary detector might be written as

$$f' = \frac{v_{wd}}{v_{wd} \mp v_{eR}} f_{we} \quad \begin{array}{l} \text{emitter moving detector stationary in R frame} \\ \text{detector moving emitter stationary in R frame} \end{array} \quad (10.11)$$

We could, of course have both the detector and emitter moving in the R frame. This would combine both of our previous scenarios

$$f_{wd} = \frac{v_{sound} \pm v_{dR}}{v_{sound} \mp v_{eR}} f_{we} \quad (10.12)$$

where we use the top sign for the speed when the mover is going toward the non-mover.

With this view, that we are neither in the e frame nor the d frame but in a separate R frame, we could even drop the R subscript without too much confusion

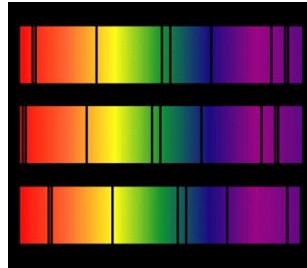
$$f_{wd} = \frac{v_{sound} \pm v_d}{v_{sound} \mp v_e} f_{we} \quad (10.13)$$

where now v_d is the speed of the detector in the reference frame of the observer, and v_e is the speed of the emitter in the reference frame of the observer. The quantity f_{we} is still the frequency of the wave as seen by the emitter and f_{wd} is still the frequency of the wave recorded by the detector.

10.2.2 Doppler effect in light

Light is also a wave, and so we would expect a Doppler shift in light. Indeed we do see a Doppler shift when we look at moving objects. Here is an optical spectrum of the Sun on the top and a spectrum of a similar star moving away

from us in the middle. The final spectrum is for a star moving toward us.

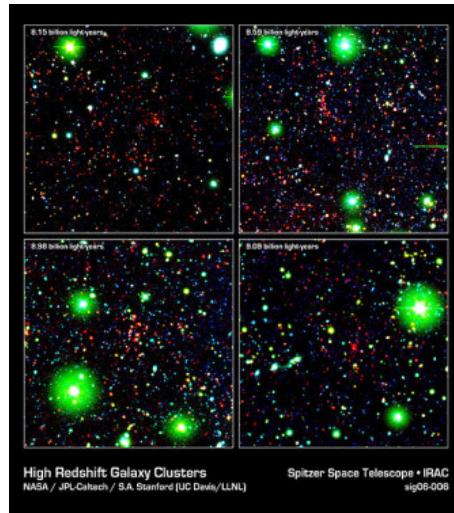


Top: Normal ‘dark’ spectral line positions at rest. Middle: Source moving away from observer. Bottom: Source moving towards observer. (Public domain image courtesy NASA: <http://www.jwst.nasa.gov/education/7Page45.pdf>)

Note that the wavelength of the lines is shifted toward the red part of the spectrum when the glowing object moves away from us. This is equivalent to lowering of the frequency of a truck engine noise as it goes away from us. The larger wavelengths indicate a lower frequency of light because

$$f = \frac{c}{\lambda}$$

This gives us a way to determine if distant stars and galaxies are moving toward or away from us. We look for the chemical signature pattern of lines, then see whether they are shifted to the red (moving away from us) or blue (moving toward us) compared to the position in their spectrum of the Sun. This photo is of some of the most distant galaxies that are moving very fast away from us. Their redshift is very large.



High Redshift Galaxy Cluster shown here in false color from the Spitzer Space Telescope. (Public domain image courtesy NASA/JPL-Caltech/S.A. Stanford (UC Davis/LLNL)

Deriving the Doppler equation for light is more tricky because the speed of light is constant and the same in every reference frame. We really tackle this in our PH279 class. So I will just quote the result here.

Chapter 11

11 Shock waves and Non-Sinusoidal Waves

1.17.8

Fundamental Concepts

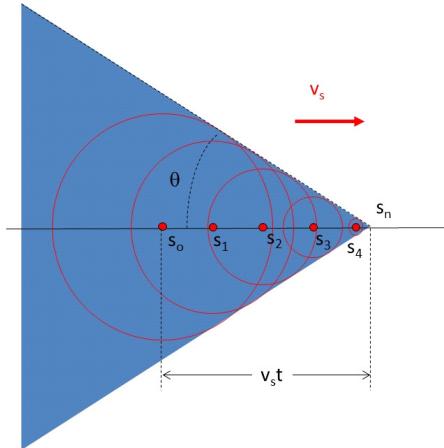
- If the source travels faster than the wave speed in the medium you get shock waves
- Actual waves can be viewed as sums of sinusoidal waves

11.1 Shock Waves

What happens when the speed of the source is greater than the wave speed?

Remember that the wave speed depends only on the medium. Let's call the crests of a wave the *wave front*. In the picture below, a point source is generating a wave and the red lines are the wave fronts.

When $v_e = v_{sound}$ the waves begin to pile up. If we allow $v_e > v_{sound}$ then the wave fronts are no longer generated within each other.



The leading edge of the wave fronts “build up” to form a cone shape. We recognize this as a superposition of the waves. This is constructive interference. The half angle of this cone is called the *Mach angle*

$$\sin \theta = \frac{v_{sound} t}{v_e t} = \frac{v_{sound}}{v_e} \quad (11.1)$$

This ratio v_e/v_{sound} is called the Mach number

$$M = \frac{v_e}{v_{sound}}$$

and the conical wave front is called a shock wave.

$$\sin \theta = \frac{1}{M}$$

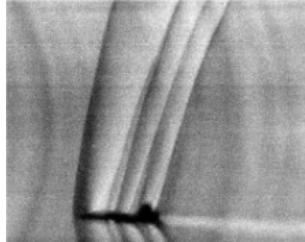
We see these conical waves often in water (where we call them a wake)



Boat wakes as a Doppler cone. Image courtesy US Navy. Image is in the Public Domain.

and hear them when jet aircraft go supersonic. In the next figure we can see a picture of a T-38 breaking the sound barrier. You can see the Mach cones, but notice that there are several! Remember that a disturbance creates a wave.

There are disturbances created by the nose of the plane, the rudder, and the wings, and perhaps the cockpit in this Schlieren photograph.



Dr. Leonard Weinstein's Schlieren photograph of a T-38 Talon at Mach 1.1, altitude 13,700 feet, taken at NASA Langley Research Center, Wallops in 1993. Image Courtesy NASA, image is in the Public Domain.

Superposition is also the basis of making music. We have talked about resonance making music louder. But there is more to music and superposition. Let's take this topic on next.

11.2 Non Sinusoidal Waves

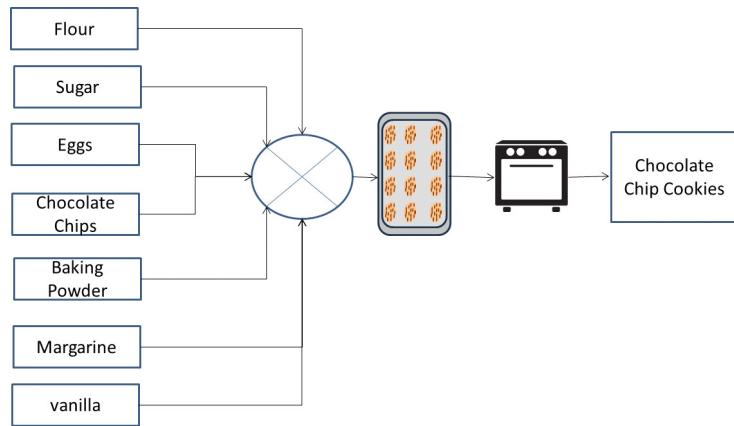
We have dealt with pulses, and with sinusoidal waves. But this seems kind of restrictive. Are all waves either pulses or sinusoidal? Can the universe really be described by such simple mathematics? The answer is both no, and yes. There are non-sinusoidal waves, in fact, most waves are not sinusoidal. But it turns out that we can use a very clever mathematical trick to make any shape wave out of a superposition of many sinusoidal waves. So our mathematics for sinusoidal waves turns out to be quite general.

11.2.1 Music and Non-sinusoidal waves

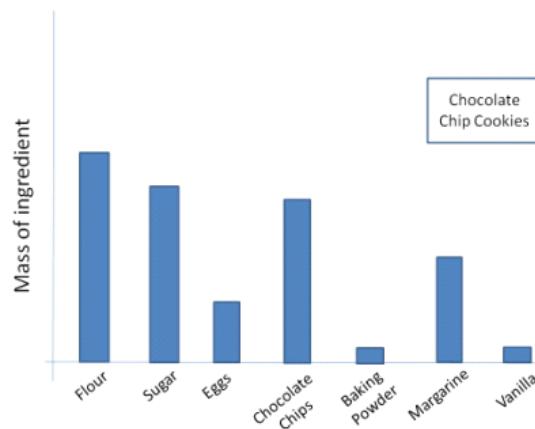
From our example of standing waves on strings, we know that a string can support a series of standing waves with discrete frequencies—the harmonic series. We have also discussed that usually we excite more than one standing wave at a time. The fundamental mode tends to give us the pitch we hear, but 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.

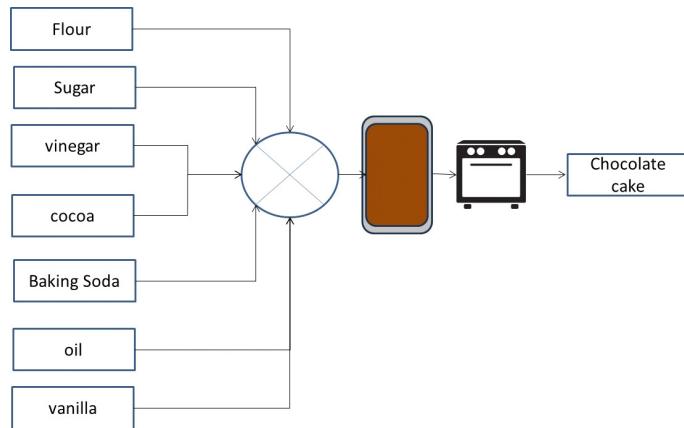


The recipe is a list of ingredients, and a symbolic instructions to mix and bake. The product is chocolate chip cookies. Of course we need more information. We need to know how much of each ingredient to use.



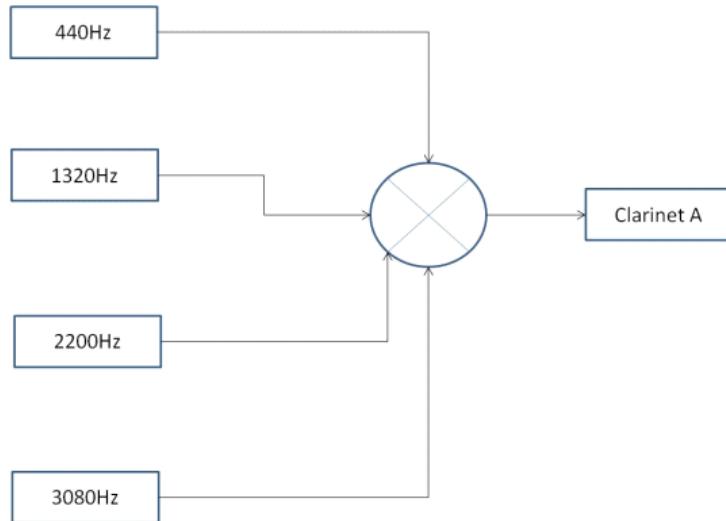
This graph gives us the amount of each ingredient by mass.

Now suppose we want chocolate cake.

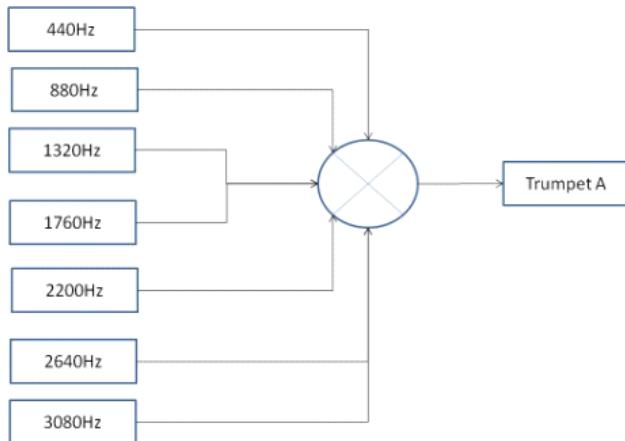


The predominant taste in each of these foods is chocolate. But chocolate cake and chocolate chip cookies don't taste exactly the same. We can easily see that the differences in the other ingredients make the difference between the "cookie" taste and the "cake" taste that goes along with the "chocolate" taste that predominates.

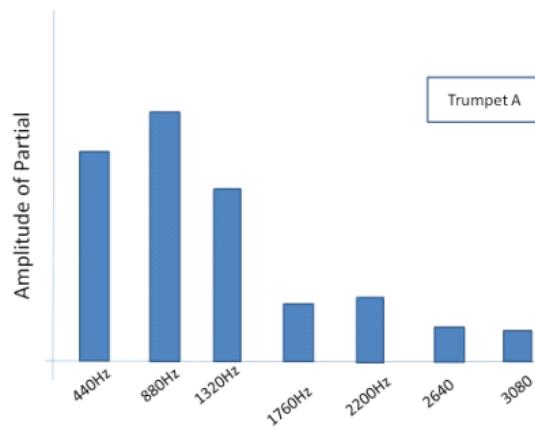
The sound waves produced by musical instruments work in a similar way. Here is a recipe for an "A" note from a clarinet.



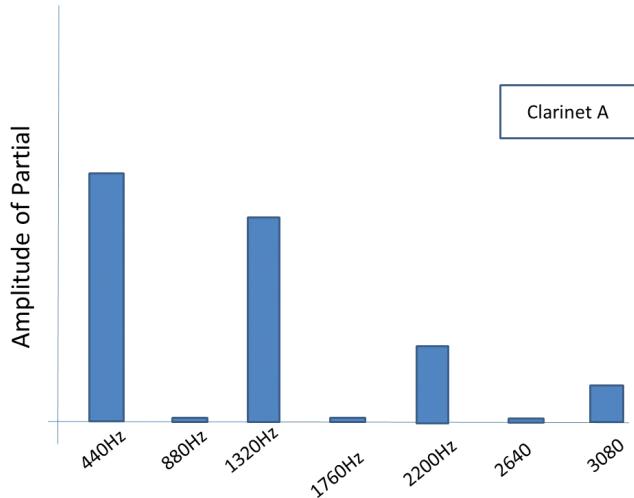
and here is one for a trumpet playing the same "A" note.



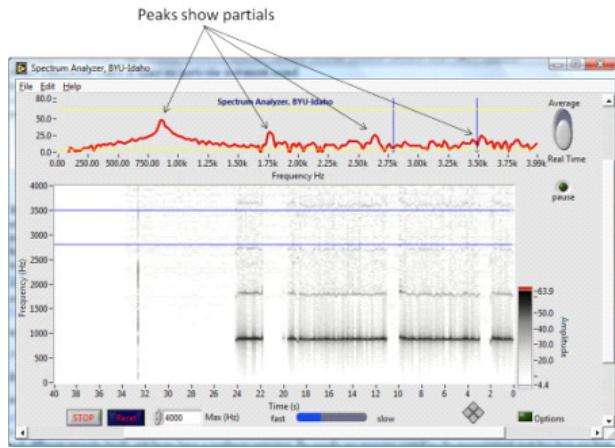
A trumpet sounds different than a clarinet, and now we see why. There are more harmonics involved with the trumpet sound than the clarinet sound. These extra standing waves make up the “brassiness” of the trumpet sound. As with our baking example, we need to know how much of each standing wave we have. 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. For a clarinet we might have something more like this

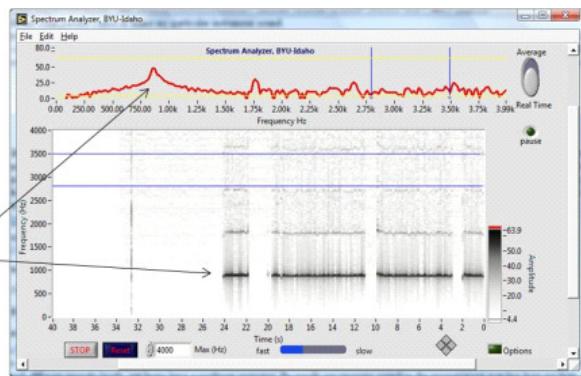


We have a tool that you can download to your PC to detect the mix of harmonics of musical instruments, or mechanical systems. In music, the different harmonics are called *partials* because they make up part of the sound. A graph that shows which harmonics are involved is called a *spectrum*. The next figure is the spectrum of a six holed bamboo flute. Note that there are several harmonics involved.

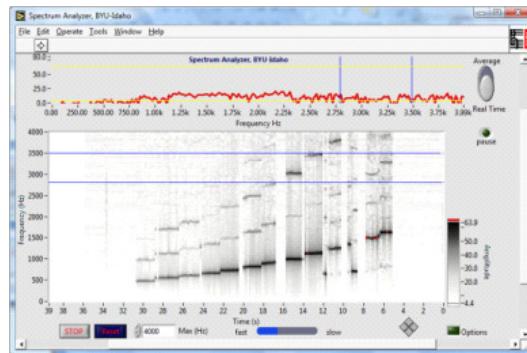


Note that our graph has two parts. One is the instantaneous spectrum, and one is the spectrum time history.

Notice that there
are non-harmonic
partials in this
instrument too!

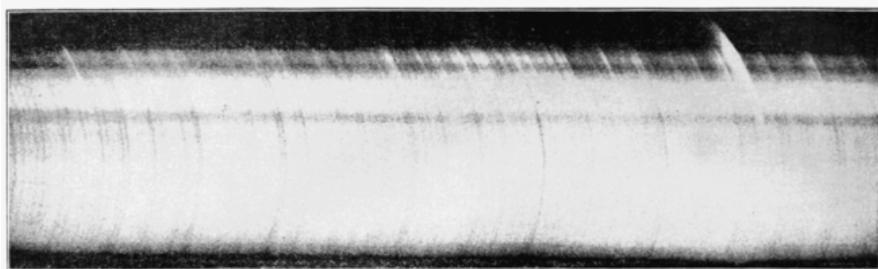


By observing the time history, we can see changes in the spectrum. We can also see that we don't have pure harmonics. The graph shows some response off the specific harmonic frequencies. This six holed flute is very "breathy" giving a lot of wind noise along with the notes, and we see this in the spectrum. 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

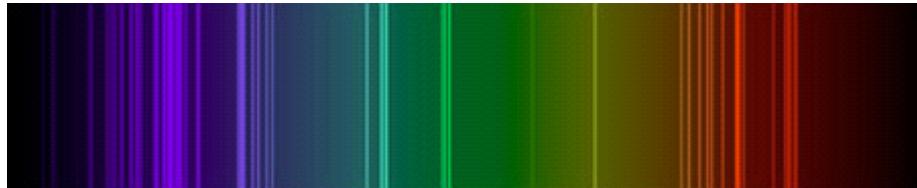


Solar coronal spectrum taken during a solar eclipse. The successive curved lines are each different wavelengths, and the dark lines are wavelengths that are absorbed.

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)

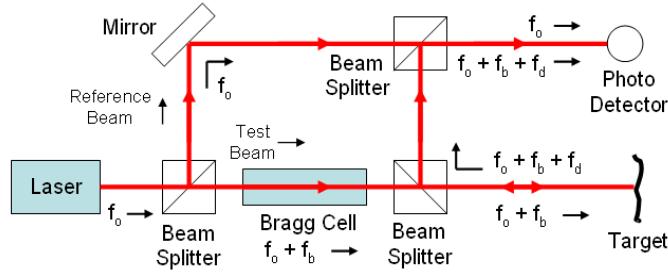
The lines in this graph show the amplitude of each harmonic component of the light. Darker lines have larger amplitudes. The harmonics come from the excitation of electrons in their orbitals. Each orbital is a different energy state, and when the electrons jump from orbital to orbital, they produce specific 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)

11.2.2 Vibrometry

Just like each atom has a specific spectrum, and each instrument, each engine, machine, or anything that vibrates has a spectrum. We can use this to monitor the health of machinery, or even to identify a piece of equipment. Laser or acoustic vibrometers are available commercially.



Laser Vibrometer Schematic (Public Domain Image from Laderaranch:
http://commons.wikimedia.org/wiki/File:LDV_Schematic.png)

They provide a way to monitor equipment in places where it would be dangerous or even impossible to send a person. The equipment also does not need to be shut down, a great benefit for factories that are never shut down, or for a satellite system that cannot be reached by anyone.

Fourier Series: Mathematics of Non-Sinusoidal Waves

We should take a quick look at the mathematics of non-sinusoidal waves.

Let's start with a superposition of many sinusoidal waves. 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.

Example: Fourier representation of a square wave.

For example, we could represent a function $f(x)$ with the following series

$$f(x) = C_0 + C_1 \cos\left(\frac{2\pi}{\lambda}x + \varepsilon_1\right) \quad (11.2)$$

$$+ C_2 \cos\left(\frac{2\pi}{\frac{\lambda}{2}}x + \varepsilon_2\right) \quad (11.3)$$

$$+ C_3 \cos\left(\frac{2\pi}{\frac{\lambda}{3}}x + \varepsilon_3\right) \quad (11.4)$$

$$+ \dots \quad (11.5)$$

$$+ C_n \cos\left(\frac{2\pi}{\frac{\lambda}{n}}x + \varepsilon_n\right) \quad (11.6)$$

$$+ \dots \quad (11.7)$$

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

$$C_m \cos(mkx + \varepsilon_m) = A_m \cos(mkx) + B_m \sin(mkx) \quad (11.8)$$

where k is the wavenumber

$$k = \frac{2\pi}{\lambda} \quad (11.9)$$

and λ is the wavelength of the complicated but still periodic function $f(x)$. Then we identify

$$A_m = C_m \cos(\varepsilon_m) \quad (11.10)$$

$$B_m = -C_m \sin(\varepsilon_m) \quad (11.11)$$

then

$$f(x) = \frac{A_o}{2} + \sum_{m=1}^{\infty} A_m \cos(mkx) + \sum_{m=1}^{\infty} B_m \sin(mkx) \quad (11.12)$$

where we separated out the $A_o/2$ term because it makes things nicer later.

Fourier Analysis

The process of finding the coefficients of the series is called *Fourier analysis*. We start by integrating equation (11.12)

$$\int_0^\lambda f(x) dx = \int_0^\lambda \frac{A_o}{2} dx + \int_0^\lambda \sum_{m=1}^{\infty} A_m \cos(mkx) dx + \int_0^\lambda \sum_{m=1}^{\infty} B_m \sin(mkx) dx \quad (11.13)$$

We can see immediately that all the sine and cosine terms integrate to zero (we integrated over a wavelength) so

$$\int_0^\lambda f(x) dx = \int_0^\lambda \frac{A_o}{2} dx = \frac{A_o}{2} \lambda \quad (11.14)$$

We solve this for A_o

$$A_o = \frac{2}{\lambda} \int_0^\lambda f(x) dx \quad (11.15)$$

To find the rest of the coefficients we need to remind ourselves of the orthogonality of sinusoidal functions

$$\int_0^\lambda \sin(akx) \cos(bkx) dx = 0 \quad (11.16)$$

$$\int_0^\lambda \cos(akx) \cos(bkx) dx = \frac{\lambda}{2} \delta_{ab} \quad (11.17)$$

$$\int_0^\lambda \sin(akx) \sin(bkx) dx = \frac{\lambda}{2} \delta_{ab} \quad (11.18)$$

where δ_{ab} is the Kronecker delta.

To find the coefficients, then, we multiply both sides of equation (11.12) by $\cos(lkx)$ where l is a positive integer. Then we integrate over one wavelength.

$$\int_0^\lambda f(x) \cos(lkx) dx = \int_0^\lambda \frac{A_o}{2} \cos(lkx) dx \quad (11.19)$$

$$+ \int_0^\lambda \sum_{m=1}^{\infty} A_m \cos(mkx) \cos(lkx) dx \quad (11.20)$$

$$+ \int_0^\lambda \sum_{m=1}^{\infty} B_m \sin(mkx) \cos(lkx) dx \quad (11.21)$$

which gives

$$\int_0^\lambda f(x) \cos(mkx) dx = \int_0^\lambda A_m \cos(mkx) \cos(mkx) dx \quad (11.22)$$

that is, only the term with two cosine functions where $l = m$ will be non zero.
So

$$\int_0^\lambda f(x) \cos(mkx) dx = \frac{\lambda}{2} A_m \quad (11.23)$$

solving for A_m we have

$$A_m = \frac{2}{\lambda} \int_0^\lambda f(x) \cos(mkx) dx \quad (11.24)$$

We can perform the same steps to find B_m only we use $\sin(lkx)$ as the multiplier. Then we find

$$B_m = \frac{2}{\lambda} \int_0^\lambda f(x) \sin(mkx) dx \quad (11.25)$$

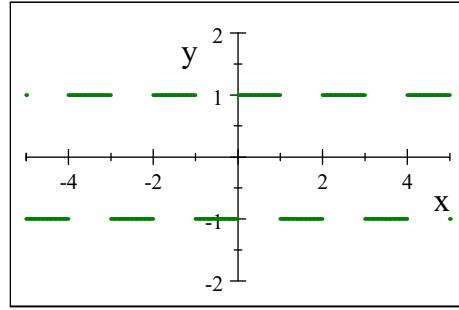
11.2.3 Square wave

Let's find the series for a square wave using our Fourier analysis technique.

Let's take

$$\lambda = 2 \quad (11.26)$$

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < \frac{\lambda}{2} \\ -1 & \text{if } \frac{\lambda}{2} < x < \lambda \end{cases} \quad (11.27)$$



since $f(x)$ is odd, $A_m = 0$ for all m . We have

$$B_m = \frac{2}{\lambda} \int_0^{\frac{\lambda}{2}} (1) \sin(mkx) dx + \frac{2}{\lambda} \int_{\frac{\lambda}{2}}^{\lambda} (-1) \sin(mkx) dx \quad (11.28)$$

so

$$B_m = \frac{1}{m\pi} (-\cos(mkx)|_0^{\frac{\lambda}{2}} + \frac{1}{m\pi} (\cos(mkx)|_{\frac{\lambda}{2}}^{\lambda}) \quad (11.29)$$

Which is

$$B_m = \frac{1}{m\pi} \left(1 \cos\left(m \frac{2\pi}{\lambda} x\right)|_0^{\frac{\lambda}{2}} + \frac{1}{m\pi} \left(\cos\left(m \frac{2\pi}{\lambda} x\right)|_{\frac{\lambda}{2}}^{\lambda} \right) \right) \quad (11.30)$$

so

$$B_m = \frac{1}{m\pi} \left(\left(-\cos\left(m \frac{2\pi}{\lambda} \frac{\lambda}{2}\right) \right) + \cos\left(m \frac{2\pi}{\lambda} (0)\right) \right) \quad (11.31)$$

$$+ \frac{1}{m\pi} \left(\left(\cos\left(m \frac{2\pi}{\lambda} \lambda\right) - \cos\left(m \frac{2\pi}{\lambda} \frac{\lambda}{2}\right) \right) \right) \quad (11.32)$$

which is

$$B_m = \frac{2}{m\pi} (1 - \cos(m\pi)) \quad (11.33)$$

Our series is then just

$$f(x) = \sum_{m=1}^{\infty} \frac{2}{m\pi} (1 - \cos(m\pi)) \sin(mkx) \quad (11.34)$$

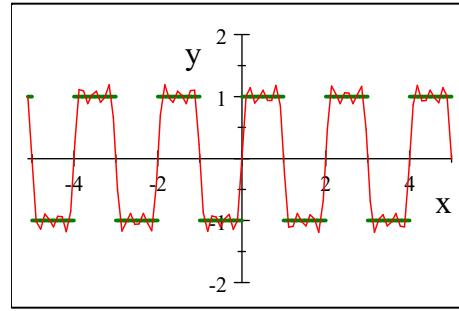
and we can write a few terms

Term	
1	$\frac{4}{\pi} \sin(kx)$
2	0
3	$\frac{4}{3\pi} \sin(3kx)$
4	0
5	$\frac{4}{5\pi} \sin(5kx)$

(11.35)

then the partial sum up to $m = 5$ looks like

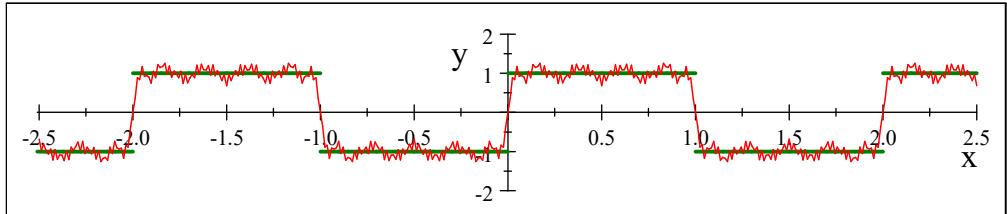
$$f(x) = \frac{4}{\pi} \sin(kx) + \frac{4}{3\pi} \sin(3kx) + \frac{4}{5\pi} \sin(5kx) \quad (11.36)$$



If we take many terms,

$$\begin{aligned} f(x) = & \frac{4}{\pi} \sin(kx) + \frac{4}{3\pi} \sin(3kx) + \frac{4}{5\pi} \sin(5kx) + \frac{4}{7\pi} \sin(7kx) + \frac{4}{9\pi} \sin(95kx) \\ & + \frac{4}{11\pi} \sin(11kx) + \frac{4}{13\pi} \sin(13kx) + \frac{4}{15\pi} \sin(15kx) + \frac{4}{17\pi} \sin(17kx) + \frac{4}{19\pi} \sin(19kx) \end{aligned} \quad (11.37)$$

We see the function get closer and closer to a 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. For example, oil based compounds all have specific spectral signatures in the wavelength range between 3 – 5 micrometers. If you wish to tell the difference between gasoline and crude oil, you can restrict your study to these wavelengths alone.

Part II

Optics

Chapter 12

12 Nature of Light and Reflection 3.1.1, 3.1.2

In the movie *Star Wars* inter-galactic star ships blast each other with laser cannons. The laser beams streak across the screen. This is dramatic, but not realistic. For us to see the light, some of the light must get to our eyes. The light must either travel directly to our eyes from the source, or it must bounce off of something. We can make a laser beam visible by providing dust for the light to bounce off of so it travels to our eyes. But unless that galaxy far far away is really dusty, Hollywood doesn't understand light waves very well. Let's start our study of geometric optics with a few comments on what light is. Then let's study reflection, the bouncing of light off of an object (like dust, or a mirror).

Question 223.13.4

Fundamental Concepts

- Light is a wave in the electromagnetic field
- Light travels fast, with a speed of $c = 3 \times 10^8$ m/s
- Law of Reflection

12.1 What is Light?

Before the 19th century (1800's) light was assumed to be a stream of particles. Newton was one of the chief proponents of this theory. The theory was able to explain reflection of light from mirrors and other objects and therefore explain vision. In 1678 Huygens showed that wave theory could also explain reflection and vision.

In 1801 Thomas Young demonstrated that light had attributes that were best explained by wave theory. We will study Young's experiment later. The crux of his experiment was to show that light displayed constructive and destructive

interference—clearly a wave phenomena! The theory of the nature of light took a dramatic shift

In 1805 Joseph Smith was born in Sharon, Vermont.

In September of 1832 Joseph Smith received a revelation that said in part :

For the word of the Lord is truth, and whatsoever is truth is light, and whatsoever is light is Spirit, even the Spirit of Jesus Christ. And the Spirit giveth light to every man that cometh into the world; and the Spirit enlighteneth every man through the world, that hearkeneth to the voice of the Spirit. (D&C 84:45-46)

In December of 1832 Joseph Smith received another revelation that says in part:

This Comforter is the promise which I give unto you of eternal life, even the glory of the celestial kingdom; which glory is that of the church of the Firstborn, even of God, the holiest of all, through Jesus Christ his Son—He that ascended up on high, as also he descended below all things, in that he comprehended all things, that he might be in all and through all things, the light of truth; which truth shineth. This is the light of Christ. As also he is in the sun, and the light of the sun, and the power thereof by which it was made. As also he is in the moon, and is the light of the moon, and the power thereof by which it was made; as also the light of the stars, and the power thereof by which they were made; and the earth also, and the power thereof, even the earth upon which you stand. And the light which shineth, which giveth you light, is through him who enlighteneth your eyes, which is the same light that quickeneth your understandings; which light proceedeth forth from the presence of God to fill the immensity of space—the light which is in all things, which giveth life to all things, which is the law by which all things are governed, even the power of God who sitteth upon his throne, who is in the bosom of eternity, who is in the midst of all things. (D&C 88:5-12)

Light, even real, physical light, seems to be of interest to members of the Restored Church.

In 1847 the saints entered the Salt Lake Valley.

In 1873 Maxwell published his findings that light is an electromagnetic wave (something we will try to show before this course is over!).

Planck's work in quantization theory (1900) was used by Einstein In 1905 to give an explanation of the photoelectric effect that again made light look like a particle.

Current theory allows light to exhibit the characteristics of a wave in some situations and like a particle in others. We will study both before the end of our optics unit.

The results of Einstein's work give us the concept of a *photon* or a quantized unit of radiant energy. Each "piece of light" or photon has energy

$$E = hf \quad (12.1)$$

where f is the frequency of the light and h is a constant

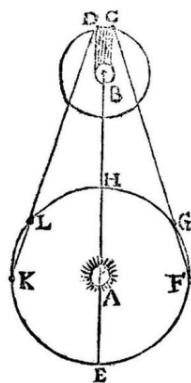
$$h = 6.63 \times 10^{-34} \text{ Js} \quad (12.2)$$

The nature of light is fascinating and useful both in physical and religious areas of thought.

12.2 Measurements of the Speed of Light

One of the great foundations of modern physical theory is that the speed of light is constant in a vacuum. Galileo first tried to measure the speed of light. He used two towers in town and placed a lantern and an assistant on each tower. The lanterns had shades. The plan was for one assistant to remove his shade, and then for the assistant on the other tower to remove his shade as soon as he saw the light from the first lantern. Back at the first tower, the first assistant would use a clock to determine the time difference between when the first lantern was un-shaded, and when they saw the light from the second tower. The light would have traveled twice the inter-tower distance. Dividing that distance by the time would give the speed of light. You can probably guess that this did not work. Light travels very quickly. The clocks of Galileo's day could not measure such a small time difference. Ole Rømer was the first to succeed in measuring the speed of light.

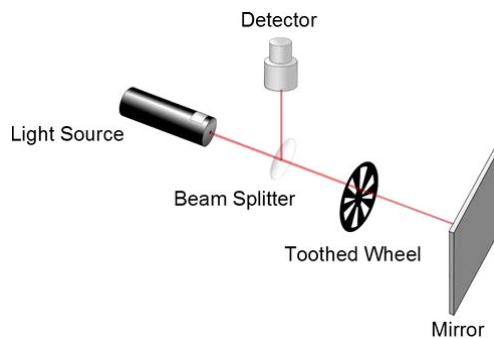
12.2.1 Rømer's Measurement of the speed of light



A diagram illustrating Rømer's determination of the speed of light. Point A is the Sun, point B is Jupiter. Point C is the immersion of Io into Jupiter's shadow at the start of an eclipse

Rømer performed his measurement in 1675, 269 years before digital devices existed!. He used the period of revolution of Io, a moon of Jupiter, as Jupiter revolved around the sun. He first measured the period of Io's rotation about Jupiter, then he predicted an eclipse of Io three months later. But he found his calculation was off by 600 s. After careful thought, he realized that the Earth had moved in its orbit, and that the light had to travel the extra distance due to the Earth's new position. Given Rømer's best estimate for the orbital radius of the earth and his time difference, Rømer arrived at a estimate of $c = 2.3 \times 10^8 \frac{\text{m}}{\text{s}}$. Amazing for 1675!

12.2.2 Fizeau's Measurement of the speed of light



Hippolyte Fizeau measured the speed of light in 1849 using the apparatus indicated in the figure above. He used a toothed wheel and a mirror and a beam of light. The light passed through the open space in the wheel's teeth as the wheel rotated. Then was reflected by the mirror. The speed would be

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

We just need Δx and Δt .

It is easy to see that

$$\Delta x = 2d$$

because the light travels twice the distance to the mirror (d) and back. If the wheel turned just at the right angular speed, then the reflected light would hit the next tooth and be blocked. Think of angular speed

$$\omega = \frac{\Delta \theta}{\Delta t}$$

so the time difference would be

$$\Delta t = \frac{\Delta \theta}{\omega}$$

we find $\Delta\theta$ by taking 2π and dividing by the number of teeth on the wheel.

$$\Delta\theta = \frac{2\pi}{N_{teeth}}$$

Then the speed of light must be

$$\begin{aligned} c &= v = \frac{2d}{\underline{\omega}} \\ &= \frac{2d\omega}{\Delta\theta} \\ &= \frac{2d\omega N_{teeth}}{2\pi} \\ &= \frac{d\omega N_{teeth}}{\pi} \end{aligned}$$

then if we have 720 teeth and ω is measured to be $d = 7500\text{ m}$

$$\begin{aligned} c &= \frac{(7500\text{ m})(172.79\text{ Hz})(720)}{\pi} \\ &= 2.97 \times 10^8 \frac{\text{m}}{\text{s}} \end{aligned}$$

which is Fizeau's number and it is pretty good!

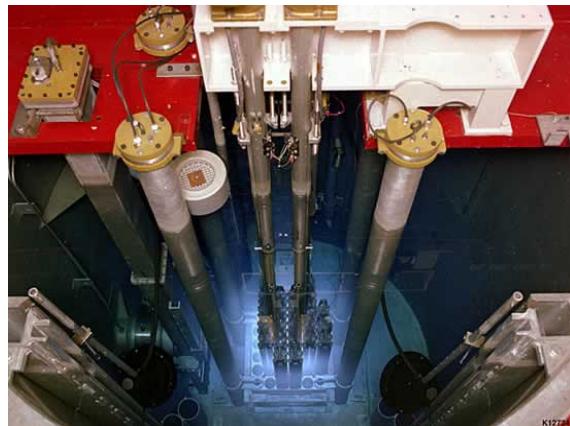
Modern measurements are performed in very much the same way that Fizeau did his calculation. The current value is

$$c = 2.9979 \times 10^8 \frac{\text{m}}{\text{s}} \quad (12.3)$$

12.2.3 Faster than light

The speed of light in a vacuum is constant, but in matter the speed of light changes.

We will study this in detail when we look at refraction. But for now, a dramatic example is Cherenkov radiation. It is an eerie blue glow around the core of nuclear reactors. It occurs when electrons are accelerated past the speed of light in the water surrounding the core. The electrons emit light and the light waves form a Doppler cone or a light-sonic boom! The result is the blue glow.

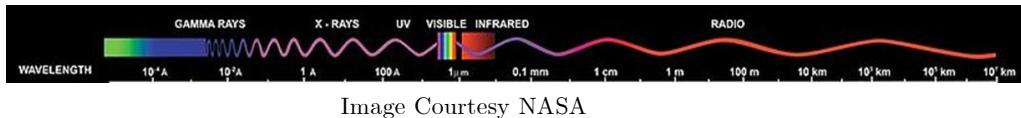


Cherenkov radiation (United States Department of Energy, image in the public domain)

This does bring up a problem in terminology. What does the word “medium” mean? We have used it to mean the substance through which a wave travels. This substance must have the property of transferring energy between its parts, like the coils of a spring can transfer energy to each other, or like air molecules can transfer energy by collision. For light the wave medium is the electromagnetic field. This field can store and transfer energy—but we won’t study it until Principles of Physics III (PH220). But many books on physics call materials like glass a “medium” through which light travels. The water in our last example is a “medium” in this sense. Are glass and water wave mediums for light? The answer is an emphatic NO! Light does not need any matter to form its wave. The wave medium is the electromagnetic field. So we will have to keep this in mind as we allow light to travel through matter. I will try to say that the light enters a new “material” to describe something like light entering a piece of glass. But some books may call the matter a “medium.” You must remember that by this they don’t mean the wave medium.

Light can’t be a wave in some type of matter. We know this because light travels through empty space. But in Principles of Physics III we will find that space isn’t exactly empty. It has an electromagnetic field in it. And it turns out that light seems to be a wave in this electromagnetic field. It will take us a while to fully understand this concept—it is done in PH220—but don’t worry. Physicists knew that light was a wave for almost 80 years before the electric field was shown to be the medium. We can do a lot just knowing light behaves like a wave.

This makes the light we see just one small part of a whole class of waves that are possible in this electromagnetic field medium. Radio waves, and microwaves, and x-rays are all just different types of electromagnetic waves. The next figure shows where all of these electromagnetic waves fit ordered by wavelength (and frequency).



There is something very unique about this electromagnetic field medium. The waves in this medium travel at a constant speed no matter what frame of reference we are in. This fact lead to the formation of the Special Theory of Relativity and the famous equation

$$E = mc^2$$

where c is this speed of light

$$c = 299792458 \frac{\text{m}}{\text{s}}$$

Light does slow down when it enters a material, like glass, or even air. This is not really because it moves slower. what happens is that light is absorbed by the electrons in the atoms of the material substance. The electron temporarily takes up all the energy from a bit of the light wave. But only temporarily. It eventually has to give up the energy and a light wave is reformed coming from the atoms. There is really a superposition of what is left of the original light wave, and the new light wave generated by the atoms. And the combined wave moves slower. Think of our last lecture on non-sinusoidal waves. The combined wave is made of two sinusoidal waves both traveling at the speed of light. But the combined wave will appear to travel at a slower speed. How much slower depends on how long the electrons in the atoms can hang on to the light and the geometry of the situation. Each substance is different.

We can devise a way to express how much slower light will appear to go in a substance using the ratio

$$\frac{c}{v}$$

the ratio of the speed of light, c , to the average speed in the substance, v . This ratio is so useful that we give it a name, the *index of refraction*.

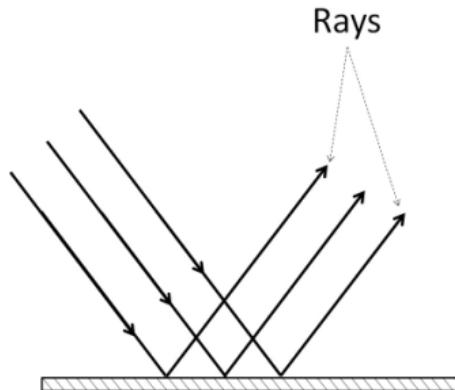
$$n = \frac{c}{v}$$

Let's look at some a simple interaction of light with a material, but let's not look at the wave and atomic re-emission level. Let's treat our combined wave as just one resultant wave that travels in a specific direction. We can call this wave a *ray* (as in a ray of light!).

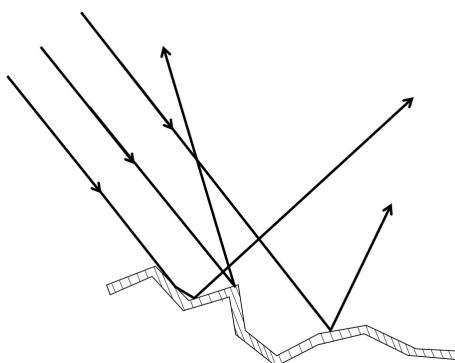
12.3 Reflection

Using the ray approximation we wish to find what happens when a bundle of rays reaches a boundary between materials. If the material boundary is very

smooth, then the rays are reflected (bounce off) in a uniform way. This is called *specular* reflection



If the surface is not smooth, then something different happens. The rays are reflected, but they are reflected randomly



This is called *diffuse* reflection

This difference can be seen in real life. In the next figure, the surface on the left is a specular reflector and you can see the reflected light beam. But the surface on the right photo is diffuse, no reflected beam is seen.



We said the surface must be smooth for there to be specular reflection. What does smooth mean? Generally the size of the rough spots must be much smaller than a wavelength to be considered smooth. So suppose we have a red laser. How small do the surface variations have to be for the surface to be considered smooth? The wavelength of a *HeNe* laser is

$$\lambda_{HeNe} = 633 \text{ nm}$$

This is very small. Modern optics for remote sensing are often manufactured to 1/10 of a wavelength, which would be 63 nm. Mirrors are smooth for visible light.

How about a microwave beam of light like your cell phone uses?

$$\begin{aligned} c &= \lambda f \\ \lambda &= \frac{c}{f} = \frac{3 \times 10^8 \frac{\text{m}}{\text{s}}}{1 \text{ GHz}} \\ &= 0.3 \text{ m} \end{aligned}$$

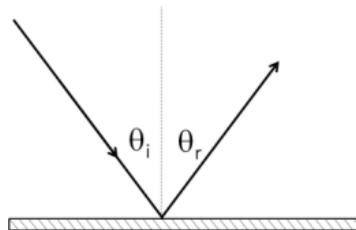
Rough brick walls are smooth for cell phone light!

We can see that we must be careful in our definition of “smooth.”

12.4 Law of reflection

Experience shows that if we do have a smooth surface, that light bounces much like a ball. This is why Newton thought light was a particle. Suppose we take a flat surface and we shine a light on it. We have a ray that approaches at an angle θ_i measured from the normal. Then the reflected ray will leave the surface with an angle θ_r measured from the normal such that

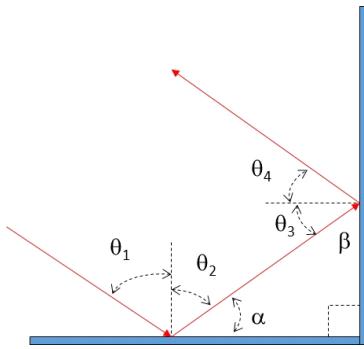
$$\theta_r = \theta_i$$



This is called the *law of reflection*. We will look at how superposition makes this happen later. But from experiment we can see that this law of reflection works.

12.4.1 Retroreflection

Let's take an example



Let's take our system to be two mirrors set at a right angle. We have a beam of light incident at angle θ_1 . By the law of reflection, it must leave the mirror at $\theta_2 = \theta_1$. We can see that α must be $90^\circ - \theta_2$ and it is clear that $\theta_3 = \alpha$. By the law of reflection, $\theta_3 = \theta_4$. Then, since

$$90^\circ = \theta_2 + \alpha$$

But now look at the triangle in the corner (that has angles α and β). Its interior angles have to sum to 180° , and one of the angles is 90° right in the corner. So

$$180^\circ = 90^\circ + \alpha + \beta$$

Then

$$90^\circ = \alpha + \beta$$

so we can see that β has to be equal to θ_2

$$\beta = \theta_2 = \theta_1$$

and

$$90^\circ = \theta_1 + \theta_4$$

$$\begin{aligned} 90^\circ &= \beta + \theta_3 \\ &= \theta_1 + \theta_3 \end{aligned}$$

and of course by the law of reflection $\theta_3 = \theta_4$ then the total angular change is

$$\theta_1 + \theta_2 + \theta_3 + \theta_4$$

but if we rearrange this we have

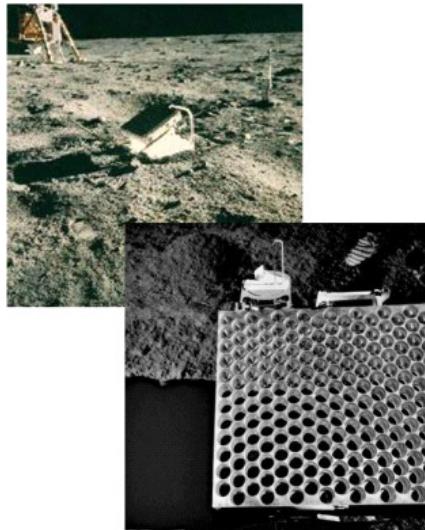
$$\begin{aligned} &\theta_1 + \theta_3 + \theta_2 + \theta_4 \\ &= 90^\circ + 90^\circ = 180^\circ \end{aligned}$$

or the outgoing ray is sent back toward the source! If we do this in three dimensions we have a corner cube. Here is a picture of a radar corner cube array.



Radar retroreflector tower located in the center of Yucca Flat dry lake bed. Used as a radar target by maneuvering aircraft during "inert" contact fusing bomb drops at Yucca Flat. Sandia National Laboratories conducted the tests on the lake bed from 1954 to 1956. (Image in the Public Domain in the United States)

And we left visible corner cube arrays on the Moon so we can bounce laser beams off of them and monitor the Earth-Moon distance.



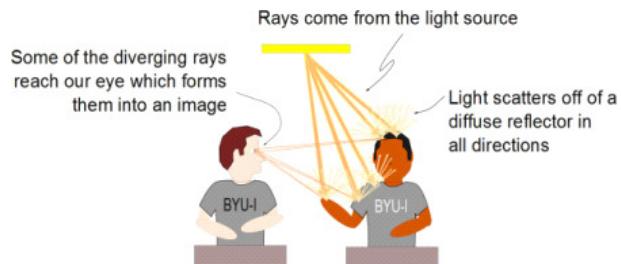
Apollo Retroreflector (Images in the Public Domain courtesy NASA)

Question 223.13.9

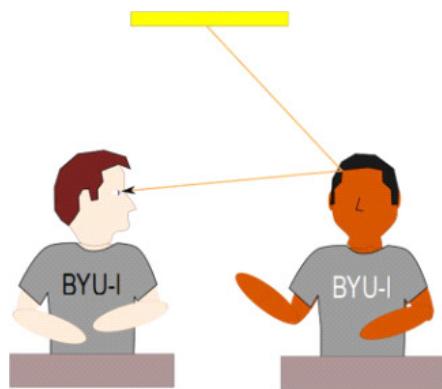
12.5 Reflections, Objects, and seeing

Armed with the law of reflection, we can start to understand how we see things. Using the ray concept, we can say that a ray of light must leave the light source. That ray then reflects from something. Suppose you look at the person sitting next to you in class. We should wonder, how is it that we can see them? We can only detect (see) light that gets to our eyes. Let's trace the light from it's source (the light fixture in our room) to our eyes to see how we see our neighbors.

Light comes from the ceiling lights and goes in all directions. Some of that light hits our neighbor. And some of that light that hits our neighbor will reflect. But is the person a specular or diffuse reflector? Once again, we can only give an answer relative to the wavelength of light. For visible light, your neighbors do not look like mirrors. They are diffuse reflectors. Light bounces off of them in every direction. Some of that light that reflects from your neighbor reflects in the direction of your eyes. That light can be detected. Your eye is designed to take this diverging set of rays and condense it into a picture of the person that your brain can interpret.

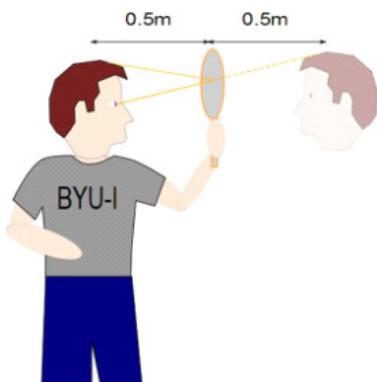


We tend to not draw the rays that bounce off the diffuse reflector but that don't get to our eyes, because we don't see them. So a ray diagram is usually much simpler.



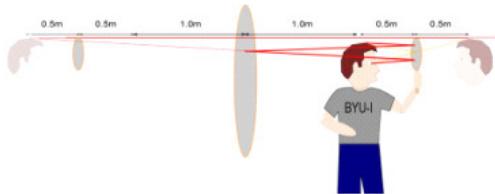
This is easy to understand, but we must keep in mind the wildly fluctuating superposition of waves that is masked by our macroscopic view.

We can use the idea of a ray diagram to solve problems. Suppose you hold a mirror half a meter in front of you and look at your reflection. Where would the reflection appear to be?



Knowing that rays travel in straight lines and that our mind interprets rays as going in straight lines, then we can use rays to see where the light appears to

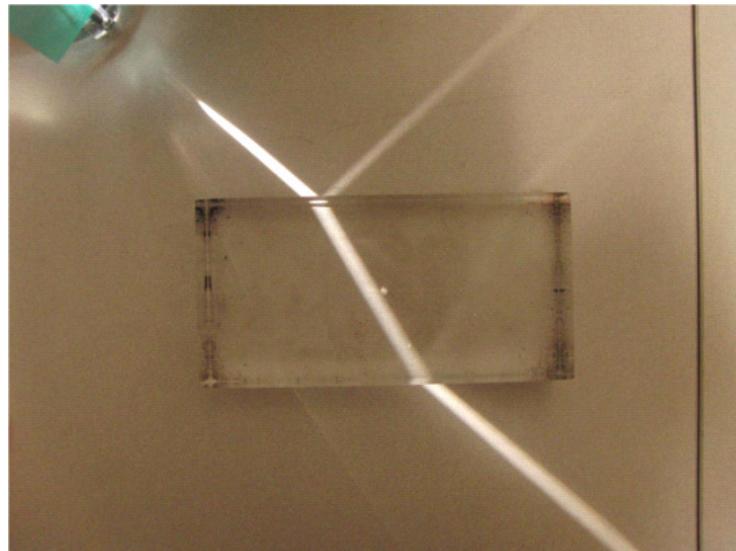
be from. The image is half a meter behind the mirror. Now suppose we look at an image of that image in a mirror behind us.



The ray diagram makes it easy to see that the image will appear to be 2 m behind the big mirror. You might not feel that this was so easy, but you might find it is not so bad in a problem in your near future.

Chapter 13

13 Refraction 3.1.3, 3.1.4



We have studied reflection of light. Now let's consider what happens when light enters a material like glass or water.

Fundamental Concepts

- Refraction
- Total internal reflection

In the last lecture we split up electromagnetic waves into different kinds of light. We categorized the different kinds of light by noting they have different

wavelength ranges. We found visible light, of course, but also infrared and ultraviolet and x-rays and radio waves. Not all surfaces reflect all the light. Some, like the lenses shown below, reflect some light at visible wavelengths, but are transparent so most of the light travels through them.



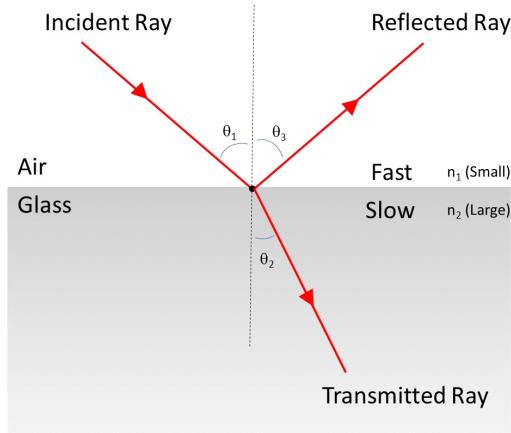
We need a way to deal with transparent materials. This is tricky, because different wavelengths of light penetrate different materials in different ways. As an example, this is also a lens



Infrared lens, but visible mirror (Image courtesy US Navy, image in the public domain)

but it clearly is not transparent at visible wavelengths. It is transparent in the infrared. So what might be transparent at one wavelength might not be at another.

When light travels into a material, we say it is transmitted. This is nothing new. Light is a wave, so the part that goes through is called the transmitted part. The situation is shown schematically below.



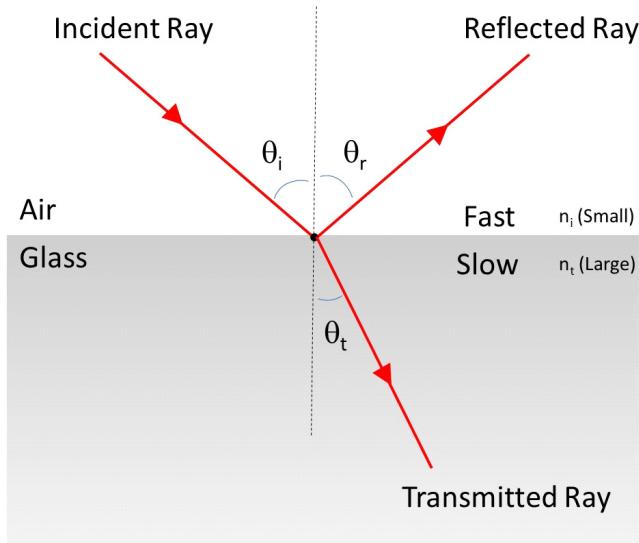
In the figure we see a ray incident on an air-glass boundary. Some of the light is reflected just as we saw before. But some passes into the glass. Notice that the angle between the normal and the new transmitted ray is *not* equal to the incident ray. We say the ray has been bent or *refracted* by the change in material. Many experiments were performed to find a relationship between the incident and the refracted angles. It was found that

$$\frac{\sin(\theta_2)}{\sin(\theta_1)} = \frac{v_2}{v_1} = \text{constant} \quad (13.1)$$

Many optics books write this as

$$\frac{\sin(\theta_t)}{\sin(\theta_i)} = \frac{v_t}{v_i} = \text{constant} \quad (13.2)$$

where the subscript *i* stands for “incident” and the subscript *t* stands for “transmitted.”

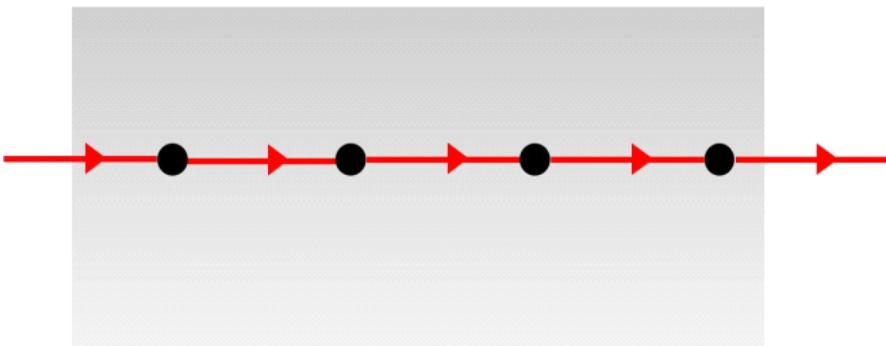


Note that we are using the fact that the speed of light changes in a material. We should probably recall why this should occur

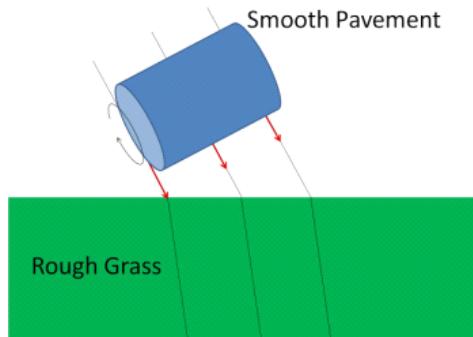
13.1 Speed of light in a material

In a vacuum, light travels as a disturbance in the electromagnetic field with nothing to encounter. In a material (like glass) the light waves continually hit atoms. We have not studied antennas, but I think many of you know that an antenna works because the electrons in the metal act like driven harmonic oscillators. The incoming radio waves drive the electron motion. Here each atom has electrons, and the atoms act like little antennas, their electrons moving and absorbing the light. But the atom cannot keep the extra energy¹, so it is readmitted. It travels to the next atom and the process repeats. Quantum mechanics tells us that there is a time delay in the re-emission of the light. This re-emitted light mixes with what is left of the original wave. And this causes the propagation of the combined light wave to slow down. Thus the speed of light is slower in a material.

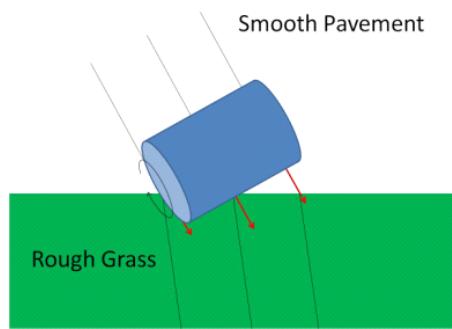
¹If you get to take Modern Physics (PH279) and Quantum Physics (PH433) you will learn how this works.



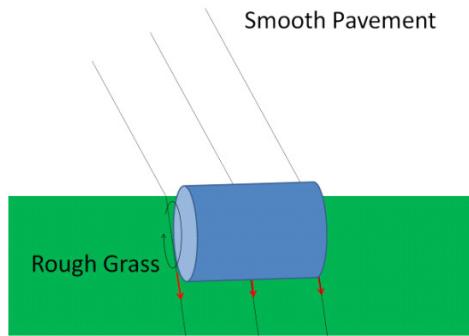
As a mechanical analog, consider a rolling barrel.



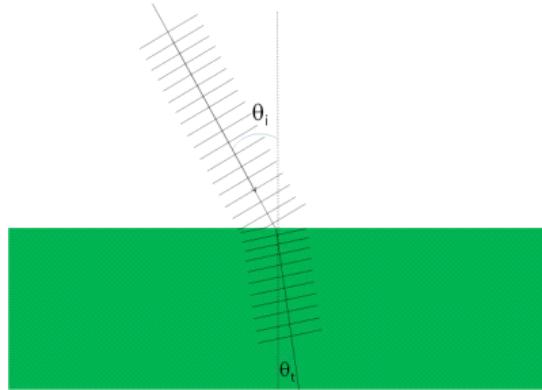
As the barrel rolls from a flat low-friction concrete to a higher-friction grass lawn, the friction slows the barrel. If the barrel hits the lawn parallel to the boundary (so its velocity vector is perpendicular to the boundary), then the barrel continues in the same direction at the slower speed. But if it hits at an angle, the leading edge is slowed first.



This makes the trailing edge travel faster than the leading edge, and the barrel turns slightly.



We expect the same behavior from light.



We can see that the left hand side of the wave hits the slower (green) material first and slows down. The rest of the wave front moves quicker. The result is the turning of the wave.²

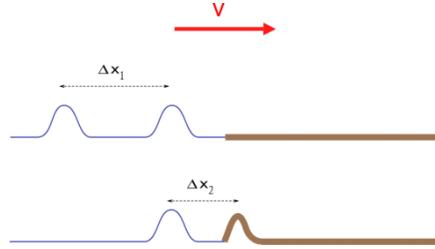
13.2 Change of wavelength

We have found that when a wave enters a material, its speed may change. But we remember from wave theory

$$v = \lambda f \quad (13.3)$$

But it is time to review: does λ change, or does f change? If you will recall, we found that the change in speed at the boundary changes the wavelength. Recall that if we go from a fast material to a slow material, the forward part of the wave slows and the rest of the wave catches up to it.

²Once again this is a bit of a simplification, but it will do for now. If you are lucky enough to take a junior level optics class, you will revisit this.



This will compress pulses, and lower the wavelength. Now that we know more about light we can also argue that f cannot change because

$$E = hf$$

If f changed, then we would either require an input of energy or we would store energy at the boundary because

$$\Delta f = \frac{\Delta E}{h}$$

This can't be true. If the wavelength changes, there is no change in energy.

And since

$$v_i = \lambda_i f$$

and

$$v_t = \lambda_t f$$

then the ratio

$$\frac{v_i}{v_t} = \frac{\lambda_i}{\lambda_t}$$

and we have a solution for the wavelength in the material

$$\lambda_t = \lambda_i \frac{v_t}{v_i}$$

which agrees with our previous analysis.

13.3 Index of refraction and Snell's Law

Because the equation

$$\frac{\sin(\theta_i)}{\sin(\theta_t)} = \frac{v_t}{v_i} = \text{constant} \quad (13.4)$$

has a constant ratio of velocities, it is convenient to define a term that represents that ratio. We already have a concept that can help. The *index of refraction* is just such a term. It assumes that one speed is the speed of light in vacuum, c .

$$n \equiv \frac{c}{v} \quad (13.5)$$

Then for our example

$$\frac{\sin(\theta_i)}{\sin(\theta_t)} = \frac{1}{n} \quad (13.6)$$

Suppose we don't have a vacuum (or air that is close to a vacuum). We can write our formula as

$$n_i \sin(\theta_i) = n_t \sin(\theta_t) \quad (13.7)$$

where we have

$$n_i \equiv \frac{c}{v_i} \quad (13.8)$$

and

$$n_t \equiv \frac{c}{v_t} \quad (13.9)$$

This is called *Snell's law of refraction*.

Using the index of refraction we can write our equation relating the ratio of velocities and wavelengths as

$$\frac{v_i}{v_t} = \frac{\lambda_i}{\lambda_t} = \frac{\frac{c}{n_i}}{\frac{c}{n_t}} = \frac{n_t}{n_i} \quad (13.10)$$

which gives

$$\lambda_i n_i = \lambda_t n_t \quad (13.11)$$

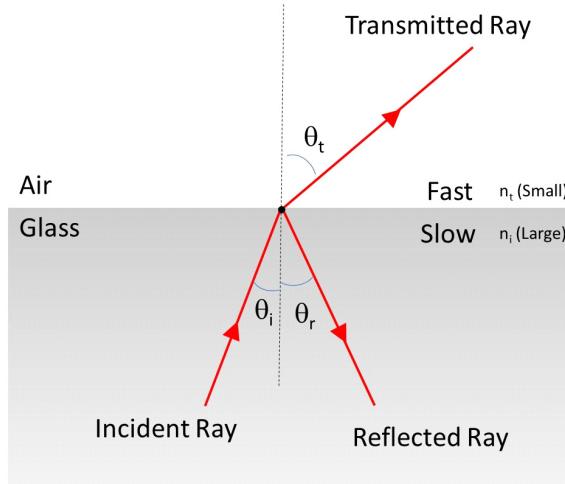
and if we have vacuum and a single material we can find the index of refraction from

$$n = \frac{\lambda}{\lambda_{in}} \quad (13.12)$$

where $\lambda_{in} = \lambda_t$ is the wavelength in the material and $\lambda = \lambda_i$ is the wavelength outside the material.

13.4 Total Internal Reflection

Up to now we have assumed that light was coming from a region of low index of refraction (air) into a region of high index of refraction (water or glass). We should pause to look at what can happen if we go the other way.



We start with Snell's law

$$n_i \sin \theta_i = n_t \sin \theta_t \quad (13.13)$$

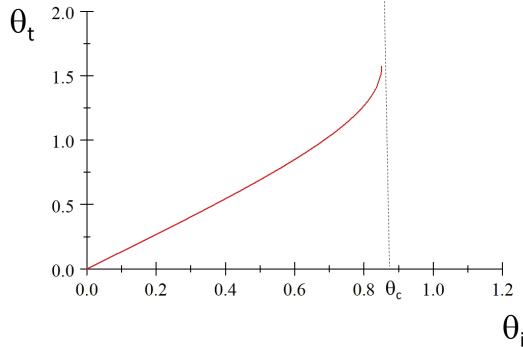
but this time $n = n_i$ and $n_t = 1$ so

$$n \sin \theta_i = \sin \theta_t \quad (13.14)$$

which gives

$$\theta_t = \sin^{-1} (n \sin \theta_i) \quad (13.15)$$

If we take $n = 1.33$ (water) we can plot this expression as a function of θ_i



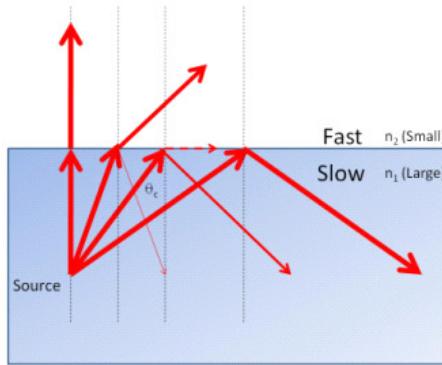
we see that at $\theta_i = 0.850\,91 \text{ rad}$ (48.754°) the curve becomes infinitely steep.
If we use this value in our equation this gives

$$\theta_t = \sin^{-1} (n \sin (0.850\,91)) \quad (13.16)$$

$$= 1.570\,8 \text{ rad} \quad (13.17)$$

$$= 90^\circ \quad (13.18)$$

The light skims along the edge of the water!



We can find the value of θ_i that makes this happen without graphing. Set $\theta_t = 90^\circ$ then

$$\theta_i = \theta_c \equiv \sin^{-1} \left(\frac{1}{n} \right) \quad (13.19)$$

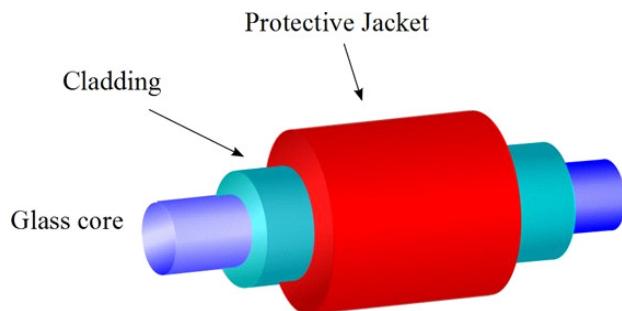
We give this value of θ_i a special name. It is the *critical angle* for internal reflection. But what happens if we go farther than this ($\theta_i > \theta_c$). We will no longer have a transmitted ray. But the wave energy has to go somewhere! The ray will be reflected. This is why when you dive into a pool and look up, you see a region of the roof of the pool area (or sky) but off to the side of the pool the surface looks mirrored. It is also why you sometimes see the sides of a fish tank appear to be mirrored when you look through the front.



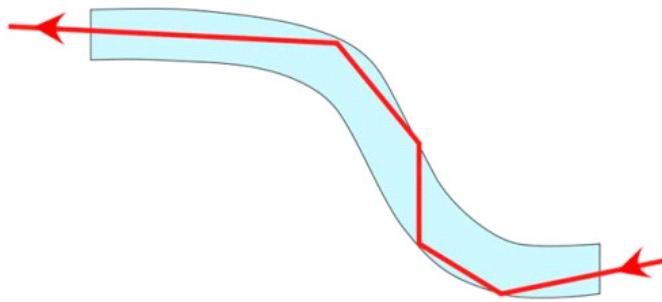
More importantly, it is why cut gems (like diamonds) sparkle. They capture the light with facets that are cut at angles that create total internal reflection. The light that enters the gem comes back out the front.

13.5 Fiber Optics

Beyond pretty pebbles, this effect is very useful! It is the heart and soul of fiber optics.



An interior material with a lower index of refraction is inclosed in a cladding with a higher index. This creates a light pipe that traps the light in the fiber.



Modern fibers don't always have a hard boundary. The fibers have a gradual change in index of refraction that changes the direction of the light gradually. This keeps the light in the fiber but tends to direct along the fiber so the beam is not crisscrossing as it goes.

The cutting edge of fiber design today uses hollow fibers or fibers filled with different index material.



Hollow-Core Fiber (Courtesy Defense Advanced Research Projects Agency (DARPA), image in the public domain)

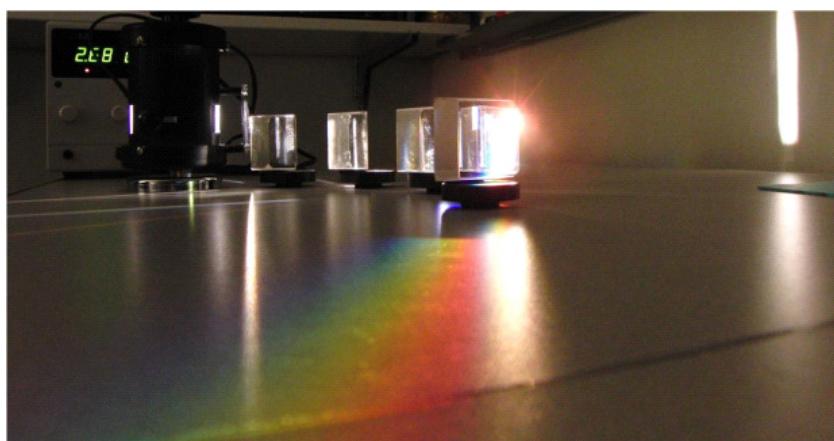
Chapter 14

14 Dispersion and Huygen's Principle 3.1.5 3.1.6

Fundamental Concepts

- The index of refraction is wavelength dependent
- That different wavelengths bend different amounts when refracted is called *dispersion*
- White light is a superposition of many other frequencies

14.1 Dispersion



Who hasn't played with a prism? We immediately recognize a rainbow. But why does the prism make a rainbow? The secret lies in the nature of the refractive index. We know that some wavelengths will pass through a material, say, a piece of glass. But some (like infrared wavelengths) won't. But there is more to this. Not all the wavelengths that do pass through the glass will have the same transmitted speed. The speed in the material depends on the wavelength! And the index of refraction depends on the speed.

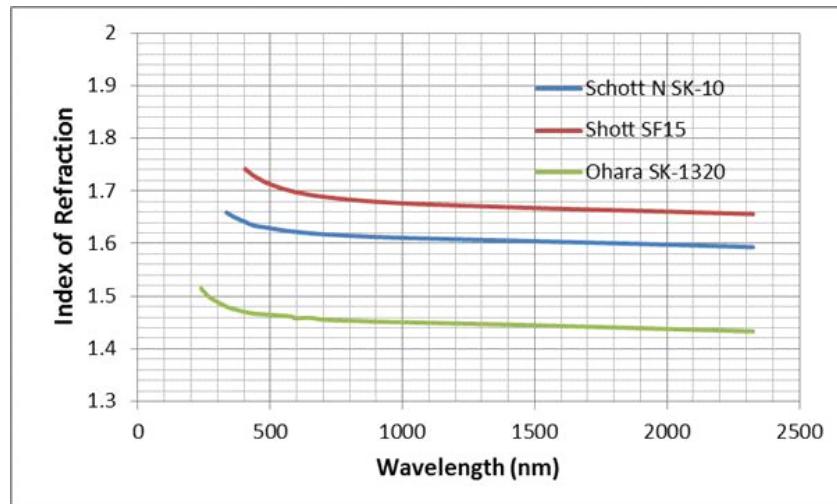
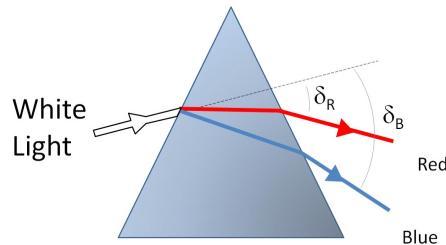


Figure 14.1: Index of refraction as a function of wavelength (Ohara optical glass <http://www.oharacorp.com/fused-silica-quartz.html> data and Schott optical glass data http://www.uqgoptics.com/materials_glasses_schott.aspx)

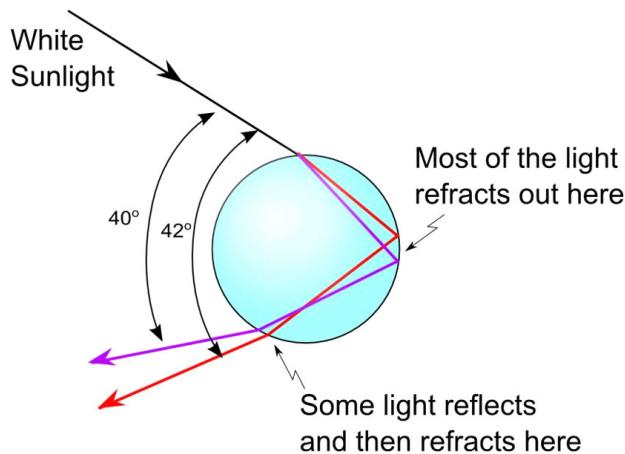
Notice that in the figure, the index of refraction depends on wavelength. This means that as light enters a material, different wavelengths will be refracted at different angles. White light is a superposition of many light waves with many different wavelengths. Thus white light is pulled apart by refraction into a rainbow. This process is called *dispersion*.

The graph tells us that blue light bends more than red light.



We call the change in direction measured from the original direction of travel, δ , the *angle of deviation*. The colors we can see are called the visible spectrum. Note that this is a second way to make a visible light spectrum. The first way used a diffraction grating and the wave nature of light. Although this dispersive element (prism) method of making a spectrum uses the ideas of geometric optics, it is still the wave nature of light that makes the waves of different wavelengths refract differently.

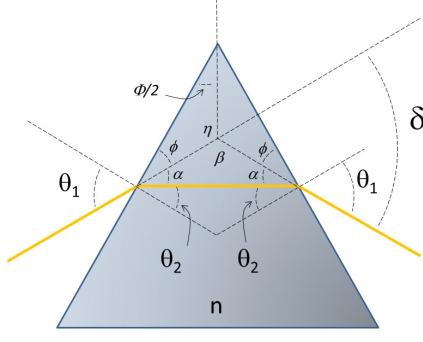
Let's look at a natural rainbow. The dispersion is caused by small droplets of water. The white sunlight enters the drop and is dispersed. It bounces off the back of the drop and then leaves the drop, again being dispersed. Red light leaves the drop at about 42° from its input direction, and blue light leaves at about 40° .



This effectively separates all the different wavelengths and we see a rainbow at angles between 40° and 42° from the incoming light direction.

14.1.1 Calculation of n using a prism

Let's do a problem using the idea of refraction in a prism. Let's find the index of refraction of a the prism material. Suppose we make a prism as shown. We know the angle Φ and can measure the exit angle δ . In terms of these two variables, what is n ?



R. Todd Lines

Using the notation indicated in the figure, we choose θ_1 such that the interior ray is horizontal.¹ This is a refraction problem, so we will want to use Snell's law.

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

Thus, we need to find θ_1 and θ_2 . Knowing Φ and δ , and realizing $n_1 = 1$, we can find θ_2 and θ_1 . Then use

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

to find n_2 . Let's see one way to do this. In geometry it is fair game to extend lines and even make some new lines of our own. By doing this we can realize that

$$\theta_1 = \theta_2 + \alpha \quad (14.1)$$

and that

$$\delta = 180 - \beta \quad (14.2)$$

and it is also true that

$$180 = \beta + 2\alpha \quad (14.3)$$

Then

$$\delta = 2\alpha \quad (14.4)$$

and

$$\alpha = \frac{\delta}{2} \quad (14.5)$$

Now also realize that

$$90 = \alpha + \theta_2 + \phi \quad (14.6)$$

¹WARNING! in the upcoming homework problem you can't make the same assumptions!

and

$$180 = \Phi + 2\alpha + 2\phi \quad (14.7)$$

or

$$90 = \frac{\Phi}{2} + \alpha + \phi \quad (14.8)$$

Then

$$\alpha + \theta_2 + \phi = \frac{\Phi}{2} + \alpha + \phi \quad (14.9)$$

$$\theta_2 = \frac{\Phi}{2} \quad (14.10)$$

We can put these in our equation for θ_1

$$\theta_1 = \theta_2 + \alpha \quad (14.11)$$

$$= \frac{\Phi}{2} + \frac{\delta}{2} \quad (14.12)$$

$$= \frac{\Phi + \delta}{2} \quad (14.13)$$

Now we can use Snell's Law

$$\sin(\theta_1) = n \sin(\theta_2) \quad (14.14)$$

$$\sin\left(\frac{\Phi + \delta}{2}\right) = n \sin\left(\frac{\Phi}{2}\right) \quad (14.15)$$

then

$$n = \frac{\sin\left(\frac{\Phi}{2}\right)}{\sin\left(\frac{\Phi + \delta}{2}\right)} \quad (14.16)$$

and since we know Φ and δ , we can find n .

14.2 Huygen's Principle

We have learned the laws of reflection and refraction. But how does the light know to bounce or bend the way it does? To understand what is going on at a fundamental level we need to remember that light is a wave. And waves can hit and splash off of objects. Sound waves could bounce off of things. So can light waves. Only the light waves that we can see are very small. So they bounce off of very small things like atoms. Suppose we send our light into a material. It will hit the atoms and bounce off. We know this, but there is more. The light bounces off the atoms but this makes a new light wave coming from the atoms. And that new wave is spherical (think that most waves start spherical). If we have many atoms, the light will hit many atoms and make many little spherical *scattered* waves. An early researcher had an idea of how to find the direction of the light using this underlying wave nature of light. We could follow

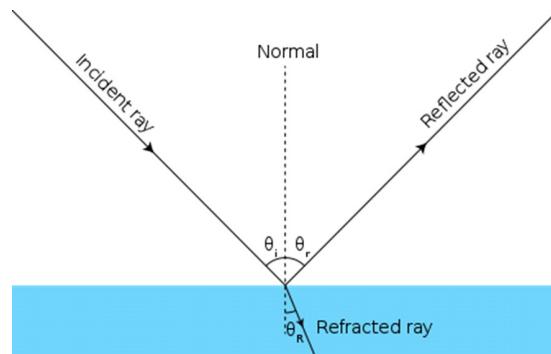
the light path by finding the little spherical scattered waves and look for where their crests line up. This would be constructive interference, and this is where the ray of light would end up!



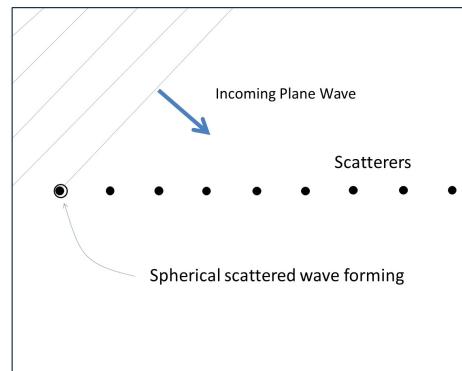
This early researcher was named Huygen, so this method of finding the new wave is called *Huygen's principle*.

We know that light should reflect off a surface.

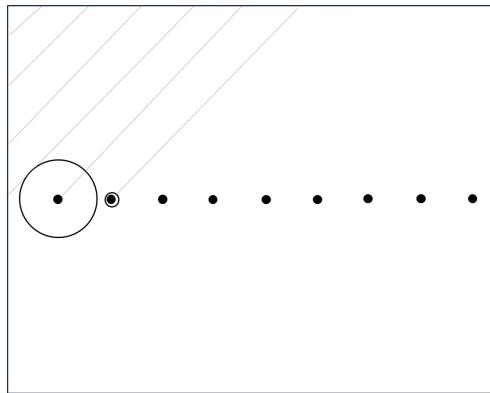
$$\theta_i = \theta_r$$



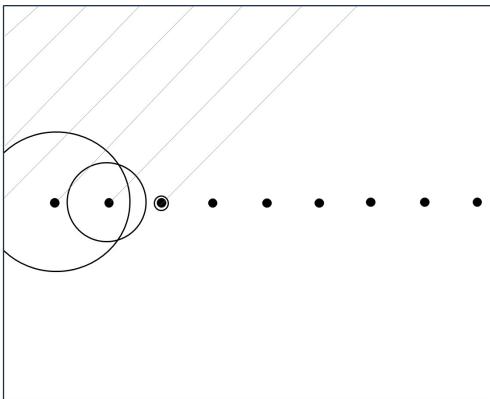
Let's see how Huygen's principle gives this law of reflection. Think about the top row (or about $\lambda/2$ thickness) of scatterers. In the next figure we have drawn such a set of scatters with an incoming plane wave.



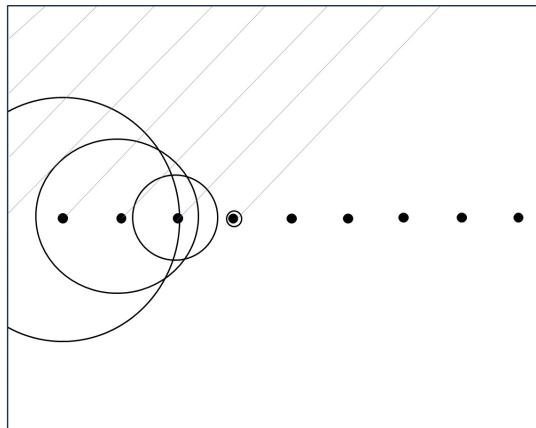
On the left, the plane wave crest is striking a scatter and has started a spherical secondary wave. One period later the situation might look like this



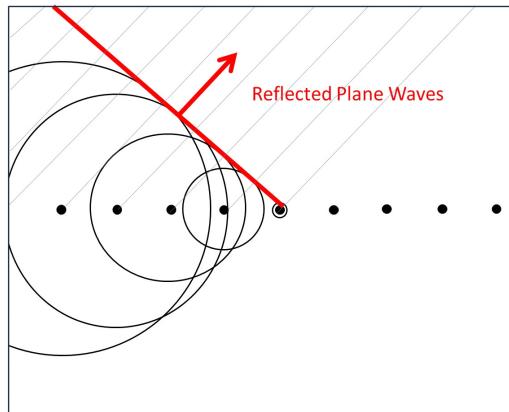
where now the first crest has moved on to the second scatterer and the second crest is at the first scatterer. Notice that the first scatterer's spherical secondary wave has grown. This process continues as the wave moves onward. In another period, we would have three scattered waves



and in another period, four.

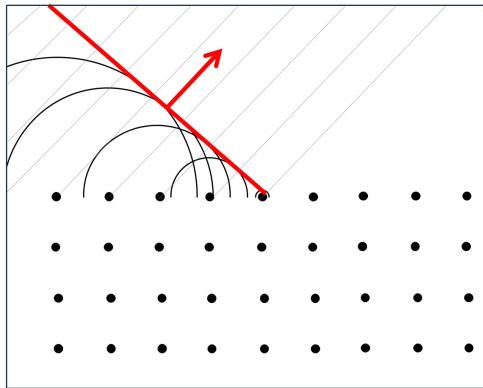


and five.

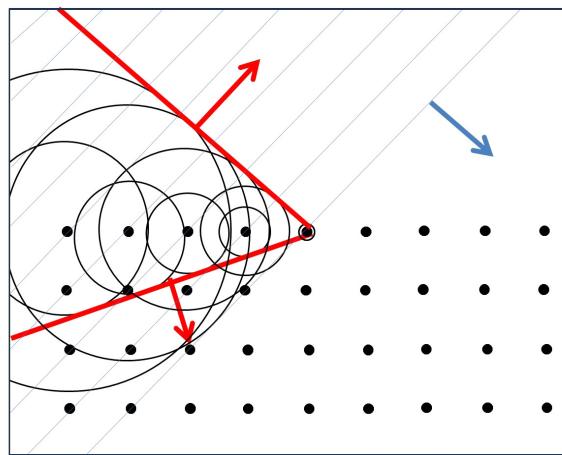


By looking for where the wave crests line up we can find the new crest of the reflected wave. By superimposing these scattered waves we the reflected wave is formed.

We also know *Snell's law of refraction*. But let's see how Huygen's principle leads to our scattering model to show this is true. In our discussion on reflection we just used the top layer of scatterers, but we know that there are more scatterers in a solid.

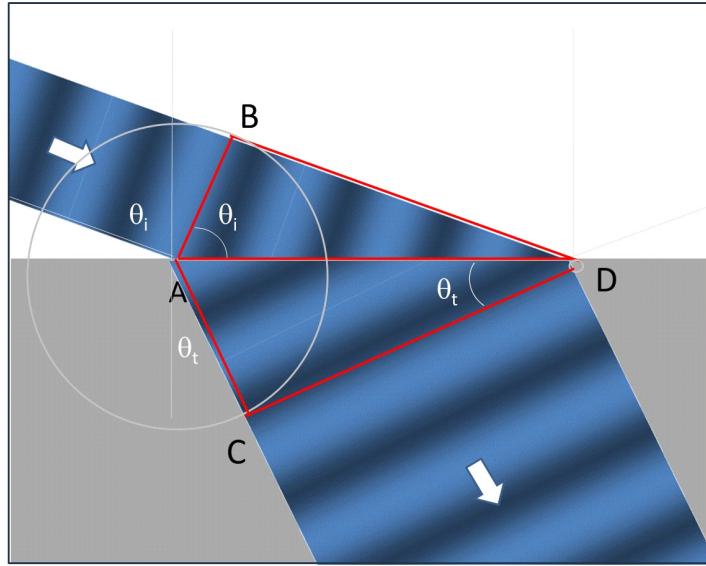


So, what happens inside the material with all the other layers of scatterers? Of course our secondary waves are kind of spherical shaped, so the situation is something like this



and we see that we are forming another wave front *inside the material*. Our scattering model predicts the transmitted wave as well as the reflected wave.

We can see that Snell's law follows from basic geometry just as did the law of reflection. Snell and others used measurements to find Snell's law. But now that we understand the geometry we can use it to find Snell's equation. In the next figure you can see an incoming wave. We have marked a wavefront \overline{AB} that is just starting to hit the interface between the two materials. And we have marked another wavefront, \overline{CD} be the transmitted wave front that has just become entirely inside the new (transmitted) material.



The time it takes the wave front to go from B to D would be

$$\Delta t_{BD} = \frac{\overline{BD}}{v_i}$$

but because the new transmitted wave is slower, the transmitted wavefront has only gone the distance \overline{AC} in the same time

$$\Delta t_{AC} = \Delta t_{BD} = \Delta t = \frac{\overline{AC}}{v_t}$$

We have two triangles. And they share a side \overline{AD} . And they are right triangles. So the two triangles must be similar triangles. Now let's look at our incident angle θ_i that of course we measure from a normal line. We are going to want the $\sin \theta_i$ for Snell's law and we can see that

$$\sin \theta_i = \frac{\overline{BD}}{\overline{AD}}$$

and also

$$\sin \theta_t = \frac{\overline{AC}}{\overline{AD}}$$

and we can solve both for $1/\overline{AD}$

$$\begin{aligned}\frac{\sin \theta_i}{\overline{BD}} &= \frac{1}{\overline{AD}} \\ \frac{\sin \theta_t}{\overline{AC}} &= \frac{1}{\overline{AD}}\end{aligned}$$

and set them equal so that

$$\frac{\sin \theta_i}{\overline{BD}} = \frac{\sin \theta_t}{\overline{AC}}$$

but

$$\begin{aligned}\overline{BD} &= v_i \Delta t \\ \overline{AC} &= v_t \Delta t\end{aligned}$$

so

$$\frac{\sin \theta_i}{v_i \Delta t} = \frac{\sin \theta_t}{v_t \Delta t}$$

And the Δt cancels from both sides

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_t}{v_t}$$

We can multiply both sides by c , the speed of light

$$c \frac{\sin \theta_i}{v_i} = c \frac{\sin \theta_t}{v_t}$$

to achieve

$$n_i \sin \theta_i = n_t \sin \theta_t$$

which is Snell's law.

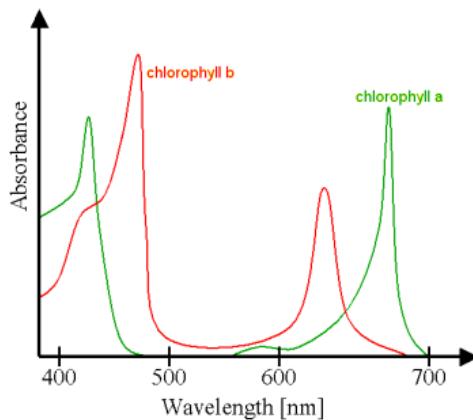
Huygen's principle is used many times in the study of optics to help us understand the geometry of an optics problem and from that form an equation that makes calculations possible!

14.3 Filters and other color phenomena

Of course, we have assumed without statement, that white light is made up of all the colors of the rainbow. We should ask why a red shirt is red, or why passing light through a green film makes the light look green as it leaves.

Both of these phenomena are examples of removing wavelengths from white light.

In the case of the red shirt, the red dye in the cloth absorbs all of the visible colors except red. The red is reflected, so the shirt looks red. The filter is much the same. The green pigment in the film causes nearly all visible colors to be absorbed except green. So only green light is transmitted. This is why leaves are green. Chlorophyll absorbs red and blue wavelengths, so the green is reflected or transmitted.



Chlorophyll Spectrum (Public Domain image courtesy Kurzon)

Chapter 15

15 Polarization and Image formation 3.1.7, 3.2.1

Fundamental Concepts

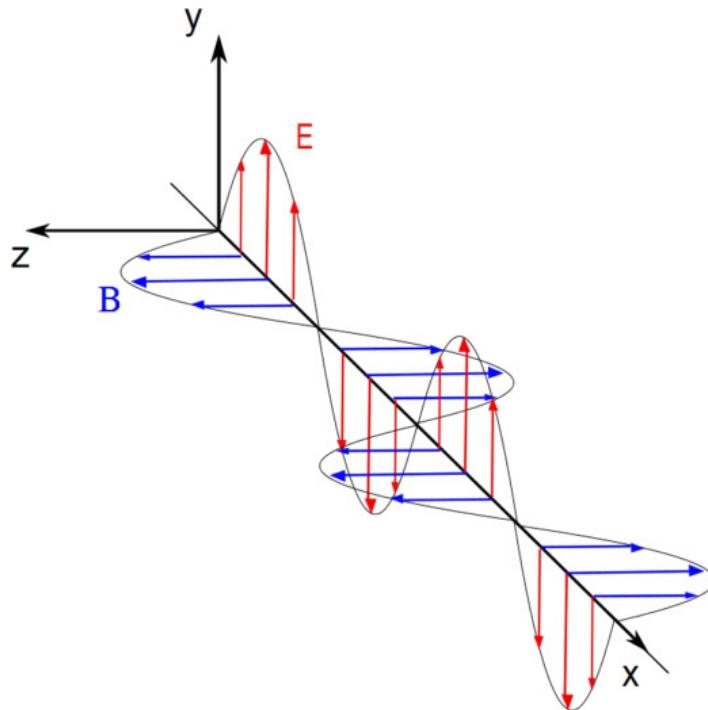
- The direction of the electric field in a plane wave is called the polarization direction.
- Natural light is usually a superposition of many waves with random polarization directions. This light is called unpolarized light.
- Some materials allow light with one polarization to pass through, while stopping other polarizations. The polaroid is one such material polaroids. will have a final intensity that follows the relationship $I = I_{\max} \cos^2(\theta)$
- Light reflecting off a surface may be polarized because of the absorption and re-emission pattern of light interacting with the material atoms.
- Scattered light may be polarized because of anisotropies in the scatterers.
- Birefringent materials have different wave speeds in different directions. This affects the polarization of light entering these materials.
- An image is a diverging set of rays in a recognizable pattern
- Images can be formed by reflection and refraction

15.1 Polarization of Light Waves

We said much earlier in our study of light that it was a transverse wave and that the medium for this wave is the electromagnetic field. Understanding electric

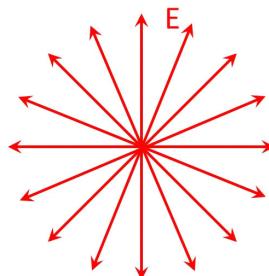
and magnetic fields is a topic for Principles of Physics III (PH220) so we can't go to deep into electric and magnetic fields here. Let's assume that all of space (the whole universe) is filled with this stuff we call electric and magnetic fields. And let's suppose that these fields can be a medium for waves. And let's identify these waves as light.

The light waves are transverse waves. So when we look at a light wave in detail we will see the pieces of the electric and magnetic fields experiencing simple harmonic motion much like the parts of our spring for waves on springs or air molecules for sound waves. It turns out that the electric field medium parts move perpendicular to the direction the wave goes, and also perpendicular to the direction the magnetic field medium parts go. This makes them transverse waves. We will now show some implications of this fact. In a course in electromagnetic theory, we often draw light as in the figure below.



and that looks a bit like our waves on strings or springs only with the electric field part going up and down and the magnetic field part going back and forth.

In what follows we will ignore the magnetic field (marked in the figure as B) because it's amplitude is very much smaller than the electric field amplitude. So, we will look at the E field and notice that it goes up and down in the figure. But we could have light in any orientation. If we look directly at an approaching beam of light we would "see" many different orientations as shown in the next figure.



When light beams have waves with many orientations, we say they are *unpolarized*. But suppose we were able to align all the light so that all the waves in the beam were transverse waves in the same orientation. Say, the one in the next figure.

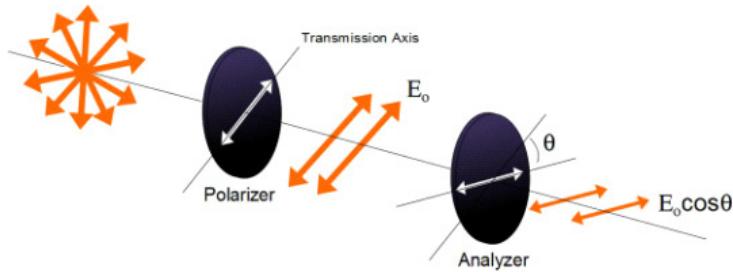


Then we would describe the light as *linearly polarized*. The plane that contains the E -field is known as the *polarization plane*.

15.1.1 Polarization by removing all but one wave orientation

One way to make polarized light is to remove all but one orientation of an unpolarized beam. A material that does this at visible wavelengths is called a *polaroid*. It is made of long-chain hydrocarbons that have been treated with iodine to make them conductive. The molecules are all oriented in one direction by stretching during the manufacturing process. The molecules have electrons that can move when light hits them. They can move farther in the long direction of the molecule, so in this direction the molecules act like little antennas. The molecules' electrons are driven into harmonic motion along the length of the molecule. This takes energy (and therefore, light) out of the beam. Little electron motion is possible in the short direction of the molecule, so light is given a preferential orientation. The light is passed if it is perpendicular to the long direction of the molecules. This direction is called the *transmission axis*.

We can take two pieces of polaroid material to study polarization.



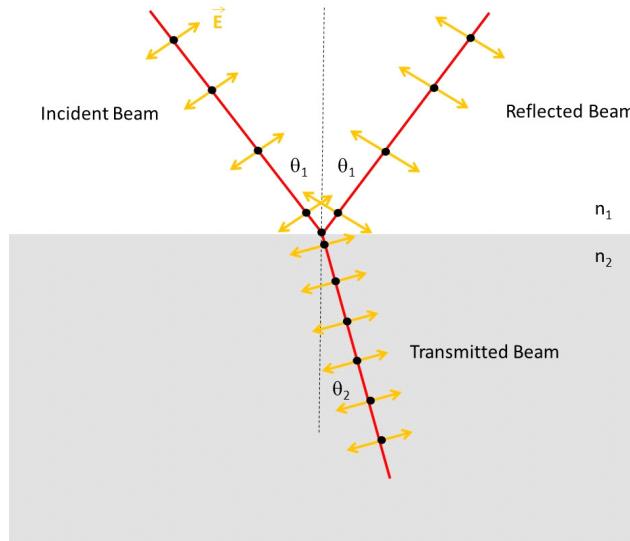
Unpolarized light is initially polarized by the first piece of polaroid called the *polarizer*. The second piece of polaroid then receives the light. This piece is called the *analyzer*. If there is an angular difference in the orientation of the transmission axes of the polarizer and analyzer, there will be a reduction of light through the system. We expect that if the transmission axes are separated by 90° no light will be seen. If they are separated by 0° , then there will be a maximum. It is not hard to believe that the intensity will be given by

$$I = I_{\max} \cos^2(\theta) \quad (15.1)$$

remembering that we must have a squared term because $I \propto E^2$.

15.1.2 Polarization by reflection

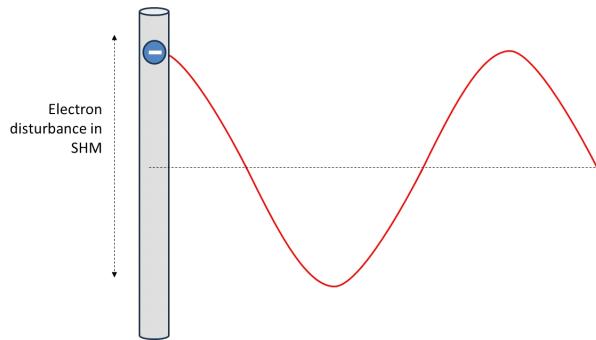
If we look at light reflected off of a desk or table through a piece of polaroid, we can see that at some angles of orientation, the reflection diminishes or even disappears! Light is often polarized on reflection. Let's consider a beam of light made of just two polarizations. We will define a plane of incidence. This plane is the plane of the paper or computer screen. This plane is perpendicular to the reflective or refractive surface in the figure below.



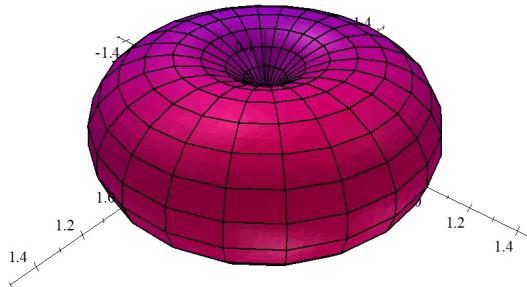
One of our polarizations is defined as parallel to this plane. This direction is represented by orange (lighter grey in black and white) arrows in the figure. The other polarization is perpendicular to the plane of incidence (the plane of the paper). This is represented by the black dots in the figure. These dots are supposed to look like arrows coming out of the paper.

When the light reaches the interface between n_1 and n_2 it drives the electrons in the medium into SHM. The perpendicular polarization finds electrons that are free to move in the perpendicular direction and re-radiate in that direction. The electron orbitals change shape and oscillate with the incoming electromagnetic wave.

The parallel ray is also able to excite SHM, but a electromagnetic analysis tells us that these little “antennas” will not radiate at an angle 90° from their excitation direction. And this is something new! Before when we used Huygens’ principle we thought of the scattered waves as little spheres. But it turns out that the light doesn’t go out in all directions equally. The disturbance for our wave is a moving charged particle like an electron. If an electron in a wire experiences simple harmonic motion there will be a wave in the electric field. But think, that wave will go out away from the electron perpendicular to the electron’s motion



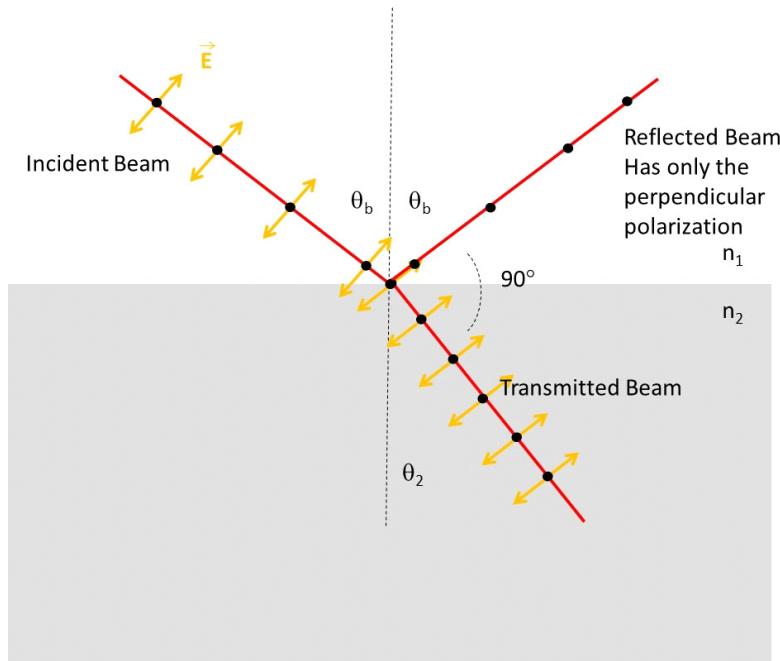
In the figure we can see the light wave won't travel up or down. The actual situation with our light striking a surface is a little more complicated. The pattern of light coming out from the little atomic oscillators looks more like a doughnut.



Angular dependence of S for a dipole scatterer.

But the point is that it's not completely spherical and none of the light wave goes up along the direction of the oscillation.

In the figure you can see that along the antenna axis, the field amplitude is zero. This means that the wave really does not go that direction. So in our case, the amount of polarization in the parallel direction decreases with the angle between the reflected and refracted rays until at 90° there is no reflected ray in the parallel direction.



The incidence angle that creates an angular difference between the refracted and reflected rays of 90° is called the *Brewster's angle* after its discoverer. At this angle the reflected beam will be completely linearly polarized.

We can predict this angle. Remember Snell's law.

$$n_i \sin \theta_i = n_t \sin \theta_t$$

Let's re-label the incidence angle $\theta_i = \theta_b$. We take $n_i = 1$ and $n_t = n$ so

$$n = \frac{\sin \theta_b}{\sin \theta_t}$$

Now notice that for Brewster's angle, we have

$$\theta_b + 90^\circ + \theta_t = 180^\circ$$

so

$$\theta_t = 90^\circ - \theta_b$$

so we have

$$n = \frac{\sin \theta_b}{\sin (90^\circ - \theta_b)}$$

ah, but we remember that $\sin (90^\circ - \theta) = \cos (\theta)$ so

$$n = \frac{\sin \theta_b}{\cos \theta_b}$$

Again we remember that

$$\tan \theta = \frac{\sin \theta}{\cos \theta}$$

so

$$n = \tan \theta_b \quad (15.2)$$

which we can solve for θ_b .

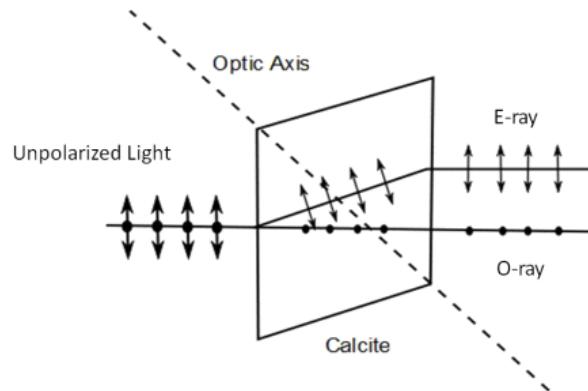
$$\theta_b = \tan^{-1} (n)$$

This phenomena is why we wear polarizing sunglasses to reduce glare.

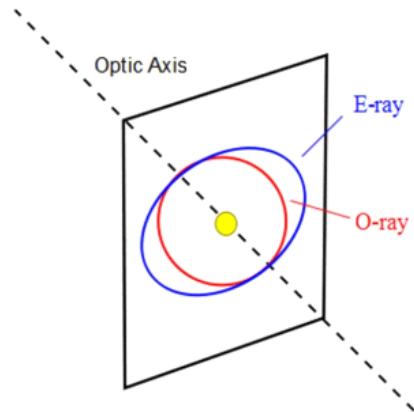
15.1.3 Birefringence

Glass is an amorphous solid—that is—it has no crystal structure to speak of. But some minerals do have definite order. Sometimes the difference in the crystal structure creates a difference in the speed of propagation of light in the crystal. This is not to hard to believe. We said before that the reason light slows down in a substance is because it encounters atoms which absorb and re-emit the light. If there are more atoms in one direction than another in a crystal, it makes sense that there could be a different speed in each direction.

Calcite crystals exhibit this phenomena. We can describe what happens by defining two polarizations. One parallel to the plane of the figure below, and one perpendicular.



With a careful setup, we can arrange things so the perpendicular ray is propagated just as we would expect for glass. We call this the *O-ray* (for *ordinary*). The second ray is polarized parallel to the incidence plane. It will have a different speed, and therefore a different index of refraction. We call it the *Extraordinary ray* or *E-ray*.



If we were to put a light source in a calcite crystal, we would see the *O*-ray send out a sphere of light as shown in the figure above. But the *E*-ray would send out an ellipse. The speed for the *E*-ray depends on orientation. There is one direction where the speeds are equal. This direction is called the *optic axis* of the crystal.



If our light entering our calcite crystal is unpolarized, then we will have two images leaving the other side that are slightly offset because the *O*-rays and *E*-rays both form images.

15.1.4 Optical Stress Analysis

Some materials (notably plastics) become birefringent under stress. A plastic or other stress birefringent material is molded in the form planned for a building or other object (usually made to scale). The model is placed under a stress, and the system is placed between two polaroids. When unstressed, no light is seen, but under stress, the model changes the polarization state of the light, and bands of light are seen.



15.1.5 Polarization due to scattering

It is important to understand that light is also polarized by scattering. It really takes a bit of electromagnetic theory to describe this. So for a moment, lets just comment that blue light is scattered more than red light. In fact, the relative intensity of scattered light goes like $1/\lambda^4$. This has nothing to do with polarization, but it is nice to know.

Now suppose we have long pieces of wire in the air, say, a few microns long. The pieces of wire would have electrons that could be driven into SHM when light hits them. If the wires were all oriented in a common direction, we would expect light to be absorbed if it was polarized in the long direction of the particles and not absorbed in a direction perpendicular to the orientation of the particles. This is exactly what happens when long ice particles in the atmosphere orient in the wind (think of the moment of inertia). We often get impressive halo's around the sun due to scattering from ice particles.

Rain drops also have a preferential scattering direction because they are shaped like oblate spheroids (not “rain drop shape” like we were told in grade school).

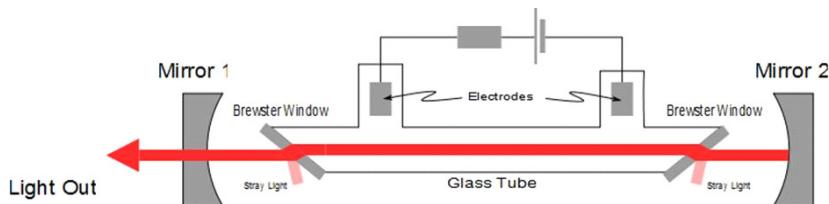
It is also true that small molecules will act like tiny antennas and will scatter light preferentially in some directions and not in others. This is called *Rayleigh scattering* and is very like small dipole antennas.

15.1.6 Optical Activity

Some substances will rotate the polarization of a beam of light. This is called being *optically active*. The polarization state of the light exiting the material depends on the length of the path through the material. Your calculator display works this way. An electric field changes the optical activity of the liquid crystal. There are polarizers over the liquid crystal, so sometimes light passes through the display and sometimes it is black.

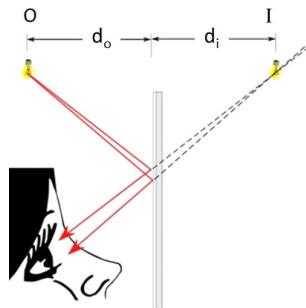
15.1.7 Laser polarization

One last comment. Lasers are usually polarized. This is because the laser light is generated in a *cavity* created by two mirrors. The mirror is tipped so light approaches it at the Brewster angle. Light with the right polarization (parallel to the plane of the drawing) is reflected back nearly completely, but light with the opposite polarization is not reflected at all. This reduces the usual loss in reflection from a mirror, because in one polarization the light must be reflected completely.



15.2 Imaging with Flat Mirrors

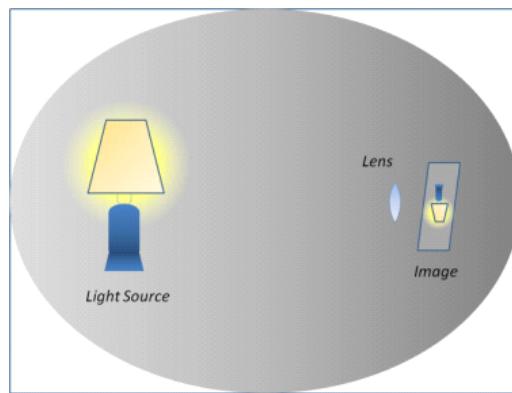
All of us have looked in a mirror at some time. We know what to expect. We see an image of ourselves. To study mirrors mathematically, we need to establish a sign convention and some standard notation.



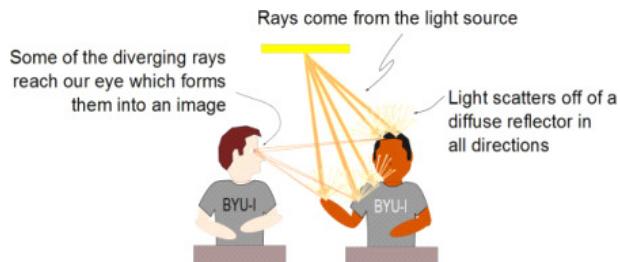
In the figure above, we have a person observing an object O in a mirror. The object is located at a distance s from the mirror. Just like for lenses, we will call this the *object distance*, d_o . The image appears to be located at a point I beyond the mirror a distance d_i . We will call this the *image distance*. Be careful, d_i looks like an incident of initial distance, but it's not. It is the distance from the mirror to the *image*. We know we see an image in the mirror, but to understand this we need to define what we mean by the word "image."

Images

Let's think about what an image is. Take a piece of paper and a lens, and hold up the lens in a darkened room that has some bright object in it. Move the lens or the paper back and forth, and at just the right distance, a miniature picture of the bright object will appear. We should think about what the word "picture" means in this sense.



We have talked about how we see objects. Remember the BYU-I guys from last time.



Our eyes gather rays that are diverging from the object because light has bounced off of the object. Our eyes intersect a diverging set of rays that form a definite pattern. That diverging set of rays forming a pattern is the picture of the object.

So when we say that the lens has formed a miniature picture of our object, we mean that the lens has somehow formed a diverging set of rays that form a pattern that looks like the pattern formed by the diverging set of rays coming from the object, itself. In other words, the object forms a diverging set of rays. And our lens forms a duplicate set of rays in the same pattern, so we see the same thing. The lens' version is smaller, upside down, and backwards, but it is still essentially the same pattern.

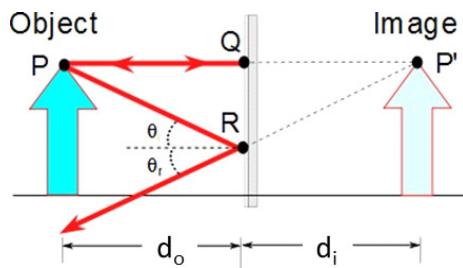
But how can a mirror form a diverging set of rays that form a pattern?

Well, the light that we see in a mirror first bounces off of something, say, a person. This makes a recognizable pattern in the light. Then the light reflects

off of the mirror. Then the light hits our eyes. So when we see something in a mirror we do see a diverging set of rays that form a pattern. But what we see makes more sense if we know how our brains processes the light signals that we see. Our eyes intercept rays of light diverging from an object. Our brain processes those rays as though they traveled only in straight lines. Our brain really believes light travels just in lines, no bouncing or reflecting. This means, if we can create a situation that makes rays diverge in the same way the object did, we will have an image of the object. But our brain will be wrong in where it thinks that image is. We will see an image as though it were behind the mirror.

15.2.1 Virtual Images

We will call We already know that mirrors often create what we call *virtual* images because the image appears to be created from diverging rays from behind the mirror. If we look behind the mirror, no rays exist there (if they did exist, they would not make it through the mirror!).



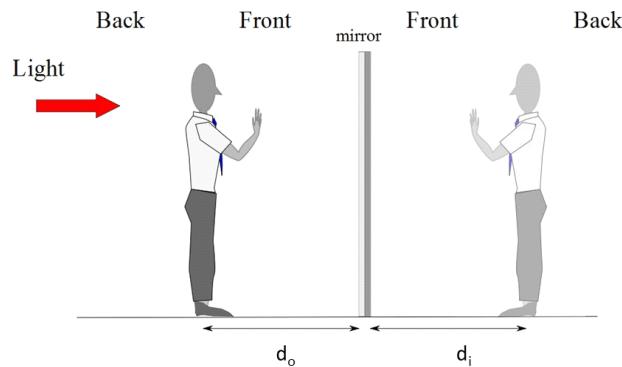
Let's look at a simple image as shown in the figure above. The object is an arrow. We could trace all the rays that diverge from this object and build a very nice representation of the arrow (like ray tracing-based computer graphics—the way movies like *Toy Story* are made) but that would take time and computation power. We only really need to use two rays, and to remember what the object looked like to figure out where the image will be.

We pick one ray from the top of the arrow that travels straight to the mirror. This ray will travel a distance s and bounce back. Using the law of reflection we know that it will bounce straight back along the path it came in on. We pick a second ray from point P that travels the path PR . This ray bounces off the mirror at an angle $\theta_r = \theta_i$. But our brain doesn't believe the light reflected. It thinks the light traveled straight paths. It traces the rays backwards along straight paths. And this makes it appear that the tip of the arrow is at position P' . The rays from the tip appear to travel the paths $P'P$ and $P'R$. The light doesn't really come from point P' . But it *appears* to come from point P' prime. We could use similar sets of two rays to find where the light from the middle of the arrow *appears* to come from, and for the bottom of the arrow, and all the

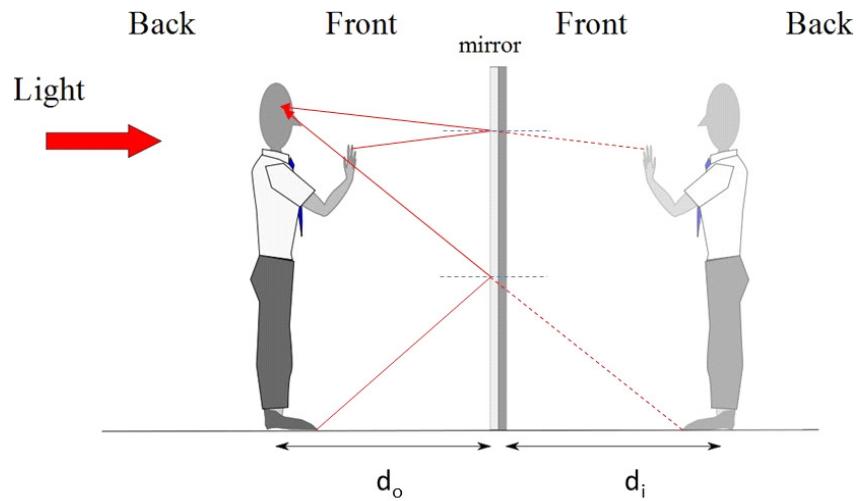
arrow parts. The collection of all these apparent locations is image location and the light that appears to come from this location is the image. But, and this is important! light doesn't come from P' . It only appears to come from P' . When an image is formed, but the light doesn't really come from the image location, we call that a *virtual image*.

15.2.2 Mirror reversal

Look into a mirror. Raise your right hand.



Your image raises what appears to be a left hand. It looks like a mirror switches the left and right sides of the image. But it might be better to note that the image raised hand is on the same side of the room as your real raised hand. Think of lying sideways on the ground in front of the mirror and raise a hand. Note that your image feet are on the same side of the room as your real feet. And note that as you put one hand up, the image hand also goes up. This is not exactly left-right reversal. What is happening?



The light from your raised hand would hit the mirror, reflect, and then be absorbed by your eye. Your brain tells you the light traveled in a straight line. So it looks like you have a raised hand in the mirror. But what has happened is that the light has traveled from your hand, to the mirror and back to your eye. This is not really left-right reversal. We need a name to describe what is really happening. The name is odd. To see why we use this name, notice that from your perspective (moving left to right across the figure) you first have the back part of your hand, then the front of your hand and then the front of the image hand, then the back of the image hand. We call this back-front reversal. It comes from the reversal in the order of parts of the object, like the front and back of your hand. So we say that a flat mirror performs a front-back reversal, not a left-right reversal.

Chapter 16

16 Images formed by Curved Mirrors and by Refraction 3.2.2, 3.2.3

Back in Newton's day, there were no electronic computers, or calculators, or mechanical adding machines. Early optics researchers did math with a pen and paper. This is one reason they liked small angle approximations. The approximation allowed them to do harder problems using easy math. And so long as the things they built worked, the approximations were good enough. We are going to use another approximation in this lecture. It is called the thin mirror approximation. It will make the math that describes mirrors, for say, telescopes, much easier. We will also introduce a way to draw curved mirrors and light that will tell you how the mirror works. We will use this drawing approach to do homework and examination problems. Professional optical designers use computer codes that do this drawing very accurately to help the optical engineer understand the system they are designing and to look for mistakes. So this drawing scheme is very useful. Because it makes optical systems so much easier to understand, let's start with the drawings and then work toward a mathematical description of thin mirrors.

Fundamental Concepts

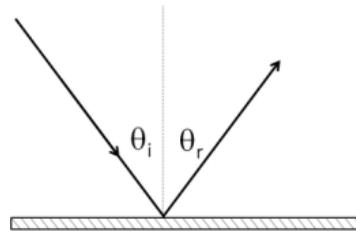
- Curved Mirrors can form images
- We can write the magnification in terms of the object and image distances
$$m = -\frac{d_i}{d_o}$$
- For “thin” mirrors $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$
- The focal length of a “thin” mirror is. $f = \frac{R}{2}$

- We can describe how a mirror operates with just three easy-to-draw rays.
- A semi-infinite bump (single sided lens) can be described by the equation

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{(n_2 - n_1)}{R}$$

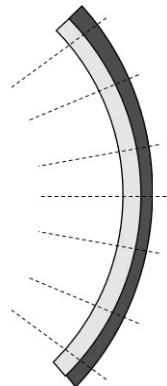
16.0.3 Concave Mirrors

Concave mirrors can form images. I'm sure you know that many telescopes are made with mirrors. We should see how this works. We recall the law of reflection

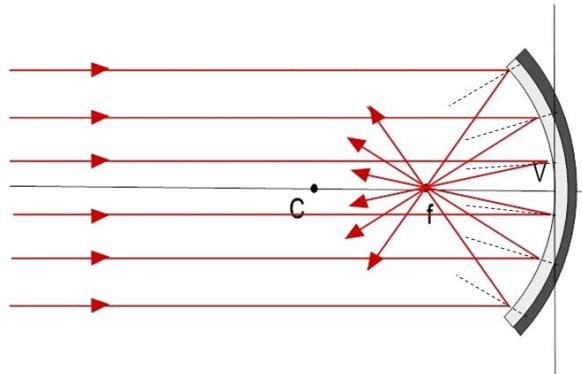


$$\theta_i = \theta_r$$

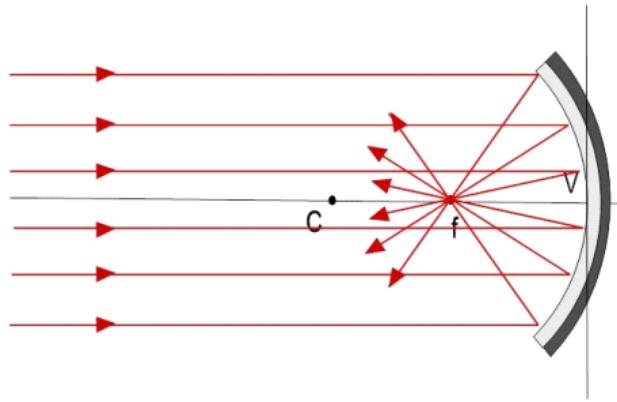
Armed with this, we can see what would happen if we curved the mirror surface. Each ray has a different normal due to the curvature of the mirror.



And the law of reflection must work at each one of these normal lines, so we get rays that reflect like this.



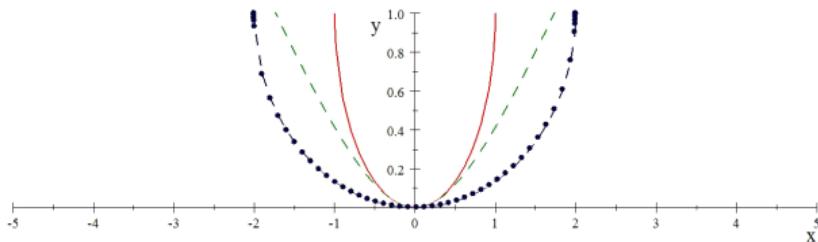
The result is that parallel rays all meet at a spot on the axis.



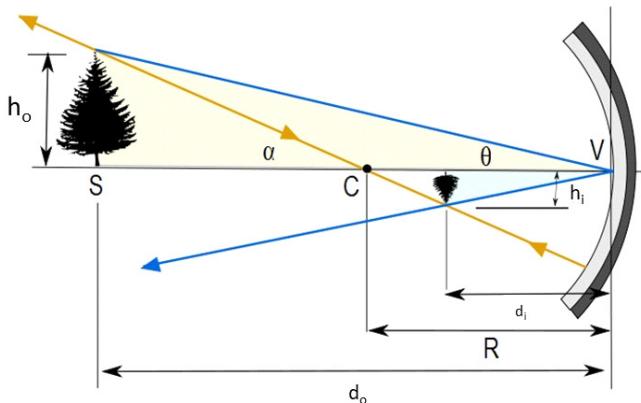
If we place something at this location (like a piece of paper), we would see it light up. With enough light we could start a fire! We have just re-invented the solar cooker. Look at the point where all the rays meet. Let's give this special place a name. We call it the focal point. It is the place where the rays come together if they originally came into the mirror parallel.

16.0.4 Paraxial Approximation for Mirrors

The correct shape of a mirror to make all the parallel rays meet at the focal point is more like a parabola, but parabolas are hard to machine or build. Spherical shapes are relatively easy to make. So we often see spherical mirrors used where we should have parabolas. This will work so long as we allow only rays that strike the mirror not too far from the principal axis. We can see why this works if we plot a sphere and a parabola (and a hyperbola). For small deviations from the center, the shape of the functions all look alike.



We would expect the reflections to be similar under these circumstances. Using only the part of a mirror where the spherical and parabolic shapes are essentially the same is called the *paraxial approximation*. So, if we meet the criteria for the paraxial approximation, our spherical mirrors should work. Note that when you need the entire mirror, say, in a communications antenna, you must do better than a spherical approximation to the correct shape for your mirror. Your satellite dish is likely not a spherical section.



Like with our flat mirror, we will measure distances from the mirror surface (from point V in the figure). We can find the image location, d_i , by taking two rays. We could use any of billions of rays. But let's try to pick rays that are easy to draw. One convenient ray is the ray that passes through the center of curvature, C . This ray will strike the mirror surface at right angles and bounce back along the same path. The incidence angle will be zero, so the reflected angle must be zero by the law of reflection. That is the yellow ray in the figure.

Another convenient ray is the ray from the tip of the object to point V . This ray will bounce back with angle θ . Right at point V the mirror surface is perpendicular to the optic axis. This makes the bounce of the blue ray just like the a bounce from a flat mirror! That is easy to draw. Where these two reflected rays cross, we will find the image of the tip of our arrow. Knowing the shape of the arrow and that the bottom is on the axis, we can fill in the rest of the image.

We can calculate the magnification for this case. We use the gold triangle to determine that

$$\tan \theta = \frac{h_o}{d_o}$$

and the light blue triangle to determine that

$$\tan \theta = \frac{-h_i}{d_i}$$

And you might object to the negative sign. But we want to indicate that the image is inverted by making its sign negative. So because the image is upside down h_i is defined to be negative. To make our tangent equation work, we need to insert another minus sign. This gives us $\tan \theta$ instead of $-\tan \theta$ when h' is inverted (when h' , itself, is negative). When we decide to make the sine of a quantity negative or positive just because we want it to be that way we say it is negative (or positive) *by convention*. And we have done this here. It is just a convention. But we will use it. So our magnification would be

$$m \equiv \frac{\text{Image height}}{\text{Object height}} = \frac{h_i}{h_o}$$

But we can write

$$\begin{aligned} h_o &= d_o \tan \theta \\ h_i &= -d_i \tan \theta \end{aligned}$$

so the magnification could be written as

$$m = \frac{h_i}{h_o} = \frac{-d_i \tan \theta}{d_o \tan \theta} = -\frac{d_i}{d_o}$$

which gives us the magnification in terms of how far away the object and image are.

16.1 Mirror Equation

We can further exploit this geometry to get a relationship between d_o , d_i , and R . Notice that

$$\tan \alpha = \frac{h_o}{d_o - R}$$

and that

$$\tan \alpha = \frac{-h_i}{R - d_i}$$

Then

$$\frac{h_o}{d_o - R} = \frac{-h_i}{R - d_i}$$

or

$$\frac{R - d_i}{d_o - R} = -\frac{h_i}{h_o}$$

We can use our magnification definition to replace h_i/h_o

$$\frac{R - d_i}{d_o - R} = \frac{d_i}{d_o}$$

And we have a useful equation, but not in the form you usually see it. Let's perform some algebra

$$\begin{aligned} (R - d_i) d_o &= d_i (d_o - R) \\ -d_i d_o + R d_o &= p d_i - R d_i \\ R d_o + R d_i &= d_o d_i + d_i d_o \\ \frac{R d_i}{R d_o d_i} + \frac{R d_o}{R d_o d_i} &= \frac{2 d_o d_i}{R d_o d_i} \\ \frac{1}{d_o} + \frac{1}{d_i} &= \frac{2}{R} \end{aligned}$$

We are close. But let's make one more change to this equation. And it will involve the focal point.

Focal Point

Remember that special place where if the light comes into the mirror parallel, the reflected rays all meet? We called that the focal point, f .

Now that we know the mirror equation, let's let d_o be very large (for example, let d_o be the distance to the Sun). Then the rays would come in parallel. And all the rays would come together at the focal point. We would get an image of the Sun right at the focal point! Mathematically this would be

$$\begin{aligned} \frac{1}{\infty} + \frac{1}{d_i} &\approx \frac{2}{R} \\ 0 + \frac{1}{d_i} &\approx \frac{2}{R} \\ \frac{1}{d_i} &\approx \frac{2}{R} \end{aligned}$$

or

$$d_i \approx \frac{R}{2}$$

This is a special image point. But we really already know what to call this special place where parallel rays come together after reflection. We call it the *focal point* and the distance from the mirror to the point f is called the *focal length*. We see that, indeed, that the length f is given by

$$f = \frac{R}{2} \tag{16.1}$$

so we can write the mirror equation as

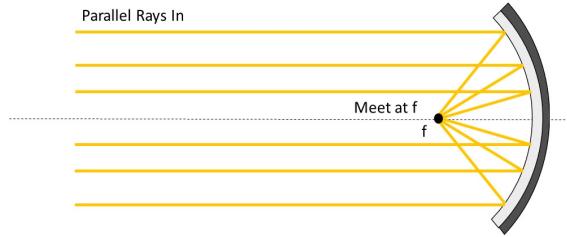
$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \tag{16.2}$$

This equation is very useful. Even when you use powerful optical design software it can form the beginning of your optical system design.

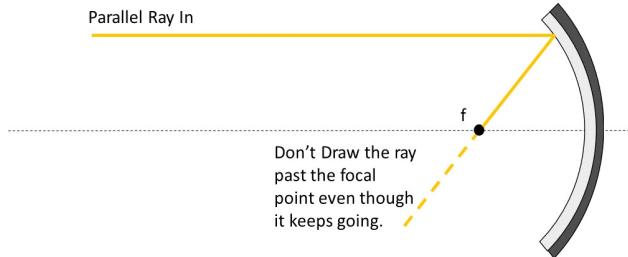
16.1.1 Ray Diagrams for Mirrors

For our mirror problems we will need to draw a diagram of the situation. And often the secret to solving the problem is in making the diagram! So, let's learn how to do this. We could draw an infinite number of rays to find where the image will be. And, indeed, optical design software does draw many rays to check our designs. But for us we need to draw just a few easy rays. But what rays are easy to draw?

Let's think about what we already know.



If parallel rays come into the mirror, they meet at the focal point. They don't really stop there unless we put something there.

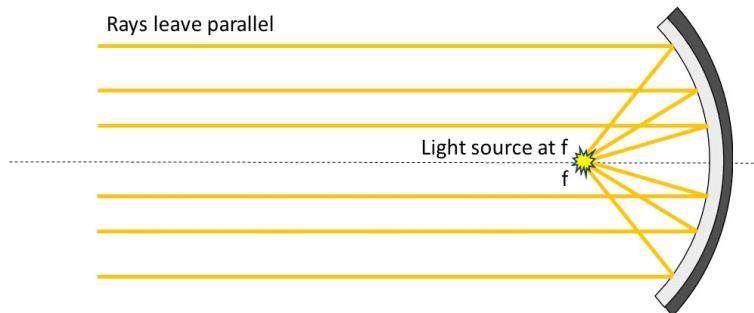


But often we will just stop our drawing at the point f . But this is easy to draw! Just a parallel line in and bounce through the focal point, f . And we know the focal distance $f = R/2$ so it is even easy to find the point f . This is one way to make a solar cooker.



so we could call any ray that comes in parallel and goes through the focal point a “solar cooker ray.” Note that it is a little weird to have a point marked f and a distance from a mirror to that point named f . But we will get used to it.

But it turns out we can use this geometry twice. Think, the law of reflection tells us that every ray that bounced off of the mirror did so with equal angles $\theta_i = \theta_r$. So what would happen if we put a light source at the focal point? The law of reflection would work the same way, and the rays would come out parallel!



This would be easy to draw! And this is also useful. Old fashioned flashlights used this to make the light from a bulb that came out in all directions go in just one direction. Big versions of this were used to send light out parallel to find enemy aircraft in World War II. These devices were called *spot lights* so we can call rays that act like this “spot light” rays.



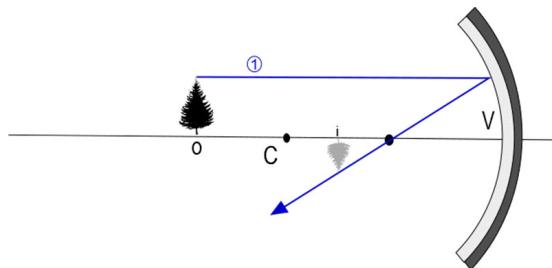
Now days, you might only see such a device used in movies about your favorite super hero (murciélagos boy) used to call on the super hero for help.

There is one more ray that is easy to draw. We have already encountered it. It is the ray that goes through the center of curvature for the spherical mirror so it hits with $\theta_i = 0$ and bounces off with $\theta_r = 0$.

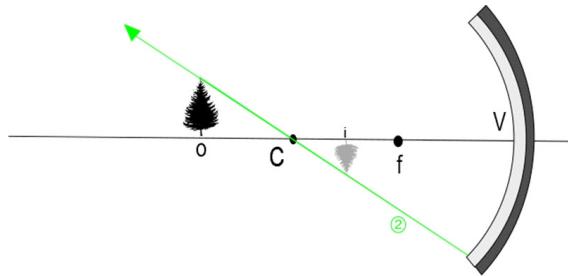
Let's put all of these together to draw ray diagrams for a spherical mirror.

Principal rays for a concave mirror:

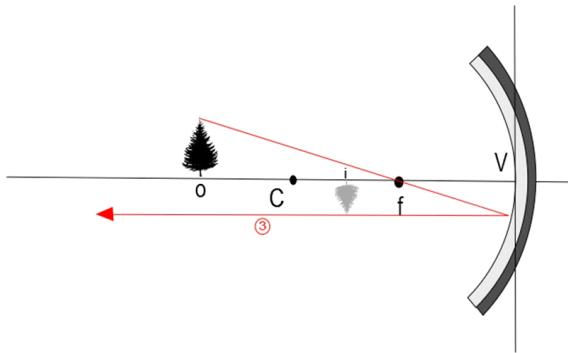
- Ray 1 is drawn from the top of the object such that its reflected ray must pass through f .



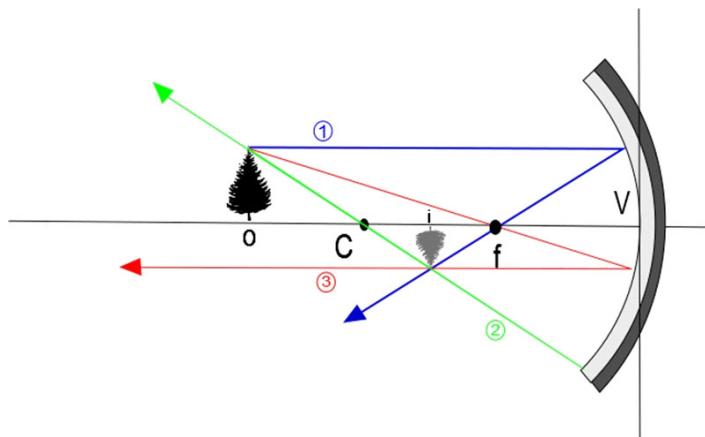
- Ray 2 is drawn from the top of the object through the center of curvature. This ray will be incident on the mirror surface at a right angle and will be reflected back on itself.



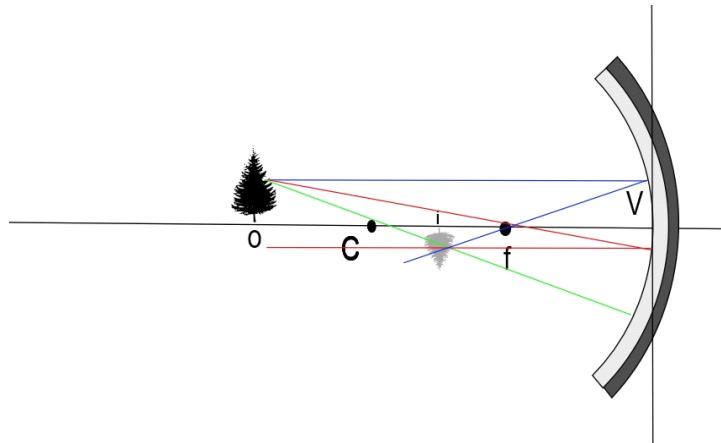
3. Ray 3 is drawn from the top of the object through the focal point to reflect parallel to the principal axis.



We can put these together, and where they meet will be where the image will form.

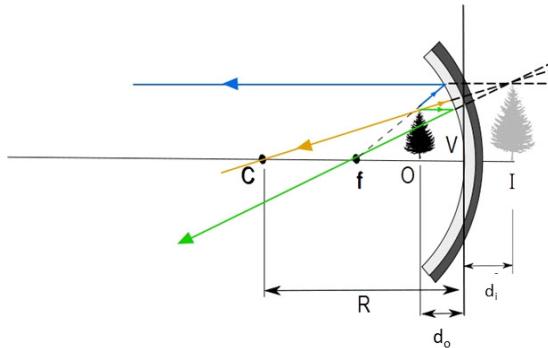


Of course, we did this for the top of the tree. We could do the same for the middle of the tree and we would get



and we can do this for every point on the tree and find the corresponding image point. But it turns out that if we do the top of the object we pretty much know where the rest of the image will be. So with just three rays we can find the image of our object!

We can do the same for an object closer than a focal length

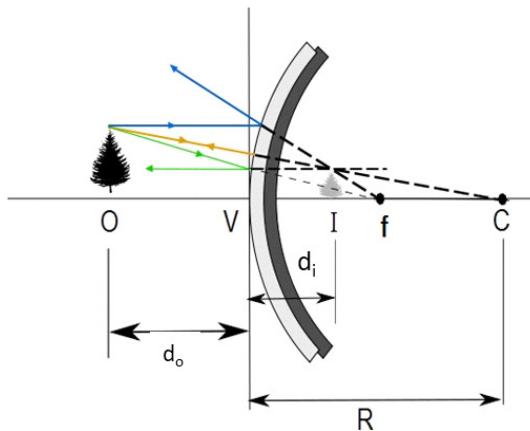


16.1.2 Principal rays for a convex mirror:

We also may have a mirror that curves, but curves the other way. We can use three rays, but this time we need to be careful.

1. Ray 1 is drawn from the top of the object such that its reflected ray appears to have come from f .
2. Ray 2 is drawn from the top of the object so that it appears to have come from the center of curvature. This ray will be incident on the mirror surface at a right angle and will be reflected back on itself.

3. Ray 3 is drawn from the top of the object to reflect parallel to the principal axis by aiming it at f .



We should tabulate our sign convention for mirrors. So far we know that if h_i is upside down it will be negative. But we can assign signs for all the other distances in our ray diagrams.

Quantity	Positive if	Negative if
Object location (d_o)	Object is in front of surface (real object)	Object is in back of surface (virtual object)
Image location (d_i)	Image is in front of surface (real image)	Image is in back of surface (virtual image)
Image height (h_i)	Image is upright	Image is inverted
Radius (R_1 and R_2)	Center of curvature is in front of surface	Center of curvature is in back of surface
Focal length (f)	In front of surface	In back of surface

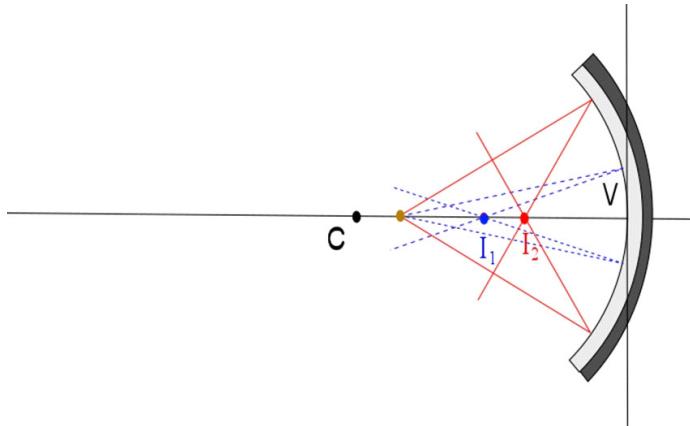
16.2 Aberrations

It's time to ask what happens when we don't use paraxial rays (when the paraxial ray approximation is not valid) but our equations are based on using only small angles or paraxial rays. You might guess that we will have problems in our imagery. That sounds bad, so let's use a special word for "problems" in imagery. That word is *aberration*.

Our first aberration comes from making the mirror the wrong shape. It should be a parabola, but instead it is a section of a sphere. We will call the problem this creates, *spherical aberration*.

Remember that reason we use spherical shapes is that spherical shapes are easier to make than parabolas or hyperbole, or other shapes. So optics manufacturers have been using spherical optics for centuries. Later, when we do lenses we will find there are ways to correct for spherical aberration.

But what does spherical aberration do? We end up with different images of our object at different image locations. The center rays (blue dashed lines in the next figure) make an image at I_1 but the outside (non-paraxial) rays focus closer at I_2 . If we put up a screen to see the image, we will have a blurry image because some of the light is out of focus at I_1 but a different bit of light is out of focus at I_2 . There is no one place where all the light is in focus.



Most of the time for this class, we will point our optics so the object is near the principal axis, so we can make the paraxial approximation that fixes this problem. But in real optical systems we can't do this.

This optical problem was the problem with the Hubble Space Telescope!



This before (left) WFPC and after (right) WFPC2 image of the core of the galaxy M100 shows the dramatic improvement in Hubble's view of the universe after the first servicing mission in December 1993. (Images Courtesy NASA)

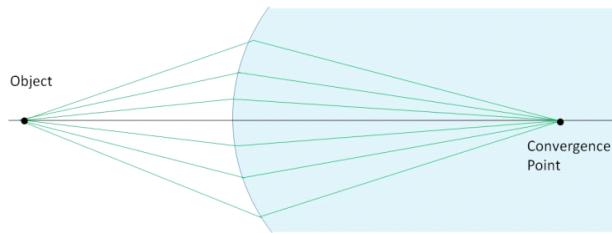
And it took some work to correct this problem. We will learn more about this as we study lenses next.

16.3 Semi-Infinite Lenses and the Semi-Infinite Lens Image Equation

We learned how to find an image location graphically for mirrors. But many optical systems including contacts and glasses, use lenses. Let's look at lenses next.

Let's start by thinking of a special case for refraction. A curved surface on a very large piece of glass. We will assume that the piece of glass is semi-infinite, but all it has to be is very large.

We can call this a semi-infinite bump of glass. And it may seem like a far fetched situation, but it turns out that our eyes are essentially a system like this because our eyes are filled with a jelly like substance, and the light in our eyes stays in this substance until it is detected in our retina. So let's look at this situation.

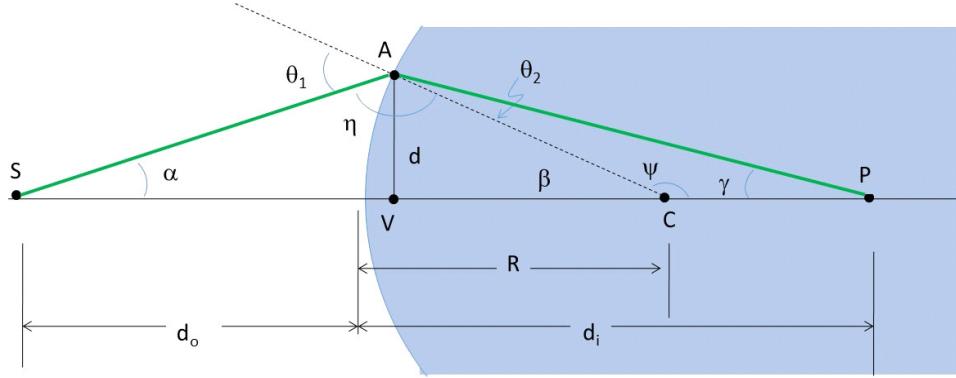


Take a point object that either glows, or has rays of light reflecting from it. The rays leave the object and reach the surface of the glass. The rays will refract at the surface. Each bends toward the normal, but because of the curvature of the glass, the rays all converge toward the center. We can identify this convergence point as the image of the point object. For mirrors we used the law of reflection to know where the light would go. But for lenses, the light is transmitted. So at the surface we can find the refracted angles using Snell's law

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

We will again use the small angle approximation. Then θ_1 and θ_2 are small and none of the rays are very far away from the axis. We know this as the paraxial approximation. Snell's law becomes

$$n_1 \theta_1 = n_2 \theta_2$$



Using the more detailed figure above with it's exaggerated angles (not so paraxial but easier to see), we observe triangles SAC and PAC . We can see that for triangle SAC the top angle labeled η , plus θ_1 must be 180° .

$$\eta + \theta_1 = 180^\circ$$

We also know that the sum of interior angles of a triangle must equal 180° . So

$$180^\circ = \alpha + \beta + \eta$$

then

$$\begin{aligned}\eta + \theta_1 &= \alpha + \beta + \eta \\ \theta_1 &= \alpha + \beta\end{aligned}$$

Likewise, from triangle PAC ,

$$\beta + \psi = 180^\circ$$

$$180^\circ = \psi + \gamma + \theta_2$$

so then

$$\beta = \theta_2 + \gamma$$

then,

$$\theta_2 = \beta - \gamma$$

and we can write our paraxial Snell's law as

$$\begin{aligned}n_1\theta_1 &= n_2\theta_2 \\ n_1(\alpha + \beta) &= n_2(\beta - \gamma) \\ n_1\alpha + n_1\beta &= n_2\beta - n_2\gamma \\ n_1\alpha + n_2\gamma &= n_2\beta - n_1\beta \\ n_1\alpha + n_2\gamma &= \beta(n_2 - n_1)\end{aligned}$$

Looking at the figure. We see that d is a leg of three different right triangles (SAV , ACV , and PAV). The ray in the figure is clearly not a paraxial ray. If we use a paraxial ray, then the point V will approach the air-glass boundary. When this happens, then $SV = d_o$, $VC = R$, and $VP = d_i$. So we can write (using the small angle approximation)

$$\begin{aligned}\tan \alpha &\approx \alpha \approx \frac{d}{d_o} \\ \tan \beta &\approx \beta \approx \frac{d}{R} \\ \tan \gamma &\approx \gamma \approx \frac{d}{d_i}\end{aligned}$$

So, our Snell's law becomes

$$\begin{aligned}n_1\alpha + n_2\gamma &= \beta(n_2 - n_1) \\ n_1\frac{d}{d_o} + n_2\frac{d}{d_i} &= \frac{d}{R}(n_2 - n_1)\end{aligned}$$

We can divide out the d 's

$$\frac{n_1}{d_o} + \frac{n_2}{d_i} = \frac{(n_2 - n_1)}{R} \quad (16.3)$$

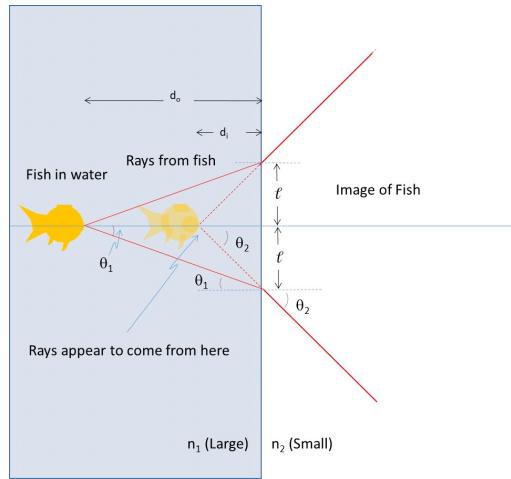
We can use this formula to convince ourselves that no matter what the angle is (providing it is small), the rays will form an image at P .

Real images will be in the glass for this special case (we will fix this with non-infinite lenses soon) so we need to change sign conventions.

Quantity	Positive if	Negative if
Object location (d_o)	Object is in front of interface surface	Object is in back of interface surface (virtual object)
Image location (d_i)	Image is in back of interface surface (real image)	Image is in front of interface surface (virtual image)
Image height (h_i)	Image is upright	Image is inverted
Radius (R)	Center of curvature is in back of interface surface	Center of curvature is in front of interface surface

We could go through the entire derivation and switch the indices of refraction. It turns out we get the same equation. The results will be different, but the equation is the same.

16.3.1 Flat Refracting surfaces



Let's consider a fish tank. The fish tank has an interface, but it is flat. Can we use our equation (16.3) to describe this?

The answer is yes, if we let $R = \infty$. This makes sense for a flat surface. If we have an infinitely large sphere, then our small part of that spherical surface that makes up the fish tank wall will be very flat.

Then

$$\frac{n_1}{d_o} + \frac{n_2}{d_i} = \frac{(n_2 - n_1)}{\infty}$$

or

$$\frac{n_1}{d_o} + \frac{n_2}{d_i} = 0$$

we see that

$$d_i = -d_o \frac{n_2}{n_1}$$

We can see from this that, since $n_2 < n_1$ for this case d_i is less than d_o . The image is in the water! It is in *front* of the water/air boundary of the tank. But light is not actually coming from that location. What has happened is that the light path has bent, but our brain image processing center thinks light travels only in straight lines. So it tells us there is a fish there, but the fish appears closer than it is. The light doesn't come from where we think the fish is. And as we know, this means the image is virtual, just like for flat mirrors!

Chapter 17

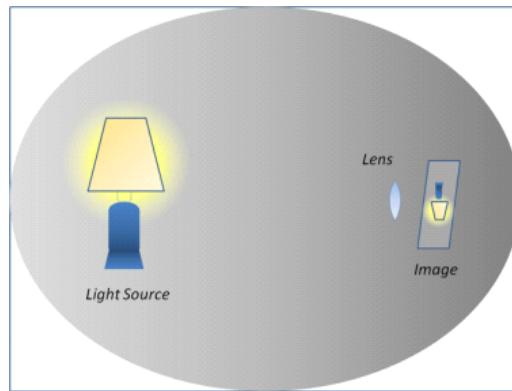
17 Thin Lenses 3.2.4, 3.2.5

Fundamental Concepts

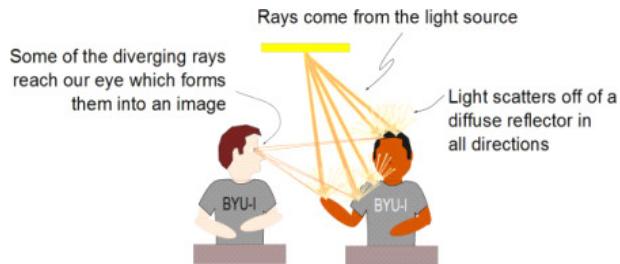
- For a thin lens, we can describe where the light goes using $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$ where $\frac{1}{f} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$
- There is a sign convention for all of these equations.
- The eye is like a semi-infinite double lens with the cornea and the crystalline lens as the two lenses
- The crystalline lens does focusing but in biology it is called accommodation
- We have three color sensors for red green and blue light
- Myopia and hyperopia are problems with the shape of the eye and are corrected with additional lenses
- Optical lenses are measured in optical power which is $1/f$ (m) with the unit of a diopter.
- Medical people measure the power of a lens in diopters. A diopter is one over the focal length with the focal length measured in meters.

17.1 Images from lenses

Let's think once again about what an image is. If you haven't done this before, take a piece of paper and a lens, and hold up the lens in a darkened room that has some bright object in it. Move the lens or the paper back and forth, and at just the right distance, a miniature picture of the bright object will appear. We should think about what the word "picture" means in this sense.



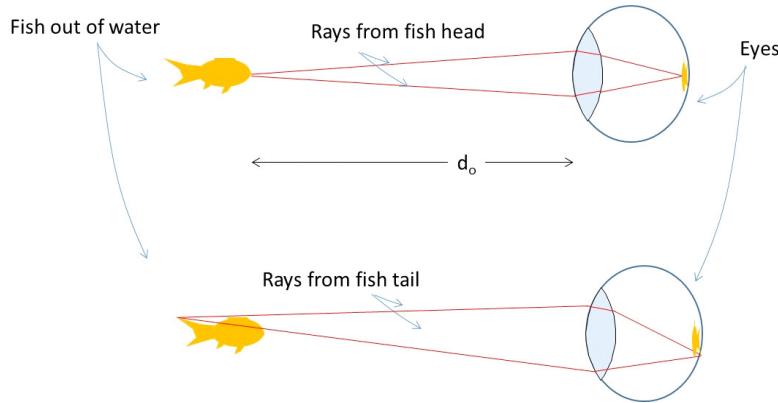
We have talked about how we see objects. Remember the BYU-I guys from last time.



Our eyes gather rays that are diverging from the object because light has bounced off of the object. Our eyes intersect a diverging set of rays that form a definite pattern. That diverging set of rays forming a pattern is the picture of the object.

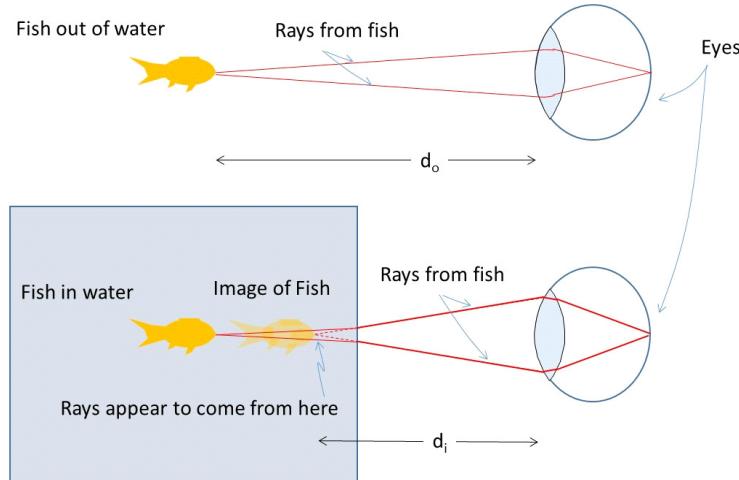
So when we say that the lens has formed a miniature picture of our object, we mean that the lens has somehow formed a diverging set of rays that form a pattern that looks like the pattern formed by the diverging set of rays coming from the object, itself. In other words, the object forms a diverging set of rays. And our lens forms a duplicate set of rays in the same pattern, so we see the same thing. The lens' version is smaller, upside down, and backwards, but it is still essentially the same pattern. We saw this with our semi-infinite lens, but it must be true for less infinite lenses as well.

As a first step to see how this works, consider our fish tank again. It would be bad on the fish, but think about looking at a fish in air. The room light would bounce off of the fish, and we would have a diverging set of rays from every point on the fish (see next figure).



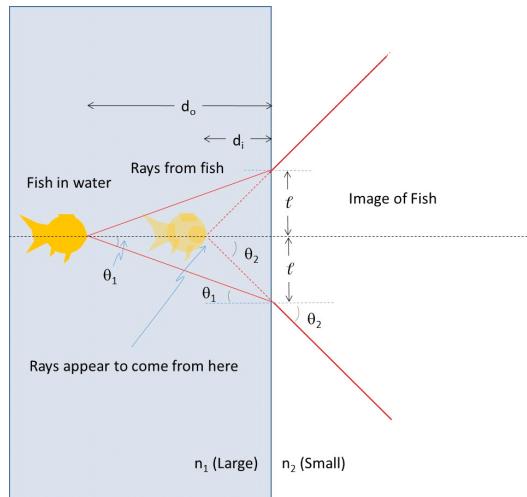
We can see that the picture is made from every point on the fish being “imaged” to a point on our retina. We collect the rays leaving every point on the fish, and bring them to corresponding points on the retina to make the picture.

It will take us a few lectures to see exactly how this is done, but as a first step, let’s consider the fish tank, itself. Put the fish back in the tank and look at it (next figure).



Rays still come from the fish. But we now know that the change from a slow material to a fast material will bend the light. These bent rays are collected by our eyes, and the picture of the fish is formed on the retina just as before. But our eyes interpret the light as though it went in straight lines with no bends (dotted lines in the last figure). Our visual processing center in our brain is designed to believe light travels in straight lines, so our mind tells us there is a fish, but that the fish head (and every other part of the fish) is closer than it really is. We call this apparent fish at the closer location an image of the fish, because this is where we think the diverging set of rays come from that form the fish pattern.

The next figure shows the details of the rays leaving a dot on the fish head (with the angles exaggerated so it's easier to see them).



A spot on the fish head is our object for this set of rays. The distance from the fish-head dot and the edge of the water/air boundary is our *object distance*, d_o .

The distance from the image of the fish-head dot to the edge of the water/air boundary is our *image distance* and is given the symbol d_i , because it appears to be where the rays come from.

We can find where the image is (d_i) knowing d_o . We can see from the figure that

$$\begin{aligned}\ell &= d_o \tan \theta_1 \\ &= d_i \tan \theta_2\end{aligned}$$

so

$$d_o \tan \theta_1 = d_i \tan \theta_2$$

or

$$d_o \frac{\tan \theta_1}{\tan \theta_2} = d_i$$

from Snell's law, we know that

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1}$$

Usually we can take the small angle approximation. This would limit our analysis to rays that are near the central axes. We have done this before with mirrors. We call this central axis the *optics axis* and the rays that are not too far away from this axis *paraxial rays*. Then for our small angles we can write

$$\tan \theta_i \approx \sin \theta_i$$

so

$$d_o \frac{\tan \theta_1}{\tan \theta_2} \approx d_o \frac{\sin \theta_1}{\sin \theta_2} = d_o \frac{n_2}{n_1} = d_i$$

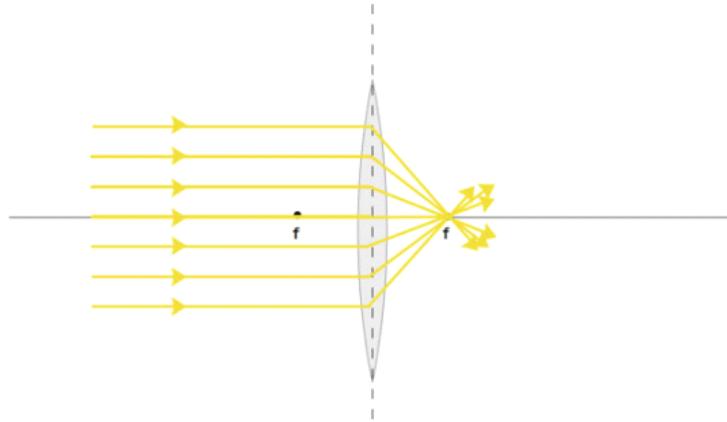
and we have the image distance

$$d_i = d_o \frac{n_2}{n_1}$$

Note that once again we have a virtual image. The place where we see the fish does not have rays of light coming from it! But our brain processing center believes the rays of light didn't bend, so it extends the rays we see backward into the fish tank and it interprets what we see as a fish being where those extended rays meet.

This is not so useful unless you have some burning need to know where fish are in a tank. But we now have the vocabulary to discuss the larger problem of how a lens works, which we will take up next time.

Before we do a lot of math to describe how lenses work, let's think about our early childhood experiences. You may have burned things with a magnifying glass¹. Using the idea of a ray diagram, here is what happens.



We know that the rays from the Sun come from so far away that they are essentially parallel. We know that these rays come together to a fine point that can start a fire. The point where these rays converge is important to us. We found such a point from our mirrors. We called this point the *focal point*.

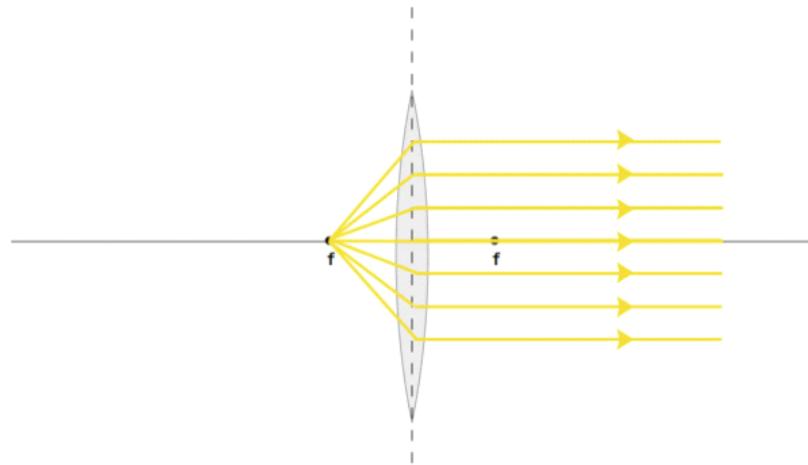
Now think about Snell's law. We have

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

The light will bend at a surface, but Snell's law doesn't care if $n_1 > n_2$ or if $n_2 > n_1$. It works both ways. The light will follow the same path either direction. We would expect that if we put a light source at the focal distance,

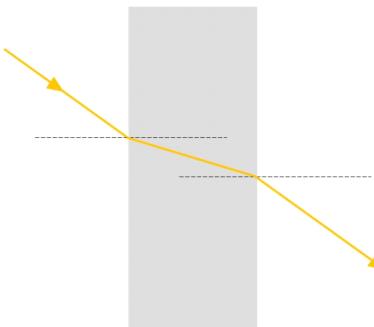
¹If you didn't do this as a child, consider trying it now. Be responsible and safe, but it is valuable to see how this works.

the rays should come out parallel.



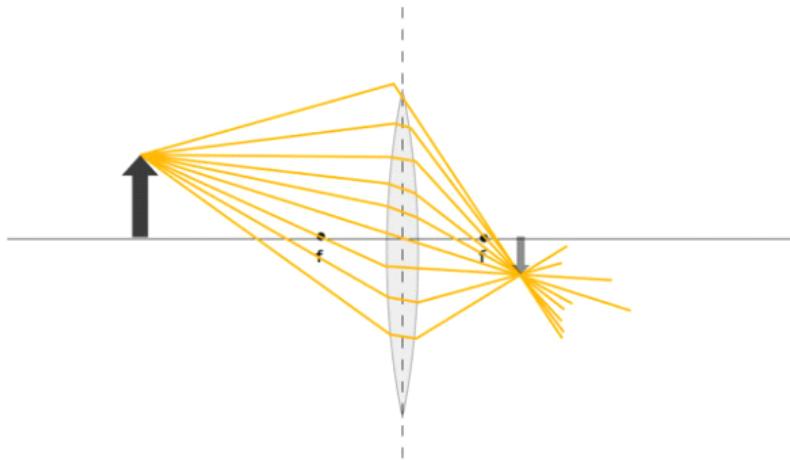
This is one way manufacturers make LED flashlights.

We need one other bit of information that we already have seen from basic refraction.

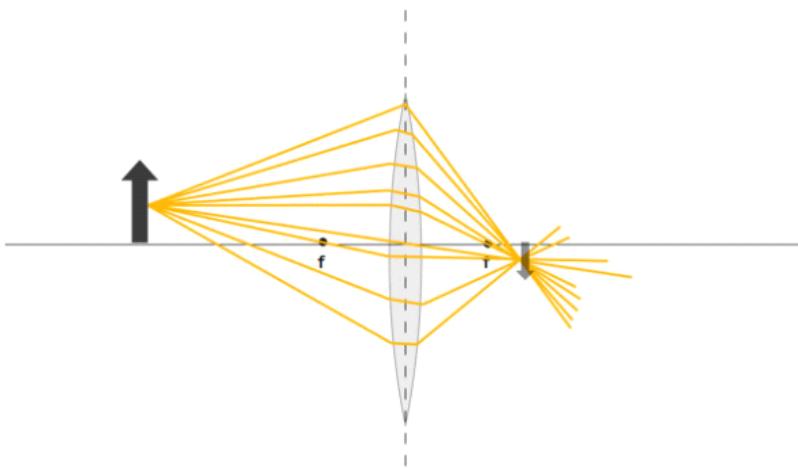


A flat block does refract the light, but when the light leaves the block it is only displaced, it retains the original direction. With these three ray scenarios, we can describe how a lens works.

We know that for every point on the object, we get millions of reflected rays that diverge. In the next figure, rays of light are reflecting off of the large, black arrow. The lens must collect these rays together to form the corresponding point on the image.



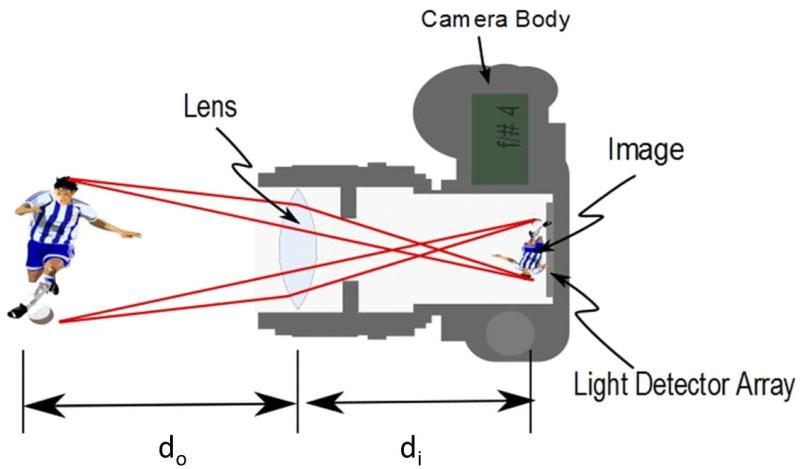
Of course, this happens for every point on the image.



But we usually pick the top of the object. If we place the bottom of the object on the *optic axis*², the bottom of the image will also be on the optic axis. So knowing where the bottom of the image is, and finding the top of the image gives a pretty good idea of where the rest of the image must be. So we will draw diagrams for the top of the object to find the top of the image.

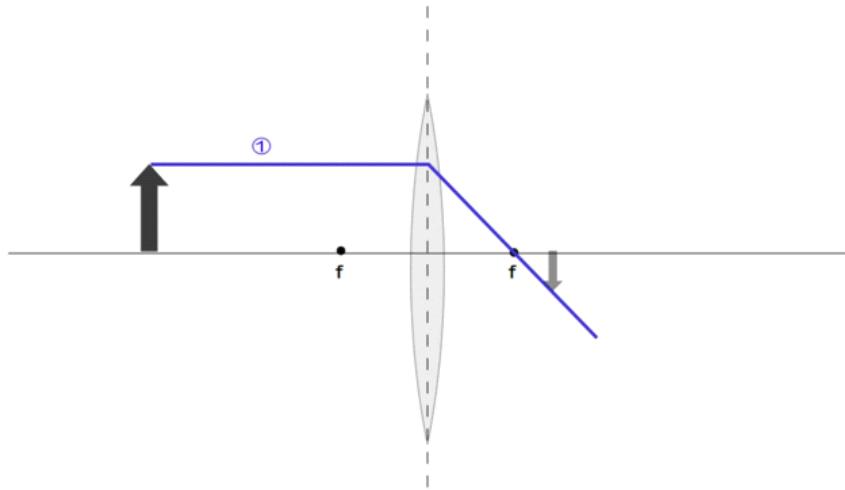
But suppose this is not true? For example, when we use a camera, we do not align the optical system on an axis before we shoot.

²The line drawn on the figure through the middle of the lens.

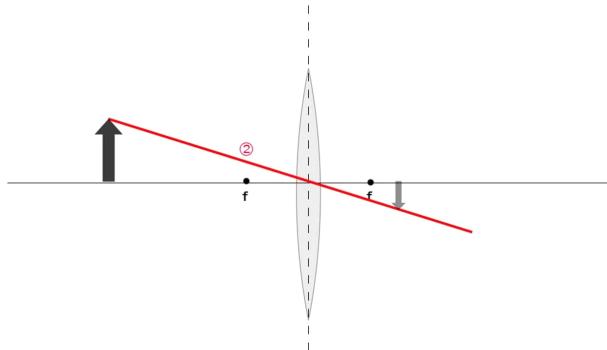


We can, of course, trace two rays for the bottom of the image in this case and find the location of the bottom of the image as well.

We know that light is a superposition of millions of waves and that when light reflects off of something like your neighbor or a fish or a soccer player there will be millions of rays. Drawing millions of rays is impractical, and fortunately, not needed. We can choose three simple rays that leave the top of the object , and see where these rays converge to form the top of the image. Let's start with a ray that travels from the top of the object and travels parallel to the optic axis. We recognize this ray as being like one of the rays from the Sun. It comes in parallel, so it will leave the lens and travel through the focal point.

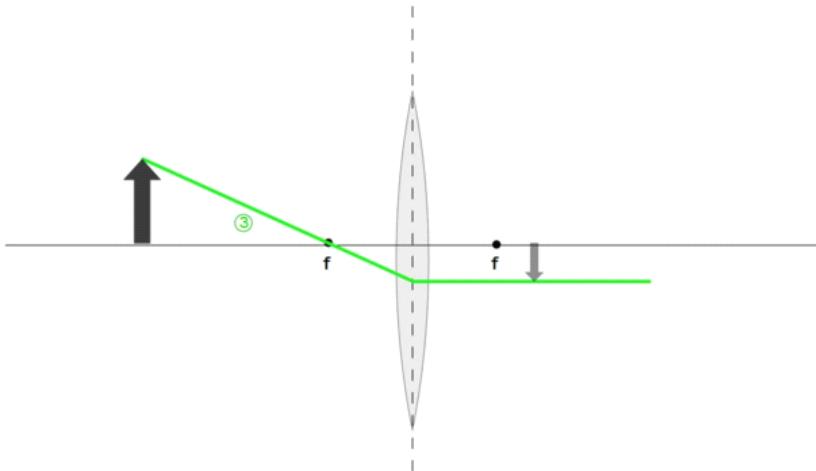


For a second easy ray, lets take the case that is like our flat block. Near the center of the lens, the sides are nearly flat. So we expect that the ray will leave in about the same direction as it was going before it struck the lens.

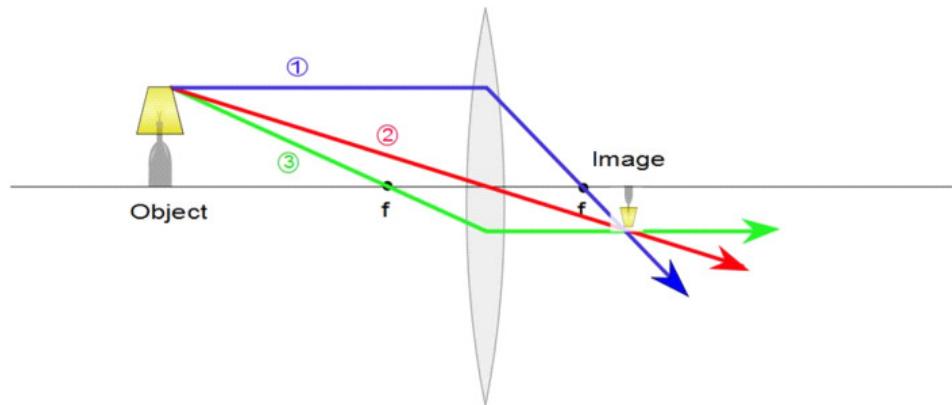


Technically this ray should be shifted. This is where our thin lens approximation comes in. If the lens is thin, then the ray going through the center of the lens won't be shifted much, and we can ignore the shift. We just draw ray 2 as a straight line.

Two rays are really enough to determine where the top of the image will be, but there is a third ray that is easy to draw, so let's draw it to give us more confidence in our answer. That ray is one that leaves the top of the object and passes through the focal point on the object side of the lens. This situation we also recognize. Since the ray goes through the focal point, it is as though the light came from that point. This is like our LED flashlight case. This ray will leave the lens parallel to the optic axis.



Where all three rays intersect, we will have the top of the image.



Notice that in this case, the image is upside down. That is normal. Also notice that it is smaller than the object. We say that the image is magnified, which may still seem a little bit strange. But in optics, a magnification of greater than one means that the image is bigger than the object. This is like a movie projector that makes a large image of a small film segment. The magnification can be equal to one, meaning the object and image are the same size. And finally the magnification can be less than one. This means that the image is smaller than the object. This is a convenient definition, because then we can use the same equation to describe all three situations.

$$m \equiv \frac{\text{Image height}}{\text{Object height}} = \frac{h_i}{h_o}$$

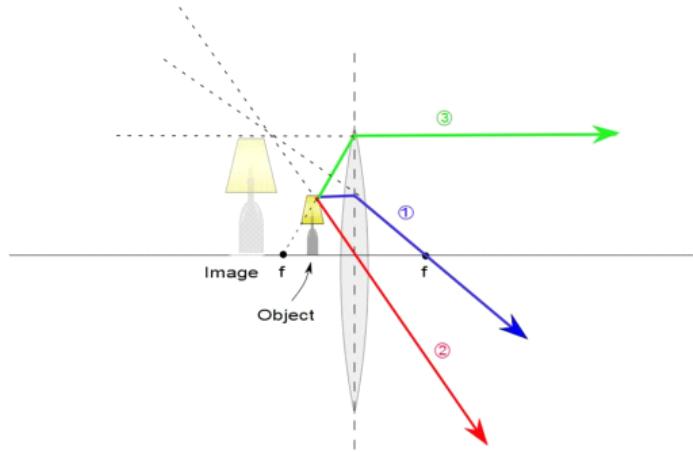
where h_o is the object height, and h_i is the image height. Notice the negative sign. It's still true that by convention (meaning physicists got together and voted on this) we say that an upside down image has a negative magnification. You just have to memorize this, there is no obvious reason for this except it is mathematically convenient.

We will find that

$$m = -\frac{d_i}{d_o}$$

17.2 Virtual Images with Lenses

Lets take another case and draw a ray diagram. This time let's place the object closer than the focal distance. This is the case when we use a lens as a magnifying glass. The rays will look like this.



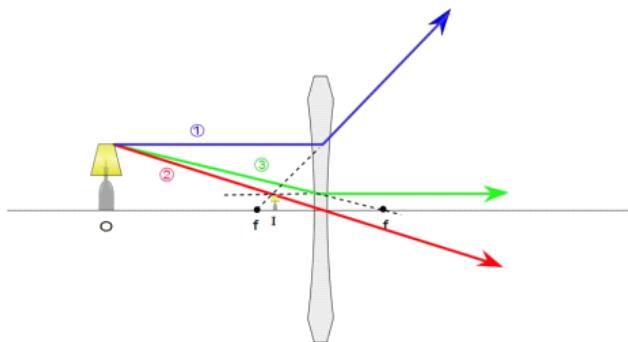
Notice that these rays never converge! We won't get an image that could project on a paper. But we know that there is an image, we can look through the lens and see it! And that is the key. The image does not really exist. There is no place where light gathers and then diverges into a pattern that we recognize. But we look through the lens, and our mind interprets the diverging rays coming from the lens as though they had only traveled in straight lines. If we extend these rays backwards along straight lines, they appear to come from a common point. This is the point they would have had to have come from if there were no lens. We believe we see an image at this location. But no light really goes there! Because this image is not really made from light diverging from this position, we call it a *virtual image*. The image we formed before that could be projected on a screen is called a *real image*. This is just like our image terminology for mirrors!

By convention, we say the distance from the lens to the virtual image is a negative value.

17.2.1 Diverging Lenses

So far our lenses have only been the sort that work as magnifying glasses. We call these *converging lenses*. These lenses are fatter in the middle and thinner on the edges. Because of this they are sometimes called *convex lenses*. By convention, we say the focal distance for this type of lens is positive. For this reason, they are often called *positive lenses*.

But what if we make a lens that is thinner in the middle and thicker on the edges. We can call this sort of lens a *concave lens*, and we will give it a negative focal length by convention, so we can also call it a *negative lens*. But what would this lens do? If we think about our three rays, ray 1 won't be bent toward the optic axis for this type of lens. In fact, if we observe an object through this lens, ray number 1 will appear to come from the focal point. Ray number 2 will still go through the middle of the lens, and if the lens is thin enough, ray 2 will pass through undeviated.



finally ray three leaves the object in the direction of the far focal point. It will hit the lens and leave parallel to the optic axis. From the figure we see that these three rays will never converge. This is like our convex mirror! And from our mirror experience we expect the rays (with help from our brain) will form a virtual image. If we extend the rays backward as shown, we see that the extensions all meet at a point. The rays leaving the lens appear to come from this point. This is the location of the top of the virtual image of the object.

You might wonder what good such a lens could do, but we will find that this type of lens is used to correct vision for nearsighted people. It's likely that many people in our class have this type of lens with them either on their eye (contacts) or on their nose (eye glasses).

17.3 Thin Lenses

Lets' find an equation for a spherical surface once more. But this time, let's let it be more practical and not make the "lens" semi-infinite. We will need to deal with two sides of the lens because (usually) both will be curved.

We found that for refraction

$$\frac{n_1}{d_o} + \frac{n_2}{d_i} = \frac{(n_2 - n_1)}{R}$$

but we did this for a spherical bump on a semi-infinite piece of glass. For this new problem let's make a few assumptions:

1. We have two spherical surfaces, with R_1 and R_2 as the radii of curvature
2. We have only paraxial rays
3. The image formed by one refractive surface serves as the object for the second surface
4. The lens is not very thick (the thickness is much smaller than both R_1 and R_2)

The answer we will get is quite simple

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \quad (17.1)$$

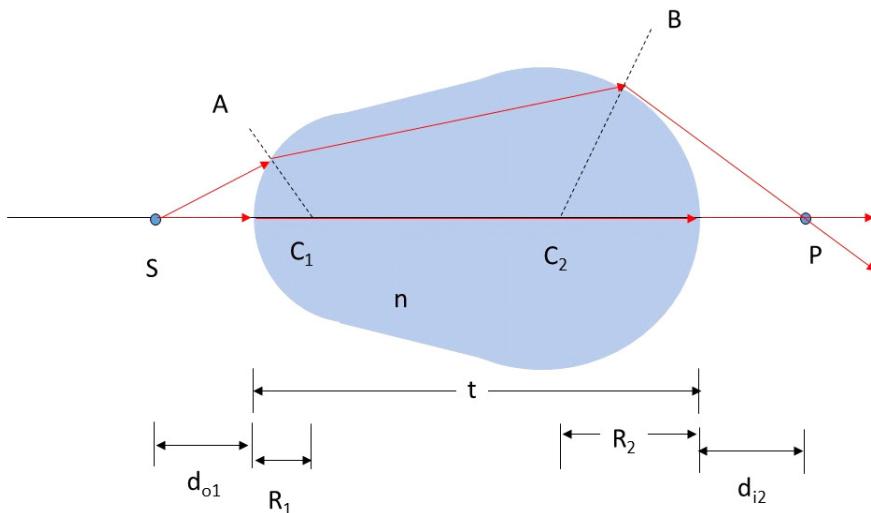
where

$$\frac{1}{f} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \quad (17.2)$$

but to appreciate what it means, lets find out where it comes from.

17.3.1 Derivation of the lens equation

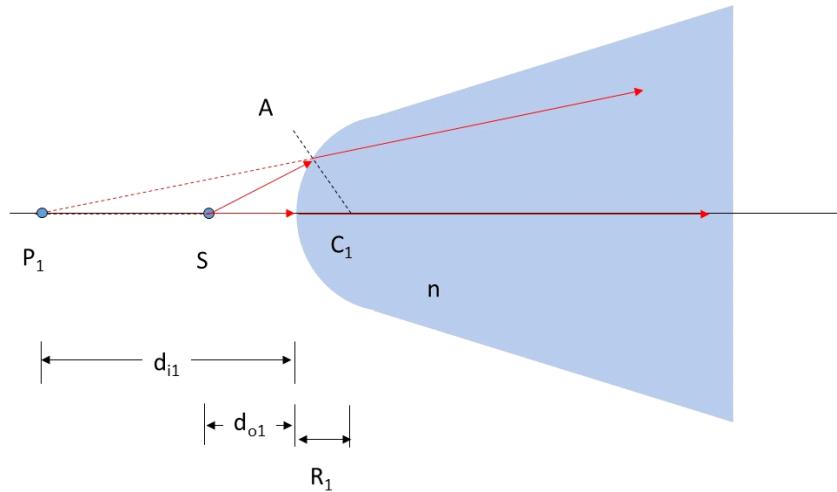
Consider the optical element in the figure below. Notice that our object is a dot, so our image will also be a dot. This is not as boring as it sounds if we consider any object can be considered as a collection of dots.



Light enters at a spherical surface on the left hand side. We use a point object located at S on the principal axis. We could put out dot object anywhere, but let's put it on the axis and trace two rays. The ray along the principal axis crosses each spherical surface at right angles, and therefore travese straight through the optic (this makes on ray very easy to trace!). The second ray hits the first spherical surface at point A . It is refracted and travels to point B . It is again refracted and travels toward the principal axis, crossing at P . The image location is the intersection of these rays, so we have an image at P .

Lets study the surfaces separately

Surface 1:



Let's treat surface 1 as though surface two did not exist. The light would bend at point A and head off into the lens material. This is just our semi-infinite bump problem so we know that

$$\frac{n_1}{d_{o1}} + \frac{n_2}{d_{i1}} = \frac{(n_2 - n_1)}{R}$$

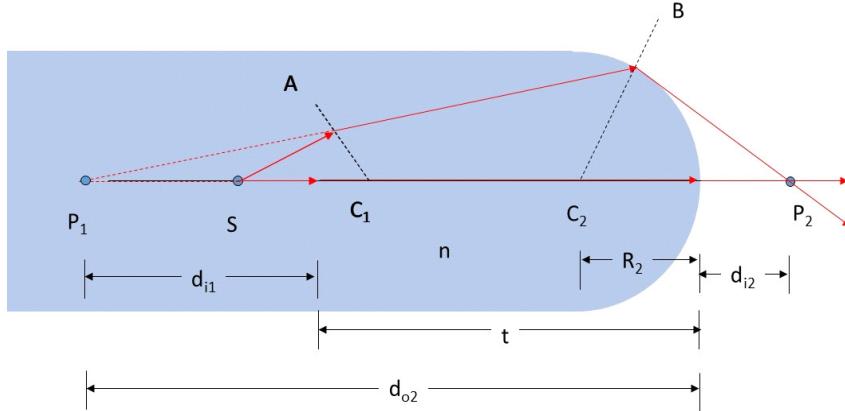
We can consider $n_1 = 1$ and $n_2 = n$ for a air-glass interface and noting that d_{i1} is negative (by convention). Then

$$\frac{1}{d_{o1}} - \frac{n}{d_{i1}} = \frac{(n-1)}{R} \quad (17.3)$$

Note that our rays are *not* converging in the glass this time. We can find the image formed by this side of our lens by tracing the diverging rays backward as we did for the fish tank. The image formed from the first side of the lens is virtual.

Surface 2

Now consider the second surface.



The second surface sees light diverging as though it came from a semi-infinite piece of glass with the object at P_1 . It's not true, the light came from S . But the second surface can't tell the difference. It only knows light is incident in a particular pattern, and that pattern appears to come from point P_1 . So we can treat the situation at surface 2 as though the object is the virtual image formed by surface 1. So

$$d_{o2} = d_{i1} + t$$

We again use our refractive equation

$$\frac{n_1}{d_{o2}} + \frac{n_2}{d_{i2}} = \frac{(n_2 - n_1)}{R}$$

but we identify $n_1 = n$ and $n_2 = 1$. We have for surface 2

$$\frac{n}{d_{o2}} + \frac{1}{d_{i2}} = \frac{(1 - n)}{R_2} \quad (17.4)$$

or

$$\frac{n}{d_{i1} + t} + \frac{1}{d_{i2}} = \frac{(1 - n)}{R_2} \quad (17.5)$$

Now we take our thin lens approximation. Let $t \rightarrow 0$. Then equations (17.3) and (17.5) become

$$\frac{1}{d_{o1}} - \frac{n}{d_{i1}} = \frac{(n - 1)}{R_1}$$

$$\frac{n}{d_{i1}} + \frac{1}{d_{i2}} = \frac{(1 - n)}{R_2}$$

We would prefer a nice equation that has how far away the object is from the lens and how far away the image is formed like our mirror equation. We can get such an equation by adding these two equations for the two surfaces

$$\frac{1}{d_{o1}} - \frac{n}{d_{i1}} + \frac{n}{d_{i1}} + \frac{1}{d_{i2}} = \frac{(n - 1)}{R_1} + \frac{(1 - n)}{R_2}$$

or

$$\frac{1}{d_{o1}} + \frac{1}{d_{i2}} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$$

This equation is very useful. It reminds us of the mirror equation (well, a little). If we again let $d_{o1} = \infty$ (put the object at ∞ so the rays enter surface 1 parallel) we find

$$\frac{1}{d_{i2}} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$$

The spot where the rays gather if the object is infinitely far away is the focal point, f . so for parallel rays we can identify $d_{i2} = f$ as the focal length of the optic

$$\frac{1}{f} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$$

which is known as the *lens makers' equation*.

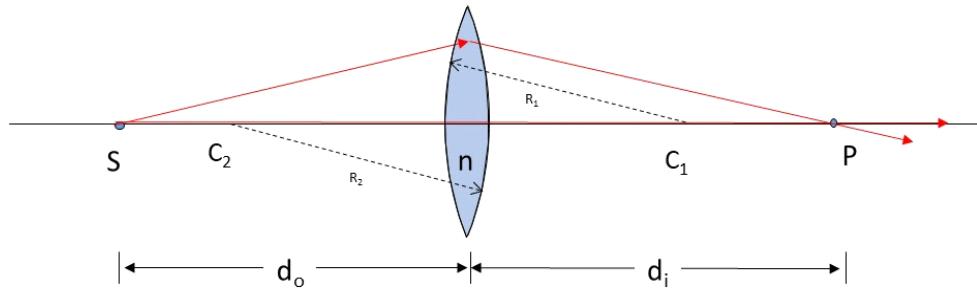
Then we have a relationship between the object distance in front of the lens, and the final image in back of the lens:

$$\begin{aligned} \frac{1}{d_{o1}} + \frac{1}{d_{i2}} &= (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \\ &= \frac{1}{f} \end{aligned}$$

We can drop the subscripts (which we can do now that we let $t = 0$ since the internal distances for the inside points are not important).

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

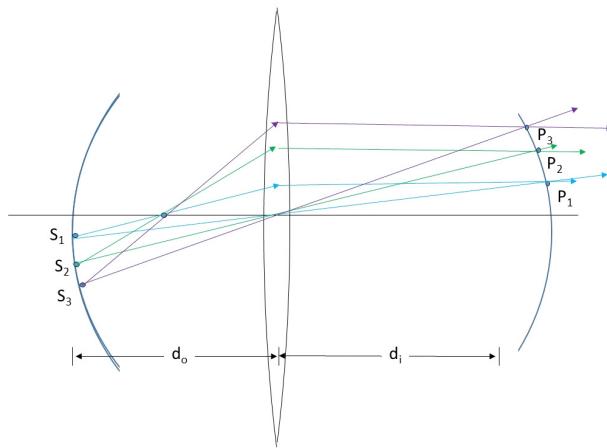
This is called the *thin lens equation*. The resulting approximate geometry is shown below.



And this is fantastic! It is the same equation as the thin mirror equation! Only we have a different equation for finding the focal length.

We did our math for just one point of light at S . Of course any real object is made of lots of points, and not all of the points are on an axis. But each

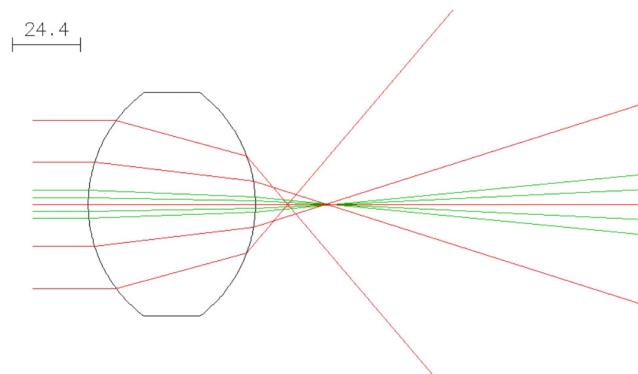
point will be imaged to a corresponding point on the image. Here is an example with three points in the object (the s_i points) and where their images are (the p_i points).



but it would work for millions of points. Our simple analysis explains the formation of actual images and not just point images.

17.3.2 Lens Aberrations

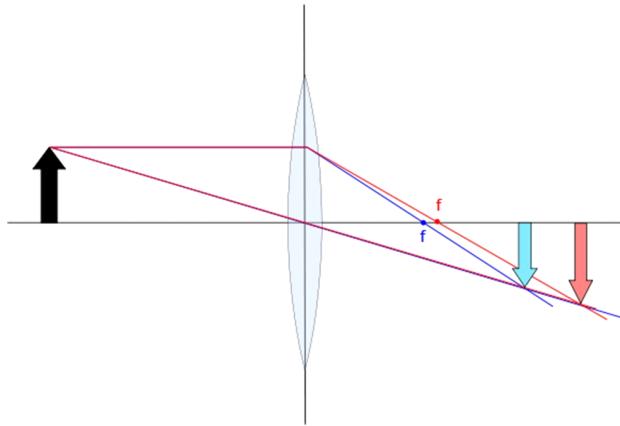
We found that for mirrors if we used a spherical shape rather than a parabola we had a problem with our images. We called the problem spiracle aberration. The reason we used a sphere is that spheres are easy to manufacture. The same problem happens with lenses if we make their sides from parts of spheres. The correct shape for a lens is a hyperbola, and hyperbole are hard to manufacture. Spheres are so much easier.



Example of Spherical Aberration for Lenses. Optical Ray Trace done in the OSLO Optical Design Package.

Again non-paraxial and paraxial rays focus at different spots because we have the wrong shape for our lens. Spherical aberration was made famous as the main problem with the Hubble Telescope.

We get another image problem or aberration with lenses. Remember that the index of refraction is different for different wavelengths of light. If we send white light into our lens, different colors will bend slightly differently (think Snell's law). So we will get a different focus point for different colors of light. In the next figure blue light makes an image at a shorter focal distance than red light. This problem is called *chromatic aberration*.



Generally mirrors can't have chromatic aberration because they don't refract the light. But Lenses always will.

There are many aberrations that come from making lenses that are easy to manufacture, but that are not the perfect shape. We won't study all these in this class. If you are curious, we cover these aberrations in PH375.

17.3.3 Sign Convention

As we have been working we used our sign convention but we need to add to our sign convention table a second radius, and the focal length.

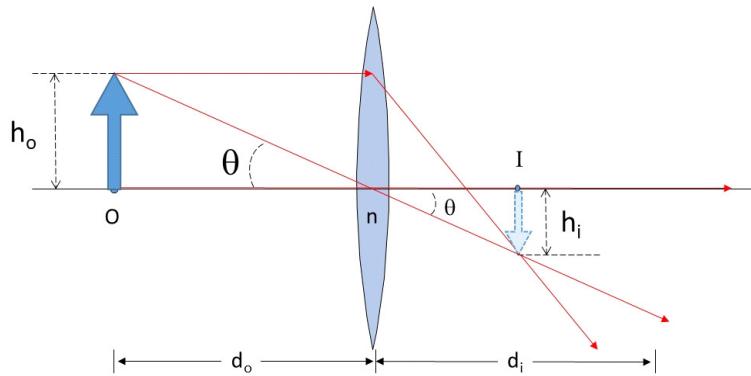
Quantity	Positive if	Negative if
Object location (d_o)	Object is in front of surface	Object is in back of surface (virtual object)
Image location (d_i)	Image is in back of surface (real image)	Image is in front of surface (virtual image)
Image height (h')	Image is upright	Image is inverted
Radius (R_1 and R_2)	Center of curvature is in back of surface	Center of curvature is in front of surface
Focal length (f)	Converging lens	Diverging lens

17.3.4 Magnification

We defined the magnification earlier as a comparison of the image height to the object height.

$$m = \frac{h_i}{h_o} \quad (17.6)$$

And we found we can put this into terms of d_o and d_i for mirrors. But we will this work for lenses?



From the diagram, we can see that

$$\begin{aligned}\tan \theta &= \frac{h_o}{d_o} \\ \tan \theta &= \frac{h_i}{d_i}\end{aligned}$$

then

$$\frac{h_o}{d_o} = \frac{h_i}{d_i}$$

or

$$\frac{h_i}{h_o} = \frac{d_i}{d_o}$$

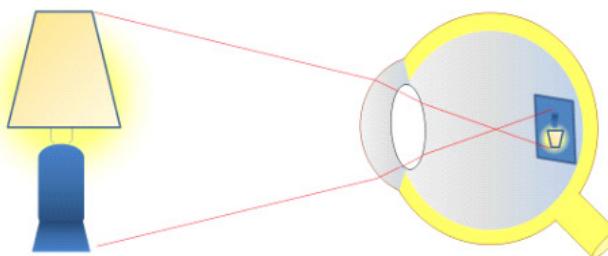
And once again we can write the magnification for a converging lens as minus the ratio of the image distance over the object distance.

$$m = -\frac{d_i}{d_o}$$

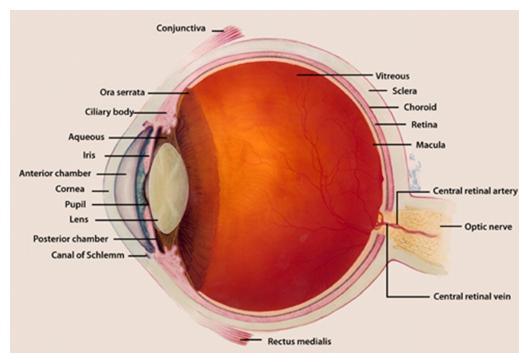
Just like before the minus sign comes from our sign convention because the image will be upside down.

We found the thin lens formula using converging lenses, but it works for diverging lenses as well, so long as the thin lens approximation is valid.

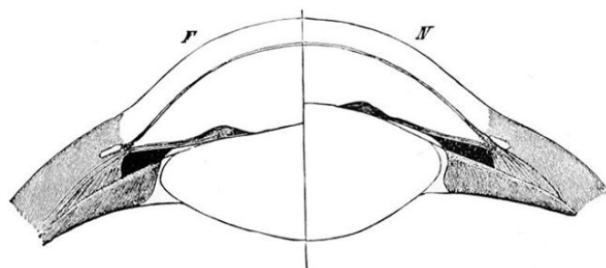
17.4 The Eye



The figure above shows the parts of the eye. The eye is like a camera in its operation, but is much more complex. It is truly a marvel. The parts that concern us are the cornea, crystalline lens, pupil, and the retina.



The Cornea-lens system refracts the light onto the retina, which detects the light. The lens is focused by a set of muscles that flatten the lens to change its focal length. The focusing process is different from a standard camera. The camera moves the lens to achieve a different image distance. Our eye can't change the distance between the lens system and the retina. So our eye changes the shape of the lens, changing its focal length.



The crystalline lens becomes thicker, and therefore more curved when the ciliary muscle flexes. Austin Flint, "The Eye as an Optical Instrument," *Popular Science Monthly*, Vol. 45, p203, 1894 (Image in the public domain)

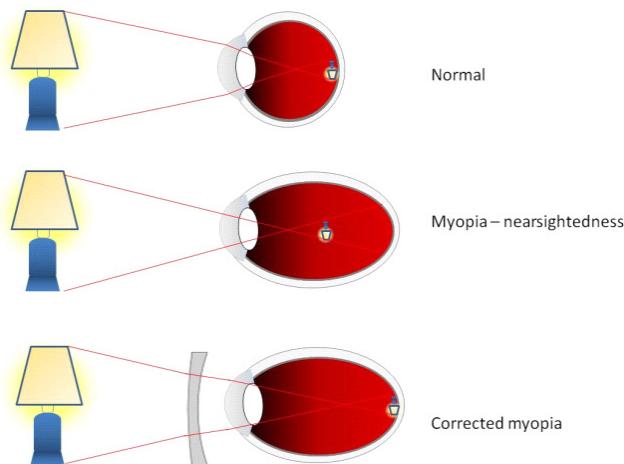
The focusing system is called accommodation. This system becomes less effective at about the time you reach an age of 40 years because the lens becomes less flexible. The closest point that can be focused by accommodation is called the near point. It is about 25 cm on average. There is, of course, no such thing as an average person, all of us are a little bit different. You young students probably have a much shorter near point than 25 cm. For those of us that are a little older, 25 cm or more is more likely.

The farthest point that can be focused is a long way away. It is called the far point. Both the near and far points degrade with years leading to bifocal glasses (and much irritation because you can't see as well, but I'm really not complaining because at least I can see).

The iris changes the area of the pupil (the aperture of the eye). The pupil is, on average, about 7 mm in diameter. This acts like the aperture adjustment of a camera.

17.4.1 Nearsightedness

In some people the cornea-lens system focuses in front of the retina. Usually this is because the shape of their eyes is not spherical but is elongated along the optic axis of the eye. This is called nearsightedness or myopia.

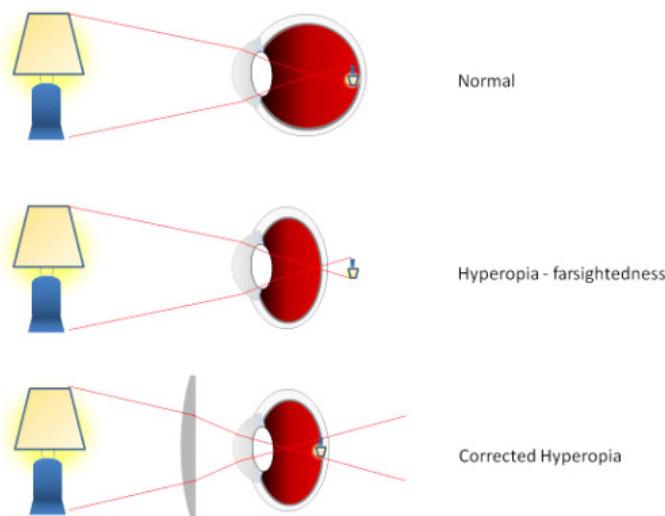


So their perfectly good cornea-lens system makes a great image in the vitreous humor (the jellylike stuff that fills the eye) and not on the retina where it would be detected. From your experience with lenses you know this would produce a blurry image. And that is what nearsighted people see much of the time. We can help nearsighted people by effectively changing the cornea-lens system of the eye by adding another lens. We want a diverging lens that makes the light more spread apart so that the lens system of the eye can make it focus where

the retina actually is. Alternately we could flatten the cornea, itself, with laser ablation.

17.4.2 Farsightedness

Sometimes the cornea-lens system focuses in back of the retina. This is usually because the eye grew too flat or “oblate.” This is called farsightedness or hyperopia. Once again we can fix the problem by adding an additional lens. This time a converging lens.



The converging lens will make the effective focal length of the system shorter, so that it can form an image where the retina actually is.

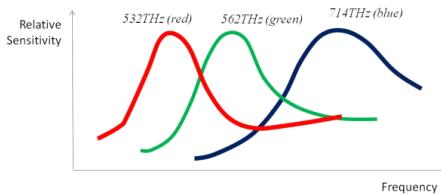
It would be convenient and very understandable if medical professionals prescribed eye glasses by telling us what focal length we needed to correct our vision. But that would not be the medical way! Eye glasses use a different unit of measure to describe how they bend light. The unit is the *diopter* and it is equal one over the focal length, but with the focal length measured in meters.

$$\text{diopter} = \frac{1}{f \text{ (m)}} \quad (17.7)$$

This measurement is called the *power* of the lens. It is just the same as giving the focal length, but less clear for science students.

17.4.3 Color Perception

The eye detects different colors. The receptors called cones can detect red, green, and blue light.



The eye combines the red, green, and blue response to allow us to perceive many different colors.

Most digital cameras also have red, green, and, blue pixels to provide color to images. The detectors in digital cameras are often have much narrower frequency bands than the eye. Likewise, television displays and monitors have red, green, and blue pixels. By targeting the eye receptors, power need not be wasted in creating light that is not detected well by the eye. The difference in band-width can cause problems in color mixing. Yellow school busses (perceived as different amounts of green and red light) may be reddish or green if the bandwidths are chosen poorly.

The science of human visual perception of imagery is called *image science*. There are many applications for this field, from forensics to intelligence gathering.

Chapter 18

18 Cameras and Magnifiers

3.2.6, 3.2.7

Fundamental Concepts

- Cameras and other imaging systems use a strange term called an f-number to tell what the intensity of the image will be

$$I \propto \frac{1}{(f/\#)^2}$$

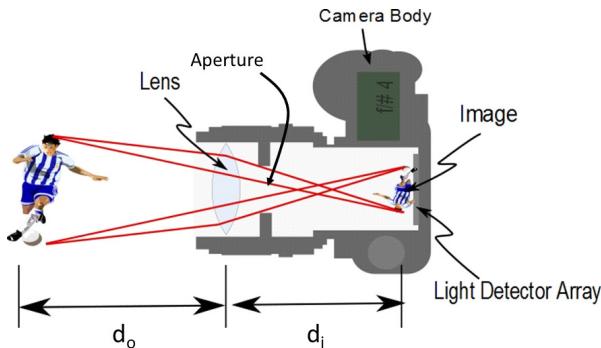
where $f/\#$ is the symbol for f-number and the f-number is given by

$$f/\# \equiv \frac{f}{D}$$

- Angular magnification $M = \frac{\theta}{\theta_o}$ describes how much bigger the image looks with an optical system compared to with just your eye.

18.1 The Camera

in 1900 George Eastman introduced the Brownie Camera. This event has changed society dramatically. The idea behind a camera is very simple.



The camera has a lens, often a compound lens made of several individual lenses, and a screen for projecting a real image created by the lens.

Let's take an example camera. Say we are at a family reunion and wish to take a picture of Aunt Sally. Aunt Sally is about 1.5 m tall. She is standing about 5 m away. Then to fit the image of Aunt Sally on our 35 mm detector, we must have

$$\begin{aligned} h_o &= 1.5 \text{ m} \\ h_i &= 0.035 \text{ m} \\ d_o &= 5 \text{ m} \\ f &= 0.058 \text{ m} \end{aligned}$$

We wish to find d_i and m . Let's do m first.

$$\begin{aligned} m &= \frac{h_i}{h_o} = \frac{-0.035 \text{ m}}{1.5 \text{ m}} \\ &= -2.3333 \times 10^{-2} \end{aligned}$$

so our image is small and inverted. The small size we wanted. But now we know that the image in our cameras is upside down. A digital camera uses its built-in computer to turn the image right side up for us on the display on the back of the camera.

Now let's find d_i .

$$\begin{aligned} d_i &= \frac{fd_o}{d_o - f} \\ &= 5.8681 \times 10^{-2} \text{ m} \\ &= 58.681 \text{ mm} \end{aligned}$$

so our detector must be 58.681 mm from the lens.

Now suppose we want to photograph a 1000 m tower from 2 km away. Then

$$\begin{aligned} m &= -\frac{0.035 \text{ m}}{1000 \text{ m}} \\ &= -3.5 \times 10^{-5} \end{aligned}$$

and

$$\begin{aligned} d_i &= \frac{(0.058 \text{ m})(2000 \text{ m})}{2000 \text{ m} - (0.058 \text{ m})} \\ &= 5.8002 \times 10^{-2} \text{ m} \\ &= 58.002 \text{ mm} \end{aligned}$$

Notice that the image distance changed, but not by very much. This is why you need a focus adjustment on the lens of a good camera. Objects far away require a different d_i value than objects that are close. Usually you twist the lens housing to make this adjustment. The lens housing has a threaded screw system that increases or decreases d_i as you twist. Cell phones and consumer cameras often have a motor that makes this adjustment for you. In some cameras you may see the lens move back and forth as someone takes a picture.

There are several things that govern whether a picture will be good. When you buy a quality manual lens, it will be marked in $f/\#$. The specification of an automatic lens will be given in terms of $f/\#$. To help us buy such lenses, we should understand what the terminology means.

Most things we want to take a snapshot of are much farther than 58 mm from the camera. For such objects we can revisit the magnification.

$$m = -\frac{d_i}{d_o}$$

but from the thin lens formula

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$$

If $d_o \gg f$ then we can say that $1/d_o \approx 0$ and so $d_i \approx f$. Then

$$m \approx -\frac{f}{d_o}$$

and we see that the size of the image is directly proportional to the focal distance. If we change the focal distance, we can change the size of the image. This is how a zoom lens works. A zoom lens is a compound lens, and the focal length is changed by increasing the distance between the component lenses (more on compound lenses soon). This is what your camera is doing when it zooms in and out when you push the telephoto button.

Remember we studied intensity

$$I = \frac{P}{A}$$

Photographic film and digital focal plane arrays detect the intensity of light falling on them. We can see that the area of our image depends on our magnification, which depends on d_i and for our distant objects it is proportional to

f . The image area is proportional to $d_i^2 \approx f^2$. So we can say that the area is proportional to f^2 . Then

$$I \propto \frac{P}{f^2}$$

The power entering the camera is proportional to the size of the aperture (hole the light goes through). A bigger aperture lets in more light. A smaller aperture lets in less light. If the camera has a circular opening, this area is proportional to the square of the diameter of the opening, D^2 so

$$I \propto \frac{D^2}{f^2}$$

This ratio is useful because it gives us a relative measure of how much intensity we get in terms of things we can easily know. Good cameras have changeable aperture sizes, and good lenses have changeable focal lengths. By using the combination of these two terms, we can ensure we will get enough light (but not too much) when we take the picture.

It would be good to give this ratio a special name. But instead, we named the ratio

$$f/\# \equiv \frac{f}{D} \quad (18.1)$$

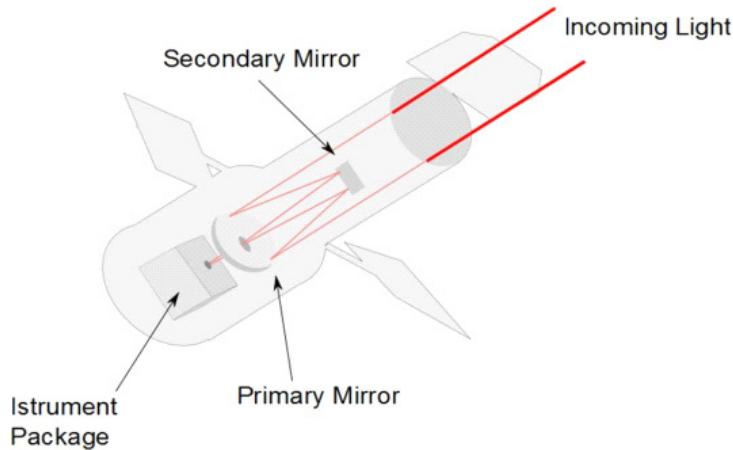
It is called the $f/\#$ (pronounced f-number) so

$$I \propto \frac{1}{(f/\#)^2} \quad (18.2)$$

So good cameras have adjustable lens systems marked in $f/\#$'s. Typical values are $f/2.8$, $f/4$, $f/5.6$, $f/8$, $f/11$, and $f/16$. Notice that since I is proportional to $(1/f/\#)^2$ these common $f/\#$ values give an increase of intensity of a factor of 2 each time you change the $f/\#$ by one marked stop.

$f/\#$ change	Intensity change factor
$f/11$ to $f/8$	2
$f/8$ to $f/5.6$	2
$f/5.6$ to $f/4$	2
$f/4$ to $f/2.8$	2

This terminology is used for telescope design as well. The Hubble telescope is an $f/24$ Ritchey-Chretien Cassegrainian system with a 2.4 m diameter aperture. The effective focal length is 57.6 m.



It is important to realize that electronic (and biological) sensors don't react instantly to what we see. The intensity is

$$I = \frac{P}{A} = \frac{\Delta E}{\Delta t A}$$

so there is a time involved. The time it takes to collect enough light to form an image on the sensor is called the *exposure time*.

$$\Delta t = \frac{\Delta E}{IA}$$

So changing our $f/\#$ changes the needed exposure time by changing the intensity. How sensitive our camera sensor is also affects the exposure time. Modern sensors have adjustable sensitivity. The photography world gives the three letters ISO for the name for this detector sensitivity. There isn't a standard for exactly what ISO setting gives what exposure. Different manufacturers use slightly different numbers. But a change in ISO settings usually are equivalent to one $f/\#$ change in exposure.

This is part of what a good photographer does in taking a picture. The photographer will adjust the $f/\#$ and the exposure time and ISO to get a photograph that is not too exposed (too light) or underexposed (too dark).

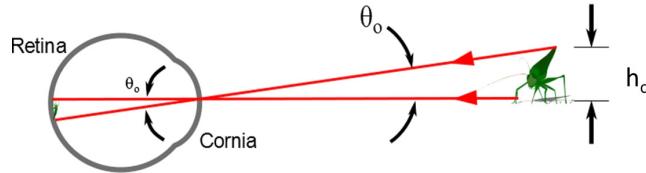
18.2 Angular Magnification

We already encountered the simple magnifier when we studied ray diagrams. But by this point in our study of optics you are probably wondering about our definition of magnification. If you are an Idahoan and are out hunting, when you look through your binoculars or scope you don't want to know how big the image is compared the actual deer, you want to know if you can see the deer better than you could with just your eyes. The magnification on your scope

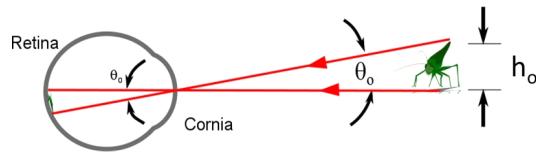
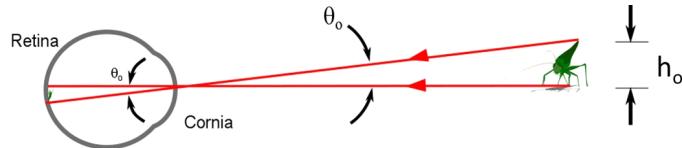
doesn't compare the image size to the object size, but it is comparison between two optical systems, one is just your eyes, the other is your eyes and the scope working together.

Let's use the simple magnifier that we know to define a new kind of magnification that does this comparison between two optical systems. We can use what we know about easy rays to draw to describe both optical systems. We usually use three principle rays, but for this analysis, let's just use one for each side of an object. Since the image is on the retina, we can see where these singular rays strike the retina and understand the size of the image on our eye light detection system. Let's choose to draw the rays that go straight through the middle of the lens of the eye (because they are the easiest ones!).

If we pick a ray from the top of our object that goes through the center of the lens, that ray won't seem to change direction at all. It will hit the retina to form the top of the image of the object. We can do the same for the bottom of the object. Then we can see from the next figure

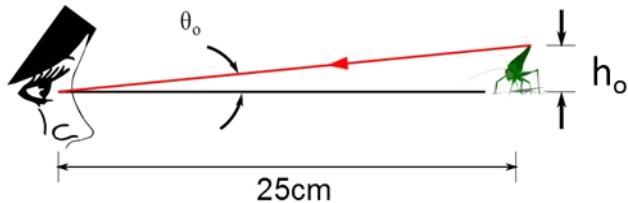


that the angle θ_o subtends¹ both the object and the image of the object. If you think about the previous figure, you will see that if the angle were to increase, so would the size of the image on the retina. We can increase this angle by, say, moving the object closer to our eyes.



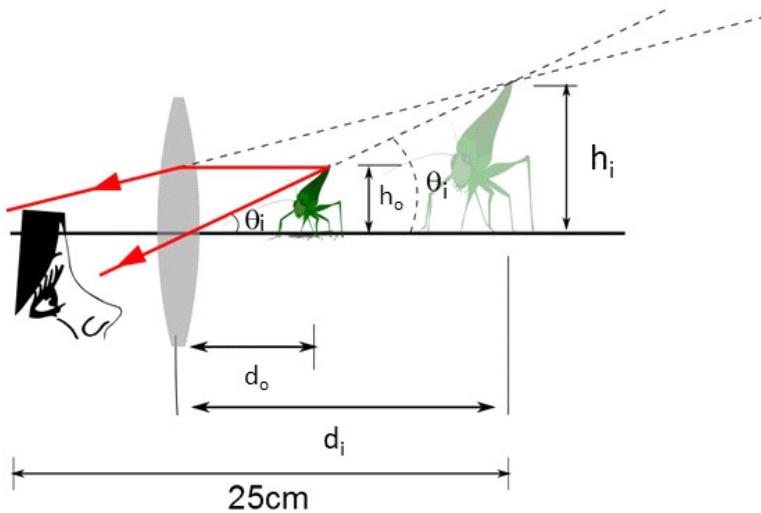
When we get to our 25 cm we reach the limit of the eye for focusing.

¹The angle that "subtends" an object is the one whose bounding lines just hit either side of the object.



If we move the object any closer, it will appear blurry. We call this position, the closest point where we can place an object and still bring it into focus with our eye, the *near point*. Thus the maximum value of θ will be at the near point for unaided viewing. We will call this maximum unaided angle θ_o .

But suppose we want to see this object in more detail. That is, we want to spread the image of the object over more of our retina so our retinal pixels do a better job of showing the details. We can use a magnifying glass to spread the light over more of our retina.

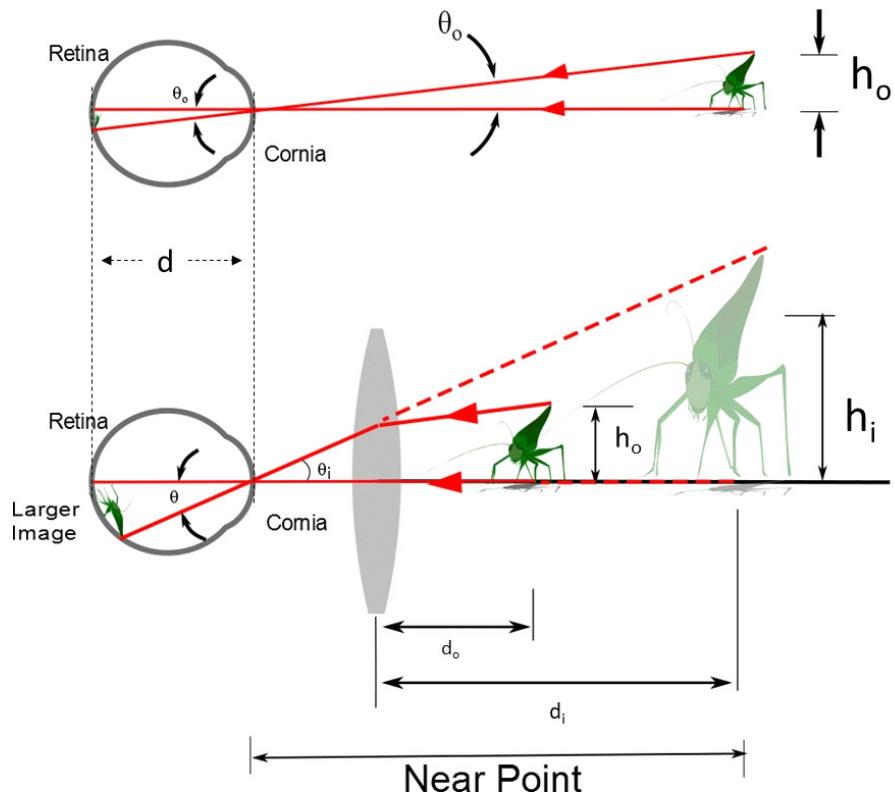


If we place the object closer to the magnifying glass than the focal distance ($d_o < f$), then we have a virtual image with magnification

$$m = \frac{-d_i}{d_o} \quad (18.3)$$

which is larger than one and positive (because d_i is negative).

But what we really want to know is how much bigger the image looks with the lens than it did without the lens. By looking at what happens to the rays when they enter our eye, we can see why the image looks bigger.



The magnifying glass has bent the light, and the bent rays make a bigger angle, θ_i , so the image on our retina is bigger. Since the image fills more of our retina, we perceive the image as being bigger.

We need a mathematical formula that will tell us how much bigger the image on our retina will be. Notice from the figure that for just our eye (no lens)

$$\tan \theta_o = \frac{h_{i,eye}}{d}$$

and for the case with the lens

$$\tan \theta_i = \frac{h_{i,lens-eye}}{d}$$

so

$$\begin{aligned} h_{i,eye} &= d \tan \theta_o \\ h_{i,lens-eye} &= d \tan \theta_i \end{aligned}$$

Then if we compare the new, larger image on the retina formed with the lens-eye system to the one formed with just the eye, we get

$$\frac{h_{i,lens-eye}}{h_{i,eye}} = \frac{d \tan \theta_i}{d \tan \theta_o} = \frac{\tan \theta_i}{\tan \theta_o}$$

and if we once again use the small angle approximation

$$\frac{h'_{lens-eye}}{h'_{eye}} \approx \frac{\theta}{\theta_o}$$

This would tell us how much bigger our object looks when viewed with the magnifying glass compared to how it looked without the magnifying glass. This is just what we want! Let's give this a new symbol

$$M = \frac{\theta_i}{\theta_o} \quad (18.4)$$

Remember, this is really different than the magnification we have studied before. The magnification we have been using compared the size of the image with the size of the object. We call M the *angular magnification*.

So, the angular magnification compares how big the object seems to be with and without a lens our lens system. It is really a comparison between the size of the area the diverging set of rays make on the retina formed with just our eye, and the one formed with the magnifier.

If the virtual image formed is farther than the near point of the eye, ($d_i > \sim 25$ cm) the image on our retina will be smaller than it would be at the near point because it is farther away. If the virtual image is closer than the near point, it will be fuzzy because the eye cannot focus closer than the near point. Thus, the value of M will be maximum when d_i for the magnifying glass is at the near point of the eye. We can find where to place the image so that we get maximum magnification. Taking just the magnifier, and placing the image at about -25 cm,

$$\begin{aligned} \frac{1}{d_o} + \frac{1}{d_i} &= \frac{1}{f} \\ \frac{1}{d_o} + \frac{1}{-25 \text{ cm}} &= \frac{1}{f} \end{aligned}$$

and so

$$\frac{1}{d_o} = \frac{-25 \text{ cm} - f}{-f(25 \text{ cm})}$$

or

$$d_o = \frac{(25 \text{ cm}) f}{25 \text{ cm} + f} \quad (18.5)$$

Using small angle approximations, we can write

$$\tan \theta_o = \frac{h_o}{25 \text{ cm}} \approx \theta_o$$

and

$$\tan \theta_i = \frac{h_o}{d_o} \approx \theta_i$$

then the maximum angular magnification is

$$\begin{aligned} m_{\max} &= \frac{\theta_i}{\theta_o} = \frac{\frac{h_o}{d_o}}{\frac{h_o}{25 \text{ cm}}} \\ &= \frac{25 \text{ cm}}{\frac{25 \text{ cm} f}{25 \text{ cm} + f}} \\ &= \frac{25 \text{ cm} + f}{f} \\ &= 1 + \frac{25 \text{ cm}}{f} \end{aligned}$$

We can also find the minimum magnification by letting d_o be at f . This gives

$$\theta_i = \frac{h}{f}$$

$$\begin{aligned} m_{\min} &= \frac{\theta_i}{\theta_o} = \frac{\frac{h}{f}}{\frac{h}{25 \text{ cm}}} \\ &= \frac{25 \text{ cm}}{f} \end{aligned}$$

We use the idea of a simple magnifier in combination with other lenses to make the magnification happen in telescopes, microscopes, and other instruments that magnify. So all of these systems can be described using the idea of an angular magnification.

Angular magnification is important. Many optical systems that we buy are really designed to make things seem bigger (or smaller) than they seem with just our eyes. Things like telescopes and microscopes and binoculars. We will see how to apply this idea of angular magnification to some such systems in our next lecture.

Chapter 19

19 Optical Systems and back to waves 3.2.8, 3.3 1

So far we have looked at single lenses and how they work. But we hinted that actual camera lenses and even our eyes are have lens systems that are made of more than one lens. We should look at how lens systems work.

We should also return to our study of waves because it turns out that the wave nature of light gives us a fundamental limit on how good our imagery can be from our cameras we design.

Fundamental Concepts

- The magnification of a two-lens system is just the product of the magnifications of the individual lenses $m_{\text{combined}} = m_1 m_2$
- If the two lenses in a two-lens system are placed so the distance between them is essentially zero, then the focal length of the two-lens system is given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}$$

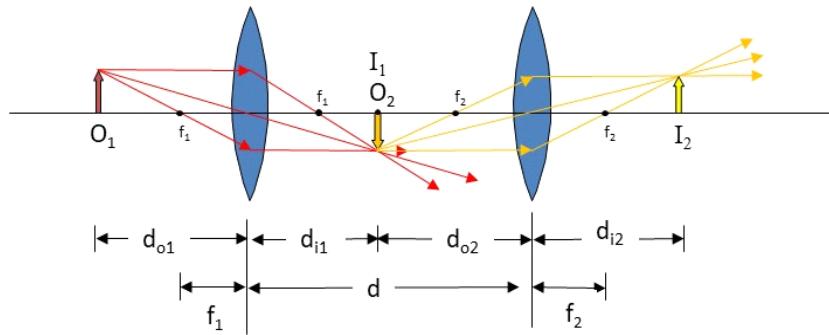
where f_1 and f_2 are the focal lengths of the individual lenses

- If the two lenses in a two-lens system are not placed so the distance between them is essentially zero, the previous equation is not valid!
- Telescopes and Microscopes are double lens systems
- Light waves experience constructive and destructive interference like other waves
- Young's double slit experiment

19.1 Combinations of Lenses

We found that to correct chromatic aberration we used two lenses. Together they are called an “achromat” and all good cameras use achromats to fix chromatic aberration. We can do something similar to correct for spherical aberration. But in each case, we are combining two lenses (or even more!). And now we have microscopes that have two lenses. How do we predict what the lens system will do? It turns out that we have all we need to know to combine lenses already.

To combine lenses, we do the same thing we did for the two surfaces of a thin lens. We form the image from the first lens as though the second lens is not there. Then we use the image from the first lens as the object for the second lens. Suppose we take two lenses of focal lengths f_1 and f_2 and place them a distance d apart.



Because this system would use a magnified image as the object for lens 2, the final magnification is the product of the two lens magnifications

$$m_{\text{combined}} = m_1 m_2 \quad (19.1)$$

For the first lens we have

$$\frac{1}{d_{o1}} + \frac{1}{d_{i1}} = \frac{1}{f_1} \quad (19.2)$$

where d_{i1} is our first lens image distance. We can solve for d_{i1}

$$d_{i1} = \frac{d_{o1} f_1}{d_{o1} - f_1} \quad (19.3)$$

We then take as the second object distance

$$d_{o2} = d - d_{i1}$$

and we use the lens formula again.

$$\frac{1}{d_{o2}} + \frac{1}{d_{i2}} = \frac{1}{f_2}$$

and again find the image distance

$$d_{i2} = \frac{d_{o2}f_2}{d_{o2} - f_2}$$

but we can use our value of d_{o2} to find

$$\begin{aligned} d_{i2} &= \frac{(d - d_{i1})f_2}{(d - d_{i1}) - f_2} \\ &= \frac{(d - d_{i1})f_2}{d - d_{i1} - f_2} \end{aligned}$$

We have an expression relating the image distances, d , d_{i1} , d_{i2} and f_2 . But we would really like to have an expression that relates d_{o1} and d_{i2} . Lets use

$$d_{i1} = \frac{d_{o1}f_1}{d_{o1} - f_1}$$

and substitute it into our expression for d_{i2}

$$d_{i2} = \frac{\left(d - \frac{d_{o1}f_1}{d_{o1} - f_1}\right)f_2}{d - \frac{d_{o1}f_1}{d_{o1} - f_1} - f_2}$$

This looks messy, but we can do some simplification

$$d_{i2} = \frac{df_2 - \frac{d_{o1}f_1f_2}{d_{o1} - f_1}}{d - f_2 - \frac{d_{o1}f_1}{d_{o1} - f_1}} \quad (19.4)$$

Well, it is still a little messy, but we have achieved our goal. We have d_{i2} in terms of the focal lengths, d , and d_{o1} .

Suppose we let $d \rightarrow 0$. That would mean the two lenses are right next to each other, like for an achromat. Then

$$\begin{aligned} d_{i2} &= \frac{-\frac{d_{o1}f_1f_2}{d_{o1} - f_1}}{-f_2 - \frac{d_{o1}f_1}{d_{o1} - f_1}} \\ &= \frac{\frac{d_{o1}f_1f_2}{d_{o1} - f_1}}{\frac{f_2(d_{o1} - f_1)}{d_{o1} - f_1} + \frac{d_{o1}f_1}{d_{o1} - f_1}} \\ &= \frac{d_{o1}f_1f_2}{f_2d_{o1} - f_2f_1 + d_{o1}f_1} \\ &= \frac{d_{o1}f_1f_2}{d_{o1}(f_2 + f_1) - f_2f_1} \end{aligned}$$

So

$$d_{i2} = \frac{d_{o1}f_1f_2}{d_{o1}(f_2 + f_1) - f_2f_1}$$

Lets undo the math that brought us d_{i2} in the first place

$$\begin{aligned}\frac{1}{d_{i2}} &= \frac{d_{o1}(f_2 + f_1) - f_2 f_1}{d_{o1} f_1 f_2} \\ &= \frac{d_{o1}(f_2 + f_1)}{d_{o1} f_1 f_2} - \frac{f_2 f_1}{d_{o1} f_1 f_2} \\ &= \frac{(f_2 + f_1)}{f_1 f_2} - \frac{1}{d_{o1}}\end{aligned}$$

or

$$\frac{1}{d_{i2}} + \frac{1}{d_{o1}} = \frac{(f_2 + f_1)}{f_1 f_2}$$

Which looks very like the lens formula with

$$\frac{1}{f} = \frac{(f_2 + f_1)}{f_1 f_2}$$

If we unwind this expression, we find

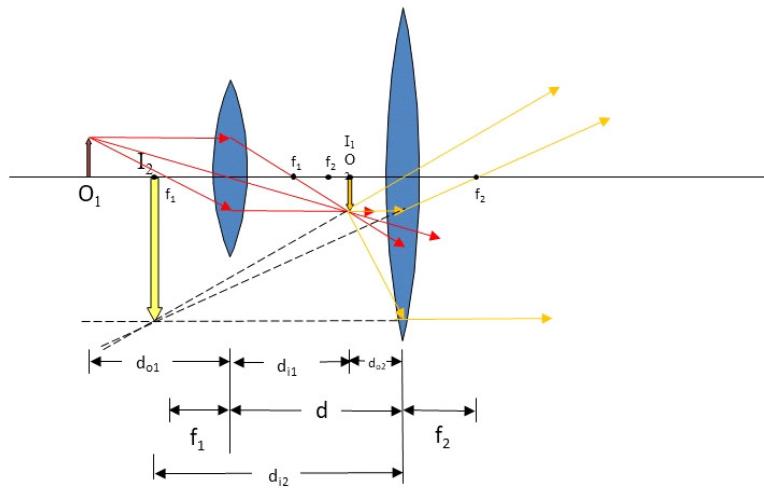
$$\begin{aligned}\frac{1}{f} &= \frac{f_2}{f_1 f_2} + \frac{f_1}{f_1 f_2} \\ \frac{1}{f} &= \frac{1}{f_1} + \frac{1}{f_2}\end{aligned}\tag{19.5}$$

This is how we combine thin lenses. We see that the two lenses are equivalent to a single lens with focal length f as long as they are close together.

Of course, we had to place our lenses right next to each other for this to work. This is not the case for a telescope or microscope. We should look at such a case. There is no need for more math. We can go back to equation (19.4).

$$d_{i2} = \frac{df_2 - \frac{d_{o1}f_1f_2}{d_{o1}-f_1}}{d - f_2 - \frac{d_{o1}f_1}{d_{o1}-f_1}}$$

But let's look at a case using ray diagrams. For this case, let's take two lenses, and let's have the first lens make a real image. Once again, let's have that image be the object for the second lens. But this time, let's move the second lens so that the image from the first lens (object for the second lens) is closer to the second lens than f_2 . If that is the case, the second lens works like a magnifier. The final image is enlarged.

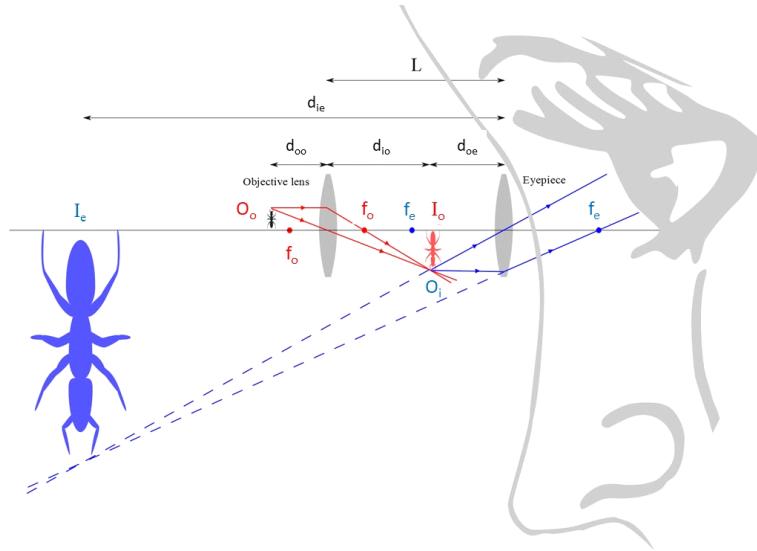


This is the basis of both microscopes and telescopes!

19.2 The Microscope

Microscopes are designed to allow us to see very small things. They go beyond a simple magnifier. But it turns out we don't have to go very far beyond a magnifying glass. To see things that are very small, we add another lens to our simple magnifier. We will place this lens near the object and call this lens the *objective* because it is next to the object. We will keep a simple magnifier and place it near the eye. This lens will be called the *eyepiece* because it is near your eye.

The objective will have a very short focal length. The eyepiece will have a longer focal length (a few centimeters).



We separate the lenses by a distance L where

$$\begin{aligned} L &> f_o \\ L &> f_e \end{aligned}$$

We place the object *just outside* the focal point of the objective lens. The image formed by the objective lens is then real and inverted. Then (and here is the trick we just learned!) we use this real image formed by the objective lens as the new object for the eyepiece. For a real image, there is really light there. So a real image can act like light bouncing off of an object. We can use the real image as an object for the eyepiece.

The image formed by the eyepiece is upright and virtual, but it looks upside down because the object for the eyepiece (the image made by the objective) is upside down.

To say it once more, what we are doing is forming a real image using the objective lens and then using the eyepiece lens as a simple magnifier to make a larger virtual image of the real image made by the objective lens. We could find d_{ie} using equation (19.4). But this is a microscope. We want to know how much bigger our final image is.

Of course, we really want to know the magnification of the system. It's not just the eyepiece that will magnify. We know what the eyepiece will do because it is being used as just a simple magnifier. Recall that magnifications are a factors. Think from our basic equation

$$m = \frac{h_i}{h_o}$$

tells us that

$$h_i = mh_o$$

or in words, m is the factor by which h_i is bigger (or smaller) than h_o . So magnifications are factors. For example, d_i could be 10 time bigger than d_o . And we have two lenses. Each lens will have a magnification of it's object. But how do we combine the magnifications of the two lenses?

To find the overall system magnification we start with the magnification of the eyepiece M_e , say, 20 times bigger than it would look with just our eye. so whatever the object, the eyepiece gives us a virtual image of that object that is M_e times bigger than it would look without the eyepiece. But the object in our case is the real image formed by the objective lens. And this real image may be bigger than the actual object. If you think about it, it makes sense that if the objective makes the object look 10 times bigger, and the eyepiece makes the image look 20 times bigger than if you looked at it with your eye, the system makes it look 200 times bigger. The combined magnification for the two-lens system is

$$m = m_o M_e \quad (19.6)$$

The minimum magnification of the eye piece will be roughly

$$M_{e_{\min}} \approx \frac{25 \text{ cm}}{f_e}$$

and the maximum will be

$$M_{e_{\max}} = 1 + \frac{25 \text{ cm}}{f_e}$$

But remember the object for the eyepiece is the image from the first lens. And that image is larger than the object by an amount

$$m_o = \frac{-d_{io}}{d_{oo}}$$

where d_{io} is the image distance for the objective lens and d_{oo} is the object distance for the objective lens. And let's estimate how big the magnification due to the first lens will be. Because for a microscope $d_{oo} \approx f_o$. and $d_{io} \approx L$ (very roughly)

$$m_o = \frac{-d_{io}}{d_{oo}} \approx -\frac{L}{f_o}$$

The combined magnification for the two-lens system is about

$$m_{system} = m_o M_e \approx -\frac{L}{f_o} \frac{25 \text{ cm}}{f_e} \quad (19.7)$$

this is the minimum magnification (because we used the minimum magnification formula for the eyepiece).

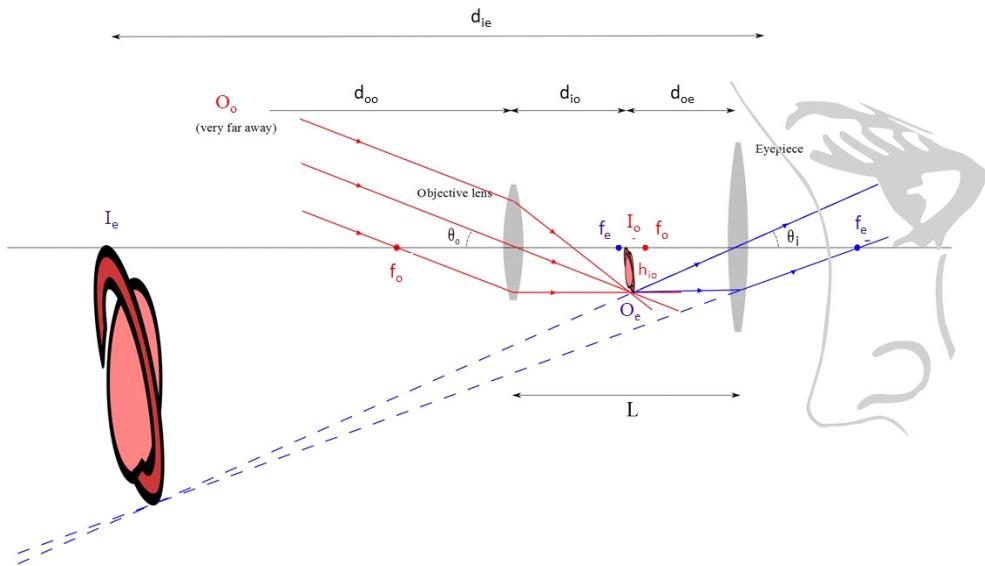
This gives us a rough idea of how a microscope works, let's try telescopes next.

19.3 Telescopes

There are two main types of telescopes *refracting* and *reflecting*. We will study refracting telescopes first.

19.3.1 Refracting Telescopes

Like the microscope, we combine two lenses and call one the objective and the other the eyepiece. The eyepiece again plays the role of a simple magnifier, magnifying the image produced by the objective.



We again form a real, inverted image with the objective. We are now looking at distant objects, so the image distance $d_{io} \approx f_o$. Once again, we use the image from the objective as the object for the eyepiece. The eye piece forms an upright virtual image (that looks inverted because the object for the eyepiece is the image from the objective, and the real image from the objective is inverted). Once again we could find d_{ie} using equation (19.4). But once again we want the magnification.

The largest magnification is when the rays exit the eyepiece parallel to the principal axis. Then the image from the eyepiece is formed at near infinity (but it is very big, so it is easy to see). This gives a lens separation of $f_o + f_e$ which will be roughly the length of the telescope tube.

The angular magnification will be

$$M = \frac{\theta_i}{\theta_o} \quad (19.8)$$

where θ_o is the angle subtended by the object at the objective (see figure above) and θ is subtended by the final image at the viewer's eye. Consider s_o is very

large. We see from the figure that

$$\tan \theta_o \approx -\frac{h_{io}}{f_o} \quad (19.9)$$

and with d_{oo} large we can use small angles.

$$\theta_o \approx -\frac{h_{io}}{f_o} \quad (19.10)$$

The angle θ_i will be the angle formed by rays bent by the lens of the eyepiece. This angle will be the same as the angle formed by a ray traveling from the tip of the first image and traveling parallel to the principal axis. This ray is bent by the objective to pass through f_e . Then

$$\tan \theta \approx \frac{h_{io}}{f_e} \approx \theta \quad (19.11)$$

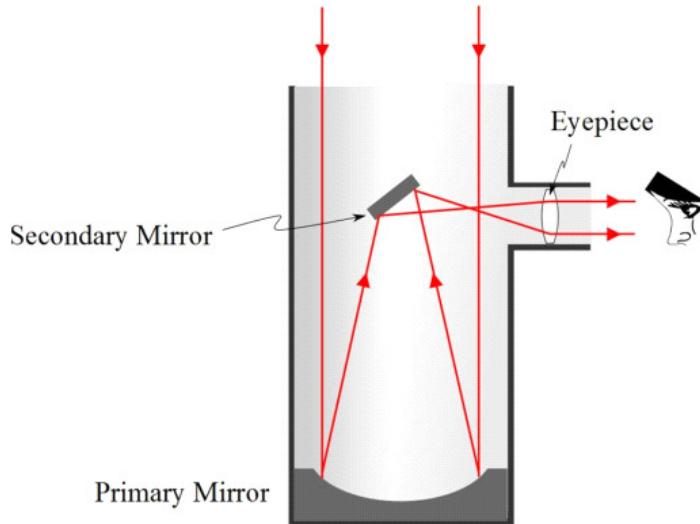
so

The magnification is then

$$M = \frac{\theta_i}{\theta_o} \approx \frac{\frac{h_{io}}{f_e}}{-\frac{h_{io}}{f_o}} \approx -\frac{f_o}{f_e} \quad (19.12)$$

19.3.2 Reflecting Telescopes

Reflecting telescopes use a series of mirrors to replace the objective lens. Usually, there is an eyepiece that is refractive (although there need not be, radio frequency telescopes rarely have refractive pieces).



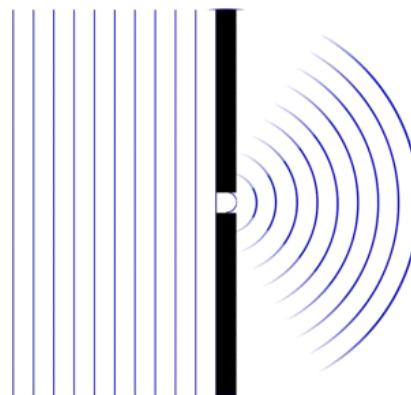
There are two reasons to build reflective telescopes. The first is that reflective optics do not suffer from chromatic aberration. The second is that large mirrors are much easier to make and mount than refractive optics. The Keck Observatory in Hawaii has a 10 m reflective system. The largest refractive system is a 1 m system. The Hubble telescope has a 2.5 m aperture and James Webb has a 6 m. Reflective telescopes are just easier to build.

The telescope pictured in the figure is a Newtonian, named after Newton, who designed this focus mechanism. Many other designs exist. Popular designs for space applications include the cassigrain telescope.

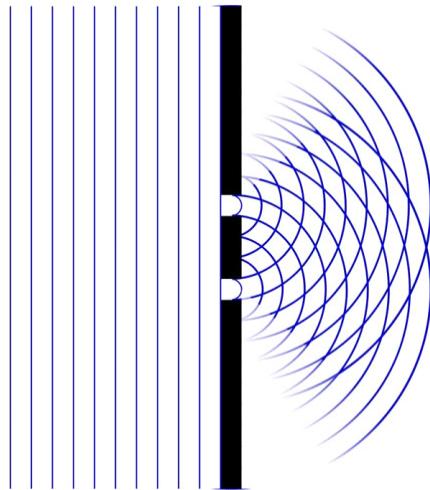
The rough design of a reflective telescope can be worked out using refractive pieces, then the rough details of the reflective optics can be formed.

19.4 Interference and Young's Experiment

Waves do some funny things when they encounter barriers. Think of a water wave. If we pass the wave through a small opening in a barrier, the wave can't all get through the small hole, but it can cause a disturbance right at the opening. We know that a small disturbance will cause a wave. But this wave will be due to a very small—almost point—source. Point sources make spherical waves. So the waves on the other side of the small opening will be nearly spherical. The smaller the opening the more pronounced the curving of the wave, because the source (the hole) is more like a point source.

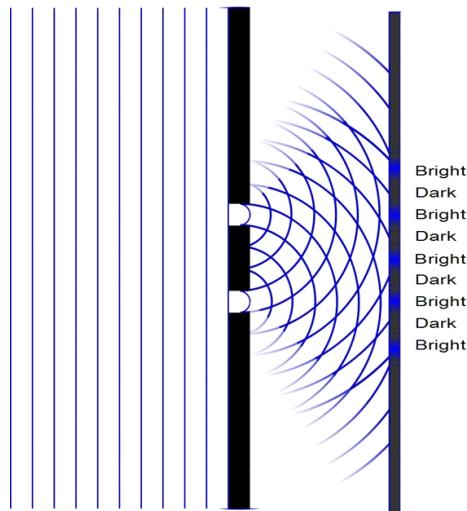


Now suppose we have two of these openings. We expect the two sources to make curved waves and those waves can interfere.



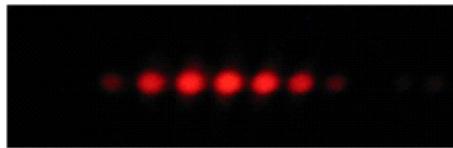
None of this is really new. We know waves generally have a spherical shape, and we have a disturbance in the wave medium where the two holes are that is experiencing simple harmonic motion. The holes are really sources for new spherical waves.

In the figure, we can already see that there will be constructive and destructive interference were the waves from the two holes meet. Thomas young predicted that we should see constructive and destructive interference in light (he drew figures very like the ones we have used to explain his idea).



Young set up a coherent source of light and placed it in front of this source a barrier with two very thin slits cut in it to test his idea.. He set up a screen beyond the barrier and observed the pattern on the screen formed by the light.

This (in part) is what he saw



We see bright spots (constructive interference) and dark spots (destructive interference). Only wave phenomena can interfere, so this is fairly good evidence that light is a wave.

But in our study of mirrors and lenses, beyond the forming of rays we just didn't have constructive and destructive interference. But from Young's experiment, clearly we should sometimes see more of the wave nature of light. It is time to return to waves and see the details of how light waves interact.

Chapter 20

20 Double Slits 3.3.2, 3.3.3

Fundamental Concepts

- Two wave mixing works for light
- Visible light wavelengths are small, we can use the small angle approximation
- Multiple slits give areas of constructive and destructive interference.

20.0.1 Constructive Interference

We can find the condition for getting a bright or a dark band if we think about it a bit. The two waves coming from the two slits are just two different waves. We can use our two wave mixing analysis with our constructive and destructive interference criteria. For constructive interference, the difference in phase must be a multiple of 2π . Let's review this, but instead of y_{\max} being the displacement of the waves, let's write them as E_{\max} because it is the electric field that is carrying the wave and it doesn't really go up or down, it just gets a higher value or lower value. Our waves will be

$$\begin{aligned} E_1 &= E_{\max} \sin(k_2 r_2 - \omega_2 t_2 + \phi_2) \\ E_2 &= E_{\max} \sin(k_1 r_1 - \omega_1 t_1 + \phi_1) \end{aligned}$$

and the resulting wave will be

$$\begin{aligned}
 E_r &= E_{\max} \sin(k_2 r_2 - \omega_2 t_2 + \phi_2) + A \sin(k_1 r_1 - \omega_1 t_1 + \phi_1) \\
 &= 2E_{\max} \cos\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) - (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \\
 &\quad \times \sin\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) + (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \\
 &= 2E_{\max} \cos\left(\frac{1}{2}[(k_2 r_2 - \omega_2 t_2 + \phi_2) - (k_1 r_1 - \omega_1 t_1 + \phi_1)]\right) \\
 &\quad \times \sin\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) + (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right)
 \end{aligned}$$

As we are now well aware, the sine part is a combined wave, and the cosine part is part of the amplitude. The amplitude can be written as

$$A = 2E_{\max} \cos\left(\frac{1}{2}[(k_2 r_2 - \omega_2 t_2 + \phi_2) - (k_1 r_1 - \omega_1 t_1 + \phi_1)]\right)$$

We should pause and think about which of our values will change. The frequency won't change, $\omega_2 = \omega_1 = \omega$. There is no slower material, so the wavelength won't change, $k_2 = k_1 = k$. We want the waves to mix at the same time so $t_2 = t_1 = t$. And the new waves are created from the old wave hitting the slits. As long as the original wave hits both slits at once, then $\phi_2 = \phi_1 = \phi_o$. We are left with

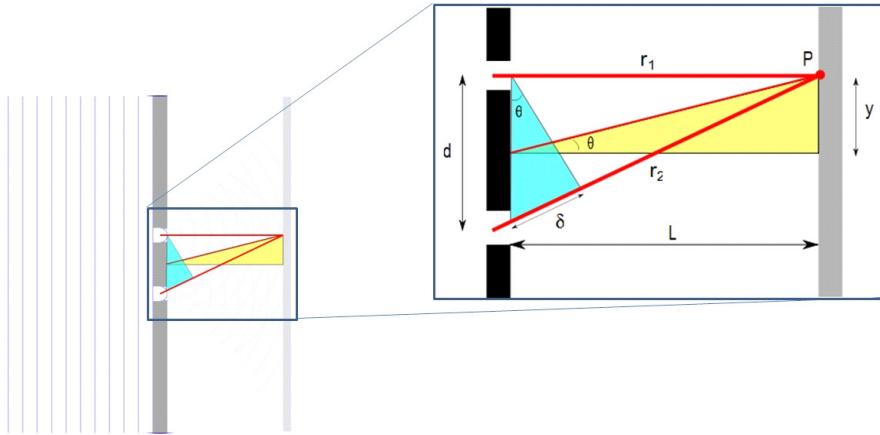
$$\begin{aligned}
 A &= 2E_{\max} \cos\left(\frac{1}{2}[(kr_2 - \omega t + \phi_o) - (kr_1 - \omega t + \phi_0)]\right) \\
 &= 2E_{\max} \cos\left(\frac{1}{2}[(kr_2) - (kr_1)]\right) \\
 &= 2E_{\max} \cos\left(\frac{1}{2}k(r_2 - r_1)\right) \\
 &= 2E_{\max} \cos\left(\frac{1}{2}k\Delta r\right)
 \end{aligned}$$

That means that to have constructive interference the path difference between the two slit-sources must be an even number of wavelengths. We have been calling the path difference in the total phase Δx , or for spherical waves Δr , but in optics it is customary to call this path difference δ or even $\Delta\ell$. So

$$\delta = \Delta r$$

But let's write the amplitude function as

$$A = 2E_{\max} \cos\left(\frac{1}{2}k\Delta r\right)$$



so our total phase equation becomes

$$\Delta\phi = k\Delta r$$

We want to use this in our criteria for constructive or destructive interference. But first, let's express Δr in terms of geometry that is easy to measure in our experiment. In this set up, the screen is much farther away than d , the slit distance, we can say that the blue triangle is almost a right triangle, and then Δr is

$$\Delta r = r_2 - r_1 \approx d \sin \theta$$

Our wave repeats every 2π radians or every wavelength, λ , then we have constructive interference (a bright spot) when

$$\Delta\phi = k\Delta r \approx kd \sin \theta = 2\pi m \quad (m = 0, \pm 1, \pm 2 \dots)$$

where m is an integer. We can solve this for Δr

$$\begin{aligned} \Delta r &= d \sin \theta = \frac{2\pi m}{k} \quad (m = 0, \pm 1, \pm 2 \dots) \\ &= \frac{2\pi m}{\frac{2\pi}{\lambda}} \quad (m = 0, \pm 1, \pm 2 \dots) \end{aligned}$$

or

$$\Delta r = d \sin \theta = m\lambda \quad (m = 0, \pm 1, \pm 2 \dots)$$

and we can give m a name. It is called the *order number*. That is, if we are off by any number of whole wavelengths then our total phase due to path difference will be a multiple of 2π and we will have constructive interference.

If we assume that $\lambda \ll d$ we can find the distance from the axis for each fringe more easily. This guarantees that θ will be small. Using the yellow triangle we see

$$\tan \theta = \frac{y}{L}$$

but if θ is small this is just about the same as

$$\sin \theta = \frac{y}{L}$$

because for small angles $\tan \theta \approx \sin \theta \approx \theta$. So if θ is small then

$$\begin{aligned}\Delta r &= d \sin \theta \\ &= d \frac{y}{L}\end{aligned}$$

and for a bright spot or “fringe” we find

$$d \frac{y}{L} = m\lambda$$

Solving for the position of the bright spots gives

$$y_{bright} \approx \frac{\lambda L}{d} m \quad (m = 0, \pm 1, \pm 2, \dots) \quad (20.1)$$

We can measure up from the central spot and predict where each successive bright spot will be.

20.0.2 Destructive Interference

We can also find a condition for destructive interference. We know that a path difference of an odd multiple of a half wavelength will give distractive interference. so

$$\Delta\phi = k\Delta r \approx kd \sin \theta = \left(m + \frac{1}{2}\right) 2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots$$

or using $k = 2\pi/\lambda$

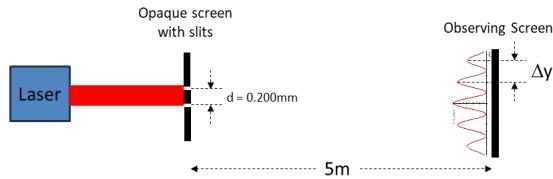
$$\Delta r = d \sin \theta = \left(m + \frac{1}{2}\right) \lambda \quad (m = 0, \pm 1, \pm 2, \dots)$$

will give a dark fringe. The location of the dark fringes will be

$$y_{dark} \approx \frac{\lambda L}{d} \left(m + \frac{1}{2}\right) \quad (m = 0, \pm 1, \pm 2, \dots) \quad (20.2)$$

Let's try an example problem.

Suppose we have a laser beam that produces pretty red light with $\lambda = 632.8 \text{ nm}$. And suppose we shine this laser on an opaque screen with two slits that are 0.020 mm apart. And suppose we let the light hit a screen that is 5.00 m away from the slits. How far apart will the bright interference fringes be on the screen?



PT: This is a two slit problem

BE

$$y_{bright} = \frac{\lambda L}{d} m \quad m = 0, \pm 1, \pm 2, \dots$$

$$\lambda = 632.8 \text{ nm}$$

$$d = 0.020 \text{ mm}$$

$$L = 5.00 \text{ m}$$

Solution

We want the space between two bright fringes. Δy , this is

$$\begin{aligned} \Delta y &= \frac{\lambda L}{d} (m+1) - \frac{\lambda L}{d} (m) \\ &= \frac{\lambda L}{d} ((m+1) - (m)) \\ &= \frac{\lambda L}{d} \end{aligned}$$

or

$$\Delta y = \frac{\lambda L}{d}$$

$$\begin{aligned} \Delta y &= \frac{(632.8 \text{ nm})(5.00 \text{ m})}{0.020 \text{ mm}} \\ &= 0.1582 \text{ m} \\ &= 15.82 \text{ cm} \end{aligned}$$

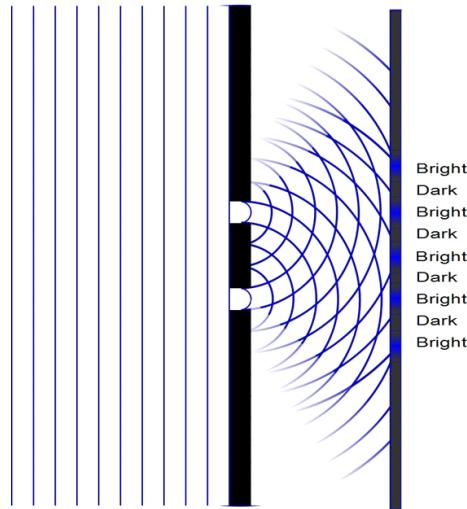
Units check

Reasonableness check

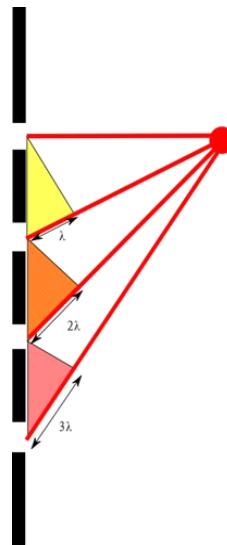
This may seem like a mostly useless thing to measure, but let's think of this problem backwards. If we have two slits and measure a pattern of 15.82 cm peak separation we can know that the slits were 0.200 mm. Light interference is used in industry for measuring very small distances on the order of tens of microns.

20.1 Multiple Slits

What happens if we have more than two slits?



At some point, two of the slits will have a path difference that is a whole wavelength, and we would expect a bright spot. But what about the other slits? If we have a slit spacing such that each of the succeeding slits has a path difference that is just an additional wavelength, then each of the slits will contribute to the constructive interference at our point, and the point will become a bright spot.



Let's look at just two slits again. The light leaves each slit in phase with the light from the rest of the slits, but at some distance L away and at some angle

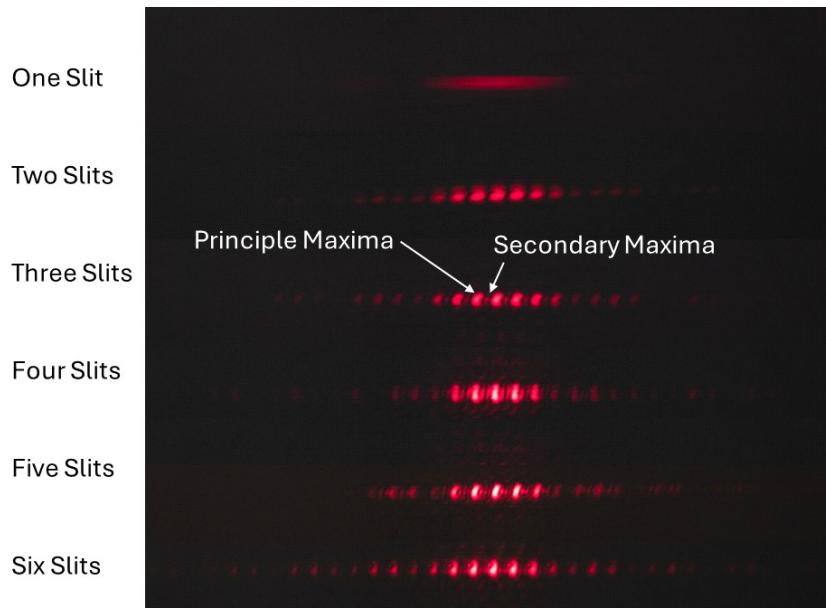
θ we will have a path difference

$$\delta = d \sin(\theta_{bright}) = m\lambda \quad m = 0, \pm 1, \pm 2, \dots \quad (20.3)$$

because the path lengths are not all the same. But we can see that if we have more than two slits, say, four slits, then this same equation would work to find a bright spot for the top two slits. And it would also work for the bottom two slits. And it would work for the middle two slits! Then for every set of two slits we get constructive interference at the same angle. That means at that spot on that angle all the constructive interference from sets of two slits add up to make a super constructive interference. This will be a very bright spot. We call this a principle maxima.

Of course we could find a spot where two of the slits were $\lambda/2$ off so there will be destructive interference from these two slits. But then a third slit would still send light at that angle. So there would be some brightness. So this isn't total destructive interference. There will be a spot less bright than the principle maxima, but not totally dark. We call these secondary maxima.

In the next figure, you can see laser light that passes through increasing numbers of slits.



Notice that as the number of slits grows, the principle maxima get brighter until they turn white because they are oversaturating the camera sensor. These patterns of light that come from passing light through holes and letting the light mix together again are called *diffraction patterns*.

Chapter 21

21 Thin Films and Light Interference 3.3.4, 3.3.5

Last lecture we reminded ourselves that light is a wave. In this lecture let's look at some examples of how the wave nature of light can be used. Applications for the using light waves stretch from radar to digital communications to atmospheric chemical analysis. But let's take on the ideas of interference from thin films and from different path lengths.

Fundamental Concepts

- Thin films can cause reflections that create interference
- There may be a phase shift when a wave reflects (the wave may invert-this is a review)
- If two waves are out of phase by half a wavelength we have total destructive interference (review)
- If two waves are out of phase by a full wavelength we have total constructive interference (review)
- Other phases provide partial constructive or partial destructive interference

21.1 Mathematical treatment of single frequency interference

Let's review our equations for interference of two wave. We start with two waves in the same medium. Only now our waves are light waves. So the medium is

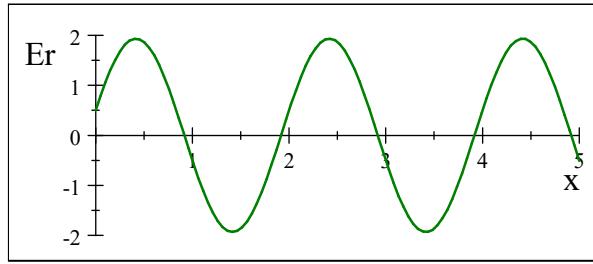
the electromagnetic field. So we can say our amplitude is E_{\max}

$$\begin{aligned} E_1 &= E_{\max} \sin(kx_1 - \omega t + \phi_1) \\ E_2 &= E_{\max} \sin(kx_2 - \omega t + \phi_2) \end{aligned}$$

Each wave has its own phase constant. Each wave starts from a different position (one at x_1 and the other at x_2). The superposition yields.

$$E_r = E_{\max} \sin(kx_1 - \omega t + \phi_1) + E_{\max} \sin(kx_2 - \omega t + \phi_2)$$

which is graphed in the next figure.



Notice that the wave form is taller (larger amplitude). Noticed it is shifted along the x axis. This graph is not surprising to us now, because we have done a case like this before. We can find the shift in general rewriting y_r . We need a trig identity

$$\sin a + \sin b = 2 \cos\left(\frac{a-b}{2}\right) \sin\left(\frac{a+b}{2}\right)$$

then let $a = kx_2 - \omega t + \phi_2$ and $b = kx_1 - \omega t + \phi_1$

$$\begin{aligned} E_r &= E_{\max} \sin(kx_2 - \omega t + \phi_2) + E_{\max} \sin(kx_1 - \omega t + \phi_1) \\ &= 2E_{\max} \cos\left(\frac{(kx_2 - \omega t + \phi_2) - (kx_1 - \omega t + \phi_1)}{2}\right) \sin\left(\frac{(kx_2 - \omega t + \phi_2) + (kx_1 - \omega t + \phi_1)}{2}\right) \\ &= 2E_{\max} \cos\left(\frac{kx_2 - kx_1}{2} + \frac{\phi_2 - \phi_1}{2}\right) \sin\left(\frac{kx_2 + kx_1 - 2\omega t + \phi_2 + \phi_1}{2}\right) \\ &= 2E_{\max} \cos\left(k\frac{x_2 - x_1}{2} + \frac{\phi_2 - \phi_1}{2}\right) \sin\left(k\frac{x_2 + x_1}{2} - \omega t + \frac{\phi_2 + \phi_1}{2}\right) \\ &= 2E_{\max} \cos\left(k\frac{\Delta x}{2} + \frac{\Delta\phi_o}{2}\right) \sin\left(k\frac{x_2 + x_1}{2} - \omega t + \frac{\phi_2 + \phi_1}{2}\right) \\ &= 2E_{\max} \cos\left(\frac{1}{2}\left(\frac{2\pi}{\lambda}\Delta x + \Delta\phi_o\right)\right) \sin\left(k\frac{x_2 + x_1}{2} - \omega t + \frac{\phi_2 + \phi_1}{2}\right) \\ &= 2E_{\max} \cos\left(\frac{1}{2}(\Delta\phi)\right) \sin\left(k\frac{x_2 + x_1}{2} - \omega t + \frac{\phi_2 + \phi_1}{2}\right) \end{aligned}$$

where the last line is just a rearrangement to match the form we got last time. As usual, let's look at the sine part first. It is still a wave. We can see that

more clearly if we define a variable

$$x = \frac{x_2 + x_1}{2}$$

and another

$$\bar{\phi} = \frac{\phi_2 + \phi_1}{2}$$

Then the sine part is just

$$\sin(kx - \omega t + \bar{\phi})$$

and we can see it is just our basic wave equation form. The rest of the equation must be the amplitude and we can clearly see that the amplitude depends on what we call the phase difference,

$$\Delta\phi = \frac{2\pi}{\lambda} (\Delta x) + \Delta\phi_o$$

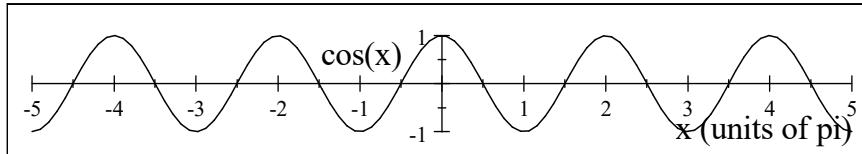
And from this we can see that for a single frequency situation there can be two sources two sources of phase difference. One can be from the two waves traveling different paths and then combining (Δx) and the other is from the two waves starting with a different phase to begin with, $\Delta\phi_o$. If the total phase difference between the two waves is a multiple of 2π , then the two waves will experience constructive interference

$$\Delta\phi = m2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots$$

Let's see that this works. Our amplitude is

$$A = 2E_{\max} \cos\left(\frac{1}{2}(\Delta\phi)\right)$$

and if we look at a cosine function we see that $\cos(\theta)$ is either 1 or -1 at $\theta = n\pi$.



So if $\Delta\phi = m2\pi$ then the amplitude is

$$\begin{aligned} A &= 2E_{\max} \left(\frac{1}{2} (m2\pi) \right) \\ &= 2E_{\max} \cos(m\pi) \\ &= \pm 2E_{\max} \end{aligned}$$

We don't really care if the amplitude function is big positively or negatively. So we get constructive interference for $\cos(m\pi)$ being either 1 or -1 . Then, this is our case for constructive interference.

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta x + \Delta\phi_o \right) = m2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots$$

How about for destructive interference? We start again with our amplitude function

$$A = 2E_{\max} \cos\left(\frac{1}{2}(\Delta\phi)\right)$$

but now we want when the cosine part needs to be zero.

$$\cos(\theta) = 0$$

Looking at our cosine graph again, that happens for cosine when $\theta = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots$. We could write this as $\theta = (m + \frac{1}{2})\pi$ for $m = 0, \pm 1, \pm 2, \pm 3, \dots$. But remember that in our amplitude function, we already have the $1/2$ in the function, so we want $\Delta\phi$ to have just the odd integer multiple of π . We could write this as

$$\Delta\phi = (2m + 1)\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots$$

So our condition for destructive interference is

$$\Delta\phi = \left(\frac{2\pi}{\lambda}\Delta x + \Delta\phi_o\right) = (2m + 1)\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots$$

We have developed a useful matched set of equations that will tell us if we mix two waves when we will have constructive and destructive interference:

$$\begin{aligned} \Delta\phi &= \left(\frac{2\pi}{\lambda}\Delta r + \Delta\phi_o\right) = m2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots && \text{Constructive} \\ \Delta\phi &= \left(\frac{2\pi}{\lambda}\Delta r + \Delta\phi_o\right) = (2m + 1)\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots && \text{Destructive} \end{aligned}$$

Let's take an example to see how this can be used.

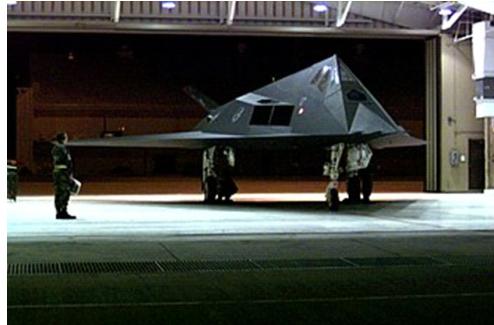
21.1.1 Example of two wave interference: Stealth Fighter

The U.S. Air Force F-117 Nighthawk Stealth Fighter is an aircraft that is designed to be mostly invisible to radar. How could this work?

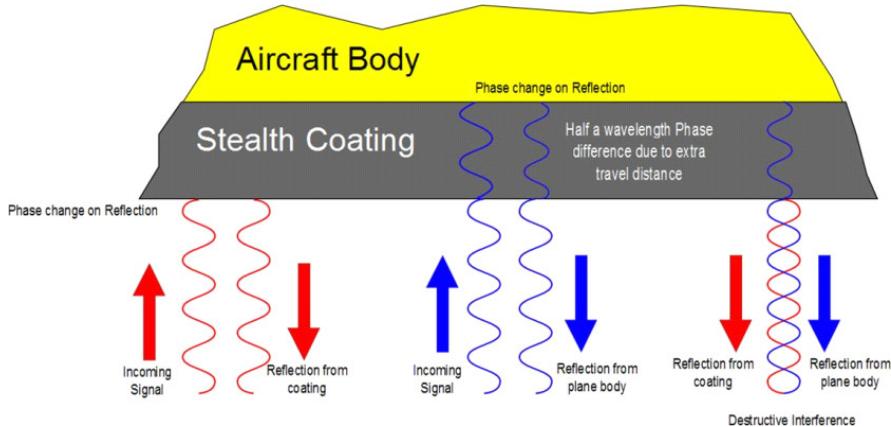


21.1. MATHEMATICAL TREATMENT OF SINGLE FREQUENCY INTERFERENCE 301

The answer is that I don't know¹. But we can make some guesses. The Stealth Fighter is kept in temperature and humidity controlled hangers.



So maybe there is a coating on the aircraft that needs protecting. Let's see if we could make an airplane "stealthy" by applying a coating.



So suppose the stealth fighter is coated with an anti-reflective polymer that is part of its mechanism for making the plane invisible to radar. Suppose we have a radar system with a wavelength of 3.00 cm. Further suppose that the index of refraction of the anti-reflective polymer is $n = 1.50$, and that the aircraft index of refraction is very large, how thick would you make the coating?

We want destructive interference, so let's start with our destructive interference condition

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta r + \Delta\phi_o \right) = (2m + 1)\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \quad \text{Destructive}$$

The radar waves all hit the plane in phase. From the figure, we see that the radar wave will reflect off of the coating. Because the index of refraction of the coating is large, this is like a fixed end of a rope. There will be an inversion.

¹I don't really know that this is true. And if I did, I wouldn't be able to use this as an example! So we are reverse engineering the stealth fighter technology. But I think it is a good guess.

But some of the wave will penetrate the polymer. This will reflect off of the plane body. The plane body has a very large index of refraction, so once again the wave will experience an inversion. The outgoing waves would then both be in phase as they leave and create constructive interference (if there were not a path difference) because

$$\Delta\phi_o = \pi - \pi = 0$$

at this point. Thus

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta r \right) = (2m + 1)\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \quad \text{Destructive}$$

But $\Delta\phi_o$ is not the only term in our equation, we also have to remember the path difference, Δr . The part of the wave that entered the polymer travels farther. If that path difference, Δr , is just right we can get destructive interference. For the $m = 0$ case

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta r \right) = (2(0) + 1)\pi = \pi$$

then the amplitude function would be

$$\begin{aligned} A &= 2E_{\max} \cos \left(\frac{2\pi}{\lambda} \Delta r \right) \\ &= 2E_{\max} \cos \left(\frac{1}{2}(\pi) \right) \\ &= 0 \end{aligned}$$

and we have destructive interference. Note that these are electromagnetic waves, so instead of y_{\max} we have used E_{\max} as the individual wave amplitude. But the important thing is that the plane cannot be seen by the radar! Of course, this works for $m = 1$ and $m = 2$, etc. as well. Any odd multiple of π will work.

$$\frac{2\pi}{\lambda} \Delta r = (2m + 1)\pi \quad m = 0, 1, 2, \dots$$

so that we are guaranteed an odd multiple of π . This is our condition for destructive interference.

But we are interested in the thickness. We realize that Δr is about twice the thickness, since the wave travels through the coating and back out. So let's let $\Delta r \approx 2t$

$$\begin{aligned} \frac{2\pi}{\lambda} 2t &\approx (2m + 1)\pi \\ 2t &\approx (2m + 1) \frac{\lambda}{2} \\ 2t &\approx \left(m + \frac{1}{2} \right) \lambda \\ t &\approx \left(m + \frac{1}{2} \right) \frac{\lambda}{2} \end{aligned}$$

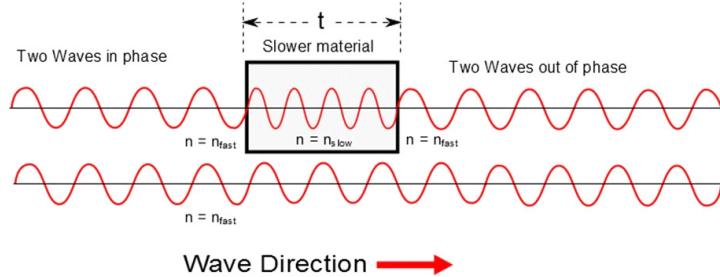
21.1. MATHEMATICAL TREATMENT OF SINGLE FREQUENCY INTERFERENCE 303

But there is a further complication. We should write our thickness equation as

$$t \approx \left(m + \frac{1}{2}\right) \frac{\lambda_{in}}{2}$$

because Δr has to provide an odd integer times the wavelength *inside the coating* for the phase to be right. After all, the wave is traveling inside the coating. We know that the wavelength will change as we enter the slower material.

To see this, consider two waves traveling to the right in the figure below. One passes through a slower medium. We expect the wavelength to shorten. We can see that, depending on the thickness t , the wave may be in phase or out of phase as it leaves. In the figure, the thickness is just right so that we have destructive interference.



We have such a wavelength shift in the coating. But we don't know the wavelength inside the coating. All we know is the radar wavelength, λ_{out} . We can fix it by writing the wavelength inside in terms of the wavelength outside. Earlier in our studies we found that the new wavelength will be given by equation (??)

$$\lambda_f = \frac{v_f}{v_i} \lambda_i$$

Let's rewrite this for our case

$$\lambda_{in} = \frac{v_{in}}{v_{out}} \lambda_{out}$$

We can express this in terms of the index of refraction

$$n = \frac{c}{v}$$

by multiplying the left hand side by c/c then

$$\lambda_{in} = \frac{cv_{in}}{cv_{out}} \lambda_{out}$$

or

$$\begin{aligned} \lambda_{in} &= \frac{\frac{c}{v_{out}}}{\frac{c}{v_{in}}} \lambda_{out} \\ &= \frac{n_{out}}{n_{in}} \lambda_{out} \end{aligned}$$

in the case of our aircraft coating the outside medium is air so $n_{out} \approx 1$

$$\lambda_{in} = \frac{1}{n_{in}} \lambda_{out}$$

This is the wavelength we need to match as the radar signal enters the medium.

Using this expression for λ_{in} in

$$t \approx \left(m + \frac{1}{2} \right) \frac{\lambda_{in}}{2} \quad m = 0, 1, 2, \dots$$

will give us the condition for destructive interference. Let's rewrite our λ_{in} equation for our case of a coating and air

$$\lambda_{in} = \lambda_{coating} = \frac{1}{n_{in}} \lambda_{out} = \frac{1}{n_{coating}} \lambda_{air}$$

thus

$$t \approx \left(m + \frac{1}{2} \right) \frac{1}{2} \left(\frac{\lambda_{air}}{n_{coating}} \right) \quad m = 0, 1, 2, \dots$$

is our condition for being stealthy.

Let's assume we want the thinnest coating possible, so we set $m = 0$. Then

$$t \approx \left(\frac{1}{4} \right) \left(\frac{\lambda_{air}}{n_{coating}} \right)$$

and our thickness would be

$$t \approx \left(\frac{1}{4} \right) \left(\frac{3.00 \text{ cm}}{1.50} \right) = 0.5 \text{ cm}$$

This seems doable for an aircraft coating!

Of course we could also make a plane that would be more visible to radar by choosing the constructive interference case. Suppose we are building a search and rescue plane. We want to enhance its ability to be seen by radar in fog. We start with the condition for constructive interference

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta r + \Delta\phi_o \right) = m2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \quad \text{Constructive}$$

It will still be true that $\Delta\phi_o = 0$.

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta r \right) = m2\pi$$

and it is still true that $\Delta r \approx 2t$.

$$\left(\frac{2\pi}{\lambda} 2t \right) \approx m2\pi$$

then

$$\left(\frac{1}{\lambda} 2t \right) \approx m$$

$$t \approx \frac{1}{2} m \lambda$$

and we still have to adjust for the coating index of refraction

$$t \approx \frac{m}{2} \left(\frac{\lambda_{air}}{n_{coating}} \right)$$

And once again we have several choices for m

$$t \approx \frac{m}{2} \left(\frac{\lambda_{air}}{n_{coating}} \right) \quad m = 0, 1, 2, \dots$$

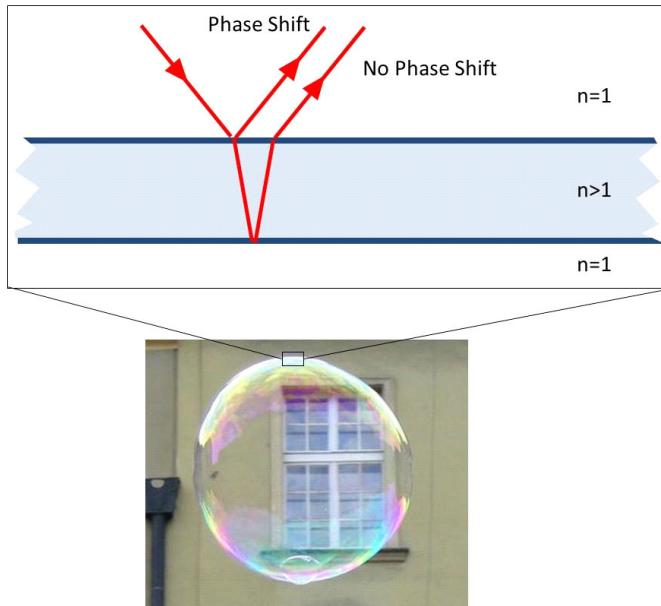
But now the coating will provide constructive interference, making it easier to track on radar from the command center. For the thinnest possibility, set $m = 1$ because the $m = 0$ case doesn't give us any thickness.

$$\begin{aligned} t &\approx m \frac{1}{2} \left(\frac{\lambda_{air}}{n_{coating}} \right) \\ &= \frac{1}{2} \left(\frac{3.00 \text{ cm}}{1.50} \right) \\ &= 1 \text{ cm} \end{aligned}$$

Note that we reasoned out these equations for the boundary conditions that we have in our problem (inversion on reflection from both the coating and the plane body). If the boundary conditions change, so do the equations.

21.1.2 Example of two wave interference: Soap Bubble

Take a soap bubble for example.



Interference from a soap bubble. (Bubble image in the Public Domain, courtesy Marcin Deregowski)

Now we have a phase shift on the first reflection, but not one on the reflection from the inside surface of the bubble because the bubble is full of air. The index of refraction of air is less than that for the bubble material. So as we leave the bubble material it is more like having a free end of a rope. As the waves leave the surface, they are half a wavelength out of phase due to $\Delta\phi_o$ because of the single inversion from the bubble outer surface. We would have destructive interference due to just this, but we also have to account for the bubble thickness. If this thickness is a multiple of a wavelength, then we are still have half a wavelength out of phase and we have destructive interference.

Here are our basic equations

$$\begin{aligned}\Delta\phi &= \left(\frac{2\pi}{\lambda} \Delta r + \Delta\phi_o \right) = m2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \text{ Constructive} \\ \Delta\phi &= \left(\frac{2\pi}{\lambda} \Delta r + \Delta\phi_o \right) = (2m + 1)\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \text{ Destructive}\end{aligned}$$

Suppose we want want constructive interference to get our colors, so we take the first

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta r + \Delta\phi_o \right) = m2\pi$$

and this time we have

$$\begin{aligned}\Delta\phi_o &= \phi_{transmitted} - \phi_{reflected} \\ &= 0 - \pi \\ &= -\pi\end{aligned}$$

It is still true that $\Delta r \approx 2t$ so from our constructive interference equation

$$\frac{2\pi}{\lambda}2t - \pi = m2\pi$$

$$\frac{2}{\lambda}t - \frac{1}{2} = m$$

$$\frac{2}{\lambda}t = m + \frac{1}{2}$$

$$t = \frac{\lambda}{2} \left(m + \frac{1}{2} \right)$$

We again have the problem that this wavelength must be the wavelength inside the bubble material $\lambda = \lambda_{in}$. But we see the outside wavelength λ_{out} . We can reuse our conversion from outside to inside wavelength from our last problem because we are once again in air and $n_{air} \approx 1$.

$$\lambda_{in} = \frac{1}{n_{in}}\lambda_{out}$$

then

$$t = \frac{\lambda_{out}}{2n_{in}} \left(m + \frac{1}{2} \right) \quad m = 0, 1, 2, \dots$$

Or writing this with $n_{in} = n_{bubble}$ to make it clear that the inside material is the bubble solution,

$$t = \left(m + \frac{1}{2} \right) \frac{1}{2} \left(\frac{\lambda_{air}}{n_{bubble}} \right) \quad m = 0, 1, 2, \dots$$

but this was the equation for destructive interference for the plane! We can see that memorizing the thickness equations won't work. We need to start with our conditions on $\Delta\phi$ for constructive and destructive interference to be safe!

How about the dark parts of the bubble with no color (the parts we can see through). These would be destructive interference

$$\Delta\phi = \left(\frac{2\pi}{\lambda} \Delta x + \Delta\phi_o \right) = (2m + 1)\pi$$

We can fill in the pieces to obtain

$$\begin{aligned}
 \left(\frac{2\pi}{\lambda} 2t - \pi \right) &= (2m+1)\pi \\
 \frac{2}{\lambda} 2t - 1 &= (2m+1) \\
 \frac{2}{\lambda} 2t &= (2m+1) + 1 \\
 \frac{4}{\lambda} t &= (2m+1) + 1 \\
 t &= \frac{\lambda}{4} (2m+2) \\
 t &= \frac{\lambda}{2} (m+1) \\
 t &= \frac{m+1}{2} \left(\frac{\lambda_{out}}{n_{bubble}} \right)
 \end{aligned}$$

This is our condition for destructive interference for the bubble. We don't have to, but we could write $m+1 = p$ where p is an integer that starts at 1 instead of zero.

$$t = \frac{p}{2} \left(\frac{\lambda_{out}}{n_{bubble}} \right) \quad p = 1, 2, \dots$$

But this is very like the condition for constructive interference for the plane.

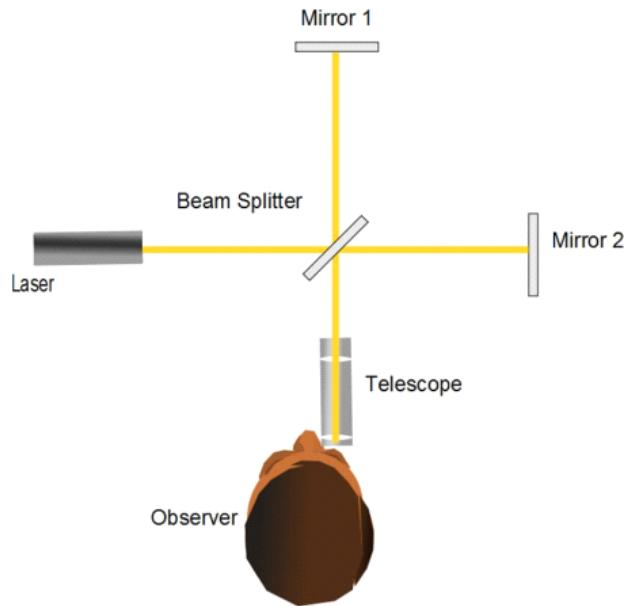
Hopefully, it is apparent that we have to start with our basic equations

$$\begin{aligned}
 \Delta\phi &= \left(\frac{2\pi}{\lambda} \Delta x + \Delta\phi_o \right) = m2\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \text{ Constructive} \\
 \Delta\phi &= \left(\frac{2\pi}{\lambda} \Delta x + \Delta\phi_o \right) = (2m+1)\pi \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \text{ Destructive}
 \end{aligned}$$

each time we attempt an interference problem because the outcome depends on both Δx and $\Delta\phi_o$. We have to construct the equation each time for the interference condition we want (constructive or destructive) finding Δx and $\Delta\phi_o$ for the boundary conditions we have.

21.2 The Michelson Interferometer

The Michelson interferometer is another device that uses path differences to create interference fringes.



The device is shown in the figure. A coherent light source is used. The light beam is split into two beams that are usually at 90° apart. The beams are reflected off of two mirrors back along the same path and are mixed at the telescope. The result (with perfect alignment) is a target fringe pattern like the first two shown below.



If the alignment is off, you get smaller fringes, but the system can still work. This is shown in the last image in the previous figure.

In the figure, we have constructive interference in the center, but if we move one of the mirrors half a wavelength, we would have destructive interference and would see a dark spot in the center. This device gives us the ability to measure distances on the order of the wavelength of the light. When the distance is continuously changed, the pattern seems to grow from the center (or collapse into the center).

Notice that if the mirror is moved $\frac{\lambda}{2}$, the path distance changes by λ because the light travels the distance to the mirror and then back from the mirror (it travels the path twice!).

Chapter 22

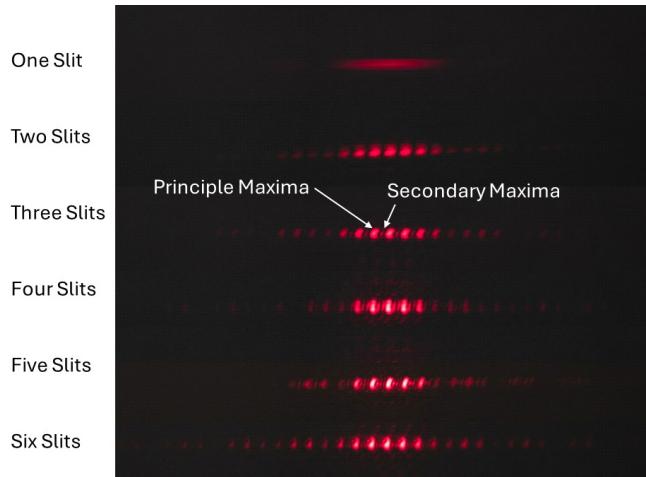
22 Slit Intensity 3.4.1, 3.4.2

Fundamental Concepts

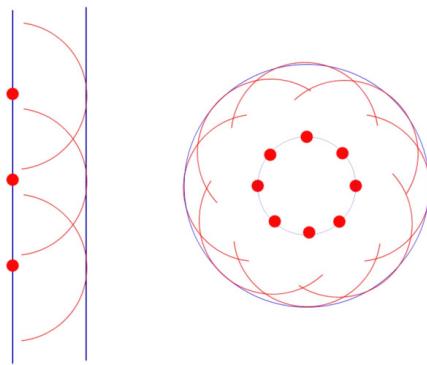
- Single slit minima
- Phasors
- Single slit intensity pattern
- The sinc function

22.1 Single Slits

We have looked at interference from two slits, and for many slits. Here is the figure again that shows how the constructive and destructive interference pattern changes as we change the number of slits or openings. Earlier we gave a name for these patterns of constructive and destructive interference. We called them diffraction patterns.



The two slits acted like two sources. We might expect that a single slit will give only a single bright spot. But if you look carefully at the last figure, you can see that there is something odd about the one slit pattern. The spot of light is stretched out. And if you look very carefully it almost looks like there are faint splotches of red light to the sides of the brighter fringe of light. let's consider a single slit very closely. To do this, let's return to the work of Huygens.¹ His idea for the nature of light was simple. He suggested that every point on the wave front of a light wave was the source (the disturbance) for a new set of small spherical waves. In optics, the crests of the waves are often called wavefronts. The next wavefront would be formed by the superposition of the little "wavelets." Here is an example for a plane wave and a spherical wave.



¹Huygens method is technically not a correct representation of what happens. The actual wave leaving the single opening is a superposition of the original wave, and the wave scattered from the sides of the opening. You can see this scattering by tearing a small hole in a piece of paper and looking through the hole at a light source. You will see the bright ring around the hole where the edges of the paper are scattering the light. But the mathematical result we will get using Huygens method gives a mathematically identical result for the resulting wave leaving the slit with much less high power math. So we will stick with Huygens in this class.

In each case we have drawn spots on the wave front and drawn spherical waves around those spots. where the wavefronts of the little wavelets combine, we have new wave front of our wave. This is sort of what happens in bulk matter. Remember that light is absorbed and re-emitted by the atoms of the material. This is why light slows down in a material. Because of the time it is absorbed, it effectively goes slower. But the light is not necessarily re-emitted in the same direction. Sometimes it is, but sometimes it is not. This creates a small, spherical wave (called a wavelet) that is emitted by that atom. So Huygens idea is not too bad.

We can use this idea for a single slit and look at what happens as the light goes through. Here is such a slit.

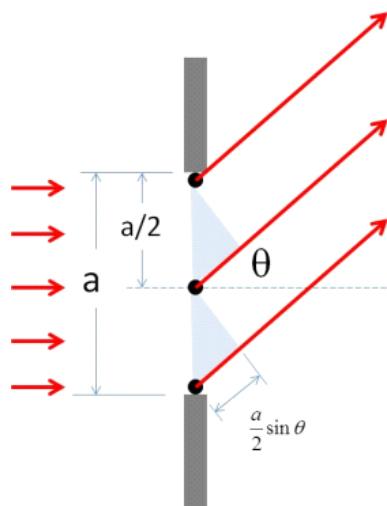
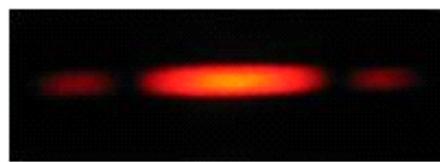


Figure 22.1:

In the figure above, we have divided a single slit of width a into two parts, each of size $a/2$. According to Huygens' principle, each position of the slit acts as a source of light rays. So we can treat half a slit as two coherent sources. These two sources should interfere. So what do we see when we perform such an experiment? Here is a brighter version of our light pattern from a single slit.



The figure shows a diffraction pattern for a thin slit. There are several terms that are in common use to describe the pattern

1. Central Maximum: The broad intense central band.
2. Secondary Maxima: The fainter bright bands to both sides of the central maxima
3. Minima: The dark bands between the maxima

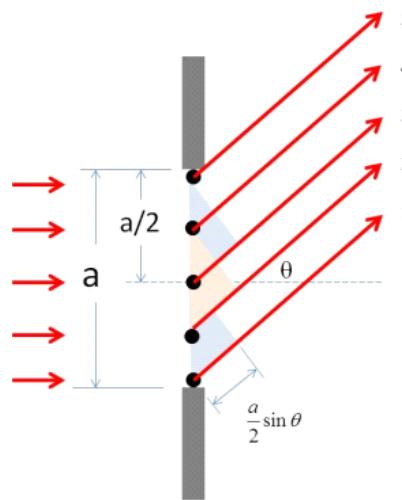
Notice, if we take two of the rays that hit one of our constructive interference spots, we get a path difference Δr that is

$$\Delta r = \frac{a}{2} \sin \theta$$

This would give a bright spot.

22.2 Narrow Slit Minima

Let's use figure 22.1 to find the *dark* minima of the single slit pattern. First we should notice that figure 22.1 could have another set of rays that contribute to the *bright* spot because they will also have a path difference of $(a/2) \sin \theta$. Let's fill these in. They are rays 2 and 4 of the next figure.



Before we started with what we are now calling rays 1 and 3. Ray 1 travels a distance

$$\Delta r = \frac{a}{2} \sin (\theta) \quad (22.1)$$

farther than ray 3. As we just argued, rays 2 and 4 also have the same path difference, and so do rays 3 and 5. If this path difference is a multiple of λ we

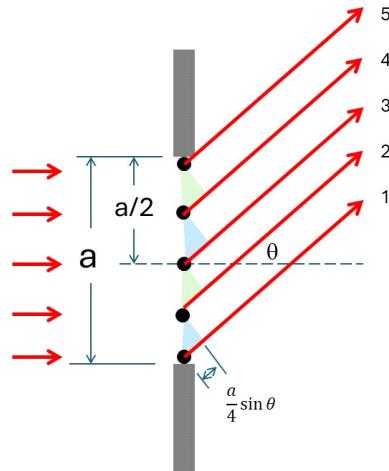
get constructive interference, but if this path difference is $\lambda/2$ then we will have destructive interference. The condition for a minima is then

$$\frac{a}{2} \sin(\theta) = \pm \frac{\lambda}{2} \quad (22.2)$$

or

$$\sin(\theta) = \pm \frac{\lambda}{a} \quad (22.3)$$

But look at our figure. We could just as well divide the slit into four equal parts



and imagine light traveling from each of those parts to our viewing screen. Then we have a path difference of

$$\Delta r = \frac{a}{4} \sin(\theta) \quad (22.4)$$

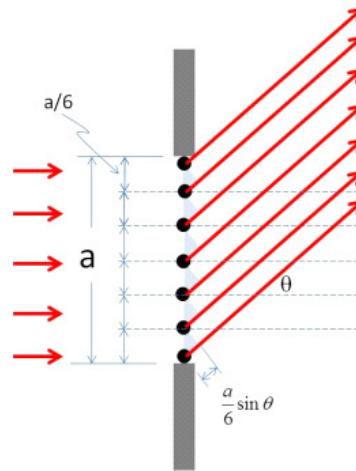
And to have destructive interference we need this path difference to be $\lambda/2$

$$\frac{a}{4} \sin(\theta) = \pm \frac{\lambda}{2} \quad (22.5)$$

or

$$\sin(\theta) = \pm \frac{2\lambda}{a} \quad (22.6)$$

But there is no reason to stop at four equal parts. Suppose we divide our opening into six equal parts.



Then we would find

$$\Delta r = \frac{a}{6} \sin(\theta) \quad (22.7)$$

and we still want a dark spot so

$$\frac{a}{6} \sin(\theta) = \pm \frac{\lambda}{2} \quad (22.8)$$

or

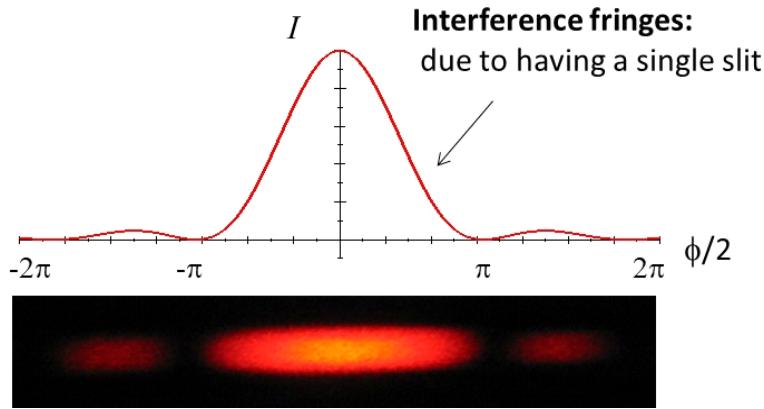
$$\sin(\theta) = \pm \frac{3\lambda}{a} \quad (22.9)$$

and in general at the angle θ that makes the sine of that angle equal to $m\lambda/a$ we get a dark spot.

$$\sin(\theta) = m \frac{\lambda}{a} \quad m = \pm 1, \pm 2, \pm 3 \dots \quad (22.10)$$

You might object that we did not find the bright spots, only the dark spots. That's fine because the bright spots have to be in between the dark spots. But we can do better. Let's look at the full intensity pattern for a single slit.

22.3 Intensity of the single-slit pattern



We have found the dark minima, now we would like to find the shape of the bright bands. We will use phasors to do this. If you have already taken Principles of Physics III (PH220) you know what a phasor is, but for the rest of us we need a quick introduction.

22.3.1 Phasors

Let's take two waves

$$E_1 = E_{o1} \sin(\omega t + \alpha_1) \quad (22.11)$$

and

$$E_2 = E_{o2} \sin(\omega t + \alpha_2) \quad (22.12)$$

where I have hidden the position dependence in the α 's

$$\begin{aligned} \alpha_1 &= kx_1 + \phi_1 \\ \alpha_2 &= kx_2 + \phi_2 \end{aligned}$$

We wish to find the resultant wave

$$E = E_{o1} + E_{o2} \quad (22.13)$$

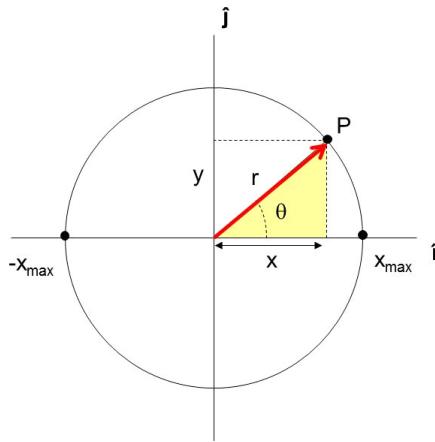
We actually know how to do the algebra to combine these waves (we have done it enough times!), but let's find another way to do the math that can extend to many many waves being added up.

Think of our equation for a wave.

$$E_1 = E_{o1} \sin(\omega t + \alpha_1) \quad (22.14)$$

We have made our wave equation look like the equation for a harmonic oscillator. And so long as we don't change x_1 or x_2 this is fine. The electric field at our position where we are mixing the waves will oscillate like a harmonic oscillator.

Way back when we studied oscillation we said the equation for a harmonic oscillator position was like the projection of circular motion onto an axis.



That gave us

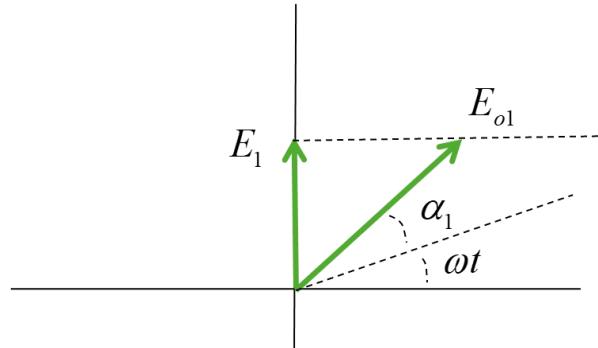
$$x(t) = x_{\max} \cos(\omega t + \phi_o)$$

or

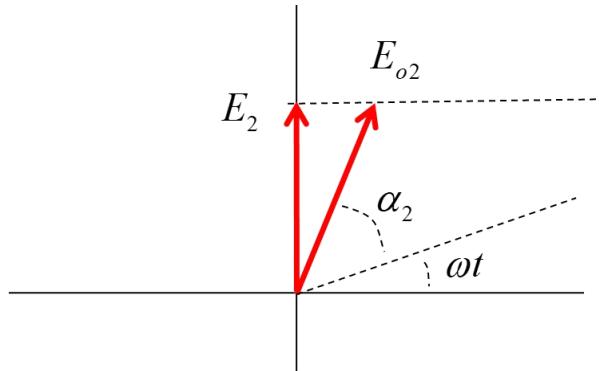
$$y(t) = y_{\max} \sin(\omega t + \phi_o)$$

as our oscillator equations. And see how much our modified wave equation looks like the second one of these oscillator equations! This gives us an idea of how to use vector addition to add up waves!

Let's represent E_{o1} as a rotating vector with angular speed ω . We get projections of E_{o1} on the x and y axis. In this discussion, we will consider the vertical (y) axis but we could choose either axis (but if we use the x -axis we would then have cosine waves).



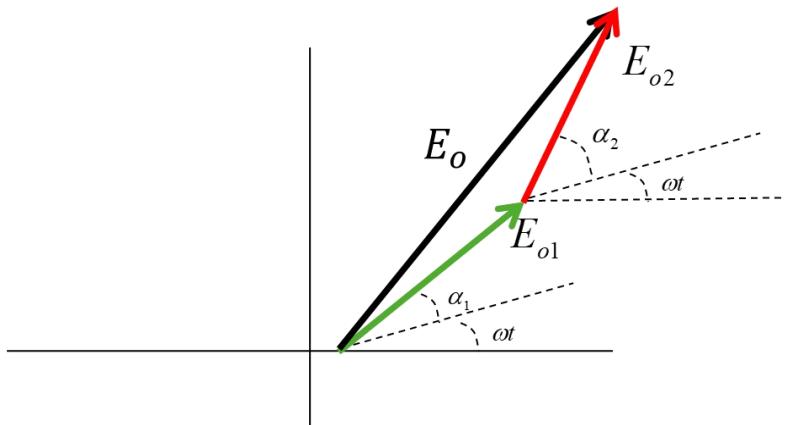
At any given time, E_1 will be the projection $E_{o1} \sin(\omega t + \alpha_1)$. We can also take a second rotating vector for E_2



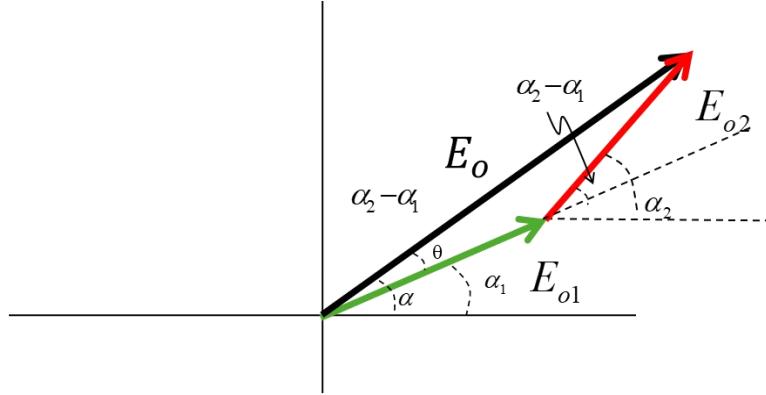
Note that E_2 is projected onto the vertical axis through an angle $\omega t + \alpha_2$. So the projection E_2 is $E_{o2} \sin(\omega t + \alpha_2)$.

These rotating vectors have a name. We call them *phasors*.

To add the phasors, we draw phasor two on the end of phasor one. It is just like adding vectors back in Principles of Physics I (PH121)!



The sum is the resultant phasor like in vector addition. Only, remember now that the phasors are rotating. But because they have the same angular frequency, ω , phasor 1 and phasor 2 rotate together as a unit. The projection of this rotating group of phasors gives the sum we are looking for. But notice the sum's value would change in time as the phasors rotate. We can choose any time to find the resulting vector, so we might as well take $t = 0$, then $\omega t = 0$ and the vectors will be rotated as shown.



Using the redrawn phasors in the diagram directly above, we can see from the law of cosines

$$E_0^2 = E_{o1}^2 + E_{o2}^2 - 2E_{o1}E_{o2} \cos(\eta) \quad (22.15)$$

but from the figure we can see that $\eta = \pi - (\alpha_2 - \alpha_1)$. That gives us

$$E_0^2 = E_{o1}^2 + E_{o2}^2 - 2E_{o1}E_{o2} \cos(\pi - (\alpha_2 - \alpha_1)) \quad (22.16)$$

or

$$E_0^2 = E_{o1}^2 + E_{o2}^2 + 2E_{o1}E_{o2} \cos(\alpha_2 - \alpha_1) \quad (22.17)$$

using the trig identity that $\cos(\pi - \gamma) = -\cos\gamma$. This is the magnitude of our resulting vector. For the case of a double slit we know that $E_{o1} = E_{o2}$ so our resulting vector would be

$$E_0^2 = E_{o1}^2 + E_{o2}^2 + 2E_{o1}E_{o2} \cos(\alpha_2 - \alpha_1) \quad (22.18)$$

$$= 2E_{o1}^2 + 2E_{o1}^2 \cos(\alpha_2 - \alpha_1) \quad (22.19)$$

We can see that when $\cos(\alpha_2 - \alpha_1) = -1$ we get zero. This is destructive interference!

We should go on to find the angle. In the figure it is called just α and it is α_1 plus a little bit more. It's not thrilling, but since we have the sides of the triangle we could just use the law of cosines again to find the little bit more angle

$$E_{o1}^2 = E_0^2 + E_{o2}^2 + 2E_0E_{o2} \cos(\theta)$$

so

$$\theta = \cos^{-1} \frac{E_{o2}^2 - E_0^2 + E_{o1}^2}{2E_0E_{o2}}$$

or when $E_{o1} = E_{o2}$

$$\theta = \cos^{-1} \frac{2E_{o1}^2 - E_0^2}{2E_0E_{o1}}$$

and

$$\alpha = \alpha_1 + \theta$$

Now we ask what the question, “what is the projection of this resultant phasor on the vertical axis?” Since we chose $t = 0$ for our picture, it is easy to pick off the fact that the resultant vector has a phase angle of α . thus

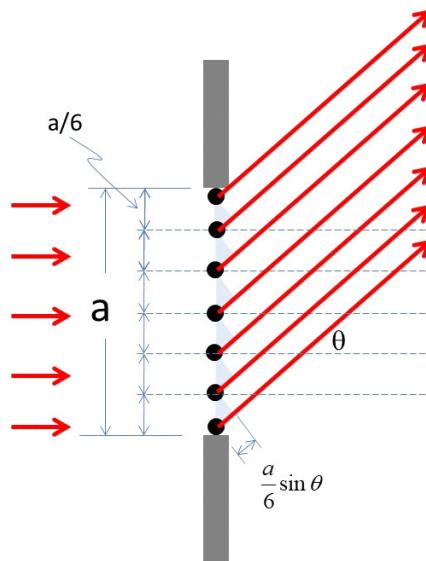
$$E_R = E_0 \sin(\omega t + \alpha) \quad (22.20)$$

This seems like a lot of work for adding just two waves together, but we will use it to add many many waves together.

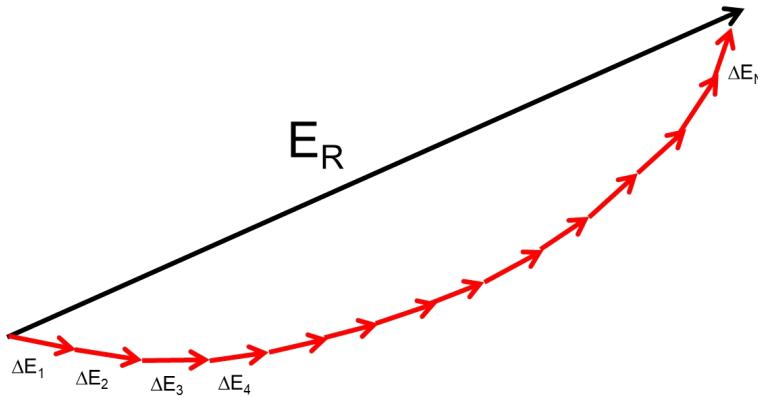
22.3.2 Using phasors to find the intensity pattern

Now that we know what a phasor is, let's try to use them to find the single slit intensity pattern.

We can consider dividing the slit into small regions Δy in length.



Each piece of the slit contributes ΔE to the electric field. We can use superposition to find the resultant wave for a given point P . The phase diagram would look something like this.



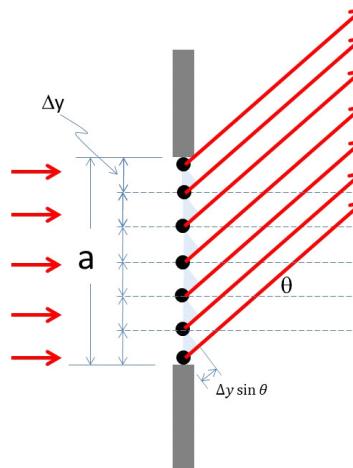
We add up all the little ΔE parts, but each has a slightly different distance, so it has a slightly different phase because kr is part of the phase. We can put the phase difference due to the different paths into the phase constant.

Our set of little waves might be like

$$\begin{aligned}\Delta E_1 &= E_o \sin(\omega t) \\ \Delta E_2 &= E_o \sin(\omega t + \phi_2) \\ \Delta E_3 &= E_o \sin(\omega t + \phi_3) \\ &\vdots\end{aligned}\tag{22.21}$$

Note that these are still waves. We have just put the kr_i part into the phase ϕ like we did in the α 's in the last section.

Note that each wave has an additional phase shift, and each has the amplitude E_o . But we need to find these phase shifts. We will call the phase shift $\Delta\phi$ to remind us that it is small. Look at the next figure.



By thinking about this figure we can see that

$$\Delta r = \Delta y \sin \theta \quad (22.22)$$

for each pair of adjacent rays. and that gives a phase difference of

$$\Delta\phi = k\Delta r = \frac{2\pi}{\lambda} \Delta y \sin \theta \quad (22.23)$$

Let's suppose that as we cross the slit we have N waves each of magnitude ΔE . Then the total phase difference between waves from the top to the bottom of the slit is

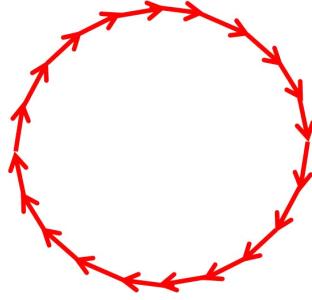
$$\phi = N\Delta\phi = N \frac{2\pi}{\lambda} \Delta y \sin \theta \quad (22.24)$$

$$= \frac{2\pi}{\lambda} a \sin \theta \quad (22.25)$$

where a is our slit width and

$$a = N\Delta y \quad (22.26)$$

We are finding dark spots in our diffraction pattern, and the first dark spot will be when the projection of our phasor sum, $E_R = 0$. And think about this, when you add vectors you put the vectors tip-to-tail, and when you are done you start at the tail of the first and draw a new vector to the tip of the last vector you added. This new vector is the sum. And to make that new vector have a length $E_R = 0$ all we have to do is to have the vectors we add make a circle. Then the distance from the tail of the first vector to the tip of the last vector is exactly zero! We are doing this with phasors, so we place our phasors tip-to-tail and find the first minimum or first dark band when $\phi = 2\pi$ or when the phasor diagram circles back on itself.



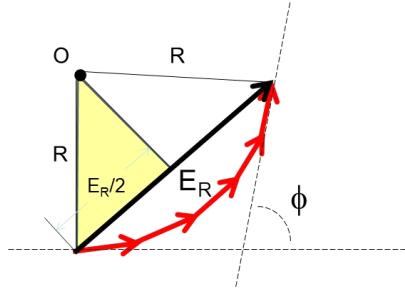
$$2\pi = \frac{2\pi}{\lambda} a \sin \theta \quad (22.27)$$

We can cancel and rearrange this to give

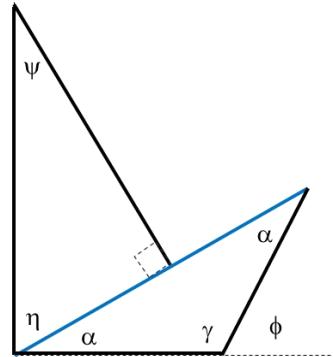
$$\frac{\lambda}{a} = \sin \theta \quad (22.28)$$

which is what we expected for the first minimum! This seems to be working. Let's now find the intensity pattern.

To find the intensity pattern on any point of the screen (not just the minima), we let Δy get very small. We add up $N \rightarrow \infty$ terms. The curve becomes smooth (not shown as smooth in the next figure, though).



The quantity E_R is the resultant from adding together all of our phasors. We can see the phasor addition in the figure. The angle at the top of the yellow triangle is $\phi/2$. One way to see this is to note that the lower triangle has angles α , γ , and α



We see that

$$\gamma = 180 - \phi \quad (22.29)$$

and

$$\phi = 2\alpha \quad (22.30)$$

Note also that

$$\eta = 90 - \alpha \quad (22.31)$$

For the upper triangle,

$$180 = \psi + \eta + 90 \quad (22.32)$$

$$90 = \psi + \eta \quad (22.33)$$

we know η , so

$$90 = \psi + 90 - \alpha \quad (22.34)$$

or

$$\psi = \alpha \quad (22.35)$$

but

$$\alpha = \frac{\phi}{2} \quad (22.36)$$

so

$$\psi = \frac{\phi}{2} \quad (22.37)$$

Then

$$\sin\left(\frac{\phi}{2}\right) = \frac{\frac{E_R}{2}}{R} = \frac{E_R}{2R} \quad (22.38)$$

where R is the radius of curvature, and from the archlength formula

$$s = R\phi \quad (22.39)$$

and since we are finding the arc length, we know it will be nearly the sum of each little phasor vector length $s = NE_o$ so

$$s = R\phi = NE_o$$

and

$$R = \frac{NE_o}{\phi}$$

then our resultant wave phasor has length

$$E_R = 2R \sin\left(\frac{\phi}{2}\right) \quad (22.40)$$

$$= 2 \frac{NE}{\phi} \sin\left(\frac{\phi}{2}\right) \quad (22.41)$$

$$= NE_o \frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \quad (22.42)$$

It is silly, but we usually define $\beta = \phi/2$ so

$$E_R = NE_o \frac{\sin(\beta)}{\beta}$$

Earlier we learned that the intensity will go as the field squared

$$I = I_{\max} \left(\frac{\sin(\beta)}{\beta} \right)^2 \quad (22.43)$$

And here is a tricky part. We get the angle ϕ from adding up the angles from each of our little ΔE pieces. Each of them has an angle

$$\Delta\phi = k\Delta r = \frac{2\pi}{\lambda} \Delta y \sin \theta \quad (22.44)$$

and N of these will be

$$\phi = N \frac{2\pi}{\lambda} \Delta y \sin \theta$$

So we can see that the angle ϕ is the same as our total phase. If we have an aperture width of a then $a = N\Delta y$ and

$$\phi = N \frac{2\pi}{\lambda} \Delta y \sin \theta = \frac{2\pi}{\lambda} a \sin \theta$$

and that makes

$$\beta = \frac{\frac{2\pi}{\lambda} a \sin \theta}{2}$$

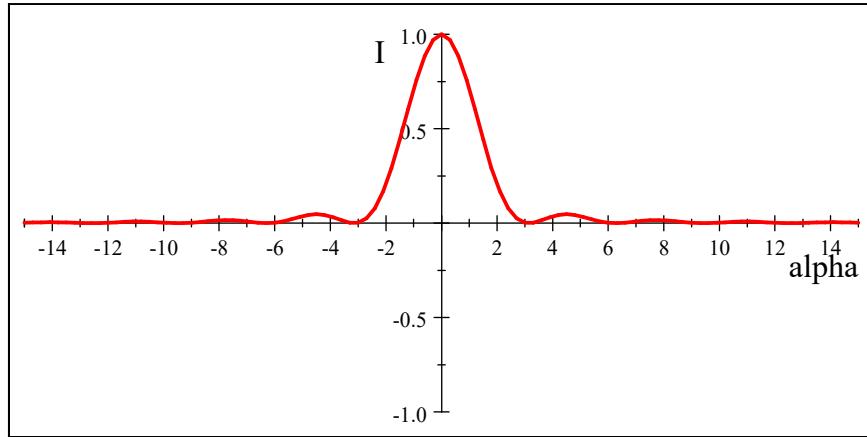
We can write our irradiance function as

$$I = I_{\max} \left(\frac{\sin \left(\frac{\frac{2\pi}{\lambda} a \sin \theta}{2} \right)}{\frac{\frac{2\pi}{\lambda} a \sin \theta}{2}} \right)^2 \quad (22.45)$$

or more simply

$$I = I_{\max} \left(\frac{\sin \left(\frac{\pi}{\lambda} a \sin \theta \right)}{\frac{\pi}{\lambda} a \sin \theta} \right)^2 \quad (22.46)$$

The function is plotted in the next figure

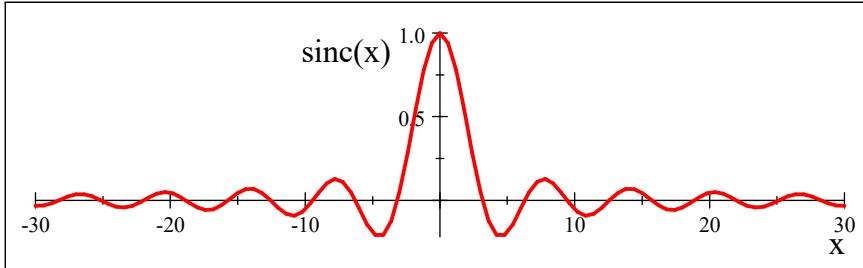


From this we can find our peaks where the pattern is bright.

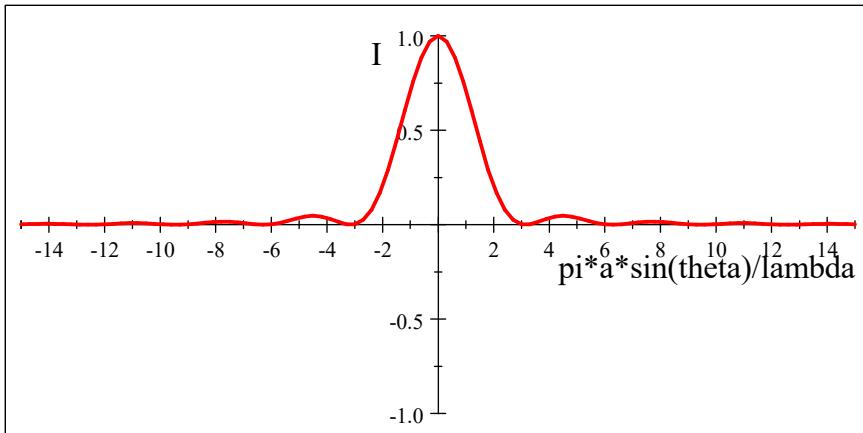
Notice that our field equation this has the form

$$\frac{\sin x}{x}$$

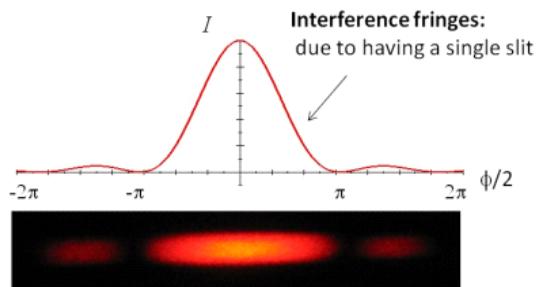
which has a distinctive shape.



this is known as a sinc function (pronounced like “sink”). It has a central maximum as we would expect. Of course our pattern has a sinc squared



You can see the central maximum and the much weaker minima produced by this function. Indeed, it seems to match what we saw very well. Putting it all together, our pattern looks like this.



You might ask, can we use our phasor approach to find the intensity pattern for double slit diffraction? Of course, the answer is yes, but I have had enough phasor fun for this class. You will see phasors again in Principles of Physics III.

Chapter 23

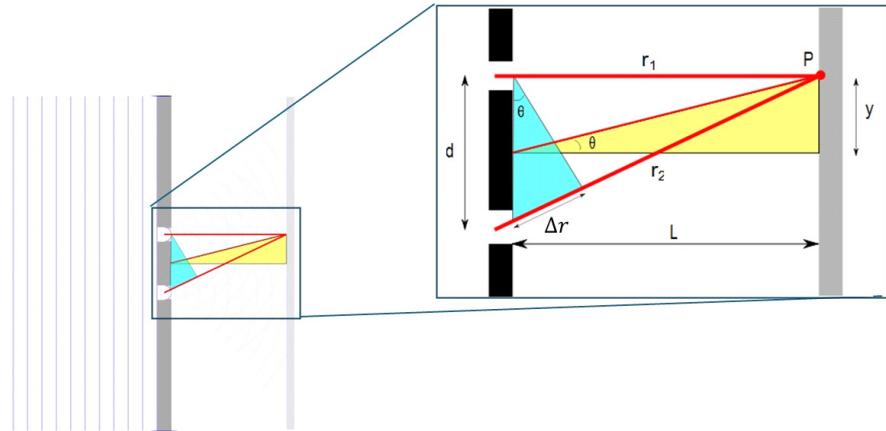
23 Double Slit and Grating intensity patterns 3.4.3, 3.4.4

In the last lecture, we found the diffraction pattern for a single slit opening. We used phasors to do this. In this lecture we want to do two or more slits. But let's return to algebraic methods.

Fundamental Concepts

- Double Slit Intensity Pattern
- Grating Intensity Pattern
- Integration over many periods to get what detectors (like eyes) see.

23.1 Double Slit Intensity Pattern



Back in Chapter (20) we found the bright and dark fringe locations for a double slit diffraction pattern. But the fringes we have seen are not just points, but are patterns that fade from a maximum intensity. In the last lecture we calculated such an intensity pattern for the light going through a single slit. We used phasors to do so. But for double slits we already have an algebraic method for finding the superposition of two waves. Let's use that method to find our intensity pattern.

Let's let a laser beam hit two parallel slits, just like we did before. The light coming through the two slits will be like two new sources of light but with the same frequency and wavelength. We can represent an electromagnetic wave using just the electric field (the magnetic field pattern is very similar and can be derived from the electric field pattern)

$$\begin{aligned} E_2 &= E_{\max} \sin(k_2 r_2 - \omega_2 t_2 + \phi_2) \\ E_1 &= E_{\max} \sin(k_1 r_1 - \omega_1 t_1 + \phi_1) \end{aligned}$$

and the resulting wave will be

$$\begin{aligned} E_r &= E_{\max} \sin(k_2 r_2 - \omega_2 t_2 + \phi_2) + E_{\max} \sin(k_1 r_1 - \omega_1 t_1 + \phi_1) \\ &= 2E_{\max} \cos\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) - (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \\ &\quad \times \sin\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) + (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \\ &= 2E_{\max} \cos\left(\frac{1}{2}[(k_2 r_2 - \omega_2 t_2 + \phi_2) - (k_1 r_1 - \omega_1 t_1 + \phi_1)]\right) \\ &\quad \times \sin\left(\frac{(k_2 r_2 - \omega_2 t_2 + \phi_2) + (k_1 r_1 - \omega_1 t_1 + \phi_1)}{2}\right) \end{aligned}$$

but now we know that we can simplify this because $\omega_2 = \omega_1 = \omega$, $k_2 = k_1 = k$, $t_2 = t_1 = t$, and so long as the light hits the slit screen perpendicular to the

screen, $\phi_2 = \phi_1 = \phi_o$. Then

$$E_r = 2E_{\max} \cos\left(\frac{1}{2}k(r_2 - r_1)\right) \sin\left(\frac{kr_2 + kr_1}{2} - \omega t + \phi_o\right)$$

We have a combined wave at point P that is a traveling wave $(\sin(k\frac{(r_2+r_1)}{2} - \omega t + \phi_o))$ but with amplitude $(2E_o \cos(\frac{1}{2}k(r_2 - r_1)))$. And notice that $r_2 - r_1 = d \sin \theta$ and $k = 2\pi/\lambda$ so we can write E_r as

$$E_r = 2E_{\max} \cos\left(\frac{1}{2}\frac{2\pi}{\lambda}d \sin \theta\right) \sin\left(k\frac{r_2 + r_1}{2} - \omega t + \phi_o\right)$$

and we have a wave with an amplitude 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, photographic film, and most detectors measure the intensity of the light. We know that

$$I = \frac{\mathcal{P}}{A}$$

Earlier we learned that the power is proportional to the square of the electric field wave amplitude.

$$\mathcal{P} \propto E^2 \quad (23.1)$$

Then the intensity must also be proportional to the amplitude of the electric field squared.

$$\begin{aligned} I &= \frac{P}{A} \propto E^2 \\ &\propto 4E_o^2 \cos^2\left(\frac{1}{2}\left(\frac{2\pi}{\lambda}d \sin \theta\right)\right) \sin^2\left(\frac{k(r_2 + r_1)}{2} - \omega t + \phi_o\right) \end{aligned}$$

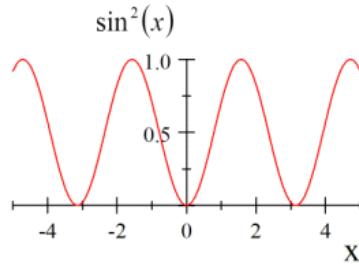
Light detectors collect power for a set amount of time. So most light detection will be a value averaged over a set *integration time*. We saw this before when we studied cameras. This means that the detector sums up (or integrates) the amount of power received over the detector time. Usually the integration time is much longer than a wave period, so we need to time-average our intensity.

$$\begin{aligned} \int_{\text{many periods}} Idt &\propto \int_{\text{many periods}} 4E_o^2 \cos^2\left(\frac{1}{2}\left(\frac{2\pi}{\lambda}d \sin \theta\right)\right) \sin^2\left(\frac{k(r_2 + r_1)}{2} - \omega t + \phi_o\right) dt \\ &= 4E_o^2 \cos^2\left(\frac{1}{2}\left(\frac{2\pi}{\lambda}d \sin \theta\right)\right) \int_{\text{many periods}} \sin^2\left(\frac{k(rx_2 + r_1)}{2} - \omega t + \phi_o\right) dt \end{aligned}$$

but the term

$$\int_{\text{many periods}} \sin^2\left(\frac{k(r_2 + r_1)}{2} - \omega t + \phi_o\right) dt = \frac{1}{2} \quad (23.2)$$

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



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

$$\bar{I} = \int_{\text{many periods}} Idt \propto 2E_o^2 \cos^2 \left(\frac{1}{2} \left(\frac{2\pi}{\lambda} d \sin \theta \right) \right) \quad (23.3)$$

where \bar{I} is the time average intensity. The important part 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. This averaging over many periods is one of the reasons why we could mostly ignore the wave nature of light when we studied mirrors and lenses!

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 before. Let's check that it works. We used the small angle approximation before. It is still valid, so let's use it again now. For small angles

$$I = I_{\max} \cos^2 \left(\frac{\pi d}{\lambda} \theta \right)$$

and for small angles

$$\tan \theta = \frac{y}{L} \approx \theta$$

so we can write I as

$$I = I_{\max} \cos^2 \left(\frac{\pi d}{\lambda} \frac{y}{L} \right)$$

Then we have constructive interference when $\cos^2 \left(\frac{\pi d}{\lambda} \frac{y}{L} \right) = 1$ or when

$$\frac{\pi d}{\lambda} \frac{y}{L} = m\pi$$

or

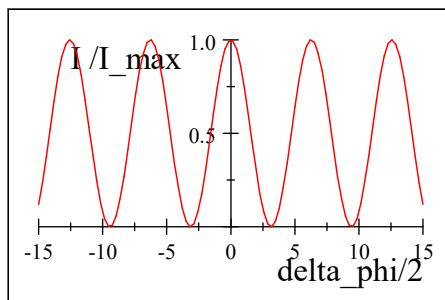
$$y = m \frac{L\lambda}{d}$$

which is what we found before!

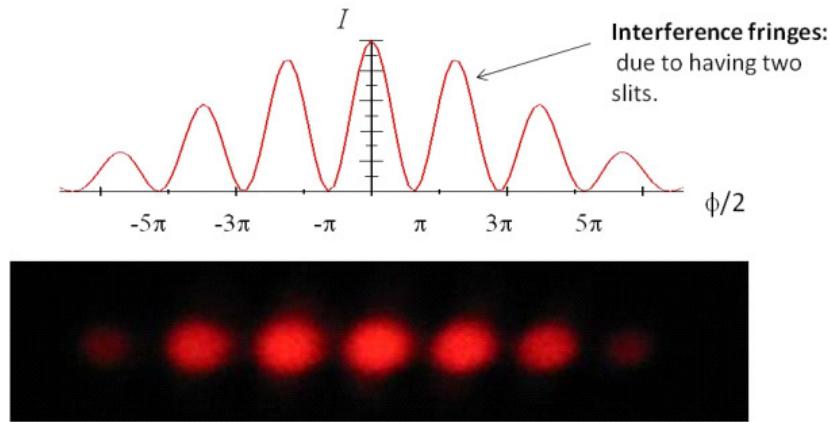
The plot of normalized intensity

$$\frac{I}{I_{\max}} = \cos^2 \left(\frac{\Delta\phi}{2} \right)$$

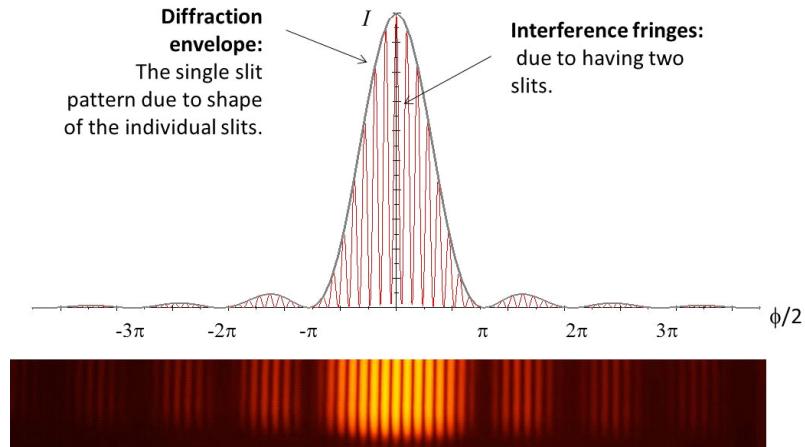
verses $\Delta\phi/2$ is given next,



This is really an interesting result. You might wonder why, when we found the two slit interference pattern, there was no evidence of the single slit fringing that we discovered in this chapter. After all, a double slit system is made from single slits. Shouldn't there be some effect due to the fact that the slits are individually single slits? The answer is that we did see some hint of the single slit pattern. Remember the figure below.



The intensity of the peaks seems to fall off with distance from the center. We dealt with only the center-most part of the pattern. If we draw the pattern for larger angles, we see the following.



It takes a bright laser or dark room to see the secondary groups of fringes easily, but we can do it. We can also graph the intensity pattern. It is the combination of the two slit and single slit pattern with the single slit pattern acting as an envelope.

$$I = I_{\max} \cos^2 \left(\frac{\pi d \sin(\theta)}{\lambda} \right) \left(\frac{\sin \left(\frac{\pi a \sin(\theta)}{\lambda} \right)}{\frac{\pi a \sin(\theta)}{\lambda}} \right)^2 \quad (23.4)$$

Note that one of the double slit maxima is clobbered by a minimum in the single slit pattern. We can find out the order of the missing maximum. Recall that

$$d \sin(\theta) = m\lambda$$

describes the maxima from the double slit. But for the single slit

$$a \sin(\theta) = \lambda$$

describes the first minimum. Dividing these yields

$$\begin{aligned} \frac{d \sin(\theta)}{a \sin(\theta)} &= \frac{m\lambda}{\lambda} \\ \frac{d}{a} &= m \end{aligned}$$

so the

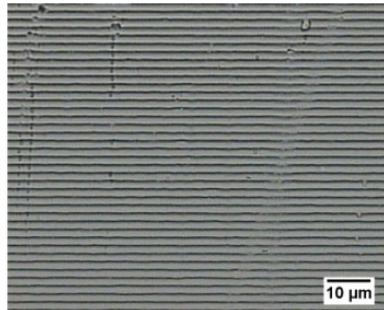
$$m = \frac{d}{a} \quad (23.5)$$

double slit maximum will be missing.

23.2 Diffraction Gratings

We now have found the single slit diffraction pattern equation and the double slit diffraction pattern. Let's look once again at many slits. If we have a whole lot of slits we give the collection of slits a new name. We call it a *diffraction grating*.

A diffraction grating is an optical element with many parallel slits spaced very close together. Here photo of part of a typical diffraction grating created by etching lines in a piece of glass. The etchings scatter the light, but the un-etched part allows the light to pass through. The un-etched parts are essentially a series of slits.



Surface of a diffraction grating (600 lines/mm). Image taken with optical transmission microscope. (Image in the public domain courtesy Scapha)

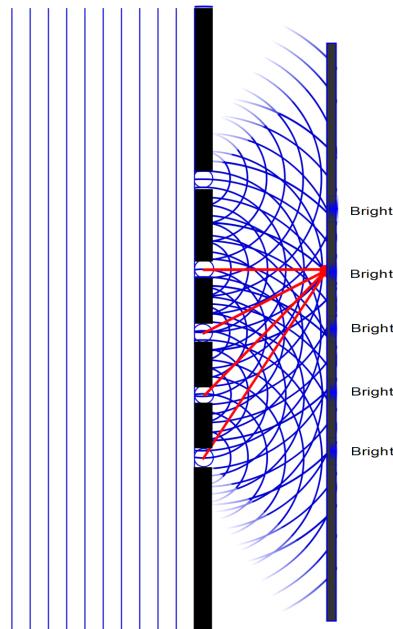
A typical grating might have 5000 slits per unit centimeter. You have probably used a diffraction grating to see rainbow colors in a beginning science class. Gratings are usually made by cutting parallel grooves in a flat surface.

If we use $5000 \frac{\text{grooves}}{\text{cm}}$ for an example, we see that the slit spacing is

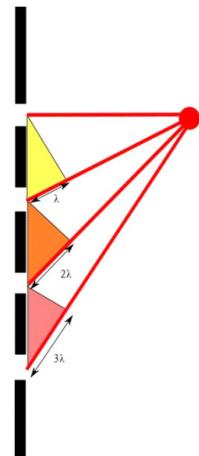
$$d = \frac{1}{5000} \text{ cm} \quad (23.6)$$

$$= 2.0 \times 10^{-6} \text{ m} \quad (23.7)$$

Take a section of diffraction grating as shown below



At some point, two of the slits will have a path difference that is a whole wavelength, and we would expect a bright spot. But what about the other slits? If we have a slit spacing such that each of the succeeding slits has a path difference that is just an additional wavelength, then each of the slits will contribute to the constructive interference at our point, and the point will become a bright spot.

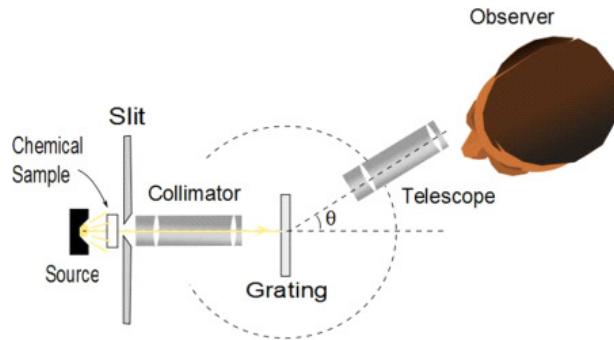


Let's look at just two slits. The light leaves each slit in phase with the light from the rest of the slits, but at some distance L away and at some angle θ we will have a path difference

$$\Delta r = d \sin(\theta_{bright}) = m\lambda \quad m = 0, \pm 1, \pm 2, \dots \quad (23.8)$$

because the path lengths are not all the same.

This equation tells us that each wavelength, λ , will experience constructive interference at a slightly different angle θ_{bright} . Different frequencies will create bright spots at different angles. We have found a way to create a spectrum with light waves. We often call a spectrum made with visible light frequencies a rainbow. Knowing d and θ allows an accurate calculation of λ . This may seem a silly thing to do, but suppose we add into our system a sample of a chemical to identify



We could then record the intensity of the transmitted light as a function of angle, which is equivalent to λ . We can again generate a spectrum. This is a traditional way to build a spectrometer and many such devices are available today.

23.2.1 Resolving power of diffraction gratings

We noticed that with two slits, we got a bright spot for a particular wavelength, but we didn't just get one bright spot. We got several. The same is true for diffraction gratings. So we expect to get a rainbow, but we really expect to get a series of rainbows. The integer m tells us which rainbow we have in the series. The integer m is called the order number.

For two nearly equal wavelengths λ_1 and λ_2 , we say that the diffraction grating can resolve the wavelengths if we can distinguish the two using the grating. The *resolving power* of the grating is defined as

$$R = \frac{(\lambda_1 + \lambda_2)}{2(\lambda_1 - \lambda_2)} = \frac{\bar{\lambda}}{\Delta\lambda} \quad (23.9)$$

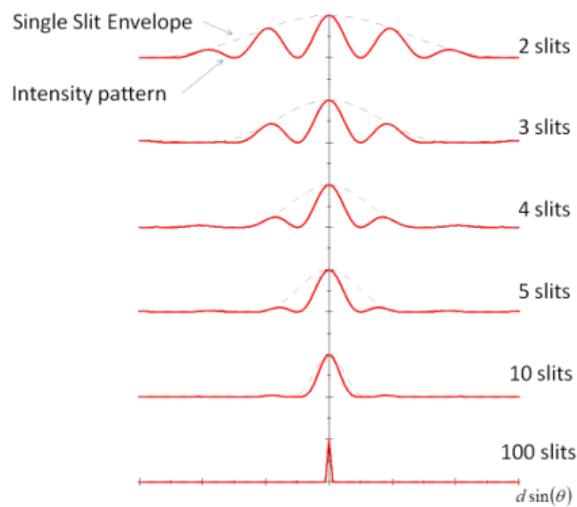
We can show that for the m -th order diffraction, the resolving power is

$$R = Nm \quad (23.10)$$

where N is the number of slits. So our ability to distinguish wavelengths increases with the number of slits and with the order (which is related to how far off-axis we look).

Note that for $m = 0$ we have no ability to resolve wavelengths. The central peak is a mix of all wavelengths and usually looks white for normal illumination.

That the resolution depends on the number of slits, N , means that we can improve our spectrometer by using more lines. Here is a representation of what happens as we increase N

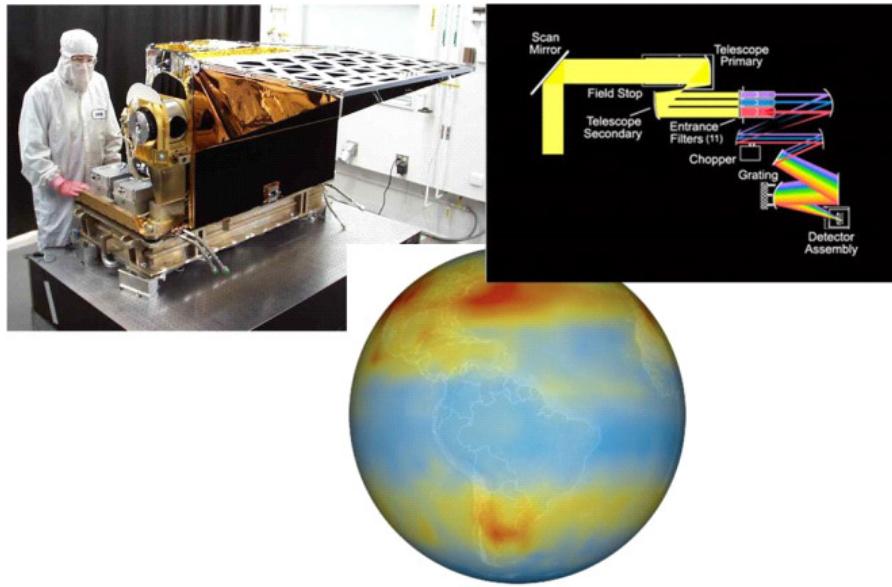


we can see that the peaks get narrower as N increases. These graphs are for a particular λ . If the peaks for a particular λ get narrower, then there will be less overlap with adjacent λ 's which means that each wavelength can more easily be resolved. We won't derive the equation for a diffraction grating intensity pattern in this class. If you are interested, you can take our junior level optics class. From that class the intensity pattern is

$$I(\theta) = I_o \left(\frac{\sin \beta}{\beta} \right)^2 \left(\frac{\sin N\alpha}{\sin \alpha} \right)^2$$

where $\alpha = \frac{dk}{2} \sin \theta$, $\beta = \frac{ka}{2} \sin \theta$ and where a is the slit width and d is the slit spacing and N is the number of slits.

Spectrometers are used in many places. One that has some public interest today is monitoring the atmosphere. Instruments like the one shown below detect the amount of special gasses in the atmosphere using IR spectrometers.



AIRS sensor, spectrometer design, and global CO₂ data. (Images in the Public Domain courtesy NASA)

The instrument shown is the AIRS spectrometer. You can see in the diagram that it uses a grating spectrometer. The picture of the Earth is a composite of AIRS data showing the northern and southern bands of CO₂.

Chapter 24

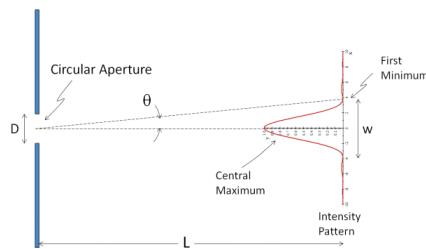
24 Resolution 3.4.5, 3.4.6, 3.4.7

Fundamental Concepts

- Circular Apertures
- Resolution is a name we give to the fundamental blurriness in geometrical optical systems due to the wave nature of light.
- Interference from circular apertures
- X-ray Diffraction
- Holography

24.1 Circular apertures

Our analysis of light going through holes has been somewhat limited by squarish holes or slits. But most optical systems, including our eyes don't have square holes. So what happens when the hole is round? The situation is as shown in the next figure.



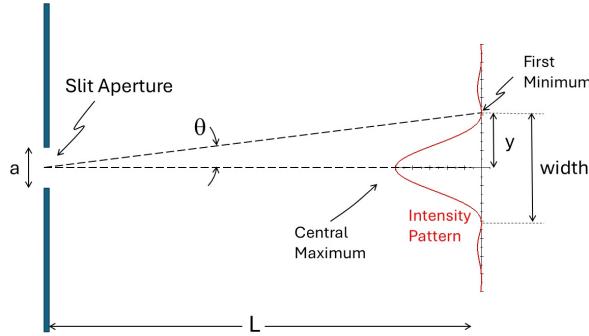
Before we discuss this situation, let's think about the width of a single slit pattern. We remember that

$$\sin(\theta) = (1) \frac{\lambda}{a}$$

for the first minima, or that

$$\theta \approx \frac{\lambda}{a}$$

for small angles. And from the figure



we can see that

$$\theta \approx \frac{y}{L}$$

so long as θ is small, then we find the position of the first minimum to be

$$y \approx \frac{\lambda}{a} L$$

This is the distance from the center bright spot peak intensity to the first dark spot. The width of the bright spot is twice this distance

$$w \approx 2 \frac{\lambda}{a} L$$

We expect something like this for our circular aperture. The derivation is not too hard, but it involves Bessel functions, which are beyond the math requirement for this course¹. So I will give you the answer

$$\theta \approx 1.22 \frac{\lambda}{D}$$

Where we have replaced a with D , the diameter of the aperture. But they are both the size of the hole.

Let's think about this last equation. Compare it to our equation for θ for a square opening. That's right, the circular aperture (hole) only adds a factor of 1.22. And as with the slit

$$\theta \approx \frac{y}{L}$$

¹We find the whole diffraction pattern for a circular aperture in PH375, *Optics*.

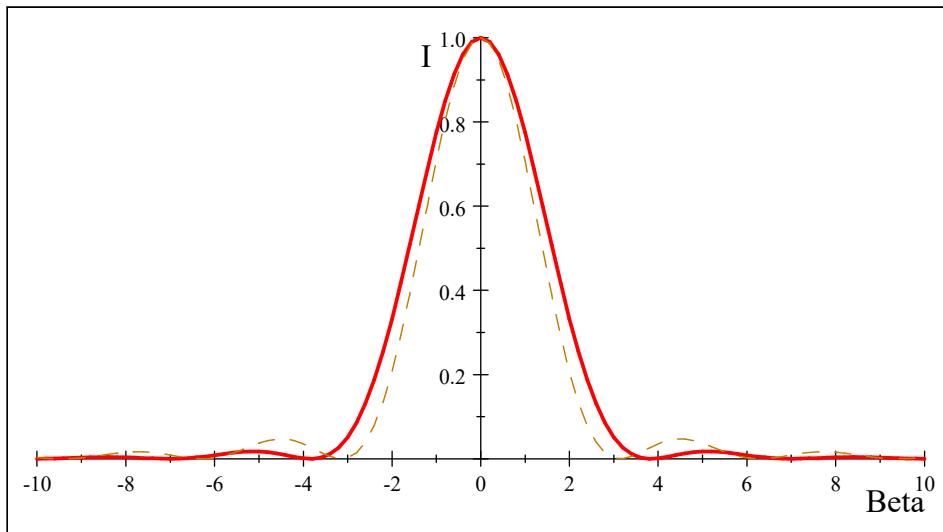
so

$$y \approx 1.22 \frac{\lambda}{D} L \quad (24.1)$$

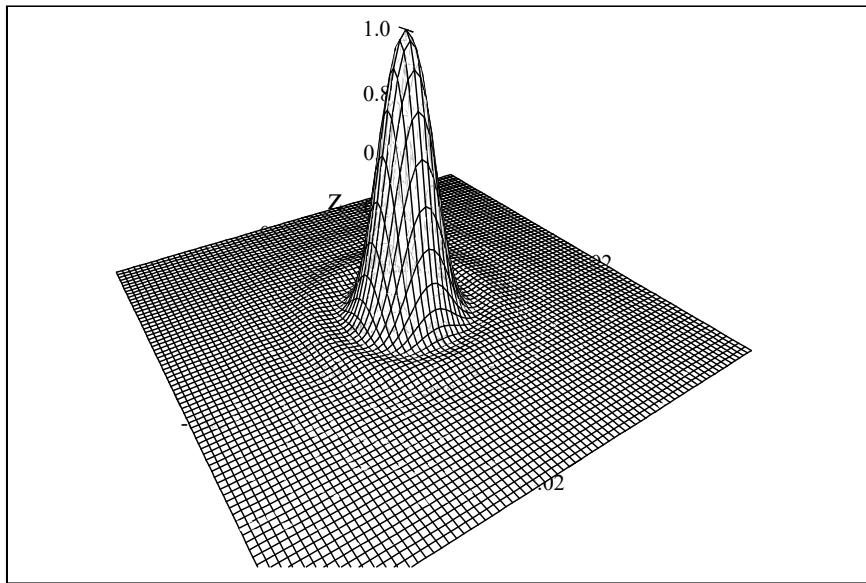
and

$$w \approx 2.44 \frac{\lambda}{D} L \quad (24.2)$$

The pattern looks a little like the slit pattern. But the secondary maxima are actually smaller for the circular aperture case. Here is a larger version in red (solid curve) with the single slit pattern in brown (dashed line).



A three dimensional version of the intensity pattern from the circular aperture.



Graph of the Airy disk for the case $a = 0.05 \text{ mm}$, $\lambda = 500 \text{ nm}$, and $R = 1 \text{ m}$.

With a bright enough laser, the pattern becomes visible.

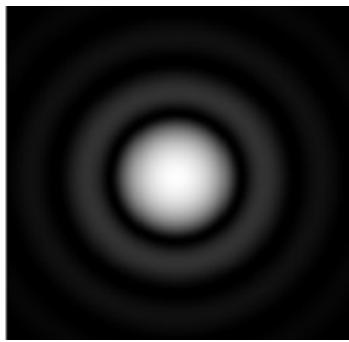


24.2 Resolution

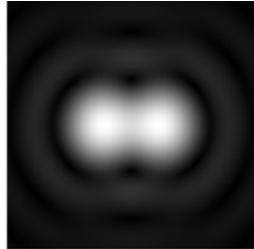
We have emphasized that an extended object can be viewed as a collection of point objects. Then the image is formed from the collection of images of those

point objects. We thought about the design of cameras and telescopes and other optical systems using ray optics. It would be great if optical systems could form images with infinite precision, but it turns out they can't. And it is the fact that light acts as a wave prevents this from being true! Our wave nature of light comes back to complicate our simple optical designs!

Because light is really a wave, the images of the point objects won't be points, themselves. And because our eyes and cameras have circular apertures, the image of the point objects will be little circular central maxima with dim concentric ring patterns.



If we have more than one point of light, we will get two such patterns.



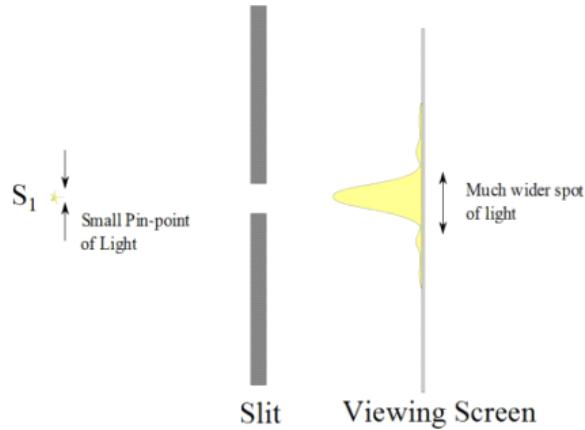
and our image of an extended object (like Aunt Sally) is made up of many, many points all reflecting light into our camera lens. We want each of those points of light to create corresponding points on our image. But instead they are creating small circles of light, and from the above figures we can see that those circles will overlap. All this makes our images fuzzy.

24.3 Resolution

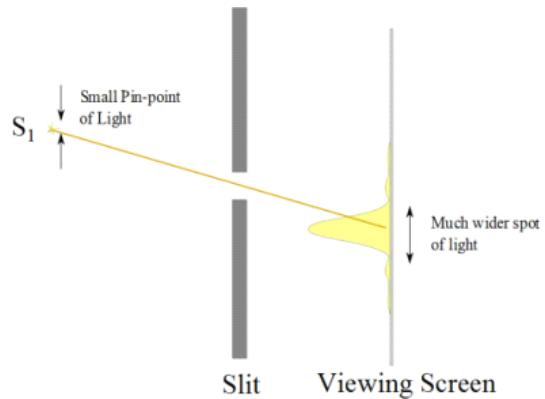
The quality of our image depends on how poorly a point object is imaged. If each point object makes a large circle of light on the screen or detector array, we get a very confusing image (it will look blurry to us). Let's review why this will happen so we can know how to minimize the effect.

Since we are using eyes or cameras, we expect circular apertures. Let's review what we know but include some ray diagrams.

We already know that if we take light and pass it through a single slit, we get an intensity pattern that has a central bright region.

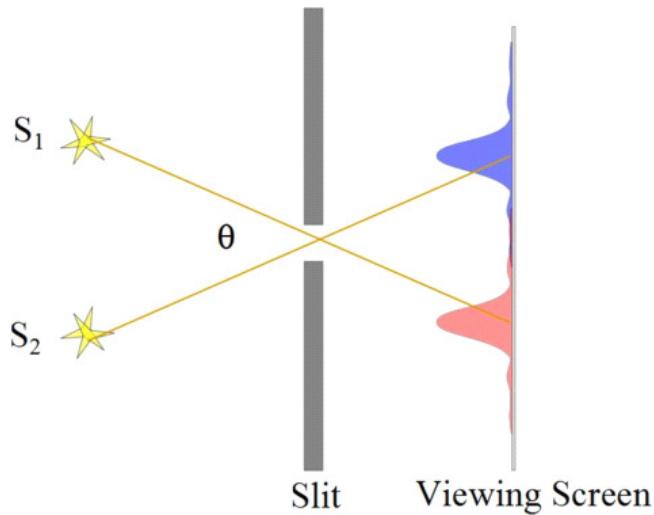


Remember that normal objects will be made up of many small points of light (either due to reflection or glowing) and each of these will form such an intensity pattern on a screen. Here is a bright point source that is not on the axis, and we see that it too makes a bright spot on the screen (and smaller bright spots or rings, depending on the shape of the aperture)

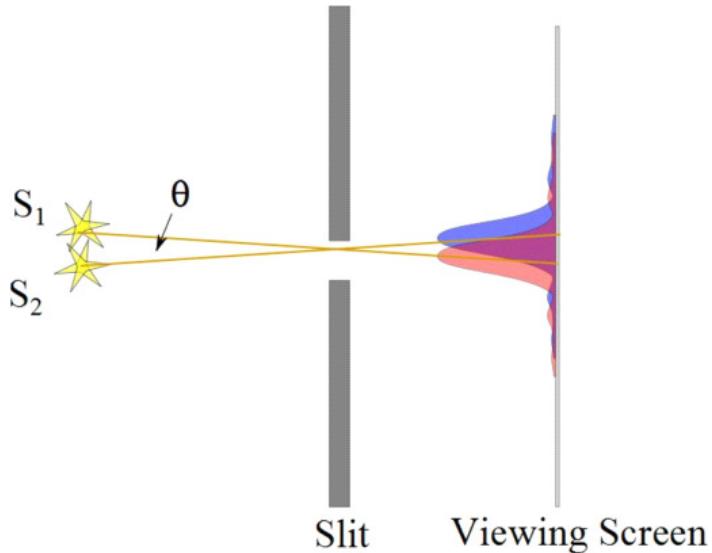


So our images will be made up of many central bright spots, each of which represents a point of light from the object. These central bright spots may overlap, (and their secondary maxima certainly will overlap).

Let's take a simple case of two points of light, \$S_1\$ and \$S_2\$. If we take a single slit and pass light from two distant point sources through the slit, we do not get two sharp images of the point sources. Instead, we get two diffraction patterns.



If these patterns are formed sufficiently far from each other, it is easy to tell they were formed from two distinct objects. Each point became a small blur, but that is really not so bad. We can still tell that the two blurs came from different sources. If our pixel size is about the same size of the blur, we may not even notice the blurriness in the digital imagery.



But if the patterns are formed close to each other, it gets hard to tell whether they were formed from two objects or one bright object. We now have a problem. Suppose you are trying to look at a star and see if it has a planet. But all you can see is a blur. You can't tell if there is one source of light or two.

Long ago an early researcher titled Lord Rayleigh developed a test to determine if you can distinguish between two diffraction patterns. When the central maximum of one point's image falls on the first minimum of another point's image, the images are said to be just resolved. This test is known as *Rayleigh's criterion*.

We can find the required separation for a slit. Remember that

$$\sin(\theta) = m \frac{\lambda}{a} \quad m = \pm 1, \pm 2, \pm 3, \dots \quad (24.3)$$

gives the minima (dark spots) for a single slit. We want the first minimum, so $m = 1$

$$\sin(\theta) = \frac{\lambda}{a} \quad (24.4)$$

Remember that

$$\tan \theta = \frac{y}{L}$$

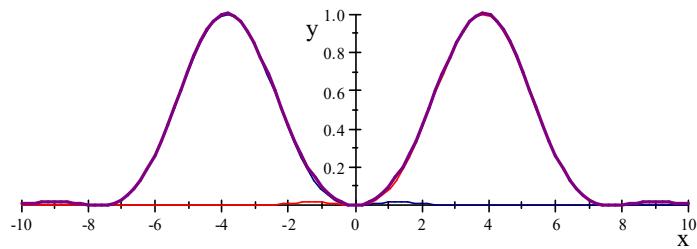
where L is the distance from the slit to the viewing screen, and y is the vertical location of the dark spot. If we place the second image maximum so it is just at this location, the two images will be just barely resolvable. In the small angle approximation, $\sin(\theta) \approx \theta$ so

$$\theta_{\min} = \frac{\lambda}{a} \quad (24.5)$$

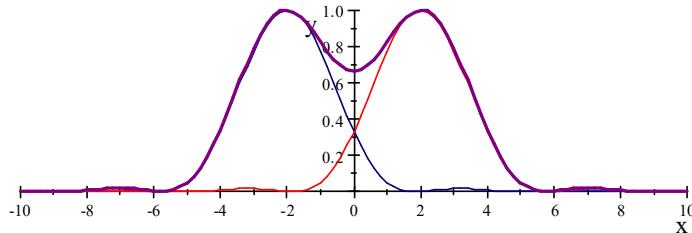
Now you may be saying to yourself that you don't often take pictures through single illuminated slits, so this is nice, but not really very interesting. But suppose, instead, that we image a circular aperture. We just pick up a factor of 1.22 and change a to D for diameter.

$$\theta_{\min} = 1.22 \frac{\lambda}{D} \quad (24.6)$$

The Rayleigh criteria tells you, based on your camera aperture size, how a point source will be imaged on the film or sensor array. If we consider extended sources (like your favorite car or Aunt Sally) to be collections of many point sources, then we have a way to tell what features will be clearly resolved on the image and what features will not (like you may not be able to see the lettering on the car to tell what model it is, or you may not be able to distinguish between the gem stones in Aunt Sally's necklace because the image is too blurry to see these features clearly).

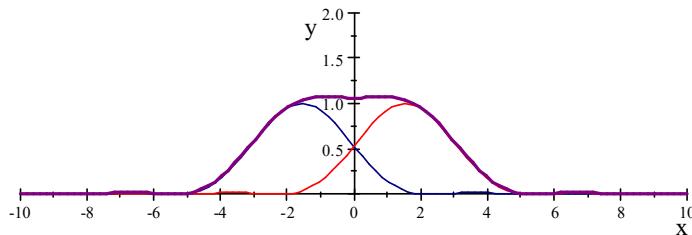


Pattern from two resolved circular slits.



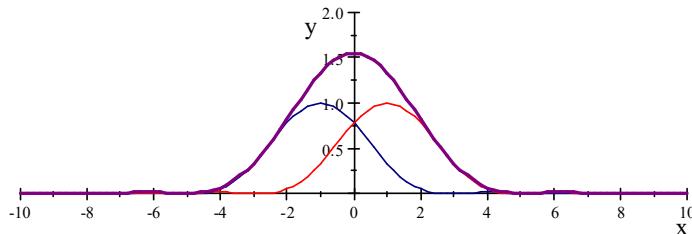
Rayleigh Criteria: Pattern from two circular sources where the sources are close enough that the maximum from one pattern is placed on the minimum of the other. Lord Rayleigh gave this as the criteria for just barely being resolved.

Astronomers sometimes use Sparrow's criteria for two sources being resolved. It is shown below.



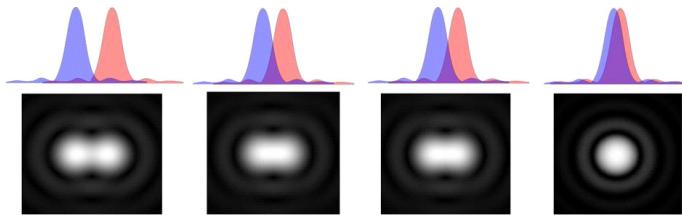
Sparrow Criteria: This is a less conservative resolution criteria than Rayleigh. When the intensity pattern is flat on the top, there must be two sources. This criterial is used in astronomy.

Since astronomers just want to know if there are two stars or one, it is enough to see that the intensity pattern went flat at the center. That must mean that there are two stars. But if the stars are any closer, the flat center becomes a peak and the stars are unresolved.



Two circular sources unresolved

Here is what the easily resolved, Rayleigh resolved, Sparrow resolved, and unresolved cases look like.

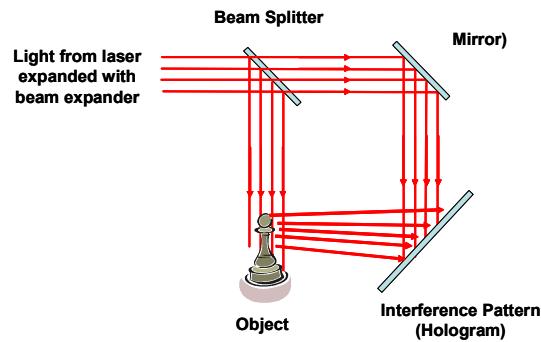


Of course, astronomers have it easy. They really image points of light. But for photos of more normal things we don't want to not be able to tell part of Aunt Sally's eye from part of her nose. Our aperture limits how good the image will look regardless of the lens that we use in the camera. Because this limit is due to diffraction patterns we call this resolution limit a *diffraction limit*. If you are designing a camera system you generally want your sensor pixel size to match the size of the central diffraction spot, because smaller pixels won't get you a less blurry image once you reach the size of the spot created by imaging a point source.

This really concludes our study of waves and optics! But let's look at two more applications of what we have learned.

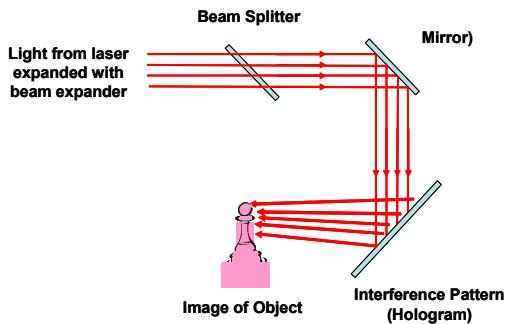
24.4 Holography

You may have seen holograms in the past. We have enough understanding of light to understand how they are generated now.



A device for generating a hologram is shown in the figure above. Light from a laser or other coherent source is expanded and split into two beams. One travels to a photographic plate, the other is directed to an object. At the object, light is scattered and the scattered light also reaches the photographic plate. The combination of the direct and scattered beams generates a complicated interference pattern.

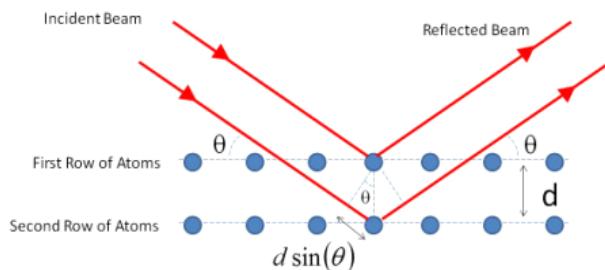
The pattern can be developed (like you develop photographic film). Once developed, it can be re-illuminated with a direct beam. The emulsion on the plate creates complicated patterns of light transmission, which combine to create interference. It is like a very complicated slit pattern or grating pattern. The result is a three-dimensional image generated by the interference. The interference pattern generates an image that looks like the original object.



24.5 Diffraction of X-rays by Crystals

If we make the wavelength of light very small, then we can deal with very small diffraction gratings. 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.



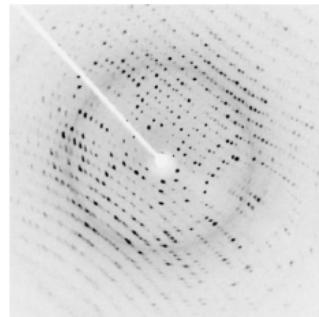
If we illuminate the crystal with x-rays, the x-rays can reflect off the top layer of atoms, or off the second layer of atoms (or off any other layer, but for now let's just consider two layers). If the spacing between the layers is d , then the path difference will be

$$\delta = 2(d \sin(\theta)) \quad (24.7)$$

then for constructive interference

$$2d \sin (\theta) = m\lambda \quad m = 1, 2, 3, \dots \quad (24.8)$$

This is known as *Bragg's law*. This relationship can be used to measure the distance between the crystal planes. As the crystal structure gets more complicated, the interference pattern gets more complicated as well. Here is an example



Diffraction image of protein crystal. Hen egg lysozyme, X-ray source Bruker I μ S,
 $\lambda = 0.154188$ nm, 45 kV, Exposure 10 s. (image in the public domain)

DNA makes an interesting diffraction pattern.



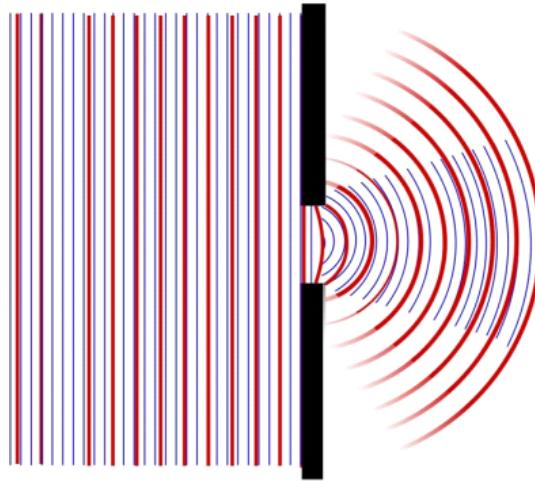
X-ray diffraction pattern of DNA (image courtesy of the National Institute of Health,
image in the public domain)

24.6 Ray Model vs. Wave Model

We have studied all the new information that we will get on waves and optics. But notice that we started our study of light using rays, but knowing that light is a wave. We did most of our work with mirrors and lenses without using wave properties of light. But then we looked at apertures and found the wave nature

of light was very important. When do we need to use the equations for light as a wave, and when can we get away with ignoring the wave nature of light?

In the next figure, two waves of different wavelengths go through a single opening. The wave representing the central maxima is shown in each case, but not the secondary maxima.



Notice that the smaller wavelength has a narrower central maxima as we would expect from

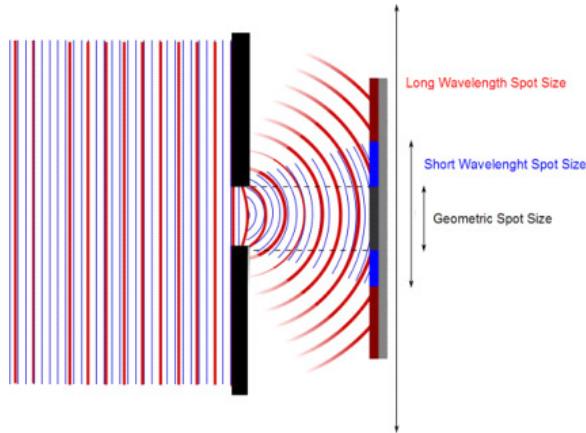
$$\sin(\theta_{dark}) = 1.22 \frac{\lambda}{D}$$

or

$$\theta_{dark} \approx 1.22 \frac{\lambda}{D}$$

we see that the ratio of the wavelength to the hole size determines the angular extent of the central maxima. The smaller the ratio, the smaller the central region. We can use this to explain why the wave nature of light was so hard to find.

The patch of light on a screen that is created by light passing through the aperture is created by the central maximum.



For the long wavelength (red) the central maximum is larger than the screen. The short wavelength spot will be wholly on the screen as shown. The geometric spot is what we would see if the light traveled straight through the opening. Notice that the short wavelength spot is closer to the size of the geometric spot. In the limit that

$$\lambda \ll a$$

or for circular openings

$$\lambda \ll D$$

then

$$\theta \approx \frac{\lambda}{a} \approx 0$$

or

$$\theta \approx \frac{\lambda}{D} \approx 0$$

and the spot size would be very nearly equal to the geometric spot size.

This is the limit we will call the *ray approximation*.

For most of mankind's time on the Earth, it was very hard to build holes that were comparable to the size of a wavelength of visible light (around 500 nm). So it is no wonder that the waviness of light was missed for so many years.

But this ray limit is very useful for apertures the size of camera lenses. So it was found for lenses and mirrors to use this small λ , large aperture approximation.

24.7 The Ray Approximation in Geometric Optics

In the last section we said that when the geometric spot size was larger about the same size as the spot due to diffraction, we could ignore diffraction and use the simpler ray model. This is usually true in our personal experiences. But this may not be true in experiments or devices we design. We should see where the crossover point is.

Intuitively, if the aperture and the spot are the same size, that ought to be some sort of critical point. That is when the aperture size is equal to the spot size. We found that for a circular aperture the spot width is

$$w \approx 2.44 \frac{\lambda}{D_{\text{aperture}}} L$$

We want the case where $w = D_{\text{aperture}}$

$$D_{\text{aperture}} = 2.44 \frac{\lambda}{D_{\text{aperture}}} L$$

This gives

$$D_{\text{aperture}} = \sqrt{2.44 \lambda L}$$

Of course this is for round apertures, but for square apertures we know we remove the 2.44. This gives about a millimeter for visible wavelengths.

$$\begin{aligned} D_{\text{aperture}} &= \sqrt{2.44 (500 \text{ nm}) (1 \text{ m})} \\ &= 1.1045 \times 10^{-3} \text{ m} \end{aligned}$$

for apertures much larger than a millimeter, we expect interference effects due to diffraction through the aperture to be much harder to see. We expect them to be easy to see if the aperture is smaller than a millimeter. But what about when the aperture is about a millimeter in size? That is a subject for PH375, and so we will avoid this case in this class. But this is not too restrictive. Most good optical systems have apertures larger than 1 mm. Cell phone cameras may be an exception (but they play tricks to beat this by having multiple apertures). Even our eyes have an aperture that varies from about 2 mm to about 7 mm, so most common experiences in visible wavelengths will work fine with what we learn. Note that for microwave or radio wave systems this may really not be true!

How about the other extreme? Suppose $\lambda \gg D$. This is really beyond our class (requires partial differential equations), but in the extreme case, we can use reason to find out what happens. If the opening is much smaller than the wavelength, then the wave does not see the opening, and no wave is produced on the other side. This is the case of a microwave oven door. If the wavelength is much larger than the spacing of the little dots or lines that span the door, then the waves will not leave the interior of the microwave oven. Of course as the wavelength becomes closer to D this is less true, and this case is more challenging to calculate, and we will save it for a 300 level electrodynamics course.

To summarize

- $\lambda \ll D$ Wave nature of light is not visible
- $\lambda \approx D$ Wave nature of light is apparent
- $\lambda \gg D$ Little to no penetration of aperture by the wave

24.8 The ray model and phase

There is a further complication that helps to explain why the wave nature of light was not immediately apparent. Let's consider a light source.



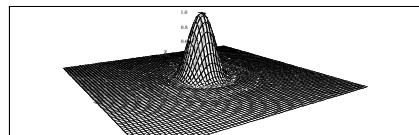
For a typical light source, the filament is larger than about a millimeter, so we should expect that diffraction should be hard to see. But the filament is made of hot metal. The atoms of the hot metal emit light because of the extra energy they have. The method of producing this light is that the atom's excited electrons are in upper shells because of the extra thermal energy provided by the electricity flowing through the filament. But the electrons eventually fall to their proper shell, and in doing so they give off the extra energy as light. It is not too hard to believe that this process of exciting electrons and having them fall back down is a random process. Each electron that moves starts a wave. The atoms have different positions, so there will be a path difference Δr between each atom's waves. There will also be a time difference Δt between when the waves start. We can model this with a $\Delta\phi_o$.

It is also true that not all of the electrons fall from the same shell. This gives us different frequencies, so we expect beating between different waves from different atoms. It is also true that we have millions of atoms, so we have millions of waves.

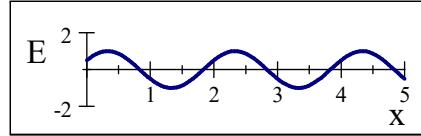
Let's look at just two of these waves

$$\begin{aligned}\lambda &= 2 \\ k &= \frac{2\pi}{\lambda} \\ \omega &= 1 \\ \phi_o &= \frac{\pi}{6} \\ t &= 0 \\ E_o &= 1 \frac{N}{C}\end{aligned}$$

$$E_1 = E_o \sin(kx - \omega t)$$



$$E_2 = E_o \sin(kx - \omega t + \phi_o)$$

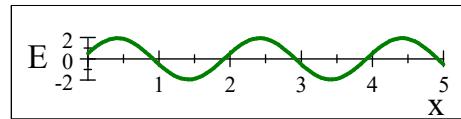


then

$$E_r = E_o \sin(kx - \omega t) + E_o \sin(kx - \omega t + \phi_o)$$

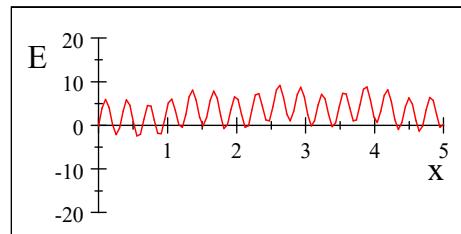
We found for these two waves

$$E_r = 2E_o \cos\left(\frac{\phi_o}{2}\right) \sin\left(kx - \omega t + \frac{\phi_o}{2}\right)$$

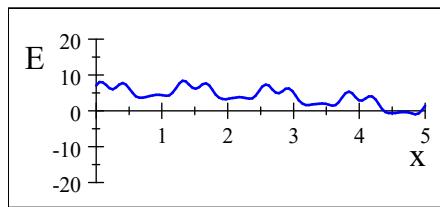


But suppose we complicate the situation by sending lots of waves at random times, each with different amplitudes and wavelengths, down the rope. If we look at a single point for a specific time, we might be experiencing interference, but it would be hard to tell. Lets try this mathematically. I will combine many waves with random phases, some coming from the right and some coming from the left.

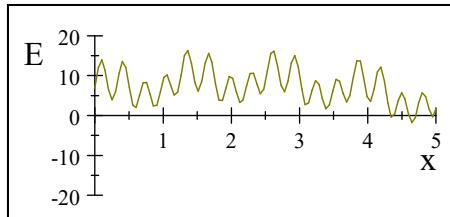
$$\begin{aligned} E_1 &= E_o \sin\left(5x - \omega t + \frac{\pi}{4}\right) + 0.5E_o \sin\left(0.2x - \omega t - \frac{\pi}{6}\right) \\ &\quad + 3.6E_o \sin\left(0.4x - \omega t + \frac{\pi}{10}\right) + 4E_o \sin\left(20x - \omega t - \frac{\pi}{7}\right) \\ &\quad + .2E_o \sin(15x - \omega t + 1) + 0.7E_o \sin(.7x - \omega t - .25) \end{aligned}$$



$$\begin{aligned}
 E_2 = & E_o \sin(0.2x + \omega t + \pi) + 2E_o \sin\left(5x + \omega t + \frac{\pi}{6}\right) \\
 & + 6E_o \sin\left(0.4x + \omega t + \frac{\pi}{3.5}\right) + 0.4E_o \sin(20x + \omega t - 0) \\
 & + E_o \sin(15x + \omega t + 1) + 0.7E_o \sin(.7x + \omega t - 4)
 \end{aligned}$$



Then $E_1 + E_2$ looks like



In this example, you could think about the superposition of E_1 and E_2 and predict the outcome, but if there were millions of waves, each with its own wavelength, phase, and amplitude, the situation would be hopeless. Note that the fluctuations in these waves are much more frequent than our original waves. With all the added waves, we get a rapid change in amplitude.

Now if these waves are light waves, our eyes and most detectors are not able to react fast enough to detect the rapid fluctuations. So if there is constructive or destructive interference that might be simple enough to distinguish, we will miss it due to our detection systems' integration times. To describe this rapidly fluctuating interference pattern that we can't track with our detectors, we just say that light bulbs emit *incoherent light*. The ray approximation assumes incoherent light.

But then light bulbs and hot ovens and most things must emit incoherent light. Does any thing emit coherent light? Sure, today the easiest source of coherent light is a laser. That is why I have used lasers in the class demonstrations so far. Really though, even a laser is not perfectly coherent. One property of the laser is that it produces light with a long *coherence length*, or it produces light that can be treated under most circumstances as being monochromatic and having a single phase across the wave for much of the beam length. Radar

and microwave transmitters emit coherent light (but at frequencies we can't see) and so do radio stations.

In the past, one could carefully create a monochromatic beam with filters. Then split the beam into two beams and remix the two beams. This would generate two mostly coherent sources if the distances traveled were not too large. This is what Young did. Now days we just use lasers.

24.9 Coherency

To be coherent,

1. A given part of the wave must maintain a constant phase with respect to the rest of the wave.
2. The wave must be monochromatic

These are very hard criteria to achieve. Most light, like that from our light bulb, is not coherent.

You may be left feeling that we have really just gotten started, and you would be right. If you are a physics major (or curious and have elective credit) you will study waves more in PH 279 (Modern Physics) and PH295 (Mathematical and Computational Physics) and you may study more optics in PH375. All of these are fantastic experiences. Even if you don't take a further class in these topics, you can study them on your own. There are many wonderful books on Optics.

Our goal this semester was to study the motion of many things and wave motion was a partial fulfilment of that goal. But not all motions of many objects are as uniform as wave motion. What if, say, the air molecules in our room instead of moving in simple harmonic motion like in a wave, they had random velocities? This is type of motion that we hinted about when we talked about thermal energy in PH121 and the hot atoms that emit incoherent light. We will take on this kind of motion next in this course.

Part III

Thermodynamics

Chapter 25

25 Fluids and Pressure

1.14.1, 1.14.2

We have spent about 2/3 of a semester studying waves. Waves are very regular motion of more than one object. But surely we can have less regular motion of more than one object!

To study such non-regular motion we are changing topics radically. PH121 taught us how individual objects move. We will use some of what we learned to address the question of how many objects move. To get started, let's review some basic properties of matter. Matter is made of many atoms. So matter is an example of many objects that might just move.

Fundamental Concepts

- Compressibility of fluids
- Density of fluids
- Pressure is a force spread over an area
- Pressure increases with depth in a fluid
- Barometers
- Manometers
- Buoyancy

25.1 Fluids

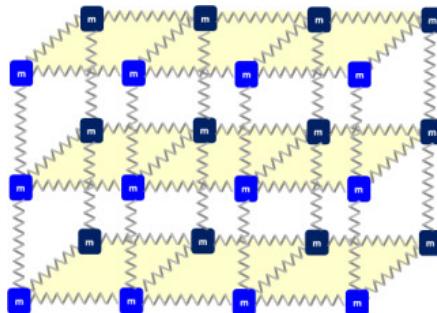
You are probably aware that there are four general states of matter

solid
liquid
gas
plasma

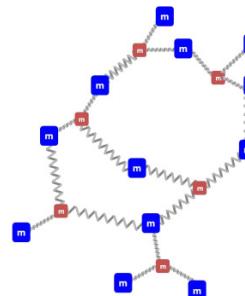
Believe it or not, plasma¹ is the most common, because stars are made of plasma. Planets are sometimes made of solids, liquids, or gases, but the great glowing stars are plasma. (I am ignoring the mysterious form of “dark matter” because so far we don’t know what it is). Plasma is a heated gas that is ionized. We will mostly ignore this state, because unless you are dealing with neon signs, fluorescent lights, or the like, on Earth you don’t encounter plasmas in every day experience.

Solids

We can view solids as having a set of forces that keep the molecules in place much as though they were attached using springs. Solids can have definite organization. If so, they are called crystals. You should observe the crystals around the Romney building if you have not already.



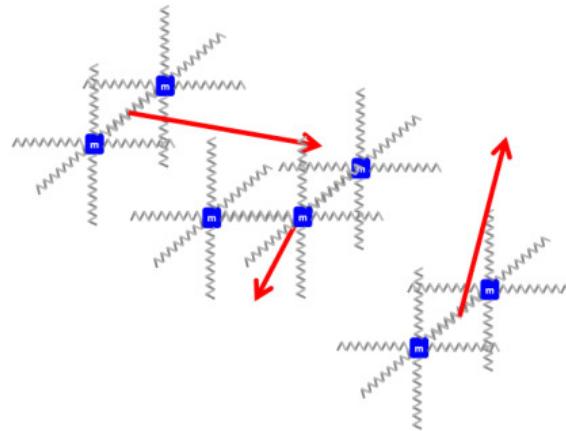
If the solid lacks definite order in its organization, it is called amorphous.



Liquids

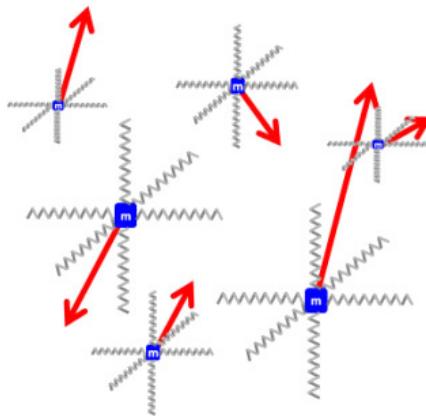
¹This is not the kind of plasma that you donate!

The molecules in a liquid are less tightly bound than those in a solid. That is why they can flow. In the next figure, the atoms are bound by one spring-like force. But the atoms are not tied together in a tight set of bonds like a solid.



Gasses

The molecules in a gas are not bound to each other—not at all.



We have an intuitive feel for what a fluid is. But let's make a more formal definition.

25.1.1 What is a fluid?

For the next few lectures we will study fluids, but what is a fluid? A **Fluid** is a collection of molecules that are randomly arranged and

are at most held together by weak cohesive forces and by forces exerted by the walls of a container.

Thus, liquids and gasses are definitely fluids. Solids are generally considered not fluids. But how about Jell-OTM? or a combination of corn starch and water (sometimes called “ooblick”)? These are non-Newtonian fluids. That is, things that are sort of solid and sort of not. But we will stick with things that are definitely fluids (at least at first), and generally fluids with negligible friction. This is a little like in PH121 when we studied frictionless surfaces. How many surfaces are truly frictionless? Very few! you might guess that there are few fluids that have no friction, and you would be right. But just like with PH121, the assumption of frictionless fluids makes the math easier, and that is good when we are starting a new topic.

25.2 Pressure

Volleyball Demo

We have already seen pressure in this course. But it's been a while, so let's review. Consider a situation where we ask six of our class members to come up and press on the ball from all directions. Suppose further that we ask each person to exert a force on the ball. And suppose we ask the person to use the area of their hand to exert the force. The motion of the ball, and even its shape depended on both the force (magnitude and direction) and the area involved in each push.

Noticing that each person exerts a force but that the force is not acting on one point, but is spread out over an area, we recognize that each person is exerting a pressure on the ball

$$P \equiv \frac{F}{A} \quad (25.1)$$

Hopefully you studied pressure in PH121 or equivalent. It is a force spread over an area. In this case it is the force of a hand spread over the area of the hand.

Now consider a ball sitting in a room surrounded by air. The air is a fluid, so its molecules are quite free to move around. Because there is some thermal energy in the room² the molecules will have some kinetic energy. So the air molecules will hit the ball. This will cause a force on the surface of the ball. And that force will be spread over the entire surface area of the ball. This is a force spread over an area. This is a pressure. We call this *air pressure*. This air force due to the colliding molecules is like having the hands pushing on the ball.

This force due to individual molecules is small and only lasts during the collision. But in the room we have many molecules, and many molecules impact the ball. The molecules also impact the walls of the room. Suppose that every time a molecule bounces back from one wall it ends up headed back to the

²Even in Rexburg.

opposite wall bounces back again toward the first wall. If the molecules keep coming, there will be a force on the wall quite a bit of the time. At least, on average there is a force, anyway. This is the force that causes air pressure. The molecules impact the walls, and the ball, and us, and everything in the room all over the surface area of each object. The result is air pressure on every object in the room.

Likewise, the water pressure in a swimming pool is caused by moving water molecules. You should convince yourself that the reason the water stays in the pool is partly because the air molecules bounce against the water surface exerting a pressure on the water!

You might notice that there are more molecules bunched together in the water than in the air. We can describe how bunched together the molecules are by giving the mass of the molecules for a unit of volume

$$\frac{m}{V} = \rho \quad (25.2)$$

We call this measure of bunched-up-ness the *density* of the material and we give it (unfortunately) the symbol ρ (the Greek letter “rho” pronounced like “row” as in “row a boat”). Sometimes it is useful to compare the density of a substance (like air) to the density of water

$$\frac{\rho_{\text{substance}}}{\rho_{\text{water}}} = \text{specific gravity}$$

This comparison is (unfortunately) called the *specific gravity*. But we need to be careful. We will find that density can depend on temperature. So let's define specific gravity as the comparison of the substance's density to the density of water at 4 °C at 1 atm of pressure. It turns out at this temperature and pressure

$$\rho_{\text{water}} = 1000 \frac{\text{kg}}{\text{m}^3}$$

which makes it convenient for comparison.

25.2.1 Working with the definition of pressure

We can work with equation 25.1 to define the force due to pressure

$$F \equiv PA$$

and can define a force due to the pressure at an element of area dA

$$dF \equiv PdA$$

Where there is a differential, expect that some time in the future we will integrate!

But before we go on, let's see what the units of pressure would be. We have a force divided by an area.

$$1 \frac{\text{N}}{\text{m}^2} = 1 \text{ Pa}$$

The symbol is the Pa, and the unit is called the *pascal*. This is the name of a famous scientist.

25.2.2 Pressure Example:

Let's do a pressure problem together,

Problem statement:

A 50.0 kg person balances on one heel of a pair of high heeled shoes. If the heel cross section is circular and has a radius of 0.5000 cm, what pressure does she exert on the floor?

Drawing



Variables

Known

M	Mass of person	$M = 50 \text{ kg}$
r	Radius of person's heel	$r = 0.500 \text{ cm}$
g	Acceleration due to gravity	$g = 9.8 \frac{\text{m}}{\text{s}^2}$

Unknown

F	Force
x	Coordinate Axis
z	Coordinate Axis
A	Area of woman's heel

Basic Equations

$$\begin{aligned} F &= ma \\ A &= \pi r^2 \\ P &= \frac{F}{A} \end{aligned}$$

Symbolic Solution

$$A = \pi r^2$$

$$F = ma = Mg$$

$$P = \frac{F}{A} = \frac{Mg}{\pi r^2}$$

Numerical Solution

$$\begin{aligned}
 P &= \frac{Mg}{\pi r^2} \\
 &= \frac{50 \text{ kg} * 9.8 \frac{\text{m}}{\text{s}^2}}{\pi (0.500 \text{ cm})^2} \\
 &= \frac{490.0 \frac{\text{m}}{\text{s}^2} \text{ kg}}{0.78540 \text{ cm}^2} \frac{(100 \text{ cm})^2}{\text{m}^2} \\
 &= \frac{490.0 \frac{\text{m}}{\text{s}^2} \text{ kg}}{0.78540} \frac{100000}{\text{m}^2} \\
 &= 6.2389 \times 10^6 \frac{\text{kg m}}{\text{s}^2 \text{m}^2} \\
 &= 6.2389 \times 10^6 \text{ Pa}
 \end{aligned}$$

$$P = 6.24 \text{ MPa}$$

Units Check

$$\frac{\text{kg} \frac{\text{m}}{\text{s}^2}}{\text{cm}^2} = \frac{\text{kg} \frac{\text{m}}{\text{s}^2}}{\text{cm}^2} \frac{(100 \text{ cm})^2}{\text{m}^2} = 10000 \frac{\text{kg} \frac{\text{m}}{\text{s}^2}}{\text{m}^2} = 10000 \text{ Pa}$$

Units

Check

Reasonableness

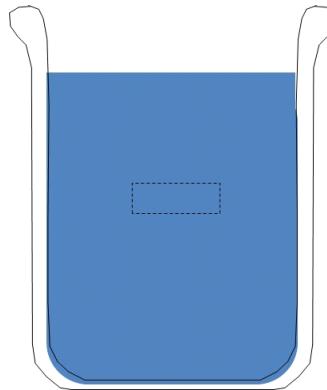
This seems like a large number, but I have had a high heeled person step on my toe, so I believe it! We can compare to average atmospheric pressure which is about $1.13 \times 10^5 \text{ Pa}$. Compared to this the toe crunch seems pretty reasonable.

25.3 Variation of Pressure with Depth

If you have been in a swimming pool, you probably have noticed that the pressure due to the water feels different at the surface than it does at the bottom of the deep end. Intuitively, we can say that the pressure at the bottom is larger than the pressure at the top. Let's see if we can show that this is true. We will do this in a few different ways.

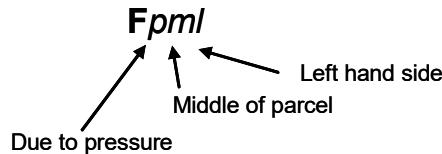
25.3.1 Pressure variation in liquids

Take a glass of water or some other liquid.

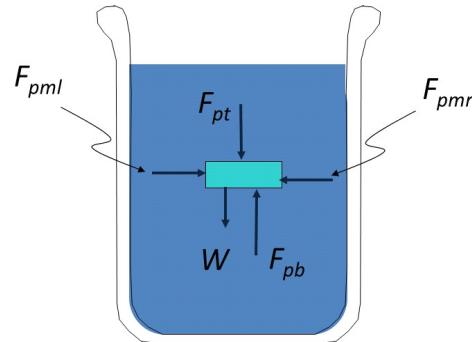


And let's look at just a part of the liquid (the section inside the box-shaped section with the dotted line around it in figure ??). The box shaped part of the liquid looks like, well, a box. And an old word for "box" is "parcel." Let's treat this "parcel of fluid" as a distinct body and look at the forces acting on it.

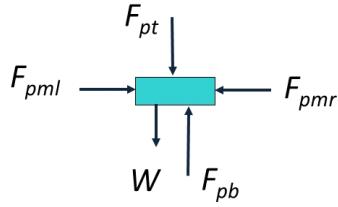
We need some way of labeling our forces. My way is kind of simple, but will work for now.



Consider Newton's second law. The sum of the forces must be equal to zero. Why?—think of the acceleration of the parcel of liquid. It's not accelerating.



And what we have in our drawing looks a lot like a free body diagram for our parcel of liquid. We could draw the free body diagram like this



where you should notice that we are drawing the forces with their points on the part of the parcel where the force acts. That is different than what we did in Principles of Physics I. This is because we are going to turn these forces into pressures using

$$P = \frac{F}{A}$$

so we want to identify the area that the force acts on. Let's write out Newton's second law for our parcel of liquid.

$$\vec{F}_{net} = m \vec{a} = 0 = \vec{F}_{pt} + \vec{F}_{pb} + \vec{F}_{pmr} + \vec{F}_{pml} + \vec{W}$$

but just like in Principles of Physics I we will break this equation into components.

$$\begin{aligned} F_{net_x} &= 0 = 0 + 0 - F_{pmr} + F_{pml} - 0 \\ F_{net_y} &= 0 = -F_{pt} + F_{pb} - 0 + 0 - W \end{aligned}$$

The x -part tells us that

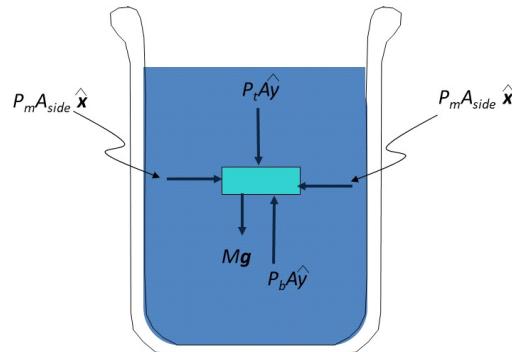
$$F_{pmr} = F_{pml}$$

which is not too much of a surprise. It seems reasonable that the side forces must be equal if the parcel doesn't accelerate. The y -part gives

$$F_{pb} - F_{pt} = W$$

Now let's use our pressure equation. The definition of pressure gives

$$F = PA$$



So, for the top of our parcel of liquid, the force on the top must be

$$F_{pt} = P_t A$$

and for the bottom the force must be

$$F_{pb} = P_b A$$

Recalling that

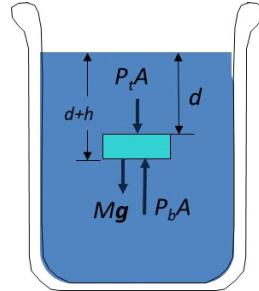
$$m = \rho V \quad (25.3)$$

where ρ is the density and V is the volume, and substituting, we get

$$\begin{aligned} F_{pb} - F_{pt} &= W \\ P_b A - P_t A &= -mg \\ (P_b - P_t) A &= -\rho V g \\ (P_b - P_t) A &= -\rho h A g \end{aligned}$$

where we have used the fact that the volume must be the area of the top or bottom, A , multiplied by the length, h , of our parcel.

$$V = Ah \quad (25.4)$$



We can solve this for the pressure at the bottom

$$P_b = P_t + \rho g h$$

Units Check

$$\text{Pa} \approx \text{Pa} + \frac{\text{kg}}{\text{m}^3} \frac{\text{m}}{\text{s}^2} \text{m}$$

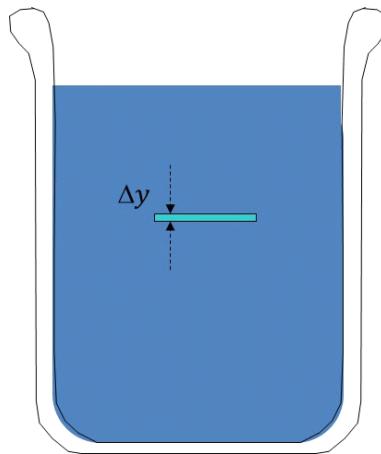
the last term can be written as

$$\left(\frac{\text{kg m}}{\text{s}^2} \right) \frac{1}{\text{m}^2} \Rightarrow \frac{\text{N}}{\text{m}^2} \Rightarrow \text{Pa}$$

so the units check.

This is a profound statement! (and a new basic equation for us). The pressure is larger at the bottom of the pool than the top. And it is larger by the amount ρgh where ρ is the liquid density, g is the acceleration due to gravity, and h is how far down we go to get to the bottom of our parcel.

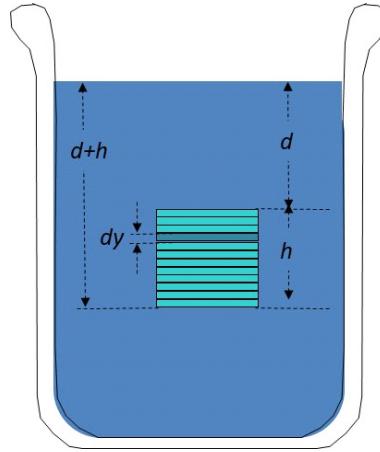
You might have thought of an objection in how we came up with this equation for the pressure at the bottom of a parcel of liquid. We said the sideways forces on the parcel were equal. But now we know the pressure increases with depth. Won't that affect the side forces so that the force on the bottom of the side is larger than the force on the top of the side? Of course the answer is yes. We could solve this problem by making the sides of the box infinitesimally small. Let's call the parcel height Δy



With our parcel of liquid so very thin now the pressure on the side can't have changed much from the top to the bottom of the parcel. So now we know the side forces cancel. We could say that the change in pressure from the top to the bottom

$$\begin{aligned} P_b &= P_t + \rho g \Delta y \\ \Delta P_{small} &= \rho g \Delta y \end{aligned}$$

And we could consider making a larger parcel to be made of many small parcels of thickness Δy .



For each of the little parcels the side forces cancel. We just have top and bottom forces. For the topmost parcel we would have

$$P_{b1} = P_t + \rho g \Delta y_1$$

And note, the bottom pressure from the topmost parcel is the top pressure for the second parcel

$$\begin{aligned} P_{b2} &= P_{t2} + \rho g \Delta y_2 \\ &= P_{b1} + \rho g \Delta y_2 \end{aligned}$$

And the bottom pressure from the second parcel is the top pressure for the third parcel

$$\begin{aligned} P_{b3} &= P_{t3} + \rho g \Delta y_3 \\ &= P_{b2} + \rho g \Delta y_3 \end{aligned}$$

To get to the bottom of the big parcel we would have

$$\begin{aligned} P_{b(total)} &= P_t + \sum_{i=1}^N \rho g \Delta y_i \\ &= P_t + \rho g \sum_{i=1}^N \Delta y_i \\ &= P_t + \rho g h \end{aligned}$$

just as we thought.

Of course we could take a limit to make $\Delta y \rightarrow dy$ and then we would have

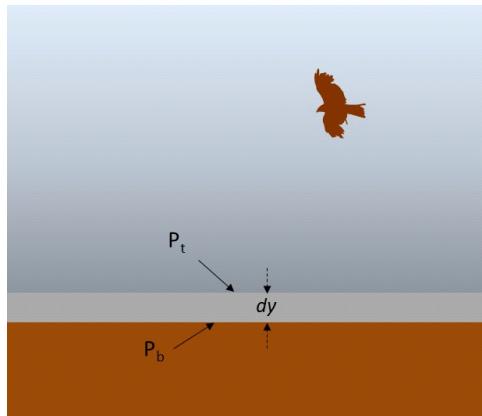
$$\begin{aligned} P_b &= P_t + \int_d^{d+h} \rho g dy \\ &= P_t + \rho g y|_d^{d+h} \\ &= P_t + \rho g ((d+h) - d) \\ &= P_t + \rho g h \end{aligned}$$

just as we said before.

Note that we used a box shaped volume, but the result is general. To see this, view any arbitrary volume as consisting of little boxes. As we make the box size small, we approximate the actual volume better and better until it can be exact. Since it works for each box that makes up the volume individually, it will work for the whole volume (it takes a minute to think about this).

25.3.2 Pressure variation in gasses

You probably noticed that in our last section we didn't let ρ change. Another way to say that the density doesn't change is to call our liquid *incompressible*. Liquids are fluids, but they are difficult to compress. So to a good approximation their density doesn't change. But gasses are also fluids, and it is easy to compress a gas. We can't make the approximation of incompressibility for gasses.



We don't know enough yet to know how ρ changes with y . But if you took high school physics or chemistry we can make an approximation using the Ideal Gas Law. If you have never heard of the Ideal Gas Law, don't despair. We will study it soon. But for now here is the equation

$$PV = Nk_B T$$

where P is the pressure, V is the volume of the air, n is a number density (how many air molecules per unit volume), $k_B = 1.3806568 \times 10^{-23} \frac{\text{J}}{\text{K}}$ is a constant, and T is the temperature. We can solve this for pressure

$$P = \frac{Nk_B T}{V}$$

We can change the form of this a bit by realizing that the number of molecules is the total mass of our gas divided by the mass of one gas molecule.

$$N = \frac{m_{total}}{m_{molecule}}$$

so that

$$P = \frac{\frac{m_{total}}{m_{molecule}} k_B T}{V}$$

then

$$P = \frac{\frac{m_{total}}{V} k_B T}{\frac{m_{molecule}}{m_{molecule}}} = \rho \frac{k_B T}{m_{molecule}}$$

and we can solve this for the density so we can see how density of the gas changes with pressure.

$$\frac{P m_{molecule}}{k_B T} = \rho$$

Now we can use our equation for a small parcel of fluid (gas this time)

$$P_b = P_t + \rho g \Delta y$$

but this time we want to know P_t at the top so let's solve for that.

$$P_t = P_b - \rho g \Delta y$$

and write it as

$$\frac{\Delta P}{\Delta y} = -\rho g = -\frac{P m_{molecule} g}{k_B T}$$

This seems pretty terrible, but we can take the limit as $\Delta y \rightarrow dy$

$$\frac{dP}{dy} = -P \left(\frac{m_{molecule} g}{k_B T} \right)$$

and we note that so long as the temperature doesn't change, all the things in the parentheses are constant. So we could give this group of constants a symbol

$$\alpha = \frac{m_{molecule} g}{k_B T}$$

Then our equation for how pressure varies with height is just

$$\frac{dP}{dy} = -\alpha P$$

and this is a nice differential equation. But rather than use our differential equations class (that we haven't taken) to solve this let's rearrange it

$$\frac{dP}{P} = -\alpha dy$$

and now let's integrate over the height of our parcel of air from $y = 0$ to $y = h$. On the pressure side we need to integrate from P_b to P_t

$$\int_{P_b}^{P_t} \frac{1}{P} dP = - \int_0^h \alpha dy$$

we get

$$\begin{aligned}\ln P \Big|_{P_b}^{P_t} &= -\alpha y \Big|_o^h \\ \ln(P_t) - \ln(P_b) &= -\alpha h\end{aligned}$$

Using one of our favorite log identities $\ln a - \ln b = \ln \frac{a}{b}$ we can rewrite this as

$$\ln \left(\frac{P_t}{P_b} \right) = -\alpha h$$

and we can un-log this by exponentiating both sides

$$\frac{P_t}{P_b} = e^{-\alpha h}$$

which gives finally

$$P_t = P_b e^{-\alpha h}$$

We could let our parcel of air be arbitrarily tall by letting the height be $h = y$ above the surface ($y = 0$) and making $P_t = P(y)$ and letting the surface pressure $P_b = P_{surface}$. Then

$$P(y) = P_{surface} e^{-ay}$$

but this can only work for small distances above the Earth's surface because in reality the temperature also changes with y . But in our classrooms this approximation is good enough (because the hvac system is keeping the whole lab at the same temperature).

25.4 Pressure Measurements

Now that we understand pressure and how it changes in fluids, we can study several pressure measurement devices. But there are some quirks in pressure measurement. Humans have measured pressure for a long time. But for most of our existence what we cared about was how much larger a pressure was than the surrounding air pressure. Think of the pressure inside a steam engine. The steam pressure has to be larger than atmospheric pressure to make the steam engine work. So humans invented pressure gages to put on things like steam engines that tell us how much higher than atmospheric pressure our particular pressure is. We call this a *gauge pressure*. But now days we understand that most of the universe is a vacuum. And that vacuum of space has almost no pressure.³ We need to calculate the *absolute* pressure, and if your gauge measured gauge pressure you need to add atmospheric pressure to your reading.

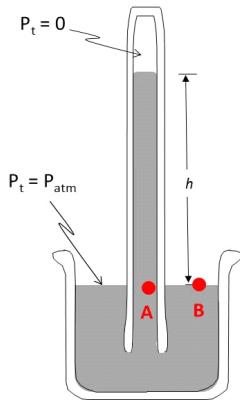
$$P_{abs} = P_{gauge} + P_{atmospheric}$$

But with this understanding, let's look at pressure measuring devices.

³You might have heard an old saying which says that nature abhors a vacuum. But since most of nature is free space, this old saying only applies to things on or in planets or stars. The universe seems to love vacuum.

25.4.1 The Barometer

The barometer is a common pressure measuring device. It consists of



1. a long tube closed at one end
2. a dish
3. Mercury or another fluid

The dish and the tube are filled with the fluid. The tube is inverted in the dish. The pressure at the top of the tube is essentially zero. This is because the weight of the fluid pulls the mass of fluid downward. If the fluid is mercury, it is massive enough to leave a vacuum behind.

The pressure at point *A* and point *B* must be the same (or fluid would flow until $P_A = P_B$). This might be a little hard to see. But remember the air has a pressure, and just because it is a gas it does not mean that pressure is lower than the liquid pressure (think of the gas pressure in a hurricane that blows over trees and buildings).

It would be useful to find out how high up the mercury (or other fluid) will stay. And we have a way to do this! We know

$$P_b = P_t + \rho gh$$

At the top we have zero pressure, so $P_t = 0$. Note, this is not atmospheric pressure, **it is vacuum**. How did we get vacuum at the top? We got it by filling the tube with mercury and then turning the tub upside down. The mercury is heavy. So heavy that it falls down the tube a bit, leaving nothing behind. So there is no air at the top of the tube, just open tube because the heavy mercury moved downward. But the mercury can only fall down so far. The air pressure on the mercury in the dish pushes, keeping the mercury from all falling out. Then for this special case

$$P_b = 0 + \rho gh \quad (25.5)$$

And we can solve for the height of the fluid

$$h = \frac{P_b}{\rho g} \quad (25.6)$$

Let's consider what P_b must be. The air outside the barometer pushes down on the liquid mercury. Remember, when we first construct a barometer, we fill the tall tube with fluid and then turn it upside down in the dish. The fluid will fall until the force due to air pressure matches the weight of the column of fluid. Then the pressure at B must be atmospheric pressure. Since the pressure at point A equals the pressure at point B (think $P_b = P_t + \rho gh$), we can use this as a convenient P_b . So

$$P_b = P_{atm}$$

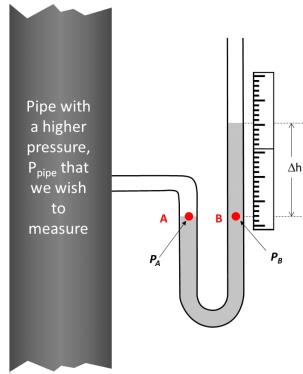
and

$$h = \frac{P_{atm}}{\rho g} \quad (25.7)$$

Thus as the atmospheric air pressure changes, the height of the mercury changes. We often hear of atmospheric pressure given in mm of Hg . This is why.

25.4.2 Manometer

The manometer finds an unknown pressure



The atmospheric pressure P_{atm} is applied on one end of the tube. In our example it is on the right side, where the tube is open at the top. The pressure to be tested is applied to the other side, in our case, the left side. The pressure at point A must equal the pressure at point B . Why? (think of their heights and ρgh). The pressure at point A is the pressure to be measured. Again

$$P_b = P_t + \rho gh$$

where we take P_A to be the test pressure and for this design the tube above B is open to the atmosphere so the top pressure will be atmospheric pressure.

$$P_A = P_B = P_{atm} + \rho gh$$

We can measure the height h , and knowing P_{atm} we can calculate P_A .

These are just two of many types of pressure measuring devices. In the next lecture we will use what we know of pressure to design some useful devices like hydraulic jacks and things that float (like boats).

Chapter 26

26 Pascal and Archimedes

1.14.3, 1.14.4

So now we know how pressure changes with depth, and we know that liquids are generally hard to compress, so we can approximate them as incompressible, let's build some things with our new knowledge.

Fundamental Concepts

1. Pascal's Principle
2. Hydraulic Presses
3. Archimedes Principle
4. Buoyant Forces

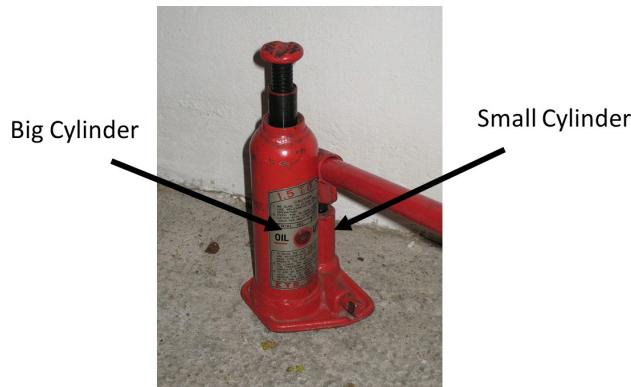
26.1 Hydraulic Press

We use fluids and their incompressibility to convert pressure into forces to smash or lift things. A hydraulic press can smash carbon. into diamonds

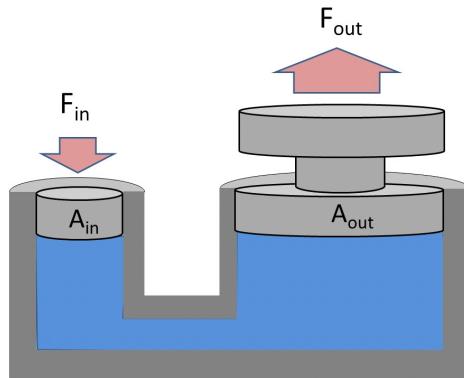


Brigham Young University Lee Library University Archives; UAP 2

A hydraulic jack can lift cars.



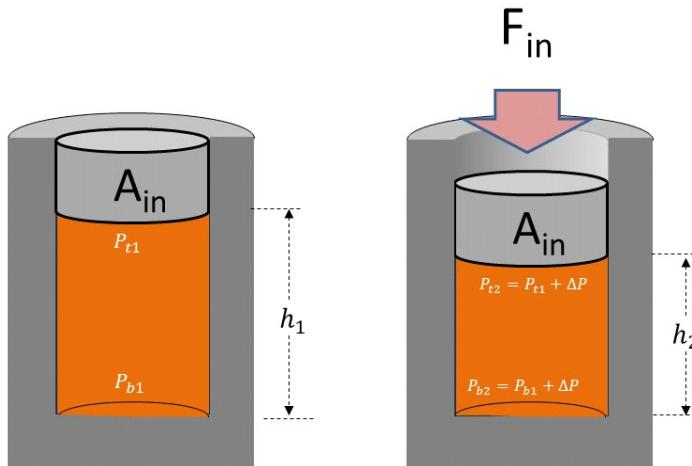
Let's take on the hydraulic jack. In the last figure, the jack is made from two cylinders filled with fluid, usually an oil. The jack in the picture is a little harder to think about. Let's draw a simplified jack:



We usually have a constant force acting on the jack area. And that force will be related to the pressure in the liquid.

$$F = PA \quad (26.1)$$

Let's look at just the left cylinder, the input of our jack. And further suppose that instead of a sensible liquid like oil, we fill the jack with a gas that is compressible



As we press down on the piston, the force is spread over the piston area. That causes an increase in pressure, ΔP . An early researcher named Pascal reasoned that at first the gas molecules that are near the piston would go flying because of the force of the piston hitting them. But they would hit other molecules (and the walls) and after a time all the gas molecules would be effected. When the system comes to equilibrium, we find that not only has the top pressure changed by ΔP so that (using the figure)

$$P_{t2} = P_{t1} + \Delta P$$

but also the bottom pressure has changed.

$$P_{b2} = P_{b1} + \Delta P$$

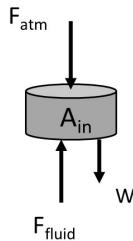
This is called Pascal's principle. We can restate it as the change in pressure in a fluid is transferred to the entire fluid and the magnitude of the change is the same everywhere.

Pascal's Law: a change in the pressure applied to a fluid is transmitted undiminished to every point of the fluid and to the walls of the container.

Now let's consider the jack again. In the jack we have a liquid, and liquids are not easily compressible. When our jack is just sitting in our garage, we have just atmospheric pressure pushing on the jack input piston,

$$F_{atm} = P_{atm}A_{in} \quad (26.2)$$

so Newton's second law for the input piston is

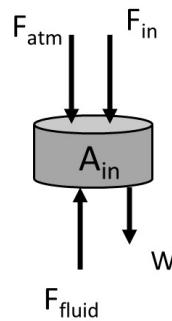


$$\Sigma F_{in} = F_{fluid} - F_{atm} - W_{in} = F_{fluid} - P_{atm}A_{in} - W_{in} = 0 \quad (26.3)$$

which tells us that

$$F_{fluid} = P_{atm}A_{in} + W_{in}$$

But we wish to add a force F_{in} to this. So on the input side,



we have, using Newton's second law

$$\Sigma F_{in} = F_{fluid} - F_{atm} - W_{in} - F_{in} = F_{fluid} - P_{atm}A_{in} - W_{in} - F_{in} \quad (26.4)$$

And once again our F_{in} will cause a change in pressure in the fluid

$$F_{in} = \Delta P A_{in} \quad (26.5)$$

But the first three forces in our Newton's second law haven't changed, and they sum to zero, so we have

$$\Sigma F_{in} = 0 - \Delta P A_{in}$$

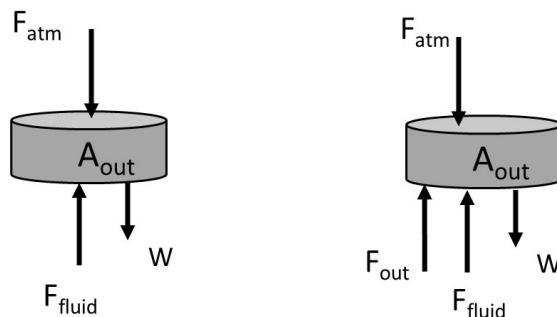
But this time we can't compress the fluid in the cylinder because it is a liquid. But we expect our ΔP to be transmitted throughout the entire fluid. By changing the pressure on the input side we have changed the pressure by ΔP on the output side (the lifting side). The only way this can happen is for the output piston to move! So the output will have an amount of force added to it. On this output side, remember before we pushed we had atmospheric pressure pushing down on the output piston.

$$\Sigma F_{out} = F_{fluid} - F_{atm} - W_{out} + F_{out} \quad (26.6)$$

Now think again about the state of our jack before we pushed on the input piston. The output piston wasn't accelerating. So before

$$\Sigma F_{out} = F_{fluid} - F_{atm} - W_{out} = 0 \quad (26.7)$$

and these forces haven't changed,



but we do have a new force $F_{out} = \Delta P A_{out}$ that we have added to our net force on the output piston.

$$\Sigma F_{out} = 0 + \Delta P A_{out} \quad (26.8)$$

Let's solve for ΔP in our two net force equations

$$\frac{\Sigma F_{in}}{A_{in}} = \Delta P = \frac{F_{in}}{A_{in}}$$

and

$$\frac{\Sigma F_{out}}{A_{out}} = \Delta P = \frac{F_{out}}{A_{out}} \quad (26.9)$$

Since ΔP in both equations is the same when the two sides are at the same elevation, then

$$\frac{F_{out}}{A_{out}} = \frac{F_{in}}{A_{in}} \quad (26.10)$$

So how does your hydraulic jack work?

$$F_{out} = F_{in} \frac{A_{out}}{A_{in}} \quad (26.11)$$

If $A_{out} > A_{in}$ a much smaller F_{in} can produce a large F_{out} (you already knew that, didn't you!).

It is important to note that we have assumed our jack fluid is not compressible. So the volume of fluid leaving the cylinder at the input side must be the same as the volume of fluid entering the output side. We could say

$$V_{out} = V_{in}$$

but

$$\begin{aligned} V_{out} &= A_{out} \Delta y_{out} \\ V_{in} &= A_{in} \Delta y_{in} \end{aligned}$$

then

$$A_{out} \Delta y_{out} = A_{in} \Delta y_{in}$$

Since $A_{out} > A_{in}$, it is clear that the output piston will travel a much smaller distance than the input piston. This is why you have to pump quite a lot on the input side of your jack to move a car a relatively small distance.

Note that our equation

$$P_b = P_t + \rho gh$$

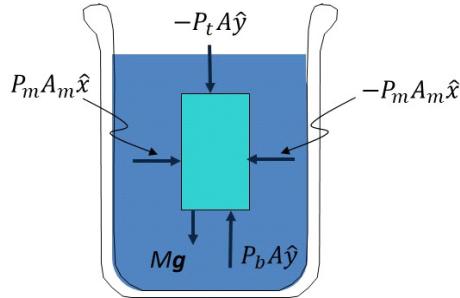
still applies! The pressure is not the same at the top and bottom of the cylinders. Pascal's principle tells us that the change in pressure ΔP was the same in the entire liquid, not that the pressure, itself, is the same. This is an important distinction! We will need to be careful with Pascal's principle.

In many cases we can sum up pressure over an area because the pressure will be the same at every point within the area. Our hydraulic jack is one of these cases.

But let's turn our attention now to pressure variation with depth again and look at things that float.

26.2 Buoyant Forces, Archimedes' Principle

Remember our parcel of liquid from last lecture.



Let's investigate the net force on our parcel and let's say our liquid is water this time. The net force must be zero, or it would be accelerating.

$$\begin{aligned}\Sigma F_x &= ma_x = 0 = P_m A_m - P_m A_m \\ \Sigma F_y &= ma_y = 0 = -P_{atm} A + P_{bottom} A - W\end{aligned}$$

But suppose I ignore the weight of the parcel of water, $W = mg$. Then there would be a net force due to all other forces in the \hat{y} direction. But our parcel isn't accelerating. That force must be matching the downward force due to gravity.

$$W = P_b A - P_t A$$

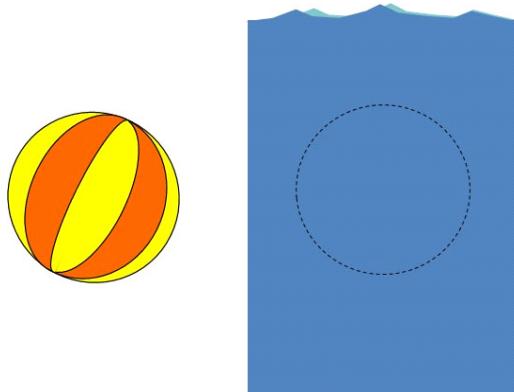
We could give this net force due to all the water pressure forces a symbol, say " B ."

$$B = P_b A - P_t A$$

Then B would be equal to the weight of the parcel of water!

$$B = W$$

Now take a beach ball sized parcel of water



The parcel is in equilibrium just as before. We know there is a force on the parcel of water due to gravity. It is downward. Just as before, the ball-shaped parcel of water is not sinking so again there is no net force. Then if we add up all the forces due to the water pressure, only, we must have a net pressure force that is upward to balance the downward force due to gravity. Again this is what we called "*B*." And again *B* is equal to the weight of the parcel of water.

The idea of a net force just due to pressure due to a fluid is so useful we give it its own name. We will call the net force due to just pressure from the fluid the *Buoyant force*.

Note that this is **not** the net force, it is just the sum of the forces due to the water pressure. It does not include gravity or any tension or spring or any other forces. It just contains the pressure forces due to the fluid.

Note also that the buoyant force is not some new kind of force. It is a name we give to the sum of the pressure forces due acting on an object in the fluid. Like we call a group of soldiers a battalion or a group of cows a heard, the sum of the group of pressure forces due to a fluid is given the name "buoyant force." But herds and battalions are not new things, they are groups of cows and soldiers. A Buoyant force is not a new force, it is the vector sum of a group of pressure forces.

Because the parcel is not sinking, we can determine that in this special case

$$B = W_{\text{fluid parcel}} \quad \text{Floating} \quad (26.12)$$

for this situation that we call floating (these are magnitudes, are the directions the same?). The term on the right is the weight of the parcel of water or fluid.

Suppose we replace this amount of water with the beach ball. The weight of the beach ball is different than the weight of the water parcel. But will any of the pressure forces be different? The two volumes are exactly the same!

It turns out that the buoyant force will be exactly the same for the beach ball as it would for the beach ball-shaped parcel of water. We can calculate the buoyant force by thinking of replacing the actual object we have with an equal volume parcel of water. The buoyant force will be equal to the weight of that equal volume of the fluid. We could say that to replace the water with the beach ball we have *displaced* a beach ball volume's worth of water. The water that was at the location of the beach ball is now somewhere else, so it is displaced. Then the buoyant force would be equal to the weight of the water that was displaced.

$$B = W_{\text{displaced fluid}} \quad (26.13)$$

Of course, the weight of the beach ball is far less than the weight of the water displaced

$$W_{\text{ball}} < W_{\text{displaced fluid}}$$

and this is why the beach ball accelerates upward.

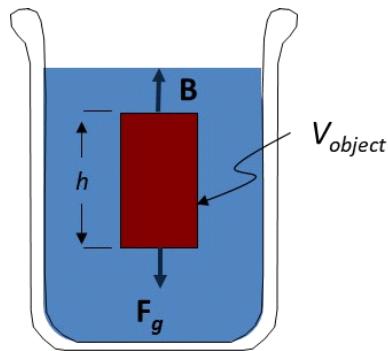
It turns out that this idea of the buoyant force being equal to the weight of an equal volume of fluid is general! This concept of the Buoyant force being equal to the weight of the water that fits inside our volume is called *Archimedes' principle*.

The magnitude of the buoyant force equals the weight of the fluid displaced by the object.

26.3 Buoyant Force: Two Cases

We can gain insight into this concept of a buoyant force by considering more floating objects. Let's consider an object that is floating submerged in the fluid (like a fish or submarine) and one that is floating at the surface and partially extends out of the fluid (like a floating log or boat).

26.3.1 Totally submerged object



Consider a solid object, a block of some material, placed in water. The weight of the solid object is Mg .

The magnitude of the buoyant force is equal to the weight of an equal volume of water. If we imagine replacing the block with a block-sized volume of water, then

$$\begin{aligned} B &= m_{fluid}g \\ &= \rho_{fluid} V_{object}g \\ &= \rho_{fluid} g V_{block} \end{aligned}$$

Think Archimedes: The weight of an equal volume of water was displaced by the block. So we have the volume of the block as part of our equation for the buoyant force.

If the object has mass M then we can write the mass of the block as

$$\begin{aligned} M &= \rho_{object} V_{object} \\ &= \rho_{block} V_{block} \end{aligned}$$

and the weight of the block is then

$$\begin{aligned} W &= Mg \\ &= \rho_{object} V_{object} g \\ &= \rho_{block} V_{block} g \end{aligned}$$

The net force is then

$$\begin{aligned} F_{net} &= B - W = \rho_{fluid} g V_{object} - \rho_{object} V_{object} g \\ &= \rho_{fluid} g V_{block} - \rho_{block} V_{block} g \\ &= g V_{block} (\rho_{fluid} - \rho_{block}) \end{aligned}$$

Consider this last equation. If the density of the block is large, larger than the density of the fluid, then $(\rho_{fluid} - \rho_{block})$ is negative, and the force will be negative. The block will accelerate downward. If the density of the block is smaller than the density of the fluid, then $(\rho_{fluid} - \rho_{block})$ will be positive and the block will accelerate upward. If the density of the block is just the same as the density of the fluid, the block will float in place (or at least sink or float upward at a constant rate). A submarine adjusts its density to dive or to surface.

We did this calculation for our block, but it could have been any object. So in general

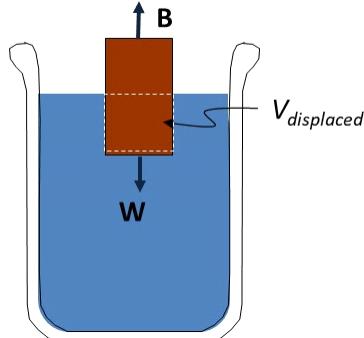
$$F_{net} = g V_{object} (\rho_{fluid} - \rho_{object}) \quad (26.14)$$

If the object has a density that is less than the fluid, the object will be accelerated upward. And if the density of the object is greater than the density of the fluid the object will be accelerated downward.

Note that the motion of a totally submerged object is determined only by the relative density! In a way, this could be confusing because we did not allow our object to change volume as it moved. If we had an object that could change size, then it changes density as it changes size. A balloon full of air released from the bottom of a swimming pool does change volume as it rises. So in that case we would need to know ΔV_{object} and how the density, ρ_{object} , changes as the balloon moves.

26.3.2 Partially submerged object

Now let's consider a partially submerged object



Let's assume $\rho_{object} < \rho_{fluid}$ (why? think of $\rho_{fluid} - \rho_{object}$). We assume static equilibrium, which means we are observing this floating object when it is just sitting still relative to the objects around it, not accelerating.

Since $a = 0$ then

$$\Sigma F = ma = 0$$

and

$$\Sigma F = B - W$$

so for an object floating on the surface

$$B = -W$$

Lets call the volume of the fluid displaced by the object V_{fluid} . Now this is not the same as V_{object} for this case! Part of the object is sticking out of the fluid, so $V_{fluid} < V_{object}$. But, it is still true that the weight of the displaced volume of fluid gives the buoyant force

$$B = \rho_{fluid}gV_{fluid}$$

Then returning to our force equation

$$\begin{aligned} B &= -W \\ \rho_{fluid}gV_{fluid} &= -\rho_{object}V_{object}g \end{aligned}$$

or

$$\frac{V_{fluid}}{V_{object}} = \frac{\rho_{object}}{\rho_{fluid}} \quad (26.15)$$

The fraction of the volume of a floating object that is below the fluid surface is equal to the ratio of the density of the object to that of the fluid.

In the case of an iceberg, not all of the ice is visible above the surface. The density of ice is about $920 \frac{\text{kg}}{\text{m}^3}$ and that of sea water is about $1000 \frac{\text{kg}}{\text{m}^3}$. So for ice

$$\frac{V_{fluid}}{V_{object}} = \frac{920 \frac{\text{kg}}{\text{m}^3}}{1030 \frac{\text{kg}}{\text{m}^3}} = 0.8932$$

$$V_{fluid} = 0.89V_{object}$$

that is the fluid displaced is 89% of the ice. Just about 10% of the ice sticks up out of the water. I



Iceberg with both beautiful blue-green submerged portion and a reflection of the surface ice and snow. Approximately 90 per cent of the iceberg is submerged. Antarctica, Palmer Peninsula, Northern area. (Dr. Mike Goebel, NOAA NMFS SWFSC, NOAA NMFS SWFSC Antarctic Marine Living Resources (AMLR) Program.)

We have gained some good insight using pressure. But think, pressure is a force spread over an area. And we know forces from Principles of Physics I (PH121). But we didn't just use forces in Principles of Physics I. We also found we could use energy to solve motion problems. In our next lecture, let's try to bring the powerful technique of conservation of energy into our study of fluid flow.

Chapter 27

27 Conservation of Energy for Fluid Flow 1.14.5, 1.14.6

So far, we have only dealt with fluids in equilibrium. The topic of fluid dynamics is a complicated mathematical field that requires differential equations to do with any exactness. There are whole upper division classes that study this called *Fluid Dynamics*. But with some simplifying assumptions, we can get a feel for how fluids flow.

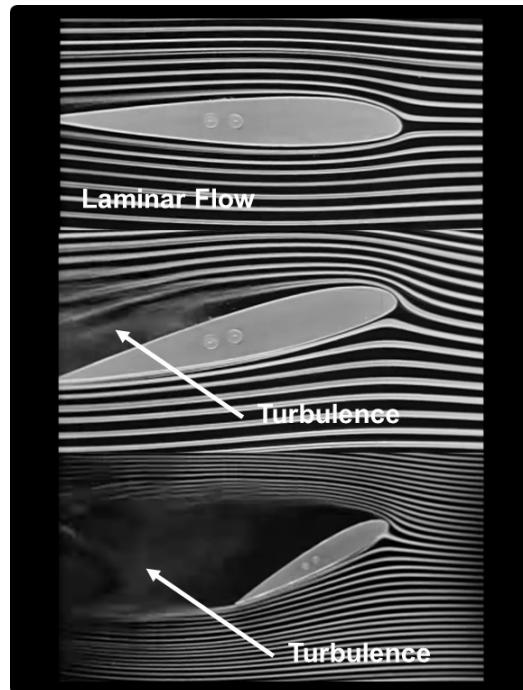
Fundamental Concepts

1. Continuity
2. Bernoulli's Equation

27.1 Moving Fluids

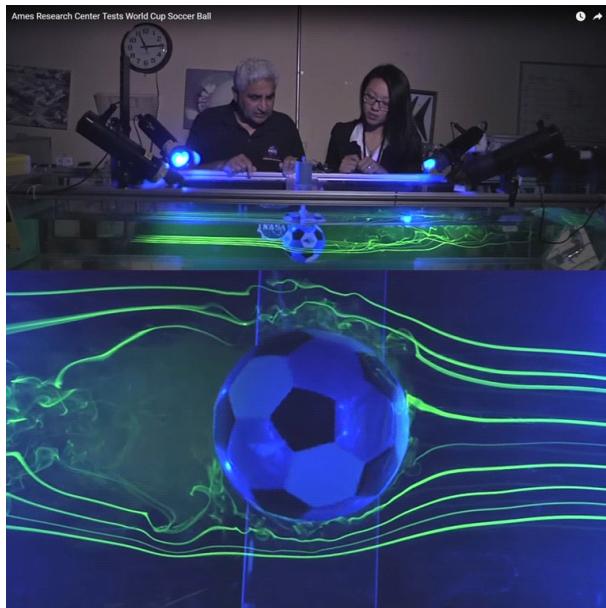
To start with, let's make some definitions.

New Definitions:



1. Laminar: Steady flow, if each particle of the fluid follows a smooth path, such that the paths of the different particles never cross each other.
2. Turbulent: non-laminar flow characterized by whirlpool or eddy regions
3. Viscosity: term describing the internal friction of a fluid (Honey is more viscous than water)
4. Irrotational: flow having no angular momentum
5. Streamline: a line indicating the path taken by a particle under steady flow

In the figure above we can see streamlines going around airfoils. The reason we can see them is that small particles are being deposited by evenly spaced sources in an air tunnel. This makes it so we can see the streamlines! Here is the same type of thing for a soccer (futbol) being tested by NASA.

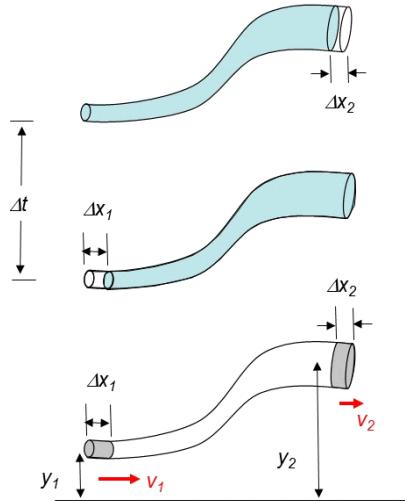


NASA Ames measuring turbulence from a soccer ball. There is a cool video at this URL! <https://www.nasa.gov/centers-and-facilities/ames/nasa-turns-world-cup-into-lesson-in-aerodynamics/>

Sadly, we will limit our study to laminar, non-viscous, irrotational, incompressible fluids. This is a little bit like saying we will only study frictionless surfaces in PH121! But to go beyond non-viscous laminar flow we need the math of partial differential equations, which is not a prereq for this class. So we will have to content ourselves with this restriction. For those of you who are interested in learning more (or are ME majors) you can take our junior level fluid dynamic course taught by the Mechanical Engineering Department.

27.2 Equation of continuity

Let's start our discussion of fluid flow with a pipe that changes diameter along its length.



And just for good measure we can have the pipe go up in position. We allow an ideal (non-compressible) fluid to flow through the pipe in laminar flow. The segment of fluid is the shaded part in the top part of figure ???. In a time Δt the fluid at the left hand side moves a distance

$$\Delta x_1 = v_1 \Delta t$$

If A_1 is the area of the pipe at the left hand side (LHS) then the mass contained in the shaded region (bottom left part of figure ???) is

$$m_1 = \rho V_1 = \rho A_1 \Delta x_1 = \rho A_1 v_1 \Delta t$$

where ρ is the density of the fluid. Likewise, for the RHS region in the bottom of figure ???

$$m_2 = \rho V_2 = \rho A_2 \Delta x_2 = \rho A_2 v_2 \Delta t$$

Unless we create, destroy, or pool mass, the mass that crosses A_2 must equal the mass that crosses A_1

$$\begin{aligned} m_1 &= m_2 \\ \rho A_1 v_1 \Delta t &= \rho A_2 v_2 \Delta t \end{aligned}$$

or

$$A_1 v_1 = A_2 v_2 \quad (27.1)$$

This is the *equation of continuity for fluids* (remember that we assumed ideal fluids).

Since we assumed that no mass was created, destroyed, or pooled as the fluid flowed, this is what we might call “conservation of mass.” And as usual, when

something is conserved, we can use it to solve problems. With this equation, knowing the diameter of a pipe and how it changes tells us about the speed of the fluid in the pipe and how it changes. Of course, there could be situations where mass is not conserved (say, a leaky pipe). So conservation of mass is an assumption that we have to check. But for now let's assume that the equation of continuity holds for our problems.

The quantity Av has a name. We call it the flow rate and for some reason I don't know we give it the symbol \mathbb{Q}

$$\mathbb{Q} = Av$$

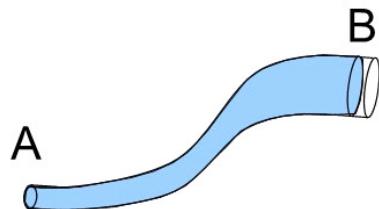
27.3 Bernoulli's Equation

It would be great if we could find the pressure as well as the speed of the fluid inside our non-uniform pipe. We would need an equation that included both pressure and speed. Let's try to find such an equation!

In PH121 we developed two ways of looking at motion problems. One was the force picture where we had to deal with the vector nature of forces and motion. The other was the energy picture where we could solve problems without considering directions (but we gave up being able to find directions). Usually the energy picture made solving problems easier. And except for the flow direction, we don't really want to know the direction of the motion of each molecule in our fluid, so giving up direction doesn't seem too bad. So, the work energy theorem is something we know from PH121. Let's start there,

$$W = \Delta K + \Delta U$$

We still need to use forces to find the work done on the parcel of water, but if we choose our axes carefully, we can keep the work equation in one dimension.

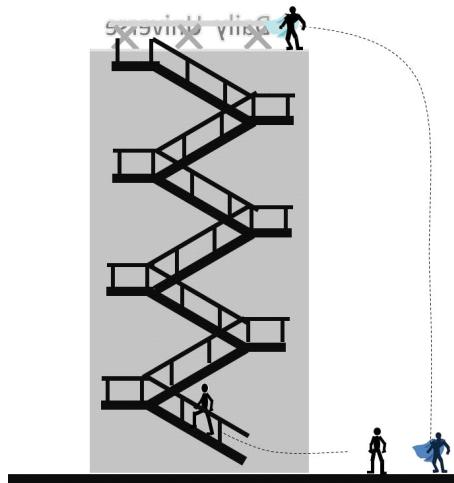


27.3.1 Work done in Δt

Let's review a little of what we learned in PH121. Suppose we have two characters, Normal Person and Super Person.



And suppose they want to get to the top of a tall building. Normal Person has to take the stairs, but Super Person can jump right up to the top.

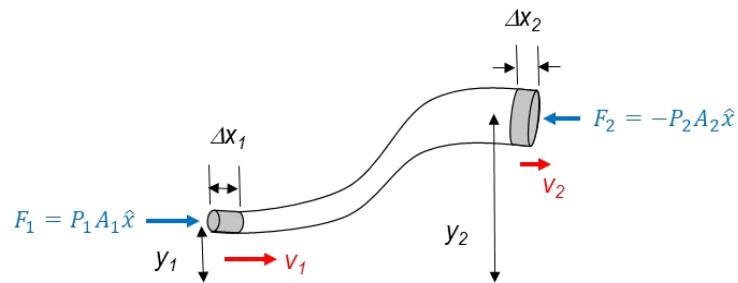


Who has done more work?

We know the change in potential energy is the same for both. And we know that

$$W = -\Delta U$$

so the amount of work is the same for both cases! Now let's return to our fluid flow.



On the left hand side (LHS) of our pipe we have a force due to pressure from the fluid in the rest of the LHS of the pipe. That fluid has molecules that will bang in to our small marked parcel of fluid. The force from the LHS fluid in the rest of the pipe will be

$$F_1 = P_1 A_1 \hat{\mathbf{x}} \quad (27.2)$$

On the right hand side (RHS) there is also more fluid in the rest of the RHS of the pipe. Our fluid will bang into that fluid, and since the fluid is non-compressible, the rest of the fluid in the RHS will push back. The force will be

$$F_2 = -P_2 A_2 \hat{\mathbf{x}} \quad (27.3)$$

From PH121 we remember that works is

$$w = \int \vec{\mathbf{F}} \cdot d\vec{\mathbf{x}}$$

and since our force is not changing this would be simply

$$w = \vec{\mathbf{F}} \cdot \vec{\Delta \mathbf{x}} \quad (27.4)$$

We need to find the work done in moving our particular parcel of fluid from the left hand side to the right hand side. As the parcel moves in the $\hat{\mathbf{x}}$ direction the amount of work will slowly change (like Normal Person going up the stairs) but really the amount of work only depends on the beginning and ending conditions. So we could ignore all the details of moving the parcel and envision the parcel jumping from the left up to the right side of the pipe (like in the Super Person case).

Then for the LHS

$$W_1 = P_1 A_1 \Delta x_1$$

and for the RHS

$$W_2 = -P_2 A_2 \Delta x_2$$

The total work will be the sum of these two works. But before we add them together, let's make some substitutions that will make our final formula for the total work more meaningful.

Lets define the volume of our fluid segment

$$V = A_1 \Delta x_1 \quad (27.5)$$

Note that this volume is NOT the volume of the entire segment. It is just the part that is Δx_1 long. Then from the same arguments that lead to the equation of continuity, we see that the volume of the marked part of the fluid must not change as it flows

$$V = A_1 \Delta x_1 = A_2 \Delta x_2$$

Then we can write the work equations as

$$W_1 = P_1 V$$

and

$$W_2 = -P_2 V$$

So finally, the total work done is

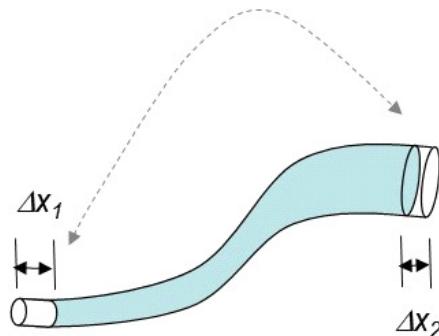
$$\begin{aligned} W &= W_1 + W_2 \\ &= P_1 V - P_2 V \\ &= (P_1 - P_2) V \end{aligned} \tag{27.6}$$

so, from the work-energy theorem

$$\begin{aligned} W &= \Delta K + \Delta U \\ (P_1 - P_2) V &= \Delta K + \Delta U \end{aligned}$$

We have an expression for the left hand side of our work-energy theorem! Now for the right hand side $\Delta K + \Delta U$

27.3.2 Kinetic Energy



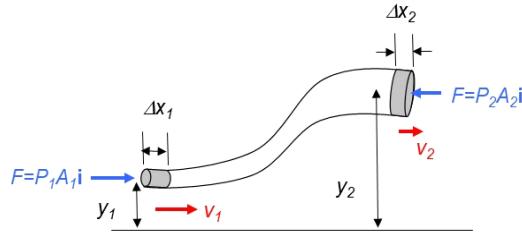
If we think about the energy of the segment, we will realize that after time Δt , it is as though most of the fluid did not move. We can't tell one part of the fluid from another. The shaded region in figure ?? is occupied before and after Δt . We can treat this problem as if we moved fluid from region 1 to region 2 (Super Person vs. Normal Person again!) and left the rest of the segment alone!

So, ignoring the rest of the fluid, we can see that if the velocity of our parcel of water changed, then the kinetic energy of the parcel must change as well

$$\Delta K = \frac{1}{2} m v_2^2 - \frac{1}{2} m v_1^2 \tag{27.7}$$

(the mass is the same, as we found before).

27.3.3 Potential Energy



Let's look at the change in potential energy for the parcel of fluid. Once again we can consider just the beginning and ending case (think of Super Person and Normal Person once again). We work as though we moved a mass $m = m_1 = m_2$ from y_1 to y_2 .

$$\Delta U = mg y_2 - mg y_1 \quad (27.8)$$

Since our pipe went upward as the parcel of water flows, we gained potential energy. But if the pipe went downward we would lose potential energy.

27.3.4 Total Work done on the system

Let's now assemble our work-energy theorem. We had found the work so far,

$$(P_1 - P_2) V = \Delta K + \Delta U$$

but now we know ΔK and ΔU . We have calculated W , ΔK , and ΔU for our parcel of water, so let's substitute in what we have found.

$$((P_1 - P_2) V) = \left(\frac{1}{2} m v_2^2 - \frac{1}{2} m v_1^2 \right) + (m g y_2 - m g y_1) \quad (27.9)$$

This equation describes our fluid flow much like the equation of continuity, but now it has pressure involved! We have achieved our goal. But we can make this equation a little easier to understand if we play a trick by dividing by V

$$((P_1 - P_2)) = \left(\frac{1}{2} \frac{m}{V} v_2^2 - \frac{1}{2} \frac{m}{V} v_1^2 \right) + \left(\frac{m}{V} g y_2 - \frac{m}{V} g y_1 \right) \quad (27.10)$$

and recalling that

$$\rho = \frac{m}{V}$$

then

$$(P_1 - P_2) = \left(\frac{1}{2} \rho v_2^2 - \frac{1}{2} \rho v_1^2 \right) + (\rho g y_2 - \rho g y_1)$$

Notice that we have subscripts 1 and 2. Let's put all the 1's on one side of our equation and all the 2's on the other side.

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2 \quad (27.11)$$

which completes our goal. We have an equation that contains the velocity *and* the pressure of the fluid.

But notice! The left hand side looks just the same as the right hand side except for the subscripts! We have a before and after picture like we did in energy and momentum problems back in Principles of Physics I. Something is being conserved. Sometimes you see this written as

$$P + \frac{1}{2}\rho v^2 + \rho g y = \text{constant} \quad (27.12)$$

This is another conservation equation! The thing that is being conserved is $P + \frac{1}{2}\rho v^2 + \rho g y$. We have not named this quantity, but whatever it is, it is not changing as the fluid flows. Our experience with conservation equations tells us that this new conservation law will help us solve problems. This equation is so useful we give it a name. It is called *Bernoulli's equation* (after the scientist that came up with it). What is really being conserved is energy. We found Bernoulli's equation from conservation of energy. But this form of conservation of energy is useful for solving pressure and fluid flow problems.

Chapter 28

28 Turbulence and Temperature 1.14.7, 2.1.1

In our last lecture we said we would ignore turbulence and viscosity. But if we are cautious we can at least get some intuition about what will happen if we have friction or non-laminar flow.

Fundamental Concepts

- Number Density
- Atomic Mass
- Atomic Mass Number
- The Mole
- Temperature

28.1 Viscosity

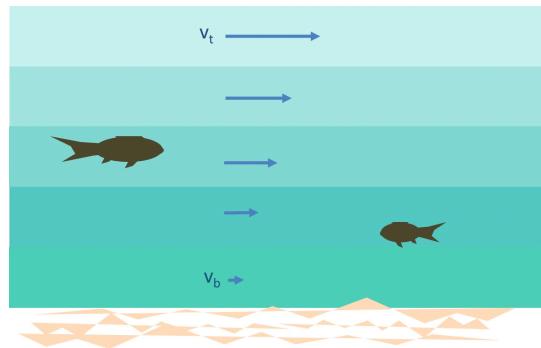
Way back in Principles of Physics I (PH121) we learned about friction. So far we haven't had any friction with our fluids. But we know this can't be true! We know there is drag force. So there must be some friction. And we can think of there being two different cases where friction can matter. There is the friction between parts of the water, itself, and there must be something like a drag force between the water and the solid earth. Let's take on the water/earth friction first.

In the past, we used two different equations for drag for fluids, one for air

$$D = \frac{1}{2} \rho C A v^2$$

where ρ is the air density, C is the drag coefficient, A is the cross sectional area, and v is the speed of an object through air. And we had a different one for a liquid like oil when we had oscillators being damped by the fluid

$$D = bv$$



where b is the damping coefficient and v is the speed. These two equations, one for a gas fluid like air and one for a thicker liquid like oil, show that friction for fluids can be quite different. And of course it is much easier to get through air than oil, so that is not a surprise. Recall that we call friction in liquids *viscosity* and we would say the oil is more viscous than the air.

There will be a drag force between the river bed and the lowest water. And that drag force will likely be more of the $D = bv$ type for a liquid. This slows the water that is right next to the river bed down to a v that is quite slow.

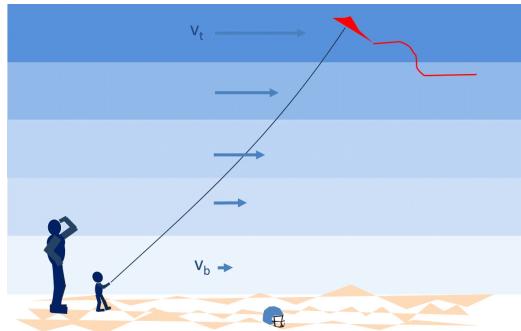
But we know fluids don't just have friction with solid surfaces. Because the fluid can move around internally there is internal friction or viscosity. When a river flows down its bed we know the drag force at the bottom causes the water to slow down. But this slower water is also viscous and so it slows the water just above it a little. So the water at the bottom of the river will be slow and the water just above it will be less slow. The water above this second layer will be faster still, and this keeps going until we reach the top of the water where there is some friction with the air, but it is much less. So the fastest water is on top. This works for air too.

You might have tried flying a kite.



Kite flown on the June 19th 2025 BUYU-I High Altitude Research Team flight as part of the high altitude balloon payload retrieval effort. The kite can carry radios that connect to the payload trackers.

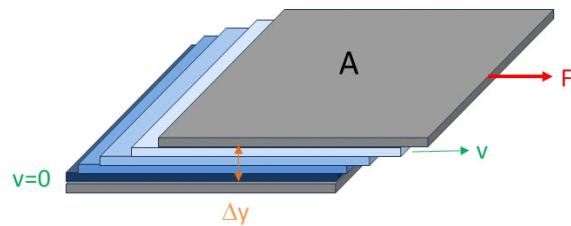
When you first start, you have to run with the kite. But once the kite is higher, the winds tend to be higher and the kit can stay up without you running.



Both cases work well as long as nothing disrupts the fluid flow. This nice well behaved flow we described as being *laminar*. In laminar fluid flow layers may have different speeds, but they flow without mixing.

All this works pretty well if we don't have turbulence and the viscosity isn't too high.

We can measure the viscosity of a fluid by placing it between two sheets



If a force, F is applied to the top sheet we have a sheer force and the sheet moves horizontally. The viscosity makes the fluid move with it. The fluid at the bottom won't move at all, the fluid at the top moves with speed v and the layers in between will have speeds in between v and 0. The viscosity η is given by

$$\eta = \frac{F\Delta y}{vA}$$

In our society we put a lot of effort to make fluids flow in pipes. Our water comes and goes in pipes. Fuel flows in pipes. Air flows in organ pipes during conference, etc. The rate at which the water flows is called a flow rate and unfortunately it is often given the symbol Q . But what makes the fluid flow? Generally is is pressure. We recall that pressure is a stress, a force spread over an area.

$$P = \frac{F}{A}$$



For our fluid in a pipe the area will be the pipe cross sectional area. We would like to know how the pressure affects the flow rate, Q . It turns out that this is fairly hard to solve (there are multivariate differential equations) so for those who are curious or need to be able to do this professionally I recommend taking a fluid dynamics class¹ but for our class let's just borrow a result.

$$Q = \frac{P_{in} - P_{out}}{R}$$

This is called Poiseuille's Law. We can see the pressure in, P_{in} and the pressure out, P_{out} . The denominator, R is the viscosity resistance. This equation won't work for fluids with no viscosity, because R would be zero and the flow rate infinite. But fluids with no viscosity are rare, so it works for most normal situations. We would not expect the flow to be uniform across the pipe. Fluid

¹At BYU-I this class is ME360 and is called Fluid Mechanics.

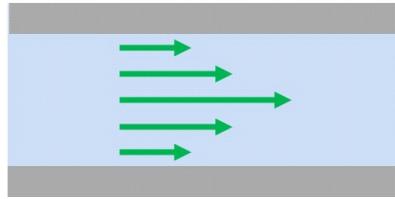
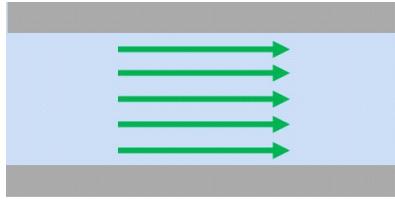
closer to the pipe walls should move slower. Fluid near the center should move faster. We can model the resistance in the pipe by borrowing from a fluid dynamics class

$$R = \frac{8\eta L}{\pi r^4}$$

where η is the viscosity of the fluid, L is the length of the pipe, and r is the radius of the pipe. We can put this simple model for R into Poiseuille's law

$$\begin{aligned} Q &= \frac{P_{in} - P_{out}}{\frac{8\eta L}{\pi r^4}} \\ &= \frac{\pi r^4 (P_{in} - P_{out})}{8\eta L} \end{aligned}$$

We can see that this gives us a dependence on the radius. The wider the pipe, the more flow we get. The more viscous the fluid, the less the flow rate. Notice also that longer pipes have lower flow. Friction will have more of an affect if the fluid interacts with the pipe longer, so this makes sense. And of course if we increase the pressure on one end, the flow changes.

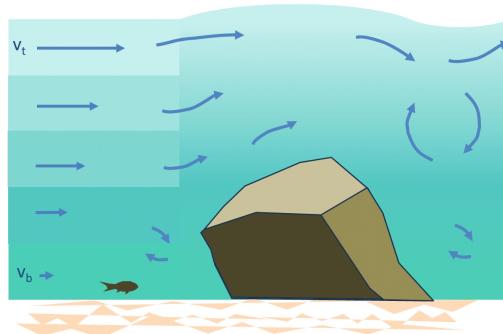


Pipe with no viscosity (top) and with more realistic viscosity (bottom)

The math that lead to this equation assumed laminar flow. But what if we don't have laminar flow?

28.1.1 Turbulence

We can disrupt laminar flow by, say, putting a rock in the way of the fluid. Here is a rock in a river.



Now the laminar flow is disrupted by the rock and the layers will mix. We call a situation like this *turbulent* flow. As you can imagine, the equations for turbulent flow are harder. Finding the flow rate for turbulent flow is really a topic for a fluid dynamics class. But we can give a measure of how turbulent our situation is.

$$Re = \frac{\rho v D}{\eta}$$

where ρ is the fluid density, v is the flow speed, D is the pipe diameter, and of course η is the viscosity of the fluid. This quantity is called the Reynolds number and if this number is high enough the flow will be turbulent. Engineering books give a table like the following for determining if the flow is still laminar.

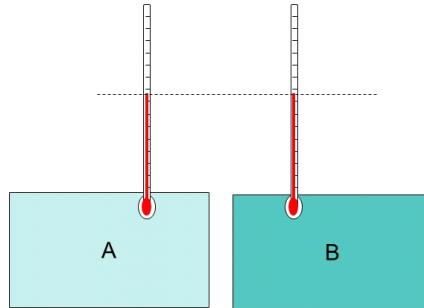
laminar	$Re < 2300$
Transient	$2300 < Re < 4000$
Turbulent	$Re > 4000$

although not all books agree on the value for the transition to turbulence. But generally if $Re < 2300$ our equations for this class work and for $Re > 2300$ we don't have laminar flow and our equations really don't work. The transient range is a transition to turbulence and for Re above 3000 to 4000 most agree that the flow is turbulent. For many engineering designs we really don't want turbulence. So this is enough to do some good work. For cool physics problems turbulence is a form of "complexity" which is a current emphasis in physics research.

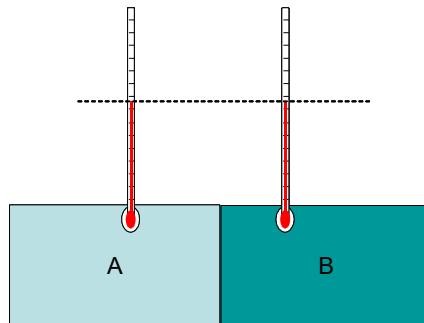
Fluid dynamics is the movement of a bulk of fluid. This is all we can do for fluid dynamics in this class. But we can still look at the motion of the molecules and atoms inside of the fluids (and solids). So lets move on to thinking about what we will call internal motion in the fluid. It turns out that this internal motion is related to how hot or cold our fluid is. We will call this the thermodynamics of the fluid!

28.2 Temperature and the Zeroth Law of Thermodynamics

We need a basis for talking about thermal systems. We sort of have an intuition for temperature, so let's use our intuition for now.



In the figure we have two objects. I measure both objects temperature. I find that they are the same. Suppose I put the two objects in “contact.”



Would you expect there to be energy transferred between the two objects? For most of us our intuition says “no.” Given this intuition, lets make some more formal definitions:

- Thermal Contact: Two objects are in thermal contact if they can exchange energy due to a temperature difference
- Thermal Equilibrium : A state where two objects would not exchange energy by heat or electromagnetic radiation if they were placed in thermal contact
- Temperature: The property that determines if an object is in thermal equilibrium with other objects (two objects in thermal equilibrium have the same temperature)

Given all this, we can state a law of thermal dynamics: **If objects A and B are separately in thermal equilibrium with a third object C, then A and B are in thermal equilibrium with each other.**

Again this fits our intuition, if we have two containers of water that are both at the same temperature as a third container of water, we expect they all have the same temperature.

In a sense, this seems obvious, so obvious that it was always assumed without statement. But in retrospect, it is a fundamental concept that deserves its place as one of the laws of thermodynamics. By the time this was acknowledged, the names “first law” and “second law” were taken, so this became the *Zeroth law of Thermodynamics*.

Before we move on, let’s review some things we know from high school chemistry.

28.3 Numbers of Atoms and Number Densities

Atoms are small. They are really incredibly small. A typical atom measures a few Angstroms across. An angstrom is 1×10^{-10} m. That is unbelievably small. Let’s try to get an understanding of this scale. Picture a meter stick. Then divide it up into 1000000000 equal units. That is how small an atom is. Or, consider scaling our meter stick to be the size of the continental US, that is about 4.0877×10^6 m. Then for our US sized meter stick, on this scale an atom would still only be about half a millimeter across! About the size of a cake sprinkle.

So it is no surprise that in a typical sample of a solid we can have something like 10^{25} atoms. We use the capital letter N to mean the number of atoms in a sample of matter.

Because these numbers are so large, it is sometimes convenient to talk about how many atoms there are in a small volume of a substance. This is often less than in the whole sample. We can say, for example, that there are 1000 atoms per cubic micrometer of material. We call such a ratio of number of atoms per volume a *number density*. In the SI system, the units for the number of atoms per unit volume are $1/\text{m}^3$. So number densities can still be quite large! Number densities are more useful for samples where the density is uniform, but we can talk about substances whose number density changes. Unfortunately you will often see a small n used to mean number density. This overuse of n is confusing, so sometimes we will write number density as N/V so it is obvious what we mean.

28.3.1 Atomic Mass

You probably ran into atomic mass in High School. The idea of atomic mass is that each nucleon (proton or neutron) has about the same mass, and the electrons don’t have much mass at all, so we can sort of count nucleons and say that an atom has a mass in units of nucleon masses. Carbon has 12 nucleons

(six protons and six neutrons) so it would have a mass of 12 u. The unit u is the atomic mas unit, the mass of one nucleon. Because of the electrons (and other effects) this is not an exact approach, and it is not good enough if we attempt to do nuclear physics. But it is close enough for us for now. So we can say that hydrogen (1H) has an atomic mass of 1 u. If we have a molecule, we can just add the atomic masses of each atom in the molecule together to find the molecule's mass. Note that to be more accurate, we can use a periodic table and find the atomic mass, or, if we need many digits of accuracy, we can go to a table of the nucleotides. We can convert these masses to SI units using

$$1 \text{ u} = 1.66 \times 10^{-27} \text{ kg}$$

28.3.2 Moles

Chances are that you are familiar with the concept of a mole as well. But experience has shown me that the mole will be new to some people. So let's start with something we know, a dozen.

What is a dozen? Well, it is 12 of something. We could have a dozen doughnuts or a dozen eggs. or a dozen large white pickup trucks. It is just an amount. It is 12 of whatever we are counting.

A mole is like a dozen, but it is a BIG dozen. Is is 6.02×10^{23} of something. I could have 6.02×10^{23} doughnuts, or 6.02×10^{23} eggs. I would need quite a few hungry friends to eat it all. But a mole is just a number of something. The mole is not a very practical amount for everyday objects like doughnuts or trucks. But think of the large number of atoms in a sample of matter. For this situation, a mole seems like a reasonable number to use.

Like we say I want to buy a dozen eggs, we could say we want to buy a mole of hydrogen atoms. Since atoms are so small, packaging them in big numbers makes sense.

This number, 6.02×10^{23} , is called *Avogadro's number*

$$N_A = 6.02 \times 10^{23}$$

To find the number of dozens of a something, we take the number of items and divide by 12.

$$n_{\text{dozen}} = \frac{N}{12}$$

To find the number of moles of a something, we take the number of items and divide by Avogadro's number

$$n_{\text{moles}} = \frac{N}{N_A}$$

usually we will use small n to be the number of moles of something. This could be a problem because n can also mean number density. So we will have to be careful to state what we mean with our variables.

28.3.3 Molar mass

We could give a mass for a dozen eggs. If each egg has the same mass (more or less) then the dozenal mass would be the mass of a dozen eggs and it would be twelve times the mass of one egg. The dozenal mass of a dozen trucks would be larger than the dozenal mass of eggs because the mass of a single truck is larger than the mass of an egg, and the mass of a dozen trucks is 12 times the mass of one truck.

We can also define a molar mass, which would be the mass of 6.02×10^{23} of something. Usually, we choose the units for this molar mass to be the mass *in grams* of a mole of something. And when we do this, the molar mass is equal to the numerical value of the atomic or molecular mass. for example, the molar mass of He, with $m = 4\text{ u}$ is $M = 4\text{ g/mol}$. The molar mass of diatomic nitrogen would be 28 g/mol because each nitrogen atom has an atomic mass of $m = 14\text{ u}$ and we add the two atomic masses together for the molecular mass. With this definition of molar mass we can see that

$$n_{moles} = \frac{M \text{ (in grams)}}{M_{molecule}}$$

Now that we have these new terms to use, let's take up measuring temperature in our next lecture.

Chapter 29

29 Measuring Temperature

2.1.2, 2.1.3

Last lecture we reviewed some basics of how materials are made and how chemists and physicists and material scientists determine how many moving parts a sample of a material might have. We also talked about measuring temperature. But didn't actually talk about what temperature is or even how temperature measuring devices work. We will start with measurement techniques in this lecture and build up to a model of what temperature is.

Fundamental Concepts

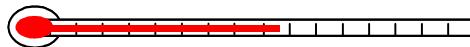
- Kelvin Temperature Scale
- volume of a liquid
- the dimensions of a solid
- the pressure of a gas at constant volume
- the volume of a gas at constant pressure
- The electric resistance of a conductor
- The color of an object
- Phase Changes and Phase Diagrams
- Thermal Expansion

29.1 How to build a thermometer

We will find that several material properties change with temperature

Things that Vary with Temperature
Volume of a liquid
The dimensions of a solid
The pressure of a gas at constant volume
The volume of a gas at constant pressure
The electric resistance of a conductor
The color of an object
The size of an object

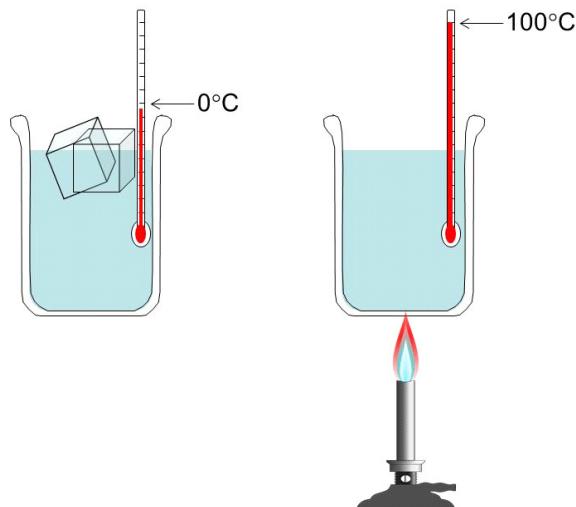
Since these things change with temperature, we can use any of these to make a temperature measuring device. Think of old fashioned thermometers, how do they work?



Old fashioned mercury thermometers (and their glycerine replacements) use the first property, volume of a liquid increases with temperature. The liquid is placed in an evacuated tube, and the amount of the tube that is filled depends on the temperature. The liquid fills more of the tube when the temperature goes up because the liquid volume gets bigger with temperature. By placing a ruled background behind the tube or even by making rule marks on the outside of the tube, we can find out how much the liquid volume changed, and calculate the temperature change.

29.1.1 Celsius temperature scale

To make our old fashioned mercury thermometer work we need to have units of temperature to mark on the tube. One of the sets of units for temperature is called the *Celsius temperature scale*. The most important thing to say about the Celsius scale is not to use it for doing thermodynamic problems (The same goes double for the Fahrenheit scale). But since it is common to see Celsius temperatures, we will talk about them.



The idea, once again, is to make marks on our thermometer tube that show how the change in liquid volume is related to temperature. To do this for the Celsius scale we put our thermometer in ice water and mark where the top of the liquid is. And we assign the value of 0°C to this mark. Then we take our thermometer and put it in boiling water. The liquid moves up the tube. Where it stops we make a mark and label this mark 100°C . This scale sets the zero point at the ice point of water and sets 100°C at the steam point of water. Then we make 100 evenly spaced divisions between these points. We can also mark the tube with the same division spacing below the 0°C mark and above the 100°C mark. But this calibration does not work well if high accuracy is required. There are often large discrepancies when temperatures are measured outside of the 0°C to 100°C range.

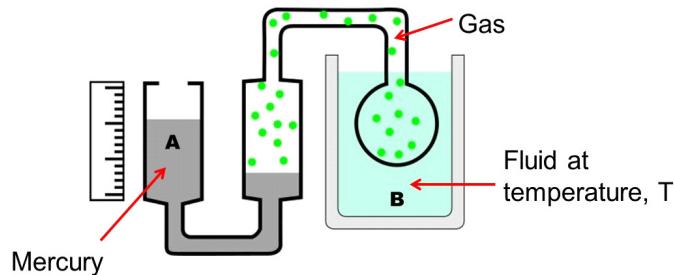
Worse yet, water only boils at 100°C at sea level, and does not always freeze at 0°C .

Even worse, there is nothing really very 0ish about 0°C . We picked it because it was a nice value for water. But we can certainly have lower temperatures than 0°C .¹ This makes 0°C more like an origin of a coordinate system. We could say that the intersection of Main St. and Center St. is our origin, but there is noting really zeroish about that intersection. But it turns out there *is* a natural zero point for temperature! A better temperature scale would start at the natural zero point. Let's see if we can make such a scale.

¹We often do in Rexburg from about December to March.

29.1.2 The Constant-Volume Gas Thermometer and the Absolute Temperature Scale

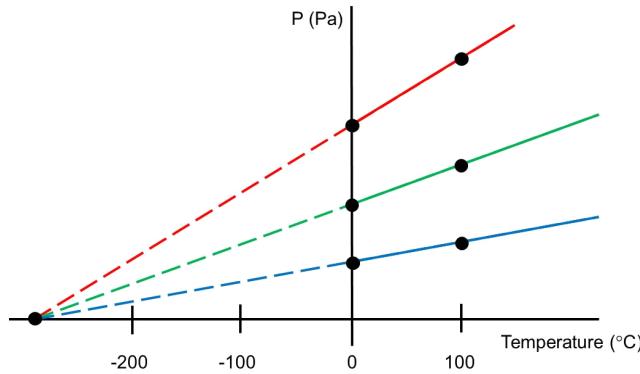
We wish to define a temperature scale that is based on a physical zero point. To do this we need a better thermometer. So here is one in the figure below.



This device uses the rise in pressure at a constant volume (the number 3 property from our list at the beginning of the lecture). It works by placing the gas bulb in thermal contact with the object who's temperature is to be measured (reservoir *B* in the figure²). A scale (like a ruler) is placed in the apparatus such that the mercury in reservoir *A* on the left is at the zero point. We could take water at its ice point, for example, to find this first part of the scale. Then the bulb is placed in thermal contact with another object by swapping out what is in reservoir *B*, say we use water at its steam point this time. The mercury will move because the pressure increases in the bulb, and the gas tries to expand. We add mercury into reservoir *A* (just when you thought you were done with U-tubes!) until the level of reservoir is again at the zero point. The height of the additional mercury we added is proportional to the change in pressure, as we now know. The Pressure for both measurements is calculated and plotted against the temperature and a linear curve is drawn between the two pressure-temperature points. This curve serves as a calibration for other temperatures. We could fill the bulb with different gasses. If we do this, we get different sloped lines. You can see this in the next figure.

But think, it seems like with this new thermometer design we should have zero temperature if we have zero gas in the bulb. We have different lines for each gas, but if we then extend these lines back to where $P = 0$ (so that we have no gas), we find they intersect with $P = 0$ at $T = 273.15^\circ\text{C}$. This is the basis we need for an absolute temperature scale! We have found a place where the temperature really should be zero (because we have no gas to have a temperature!).

²A reservoir is just a container of something.



We will take this common zero pressure point, $T = 273.15^\circ\text{C}$ as our zero point for our improved temperature scale, but we need another point. We will choose the triple point of water, the one point where water, ice and steam exist in equilibrium; this happens at 0.01°C and $4.58 \text{ mmHg} = 610.61 \text{ Pa}$. We will call this point 273.15 on our new scale which seems kind of random. But think, with this choice each unit of temperature will then be $1/273.15^{\text{th}}$ of the distance between our zero point and the water triple point. This way the degrees in this scale are the same size as the degrees in the Celsius scale (which is kind of nice). But this new scale that starts at the physical zero point we will call the Kelvin Temperature scale and the units are Kelvins (K). This is the temperature system we really should use.

29.1.3 Temperature scale conversions

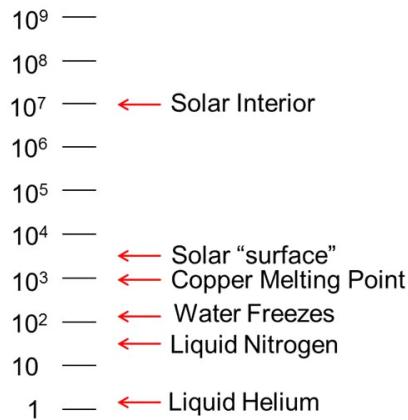
We will need some conversion factors so you can take temperatures from the Celsius and Fahrenheit scales to the Kelvin scale.

$$T_F = \frac{9}{5}T_C + 32^\circ\text{F} \quad (29.1)$$

$$\Delta T_C = \Delta T_K = \frac{5}{9}\Delta T_F \quad (29.2)$$

$$T_C = T_K - 273.15 \quad (29.3)$$

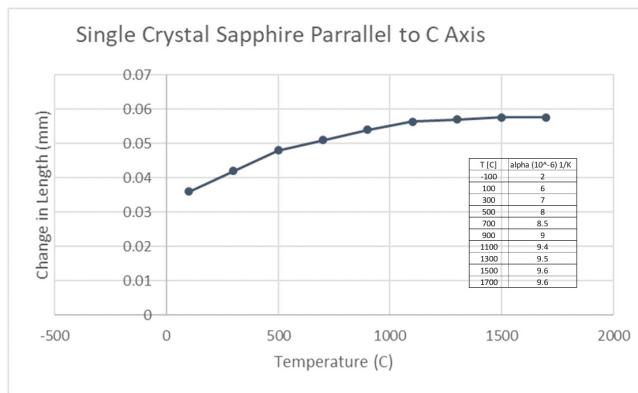
Since the Kelvin scale is not the scale your doctor or your meteorologist uses, we should try to get some feeling for the Kelvin scale. Here are some interesting temperatures placed on the Kelvin scale.



29.2 Thermal Expansion

Ring and Ball
Demo

We still only kind of know what temperature is. But we can now measure temperature, so suppose we increase the temperature of a substance. We could take the mercury in a thermometer, for example. When the temperature is higher, the mercury takes up more space. It expands. This isn't just true for liquids. Here is the change in length for a laser crystal as a function of temperature.



Notice that the change in length makes a curve that bends. We would prefer to make our thermometers out of a material that has a linear expansion. That is why we used mercury in thermometers for many years. Mercury has expansion

that is linear in temperature for many situations. When the expansion length is much less than the original dimensions of the object (L_i), we can consider the expansion to be linear in the change in temperature ΔT . Of course we could increase the temperature above this nice linear region, and we could get interesting effects like melting or exploding. But for this lecture we will limit our discussion to a linear change in length with temperature for now. Let's see how this linear region of expansion works.

29.2.1 Expansion Coefficient

What does it mean to have a linear change in length with temperature?

$$L_f = L_i + \alpha L_i \Delta T \quad (29.4)$$

It means that the final length of the object is equal to the initial length of the object plus a little bit more that depends on ΔT and not ΔT^2 or ΔT^3 or some other function of ΔT . We will get straight line plots instead of curved plots as T changes. You can see that this is the case for the equation above for thermal expansion. We can write it next to an equation for a straight line to show the linear nature of the equation

$$\begin{aligned} y &= b + mx \\ L_f &= L_i + \alpha L_i \Delta T \end{aligned}$$

If we plot L_f vs. ΔT this would make a straight line. The slope of the line would be αL_i so the amount of expansion depends on how big the object was to begin with. It also depends on what material we have. That is what the α tells us. Different substances expand at different rates even if they have the same initial length. So if you have a metal cavity of length L_i and fill it with a laser crystal of length L_i and heat them both up, you may find the crystal expands at a different rate than the metal fasteners when the temperature rises—and may need to buy a replacement for the shattered crystal.³. The coefficient α is different for every substance, because every substance expands a little differently than other substances. We should give α a name. It is called the *average coefficient of linear expansion* where from equation (29.4) we can see that

$$\begin{aligned} \alpha &\equiv \frac{\frac{\Delta L}{L_i}}{\Delta T} \\ &= \frac{\Delta L}{L_i \Delta T} \end{aligned}$$

where ΔL is the change in length due to a temperature change ΔT . And of course we already know L_i is the original length of the sample.

As long as ΔT is “not too big” α is constant. We can write this as

$$L_f - L_i = \alpha L_i (T_f - T_i)$$

³It was a very sad day when this happened to me personally.

where the subscript f means final and i means initial as usual. The units of α are inverse temperature (unfortunately it is often in $^{\circ}\text{C}$, but we would prefer to use K). Laboratories publish tables of average coefficients of linear expansion. A few are given in the table below

Material	Average $\alpha (\text{K}^{-1})$
Aluminum	25×10^{-6}
Copper	17×10^{-6}
Gold	14×10^{-6}
Lead	29×10^{-6}
Steel	11×10^{-6}
Brass	18.7×10^{-6}
quartz	0.4×10^{-6}
Glass	9×10^{-6}
Concrete	12×10^{-6}

Suppose we increase the temperature of something that has a hole in it. Does the hole increase or decrease in size? A cavity in a piece of material expands in the same way as if the cavity were filled with the material. Let's see that this must be true.

29.2.2 Volume expansion

Don't do this in
class

suppose we have a cube with sides of length L_i . What happens when we increase the cube's temperature? Each side will become larger. So what happened to the volume?

$$\begin{aligned} V_i + \Delta V &= (L_i + \Delta L)(L_i + \Delta L)(L_i + \Delta L) \\ &= (L_i + \alpha L_i \Delta T)(L_i + \alpha L_i \Delta T)(L_i + \alpha L_i \Delta T) \\ &= L_i^3 (1 + \alpha \Delta T)(1 + \alpha \Delta T)(1 + \alpha \Delta T) \\ &= L_i^3 (\Delta T^3 \alpha^3 + 3\Delta T^2 \alpha^2 + 3\Delta T \alpha + 1) \\ &= V_i (\Delta T^3 \alpha^3 + 3\Delta T^2 \alpha^2 + 3\Delta T \alpha + 1) \end{aligned}$$

Now let's divide both sides by V_i

$$\begin{aligned} 1 + \frac{\Delta V}{V_i} &= \Delta T^3 \alpha^3 + 3\Delta T^2 \alpha^2 + 3\Delta T \alpha + 1 \\ \frac{\Delta V}{V_i} &= \Delta T^3 \alpha^3 + 3\Delta T^2 \alpha^2 + 3\Delta T \alpha \end{aligned}$$

and let's make an approximation. For normal conditions, $\alpha \Delta T \ll 1$ (good for $T < 373.15 \text{ K}$ (100°C)). Then $\alpha^2 \Delta T^2$ and $\alpha^3 \Delta T^3$ are really really small. Let's ignore them.

$$\frac{\Delta V}{V_i} \approx 3\alpha \Delta T \quad (29.5)$$

We can write this as

$$\begin{aligned}\Delta V &\approx 3\alpha V_i \Delta T \\ &= \beta V_i \Delta T\end{aligned}\quad (29.6)$$

where β is called the *average coefficient of volume expansion*.

We could do the same for a two dimensional metal plate (in the approximation that a plate is 2D!)

$$\Delta A = 2\alpha A_i \Delta T \quad (29.7)$$

We need to be careful that we don't use these linear expansion equations outside of their valid range. For example, suppose we build a satellite and launch it into a low Earth orbit (LEO). The space environment for a LEO orbit (sun side to shade side) has a very large ΔT .

$$\begin{aligned}T_{\text{sun}} &= 120^\circ\text{C} = 393.15 \text{ K} \\ T_{\text{shade}} &= -100^\circ\text{C} = 173.15 \text{ K}\end{aligned}\quad (29.8)$$

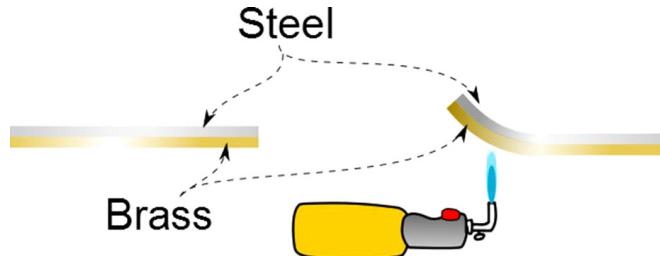
This ΔT range is too large for most materials to still be linear. Special low α materials are used to allow satellite structures to survive

But how about on Earth? Even on Earth we can have extremes

$$\begin{aligned}T_{\max} &= 57.6^\circ\text{C} = 330.75 \text{ K} \\ T_{\min} &= -89^\circ\text{C} = 184.15 \text{ K}\end{aligned}\quad (29.9)$$

This ΔT range would be a challenge to cover with most materials.⁴ So if you are going to build a device or instrument that will travel around the globe, you need to use special low α materials. Often making these materials is very expensive and the process is a tightly held industrial secret. Alternately you can air condition or heat your system. But you do have to do something to stay in the linear range.⁵

Looking at table of average linear expansion coefficients we find see different α values for different materials. Sometimes α values are quite different. Suppose we take two materials with different α 's and put them together.



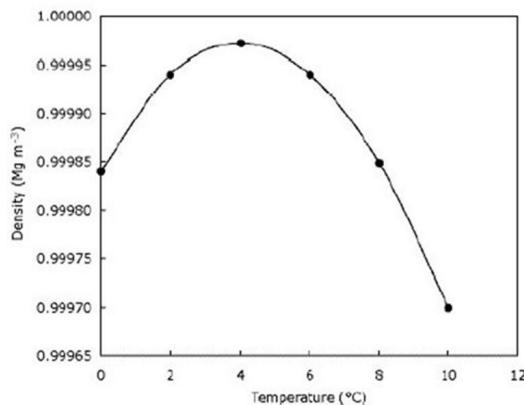
⁴ Especially metal structures and laser crystals. >sigh<

⁵ I am told by the BYU-I facilities people that the BYUIC has one huge beam across the entire auditorium that has to stay heated or it shrinks enough to let the balcony fall off the walls!

The figure shows a bimetallic strip. What would happen if we heated this bimetallic strip? The two metals will expand at different rates, forcing the bonded pair of metals to bend. And we could make a temperature measuring device that measures how much this bimetallic strip bends.

29.2.3 Water is weird (and the Lord knows what He is doing!)

Question 123.9.2



Density of water near 0 °C. (Public Domain image PD-1923, PD-Australia)

For most substances the volume increases with increasing temperature. Usually liquids increase in volume more than solids. But water is different. Look at the graph of density vs. Temperature.

Just as ice melts, it starts to *increase* in density, but then it becomes less dense with increased temperature after about 277.15 K (4 °C). We expect the behavior we see after 277.15 K (4 °C). The density is decreasing because we have the same number of water molecules, but the temperature change has increased the volume. So the density goes down.

But from 273.15 K (0 °C) to 277.15 K (4 °C) we see the volume decreases with ΔT . This is very odd. But very cool! This is why a lake freezes from the top down. The colder water near 273.15 K (0 °C) is less dense, and so before it freezes it floats to the surface. This is great, because if it became more dense just before it froze, the water would sink to the bottom and freeze, leaving the liquid water exposed to the cold atmosphere to freeze more water. The lake would fill up from the bottom with solid ice, killing fish and water life.



Lake Erie from Space. Notice the ice covering part of the lake. (Public Domain Image courtesy NASA Visible Earth,
http://visibleearth.nasa.gov/view_rec.php?id=1286

Because the water becomes less dense before freezing we don't freeze large lakes and oceans solid. And that keeps life going on Earth!

We have made progress. But we still don't know what temperature is. We know some things about what temperature does. And this can give us a clue. Temperature is making things move by expansion. And we know that there must be forces behind making things move. Forces do work. Work is energy.

Could it be that temperature has to do with exchanging energy? Question 123.9.2
Question 123.9.3

Chapter 30

30 Thermal Properties of Matter 2.1.4, 2.1.5

In this lecture we want to make our study of temperature more mathematical. And we want to consider how energy relates to temperature. The problem is that although we have a strategy for finding w , the work done as we change temperature. We will start to fill that void in this lecture.

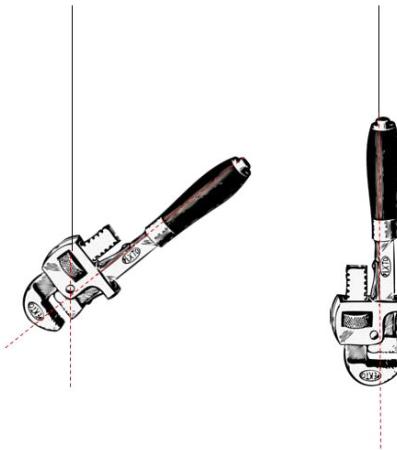
We learned in PH 121 what energy was. We studied work and potential energy, and we studied kinetic energy. But these were always the energy *of* an object. We described these energies in terms of the motion of the center of mass of an object using the particle model. Now we know that there is energy *in* an object. This is different. Particles don't have internal parts, but real things do. Our particle model was an approximation. So we need to go beyond the particle model here. We will call the energy associated with the inside parts of an object *internal energy*. And this internal energy is associated with thermal energy.

To study this, Let's refine our definition of internal energy by saying that it is all the energy of a system that is associated with its microscopic components – atoms and molecules – when viewed from a reference frame at rest with respect to the center of mass of the system. Any motion of the center of mass is our Principles of Physics I (PH121) mechanical energy. Now we want the motion of the parts of the object with respect to the center of mass. Let's review center of mass briefly.

From Principles of Physics I we know the center of mass is given by

$$r_{cm} = \frac{1}{M} \int \mathbf{r} dm \quad (30.1)$$

and we can think of the center of mass as the point an object freely spins around.



For groups of molecules, we remember

$$\vec{r}_{cm} = \frac{\sum_i m_i \vec{r}_i}{M} \quad (30.2)$$

where

$$M = \sum_i m_i \quad (30.3)$$

is the total mass of all the molecules.

You might remember this in one dimension for two masses

$$x_{cm} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \quad (30.4)$$

Now suppose we let all of the masses involved be tied together with spring-like forces. These spring-like forces would allow the masses to oscillate a little. That would change the \vec{r}_i involved in our center of mass calculation. But if the oscillations were random, the actual center of mass would not change. This is the kind of motion involved in internal energy.

By making all the little molecules move, we have changed the energy state of the object, but we have not moved it, so the mechanical energy has not changed. The energy we have described is the thermal internal energy (there is still the nuclear internal energy, but that is a subject for the class called Modern Physics (PH279)). Let's consider forms of internal energy in more detail in this lecture.

Fundamental Concepts

1. Heat capacity relates energy transfer by heat to temperature change in solids and liquids
2. Specific heat is a heat capacity per unit mass

3. A phase change is changing from solid to liquid, or from liquid to gas for a sample of a material
4. Calorimetry is the use of thermal conservation of energy to study a sample of material

30.1 Heat

Temperature, in our intuitive model, seems connected to heat. But what is heat? If we can figure this out we might be able to give a better definition of temperature. Let's give heat a symbol, Q , but in this lecture we want to give a specific definition of heat.

Heat is defined as the transfer of energy across the boundary of a system due to a temperature difference between the system and its surroundings.

When we "heat" a substance, we add energy from its surroundings. When the temperature changes, it is energy that is making the temperature change. An object could gain or lose energy *by heat*. We will use the word "heat" for the transfer of energy no matter which direction the energy is flowing. This is a little like using the word "acceleration" no matter whether we are speeding up or slowing down.

In the dim past, scientists thought heat was a fluid (called caloric). This model of heat is not correct (science changes!). We now define heat as a transfer of energy. But because of this history, we have some left-over names that are not very descriptive of the modern ideas they represent. Examples are "latent heat" and "heat capacity" (we will define these soon).

The thing to remember is that a system *has* a temperature but it can *give or receive* energy by heat.

30.1.1 Units

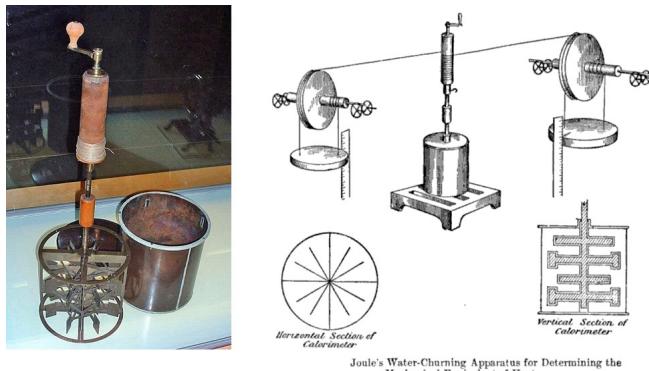
There are several units for heat. The calorie (cal) is the amount of energy transfer necessary to raise the temperature of 1 g of water from 14.5 °C (287.65 K) to 15.5 °C (288.65 K). This is somewhat arbitrary, but it works for many things, especially biological systems. The British thermal unit (Btu) is the unit used by refrigeration and heating contractors. One Btu is the amount of energy transfer necessary to raise the temperature of 1 lb of water from 63 °F (290.37 K) to 64 °F (290.93 K). This is an enormous amount of energy being transferred! If you wish to heat an entire building, this is a fine unit, but for us it is a bit large. In the SI system we already have a unit for energy, the Joule. We will try to stick to this nice, medium sized, unit. But let's see where it came from.

30.1.2 Mechanical equivalent of heat



James Prescott Joule. (Image in the Public Domain)

The name “Joule” is a person’s name, Joule was an early researcher and is credited with showing that heat transfer is a transfer of energy. Joule was able to build a device to prove that mechanical energy could be converted to heat energy. His device is shown in the picture.



Joule Apparatus (Image in the Public Domain courtesy Gaius Cornelius)

The paddle device was placed in the drum that is sitting behind it. The crank was operated by strings attached to it that were pulled by masses tied to the other end of the string (see the right hand side of the figure). When the masses were released, the paddle wheel stirred a liquid. The friction from the paddle wheel stirring the liquid raised the temperature of the liquid. The masses would lose energy equal to

$$\Delta U = 2mgh \quad (30.5)$$

assuming the masses were equal. By carefully measuring the temperature of the water, Joule found that

$$\Delta U = C \Delta T \quad (30.6)$$

with the proportionality constant approximately

$$C = 4.18 \frac{\text{J}}{\text{g}^\circ\text{C}} \quad (30.7)$$

which means that 4.18 J of mechanical energy raises the temperature of 1 g of water by 1°C . (Here C is a generic symbol for a constant). We have more precise measurements now that set this value at $4.186 \frac{\text{J}}{\text{g}^\circ\text{C}}$. But you can see that Joule did very good for working with what amounts to a fancy butter churn in the 1800's!

Notice that we are considering raising the temperature of 1 g of water by 1°C . This is the basis of the unit called the calorie. Using the same 4.5°C (287.65 K) to 15.5°C (288.65 K) change in temperature, we realize

$$1 \text{ cal} = 4.186 \text{ J} \quad (30.8)$$

which gives us a convenient way to convert from calories to Jules. But more importantly, it shows that heat is really tied to energy, not to some substance. Work can raise a temperature in a system just like “heating” the system with a fire or stove.

Because of this experiment by Joule, this idea that work can be turned into thermal energy is known as the *mechanical equivalent of heat*.

30.2 State Variables

As you know, we are at the beginning of a new topic, and we will find that physics did not develop as a cohesive whole. The new topic is called *Thermodynamics* and it developed independently of the theories of oscillation and waves and optics. Thus the language we use will be different, and sometimes old words will need new meanings.

We learned before about solids, liquids, and gases. We learned that the degree of bondedness made the difference between these states. The states of matter are often called *phases* in thermodynamics. This use of the word “phase” has nothing to do with the phase of an oscillator or wave. It is just the condition of being solid, liquid or gas.

To change from solid to liquid, or from liquid to gas, is called a *phase change*.

To describe the state of a sample of matter, we use a set of variables like mass, volume, pressure, density, and energy. We have used these throughout Principles of Physics I. Thermodynamics has its own set of descriptive variables to add to

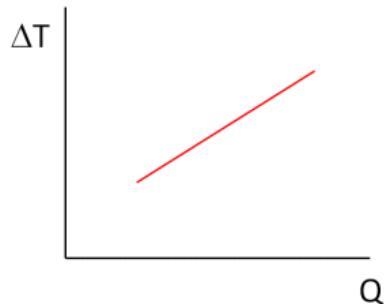
these: moles, number density, and Temperature. These variables, along with our original set, are called *state variables* and together, they describe the conditions of a particular piece of matter. We have hinted that one of these state variables will be the amount of internal energy in our sample of material, let's call that ΔE_{int} . Note here the “ Δ ” means “an amount of” and not a “difference between two amounts of” a substance. So ΔE_{int} is an amount of internal energy. We know from our study of thermal expansion that the work, w , might be such a state variable. And we also know now that energy can be transferred by heat, Q so maybe heat is a state variable.

It would be good to be able to relate these new variables. So let's try to develop a relationship between ΔE_{int} , W , and Q for our samples.

Suppose you are heating something on your stove, say, a sample of steak.



You would expect that as energy is transferred from the hot stove to the cold beef, the temperature of the steak would increase. It makes sense that the more energy you transfer, the more the temperature will rise. Gasses are more complicated, because their volume will change, but the volume of solids does not change much with temperature (we know that it changes a little) so we don't have as many complications as we would with a gas.



We could guess that the temperature increase might be linear in the amount of energy transfer we give to the solid.

$$\Delta T \propto \Delta E_{int}$$

where our only form of increasing the energy here is Q , since the change in volume of the stake sample is small enough that any work, w , is negligible. Since P and V are not changing much. So to a good approximation

$$\Delta T \propto Q$$

Experimentalists did this experiment and found that for many temperature regions this is true. Of course, we could over heat our sample



changing it's form (from stake to charcoal, and smoke), but then the sample is no longer the same substance. In this case ΔT is no longer proportional to Q . But as long as we don't change the composition of the substance (we stay in the linear regime) , we can say that ΔT is proportional to Q .

Physicists don't like proportionality signs, so they invented a constant with the right units that would make this an equality

$$\Delta T = \frac{1}{K}Q$$

But, they found that the internal structure of the solid mattered. There was a different constant, K , for every material.

They called the constant, K , the *heat capacity* of the material. The name is one of our throwbacks from the past, but for us it means the constant of proportionality between Q and ΔT that holds all the specific material properties of that substance that change the slope of the ΔT vs. Q graph.

Of course, the more material we have in our sample, the longer it takes to heat it up. If we have a roast, it takes longer to cook than a small steak. So it is customary to define a heat capacity per unit mass, this is called the *specific heat* of the material.

$$\Delta T = \frac{1}{mc}Q$$

and we often rearrange this equation to give

$$Q = mc\Delta T$$

As an example, it takes 4190 J to raise 1 kg of water by 1 K. So

$$4190 \text{ J} = (1 \text{ kg}) (c) (1 \text{ K})$$

solving for c yields

$$c_{\text{water}} = 4190 \frac{\text{J}}{\text{kg K}}$$

Which is familiar to us from Joule's work.

Here are a few more substances and their specific heat values.

Substance	$c \left(\frac{\text{J}}{\text{kg K}} \right)$	$c \left(\frac{\text{J}}{\text{mol K}} \right)$
Aluminum	900	24.3
Copper	385	24.4
Iron	449	25.1
Gold	129	25.4
Lead	128	26.5
Ice	2090	37.6
Mercury	140	28.1
Water	4190	75.4

Notice that things that heat up or cool down quickly have small specific heats. Things that heat up slowly or cool slowly have large specific heats. Water in any of its phases has large values of specific heat. This is why living next to a lake or the ocean keeps your temperatures moderate. It takes quite a lot of energy transfer to heat up or cool off the water, so the temperature of the water does not change much.

This is why Buffalo NY temperatures lower when the lake freezes. This is also the source of lake or ocean breezes. In the morning, the land changes temperature more quickly than the water. The warmer air over the land becomes less dense and rises, leaving a lower pressure area. This causes a cool breeze to form from the sea to the land.



Conversely, in a large city, the concrete and blacktop heat up much quicker, so we experience increases in daytime temperature as a city grows.

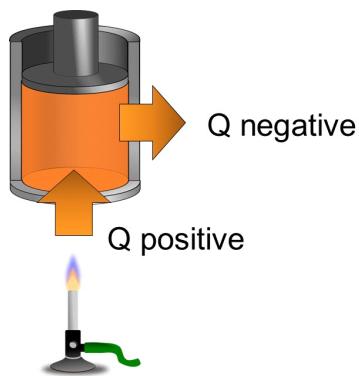
A specific heat is a per-unit-mass property, but we could just as well define specific heat as a per-unit-mole quantity. Then the specific heat would be the amount of energy it takes to increase the temperature of a mole of material. This incarnation of the idea of specific heat is called a *molar specific heat*, and we will give it the symbol C .

$$Q = nC\Delta T$$

It is worth noting that the molar heat capacities given in the last table are remarkably alike. Except for water, which is a compound, they are all about $25 \text{ J}/(\text{mol K})$. This gives us a clue towards understanding what is happening microscopically, but we will have to wait to investigate this clue until a further lecture.

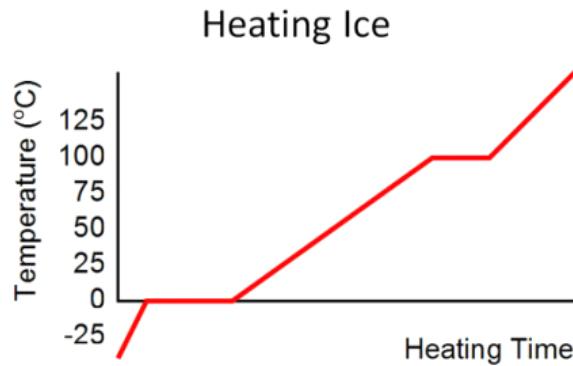
WARNING: Specific heat values are really not constant. They do vary with temperature a little. We know from our study of thermal expansion that the volume of a sample of a material will change a small amount as it is heated or cooled. After all, this is one we found to measure temperature! There is a small amount of work done in changing the volume of the material. Not all of the energy will go into changing the temperature. So our heat capacity will change a little as the volume changes. But if the change in temperature is not too big, we can treat heat capacities for solids and liquids as constants. Oh, but for gasses...we know the volume might change quite a lot as the temperature changes for a gas, so we will find that this analysis needs a few changes for gasses. We will study this is a later lecture.

We should make a sign convention about our energy transfer by heat. Let's say that if energy is added by heat, we will say Q is positive. If the energy leaves by heat, we will say that Q is negative.



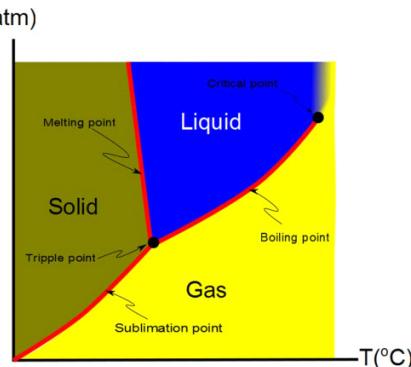
30.3 Phase Changes and Phase Diagrams

Let's consider heating ice. You take a chunk of ice from outside on a cold February day in Rexburg, Idaho. The ice has a temperature of about -25°C .¹ You place it in a pan on your stove and heat it up. Then, being a physics student, you plot the temperature as a function of time. Here is what you get.



Notice the places where you are still transferring energy by heat, but the temperature does not change. These are phase changes. A phase change is like melting and boiling. In our chart, the first is melting, the second is boiling.

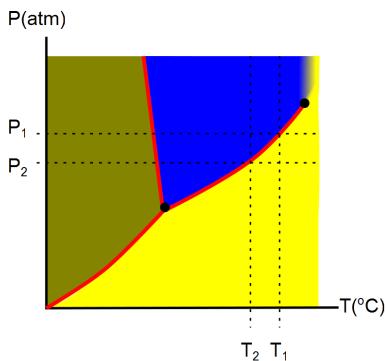
One of the things you will notice is that the melted ice does not boil at 100°C here in Rexburg. This is because we are up so high that the air pressure is less than $1\text{ atm} = 1.013 \times 10^5 \text{ Pa}$. In the next graph, I have plotted the melting, and boiling point of a substance as a function of pressure. The melting point and boiling point are not single points, but lines on this graph. This is because melting and boiling points depend on pressure. There is also a line for the sublimation point. Along this line, a solid will change directly into a gas. Dry ice does this at normal room temperature and pressure.



If we observe the boiling line, we see that as pressure goes down, the boiling

¹OK, that might be an exaggeration, it only occasionally gets that cold in Rexburg.

temperature goes down. Look at T_1 and T_2 in the next figure, where $T_2 < T_1$. Note that $P_2 < P_1$.

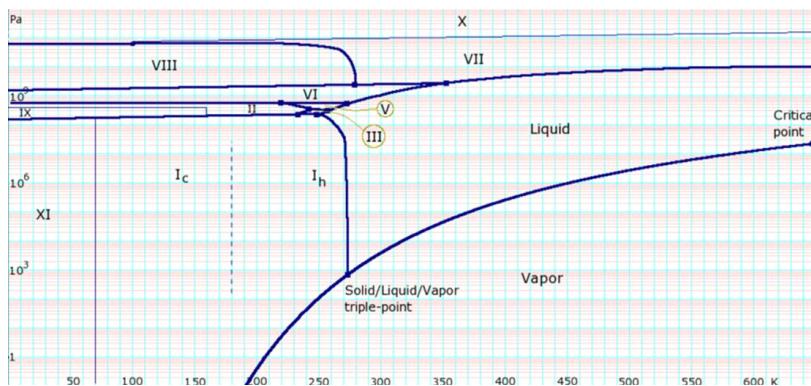


This agrees with our Rexburg boiling point experience.² We can see from the graph that melting or boiling depends on pressure, and so does sublimation point.

The triple point is a special point where all three states are possible at the same pressure and temperature. Ice, steam, and water will all exist in equilibrium, for example, at the triple point of water.

The critical point is where the clear distinction between liquid and gas stops. beyond this boiling the material is in a fluid state, but not clearly liquid or gas. Along each of the lines in our figure, we can say that the material is in phase equilibrium, meaning that both phases from either side of the line are possible and equally probable. Only at the triple point are all three phases equally probable.

Our diagram is rather schematic, and this is on purpose. An actual phase diagram for water is given in the next figure.



More Complete Phase Diagram for Water (Public Domain Image courtesy Karlhahn)

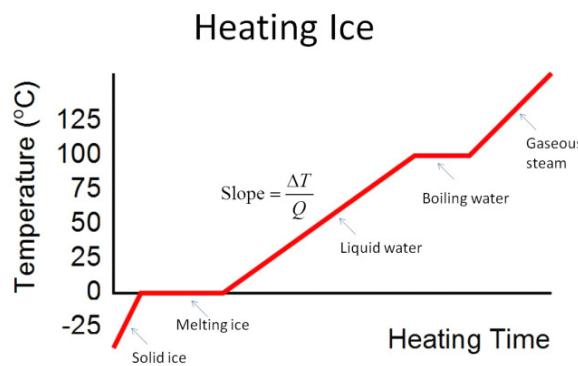
²So if we have a water boil alert here, you have to boil your water longer in Rexburg than you would in NY City in order to make it safe to drink.

You can see there are more phases for water than our simple model of solid, liquid, and gas give.

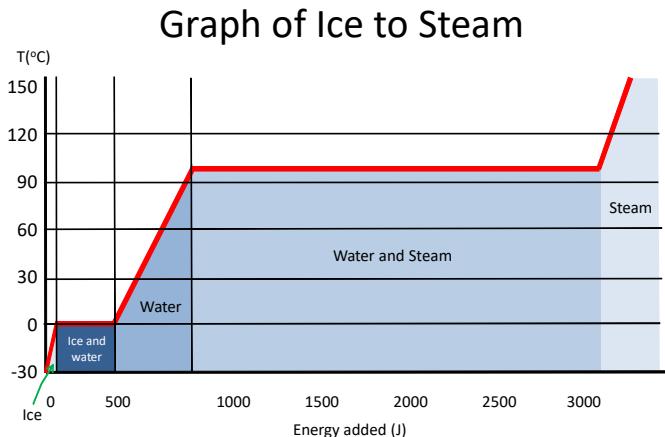
But why do we care about phase changes? Because it takes energy to make a phase change happen. Phase changes direct energy into breaking molecular bonds instead of increasing temperature. So in our graph of temperature vs. heating time the flat parts are where bonds are being broken and a phase change is happening.

30.4 Phase changes and heat of transformation

Recall the graph of temperature vs. time for heating a chunk of Rexburg ice.



Here is the graph again, but this time the horizontal axis is labeled “Energy added” because we realize that we are adding internal energy as we wait and watch the ice.



We see regions of the graph where the ice changes temperature, and we can

understand that during these times the temperature gain will be proportional to the amount of energy transferred by heat.

$$\text{Slope} = \frac{\Delta T}{Q} = \frac{1}{Mc}$$

But what about the horizontal parts of the graph? We identified these before as phase changes. The slope is zero during the melting and boiling times.

We have an amount of energy transferred by heat. You might guess that how much energy it takes for the phase change to completely change from one state to the other depends on the strength of the bonds of the material. The added energy is going into breaking those bonds. So the amount of energy it takes to, say, melt different substances will be different. From experience we know that it takes more energy to heat large amounts of material than small amounts of material. We can express this mathematically as

$$Q = \pm LM$$

where L is the symbol we give to the constant that expresses the how easily the material melts or freezes. We need at least two of these, one for melting/freezing and one for boiling/condensing. Then

$$\begin{aligned} Q &= \pm L_f M && \text{melt/freeze} \\ Q &= \pm L_v M && \text{boil/condense} \end{aligned}$$

where we call these two constants

$$\begin{aligned} L_f &\quad \text{heat of fusion} \\ L_v &\quad \text{heat of vaporization} \end{aligned}$$

generically we call each of these a *heat of transformation* but generally they are still referred to by their old title of “*Latent heat*.”

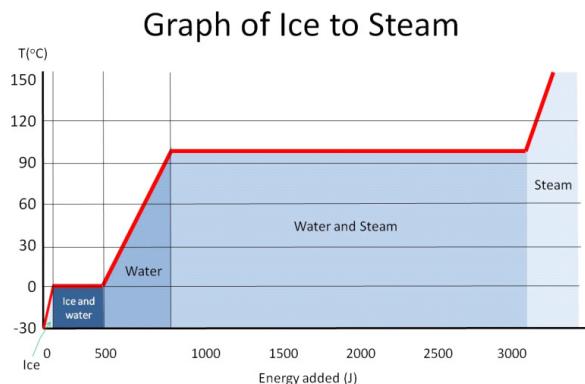
Substance	T_m (°C)	L_f ($\frac{\text{J}}{\text{kg}}$)	T_m (°C)	L_v ($\frac{\text{J}}{\text{kg}}$)
Nitrogen (N_2)	-210	0.26×10^5	-196	1.99×10^5
Ethyl alcohol	-114	1.09×10^5	78	8.79×10^5
Mercury	-39	0.11×10^5	357	2.96×10^5
Water	0	3.33×10^5	100	22.6×10^5
Lead	328	0.25×10^5	1750	8.58×10^5

Notice that our equations each have a “ \pm ” sign in them. We must supply the sign by context. We need a sign convention like back when we were studying optics!

Let's choose that energy that leaves is negative and energy that is gained by the system is positive. Also notice that so long as we warm the ice somewhat slowly, that the ice completely melts before the temperature of the water starts to rise.

Let's do a problem. Suppose we want to increase the temperature of 1 g = 0.001 kg of ice at -30.0°C and change it to steam at 120.0°C . How much energy transfer by heat is required to do this?

A graph of temperature vs. heat energy is shown. Let's take it one piece at a time.



Part A: Warming up the ice to 0°C .

This part acts as we learned in the last section. We have

$$\begin{aligned} Q &= m_i c_i \Delta T \\ &= (0.001 \text{ kg}) \left(2090 \frac{\text{J}}{\text{kg}^{\circ}\text{C}} \right) (0^{\circ}\text{C} + 30^{\circ}\text{C}) \\ &= 62.7 \text{ J} \end{aligned}$$

We can get away with using $^{\circ}\text{C}$ because we only need ΔT and the divisions of the Celsius and Kelvin scales are the same size.

Part B: Melting ice

Now the ice changes to water. We have a heat of transformation. To find out what happens we use the equation

$$\begin{aligned} Q &= m_i L_f \\ &= (0.001 \text{ kg}) \left(3.33 \times 10^5 \frac{\text{J}}{\text{kg}} \right) \\ &= 333.0 \text{ J} \end{aligned}$$

The total energy so far is $62.7 \text{ J} + 333.0 \text{ J} = 392.7 \text{ J}$. Note that during Part B the temperature didn't go up!

Part C: Warming the melted ice (water)

Again we have the normal case where we can use the specific heat of water (now that the ice has melted)

$$\begin{aligned}
 Q &= m_w c_w \Delta T \\
 &= (1 \text{ g}) \left(4190 \frac{\text{J}}{\text{kg}^\circ\text{C}} \right) (100^\circ\text{C} - 0^\circ\text{C}) \\
 &= 419.0 \text{ J}
 \end{aligned}$$

The total energy so far is $62.7 \text{ J} + 333.0 \text{ J} + 419.0 \text{ J} = 814.7 \text{ J}$

Part D: Boiling the melted ice (water)

Now we convert the water to steam.

$$\begin{aligned}
 Q &= m_w L_v \\
 &= (0.001 \text{ kg}) \left(2.26 \times 10^6 \frac{\text{J}}{\text{kg}} \right) \\
 &= 2260.0 \text{ J}
 \end{aligned}$$

The total energy so far is $62.7 \text{ J} + 333.0 \text{ J} + 419.0 \text{ J} + 2260.0 \text{ J} = 3074.7 \text{ J}$

Part E: Warming the steam

Now we have steam, and can use the specific heat of steam (we don't usually do this! we usually use the ideal gas law when it becomes a gas!)

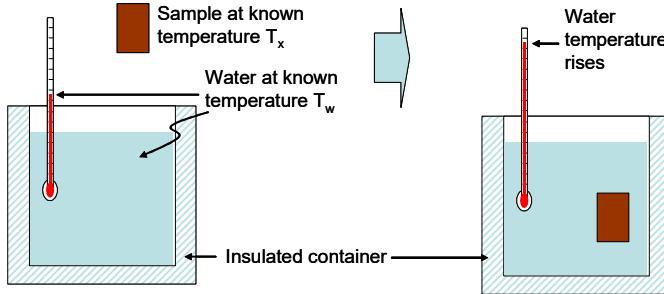
$$\begin{aligned}
 Q &= m_s c_s \Delta T \\
 &= (0.001 \text{ kg}) \left(2010 \frac{\text{J}}{\text{kg}^\circ\text{C}} \right) (120^\circ\text{C} - 100^\circ\text{C}) \\
 &= 40.2 \text{ J}
 \end{aligned}$$

The total energy transfer required is $62.7 \text{ J} + 333.0 \text{ J} + 419 \text{ J} + 2260.0 \text{ J} + 40.2 \text{ J} = 3114.9 \text{ J}$

30.5 Calorimetry

We can now calculate results for many common thermal experiences. We have all experienced eating something with a large temperature. The solution is to quickly drink something cold. This lowers the temperature. We would like to be able to quantify this. But our mouth is not a great device to use to do quantitative analysis on temperature changes. Let's consider a more suitable apparatus.

Suppose we have an insulated container with a thermometer. We can call this apparatus a *calorimeter*.



Inside of this container let's place a known mass of water and measure the temperature, T_w . Then we can introduce our sample of hot material. Suppose we know that the hot material has a temperature T_x .

Let's assume that the container walls are perfectly insulating, then knowing the sample temperature T_x and the water temperature T_w with $T_w < T_x$ we can find c or K (or even C) for our material. We recognize for this ideal system no energy can leave, because of the insulation. So

$$Q_{\text{cold}} = -Q_{\text{hot}} \quad (30.9)$$

or energy would not be conserved. Then for the water

$$Q_w = m_w c_w (T_f - T_w)$$

and for the sample

$$Q_x = m_x c_x (T_f - T_x)$$

so, using energy conservation

$$\begin{aligned} -Q_x &= Q_w \\ -m_x c_x (T_f - T_x) &= m_w c_w (T_f - T_w) \end{aligned}$$

and we can solve for c_x

$$c_x = -\frac{m_w c_w (T_f - T_w)}{m_x (T_f - T_x)} \quad (30.10)$$

Note the sign convention! If energy leaves the object, the value of Q is negative. Our denominator is, indeed, negative because for the sample T_f is less than T_x . We could use our value of c_x to identify our material by looking it up in a table of specific heat values.

Let's try a second problem. Suppose we have 2 kg of ice at 0 °C and 1 kg of water at 50 °C. What is the final temperature? This time we could have a phase change, we will need to be careful. We can look up the heat of transformation and the specific heat of ice and water:

$$\begin{aligned} L_f &= 3.33 \times 10^5 \frac{\text{J}}{\text{kg}} \\ c_w &= 4190 \frac{\text{J}}{\text{kg K}} \end{aligned}$$

$$\begin{aligned}m_i &= 2 \text{ kg} \\m_w &= 1 \text{ kg} \\T_{ii} &= 0^\circ\text{C} \\T_{iw} &= 50^\circ\text{C}\end{aligned}$$

We still don't expect to lose any energy, so

$$Q_{\text{cold}} = -Q_{\text{hot}}$$

We identify the hot thing as the water this time and the cold thing as the ice. The water only changes temperature, but the ice experiences a phase change. We need to account for all of the energy transferred so for this ice

$$Q_{\text{ice}} = Q_{\text{cold}} = Q_{\text{melt}} + Q_{\text{heat up}}$$

Then we can put this into our conservation of energy equation

$$\begin{aligned}Q_{\text{ice}} &= -Q_{\text{water}} \\Q_{\text{melt}} + Q_{\text{heat up}} &= -Q_{\text{water}} \\L_f m_i + m_i c_w (T_f - T_{ii}) &= -m_w c_w (T_f - T_{iw})\end{aligned}$$

The final temperature will be the same for both. To find that final temperature we have a little algebra to do. Let's distribute on both sides of our equation.

$$L_f m_i + m_i c_w (T_f - T_{ii}) = -m_w c_w (T_f - T_{iw})$$

$$L_f m_i + m_i c_w T_f - m_i c_w T_{ii} = -m_w c_w T_f + m_w c_w T_{iw}$$

$$L_f m_i - m_i c_w T_{ii} - m_w c_w T_{iw} = -m_w c_w T_f - m_i c_w T_f$$

$$L_f m_i - m_i c_w T_{ii} - m_w c_w T_{iw} = -(m_w c_w + m_i c_w) T_f$$

$$\begin{aligned}\frac{L_f m_i - m_i c_w T_{ii} - m_w c_w T_{iw}}{-(m_w c_w + m_i c_w)} &= T_f \\T_f &= \frac{m_i c_w T_{ii} + m_w c_w T_{iw} - L_f m_i}{(m_w c_w + m_i c_w)}\end{aligned}$$

Really we should do this in kelvin.

$$\begin{aligned}T_f &= \frac{(2 \text{ kg}) \left(4190 \frac{\text{J}}{\text{kg K}} \right) (273 \text{ K}) + (1 \text{ kg}) \left(4190 \frac{\text{J}}{\text{kg K}} \right) (273 + 50) \text{ K} - \left(3.33 \times 10^5 \frac{\text{J}}{\text{kg}} \right) (2 \text{ kg})}{\left((1 \text{ kg}) \left(4190 \frac{\text{J}}{\text{kg K}} \right) + (2 \text{ kg}) \left(4190 \frac{\text{J}}{\text{kg K}} \right) \right)} \\&= 236.68 \text{ K} \\&= -36.47^\circ\text{C}\end{aligned}$$

Now let's ask, is this reasonable?

The answer is no! We can't end up with a temperature less than the two temperatures we started with. Let's see why this went wrong.

We should have checked in advance to see if all the ice melts. It would take

$$Q = m_i L_f = (2 \text{ kg}) \left(3.33 \times 10^5 \frac{\text{J}}{\text{kg}} \right) = 6.66 \times 10^5 \text{ J}$$

to melt all the ice. Since the water in the calorimeter is losing energy, we expect that it will have a lower temperature. But it is unlikely that it will all freeze, so the lowest temperature for the water would be 0°C . That means the available energy transfer from the water is just

$$\begin{aligned} Q &= m_w C_w (T_0 - T_{wi}) \\ &= (1 \text{ kg}) \left(4190 \frac{\text{J}}{\text{kg K}} \right) (273 \text{ K} - 323.15 \text{ K}) \\ &= -2.095 \times 10^5 \text{ J} \end{aligned}$$

That is, the warmer water can provide $2.095 \times 10^5 \text{ J}$. But since $2.095 \times 10^5 \text{ J} < 6.66 \times 10^5 \text{ J}$ the warm water does not have enough energy available to melt all the ice. Once the water is all at 0°C , the ice and water are in thermal equilibrium. No energy will be transferred. So our final temperature of the mixture is 273 K or 0°C .

Of course if our system of ice and water isn't isolated, eventually all the ice will melt. But that would be due to a transfer of energy by heat from the outside environment.

This problem is a bit of a trick, but such a situation can really happen. So it is really important that we stop to consider whether our numerical answers are reasonable once we have done a calorimetry problem.

We have filled in the missing piece of information for solids and liquids. We can say that so long as we don't change the structure of the material,

$$Q = Mc\Delta T$$

or

$$Q = nC\Delta T$$

so we can find Q , w , and ΔE_{int} for solids and liquids. But we said gasses will be more difficult. We will take up the relationship between energy transfer by heat and the change in temperature for gasses in our next lecture.

Chapter 31

31 Heat Transfer and Gas Laws 2.1.6, 2.2.1

Now that we understand heat as a transfer of energy, we should look at how we can move energy around. We study this in this lecture. We will also develop a relationship between pressure, volume, and temperature for a gas. You probably know this relationship from your high school science classes. But we will look at how we got it and when it is valid.

Fundamental Concepts

- Conduction
- Convection
- Radiation and Stefan's law
- The Ideal Gas Law
- The Van der Waals Gas Law

31.1 Heat Transfer Mechanisms

In Principles of Physics I we found we could change the energy state of a system by work, w . This seems likely for internal energy as well. And we already know about Q , but how does energy transfer by heat happen?

Let's start by thinking about how to change E_{int} . The change in internal energy is the sum of all the ways we can make E_{int} become different. So far we have

$$\Delta E_{int} = Q + w$$

Work we know a lot about, but let's become more familiar with Q . We may talk about many different ways to transfer energy, then

$$\Delta E_{int} = \sum_i \Delta E_{\text{Transfer mechanism } i} \quad (31.1)$$

where work is one of those energy transfer mechanisms and there may be many different Q transfer types. Let's list a few kinds of Q .

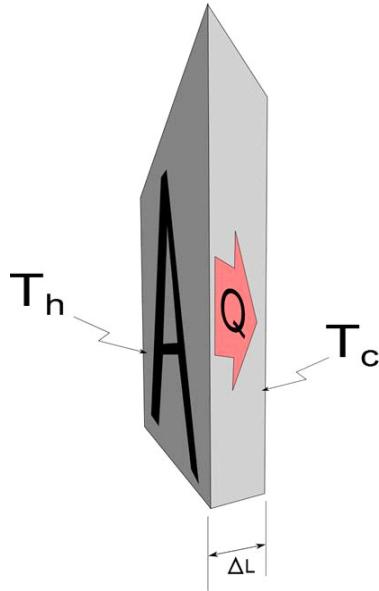
31.1.1 Thermal Conduction

This is what we have been calling heat up to this point. We define it a little more carefully as the exchange of kinetic energy between microscopic particles (molecules and atoms) through direct contact. Really we mean collisions between the atoms and molecules of the two objects.



When we add energy by heat to part of a solid, the atoms are locally displaced from their equilibrium locations. These atoms knock into their neighbors, and cause them to be displaced (they increase their kinetic energy). This process continues through the whole solid.

The rate of heat transfer will depend on the properties of the atoms that make up the substance.



We can envision a slab shaped piece of solid as in the figure. It has thickness ΔL and cross-sectional area A . Suppose one face of the slab is at T_c and the other at T_h . We find that energy Q transfers in a time interval Δt from the hotter face to the colder one. The rate

$$\mathcal{P} = \frac{E}{\Delta t} = \frac{Q}{\Delta t}$$

is proportional to A and ΔT .

$$\mathcal{P} = \frac{Q}{\Delta t} \propto A \frac{\Delta T}{\Delta L}$$

We can take ΔL very small

$$\mathcal{P} = k_{therm} A \left| \frac{dT}{dL} \right| \quad (31.2)$$

where we now have another constant k ! This k_{therm} is the *thermal conductivity* of the material and we define $\left| \frac{dT}{dL} \right|$ as the *temperature gradient*.

A gradient is a rate of change, so this gradient is the rate of change of the temperature with respect to position.

We can say that our slab is a one dimensional problem, because the energy transfer is only in one direction. If we assume that k_{therm} is not temperature dependent and the slab is uniform, then

$$\left| \frac{dT}{dL} \right| = \frac{T_h - T_c}{\Delta L}$$

and our energy transfer rate is

$$\begin{aligned}\mathcal{P}_{rod} &= \frac{Q}{\Delta t} = k_{therm} A \frac{T_h - T_c}{\Delta L} \\ &= \frac{A(T_h - T_c)}{\frac{\Delta L}{k_{therm}}}\end{aligned}\quad (31.3)$$

If we have a complicated multipart slab with n different substances we would just find the energy transferred from the first slab part and use that as the input to the next slab part. To a good approximation we could write the result as

$$\mathcal{P} \approx \frac{A(T_h - T_c)}{\sum_i^n \frac{\Delta L_i}{k_{therm_i}}} \quad (31.4)$$

The amount of energy transferred by thermal conduction is

$$Q = \mathcal{P} \Delta t \approx \frac{A(T_h - T_c) \Delta t}{\sum_i^n \frac{\Delta L_i}{k_{therm_i}}}$$

All this wasn't too bad mathematically. But notice that our slab looks a lot like a wall, and the building industry does some further simplifications to our thermal conduction equation. We define a term

$$R = \frac{\Delta L}{k} \quad (31.5)$$

as the "R-value" of a material. Then a wall or multipart solid would have

$$\mathcal{P} = \frac{Q}{\Delta t} \approx \frac{A(T_h - T_c)}{\sum_i^n R_i} \quad (31.6)$$

so that

$$Q_{conduction} \approx \frac{A(T_h - T_c) \Delta t}{\sum_i^n R_i}$$

These R-values are used in home insulation and heating design and are usually in English units

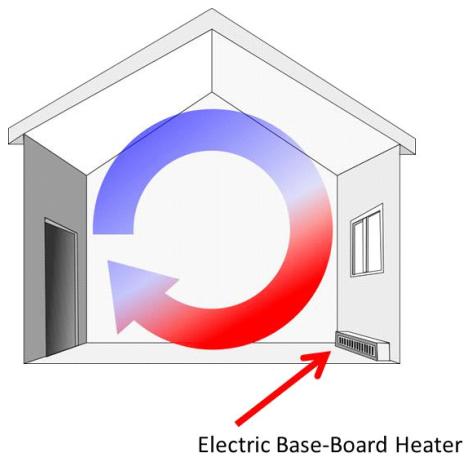
31.1.2 Convection

This is a fundamentally new type of energy transport in our study of thermodynamics, although we are all familiar with it. In the last section we used the example of an electric burner that touched our pot of vegetables. But there are

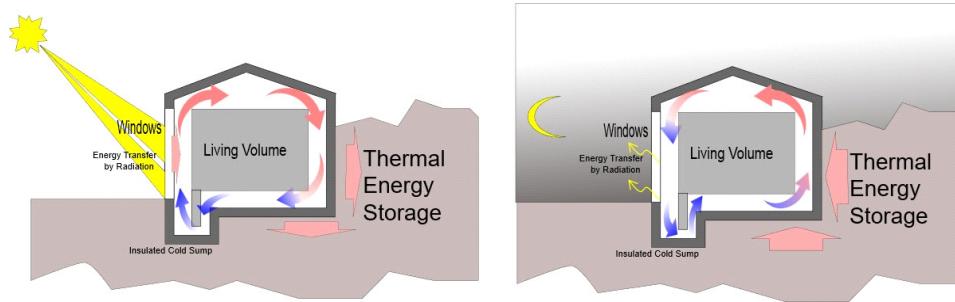
other kinds of stoves.



Convection is the transfer of energy by the movement of the atoms or molecules, themselves. Because the group of moving atoms or molecules has internal energy, moving these molecules also moves the energy. So convection is an energy transfer. This is like circulation of air in your home, or the transfer of energy from a gas flame to your cooking pot. Your student apartment might have base-board heating. But your parent's home would likely use forced convection where a fan blows the air.



For several decades houses have been made with the idea that you can heat or cool a house with passive convection. Here is an example.



This house tries to control the thermal conduction mechanisms to keep the interior of the house at nearly the same temperature year round.

31.1.3 Radiation

Let's go back to thinking about cooking. Suppose you are cooking over an old wood stove. You open the stove door and see this



And feel a blast of energy transfer by heat. But you don't feel a blast of wind. This isn't convection. And you are not touching the stove.

Light is a form of energy transfer. You will learn more about this in PH220, but we already know that light is a form of wave, and as we know, waves carry energy with them. So, light is an energy transfer. We call this energy removal or delivery by light "radiation." This is different than the use of the word "radiation" in nuclear physics.

All objects radiate energy according to Stefan's law

$$\mathcal{P} = \frac{Q}{\Delta t} = \sigma A e T^4 \quad (31.7)$$

The parts of this law are Stephen's constant $\sigma = 5.6696 \times 10^{-8} \frac{\text{W}}{\text{m}^2 \text{K}^4}$, the factor, A , is the surface area, e is the emissivity of the material. This e tells us how like a perfect absorber (called a black body) it is, and T is the temperature. Note there are other forms for these! But this is the form we will use.

The emissivity is a value from 0 to 1. If $e = 0$ then the material is a perfect reflector. This would be a perfect mirror. If the emissivity is 1, then it is a perfect absorber. This is more like blacktop in sunlight. The blacktop warms up, even in the winter, because it absorbs the sunlight. Of course, nothing has an emissivity of exactly 0 or exactly 1. Real objects are somewhere in the middle.

It may seem strange that "black" objects glow, but they do. Think again of that blacktop. The ice on the blacktop melts, because the blacktop not only absorbs the radiation from the sun, it also emits radiation.

An extreme example is the Sun, itself. Light that hits the sun from outside will scatter off the gasses in the sun. It takes a very long time for the light to reemerge, if it ever does. Most likely it will be absorbed by the gasses. So the Sun is considered a "black body!" But it glows, so it does not look black.

We can explain many common experiences using the idea of emissivity. Take this leaf for example. Why is it sinking into the snow?



The leaf has a different emissivity than the snow, so it absorbs more radiation from the sun, the temperature of the leaf rises and it melts the snow around it.

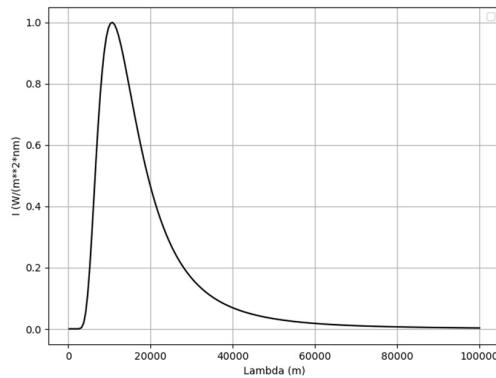
With this understanding, we can see that all objects absorb radiation as well as radiate. If an object is at temperature T_{obj} and its surroundings are at temperature T_{env} . then the net rate of energy gained or lost is

$$\mathcal{P} = \frac{Q}{\Delta t} = \sigma A e (T_{obj}^4 - T_{env}^4) \quad (31.8)$$

and the energy transfer would be

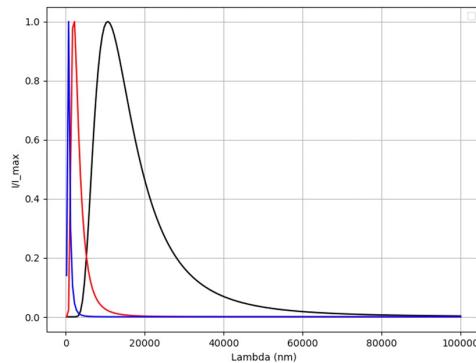
$$Q_{radiation} = \frac{Q}{\Delta t} = \sigma A e (T_{obj}^4 - T_{env}^4) \Delta t$$

Back when we were thinking of building thermometers we said that the color of an object could vary with temperature. What we meant is that the temperature determines the color of the light radiating from the object. If you get to go on in physics, you will study what wavelengths of light come from objects of different temperature (in Principles of Physics IV, called Modern Physics). But here is a graph of what wavelengths are produced by an ideal object who's temperature is 0°C .



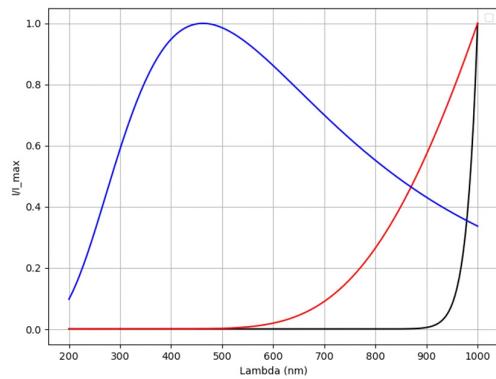
Thermal Emission Spectrum (Planck Curve)

You can see that more than one wavelength is produced but there is a peak wavelength. And if you are paying attention you can see that the peak is not in the visible spectrum. It is more like a radio wave. So we don't see this glow from a cold object. But if we increase the temperature, say, to 1200°C we would get an object that glows red. And at 6000°C we would have our object glow blue. Here are the curves for these two cases, along with our 0°C curve for reference.



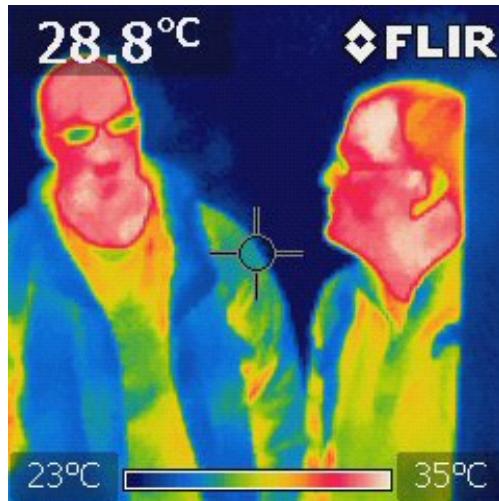
Normalized emission curves for objects with $T = 0^{\circ}\text{C}$ (black curve), $T = 1200^{\circ}\text{C}$ (red curve), and $T = 6000^{\circ}\text{C}$ (blue curve).

The curves moved down into the visible part of the spectrum. Note that in this last graph the three temperatures wouldn't give the same amount of light. Stefan's law tells us this. So the blue 6000°C curve would be much taller. In order to see the three curves on the same graph we have normalized the curves by dividing the values in a curve by their maximum. If we expand that part we see this



The blue curve is for the 6000°C object and it does peak in the blue part of the spectrum. The 1200°C curve is the red curve and it peaks off the chart, but the light it makes is all on the red side of the spectrum, so we would see red. And the black curve for 0°C doesn't rise in the visible spectrum as we already know.

But this brings us a question. Can we determine temperature by looking at the wavelength of the radiation glow coming from an object. Of course, the answer is yes. We can find the peak of the radiation, and use our curve backwards to find the temperature. Thermal cameras do exactly that.



Thermal Image of Physics Students.

Stefan's law adds up the energy from all the wavelengths that we see in our graphs to get the total power emitted.

31.1.4 Multiple transfer mechanisms

As we said before, we might have many different energy transfer mechanisms. We could have all of these (and more).

$$\Delta E_{int} = Q_{conduction} + Q_{convection} + Q_{radiation} + w$$

Let's think about an example that uses radiation, convection, and insulation to keep something cold. It is called a dewar flask.

This is the same idea as a thermos bottle, only better made (and more expensive). The flask has doubled walls with the space between evacuated to reduce convection and eliminate conduction. The walls are silvered to reduce heat transfer by radiation. The cap has a long piece of insulation attached to it to prevent conduction through the flask opening. New designs use multi-layer films which make their surfaces super Reflective.

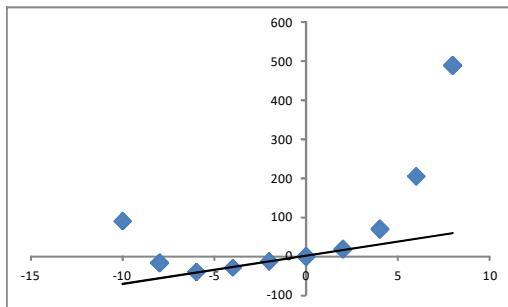
31.2 Gas Laws

We are trying to find the change in internal energy as Q and w are applied to our systems. To go further, we need to know how the pressure, temperature and volume of a gas are related. You probably studied the ideal gas law in your pre-college science classes, but before we use it in our class let's try a thought experiment.

Let's consider the following data



Figure 31.1: You might see our liquid nitrogen dewar in demonstrations. The flask is pictured on the left, and a schematic diagram of how a dewar works is given on the right.



The black straight line is a curve fit to the data. What can we say about the curve fit?

It might look useless. But look at the region from about $-6 < x < 5$. In this region, the curve fit is not too bad. Of course, the equation we got for the curve fit is not the right equation, and we know this from the regions $x < -6$ and $x > 4$. If the equation we used for the fit comes from our model for how this physical system works, we know that something is wrong with the model. We do not have an exact theory that explains the behavior of this system well. But could the curve fit be useful? If all you needed was an estimate of the value of the actual function at $x = -1$, then the curve fit might be enough to get your work done! This may not be terribly satisfying, but sometimes it is a useful way to work.

Another way to think about this is that if most of the time, under usual conditions, the phenomena we experience fall within the range $-6 < x < 5$ and the curve fit equation is much simpler, it might be good enough and convenient to use the curve fit. For example, we know that Newton's laws are not exact. We have to use General Relativity to be accurate. But in calculating our average speed on a trip to Idaho Falls, Newton's laws are plenty good enough, and *much*

more convenient than their General Relativistic forms.

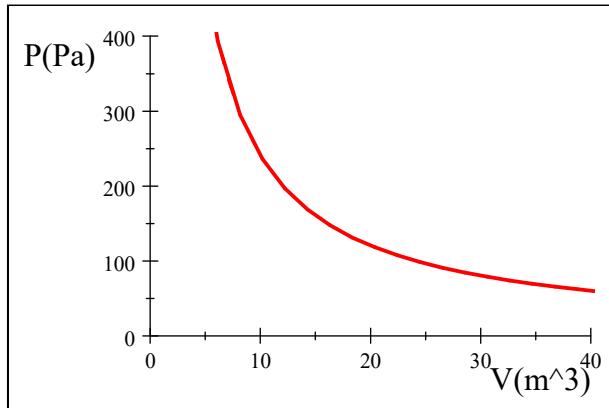
We will use the Ideal Gas Law to study the expansion of gasses as temperature changes. The Ideal Gas Law is an approximation very like our curve fit we have been considering. It is an equation that works pretty well under normal conditions, and it is much easier to use than the exact law. It allows us to gain great insight, without the mathematical difficulty of the exact relationship. Given all this, we should realize that there is no such thing as an ideal gas.

In this approximation, the atoms are so weakly bound that there is no equilibrium position between atoms. Thus there is no definite volume defined for an ideal gas. This is generally true at standard pressure and temperature for real gasses. We will let V be a variable for gases. Let's look at where the ideal gas law came from. It is much like our hypothetical experiment. The experimentalists gave us some clues on how to relate V , P , and T .

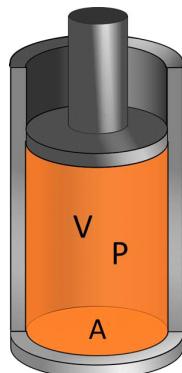
31.2.1 What lead to the Ideal Gas Law

A researcher named Boyle found that if a system is kept at a constant temperature then

$$P \propto \frac{1}{V} \text{ for constant } T \quad (31.9)$$



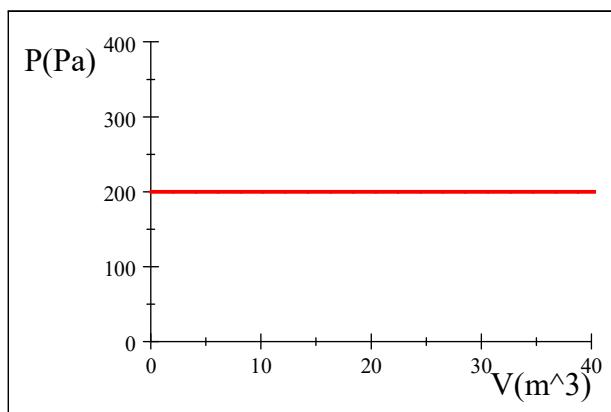
That is, if you don't let the temperature of the gas change, then a change in pressures is inversely related to a change in the volume of the gas. The more pressure, the smaller the volume. We could envision putting our gas in a cylinder, and putting a piston in the cylinder that can be used to compress our gas.



Boyle tells us that as we exert a pressure on the gas using the piston the volume of the gas changes. That seems reasonable.

Two researchers, Charles and Gay-Lussac, found that if a system is kept at constant pressure then

$$V \propto T \text{ for constant } P \quad (31.10)$$



The warmer the gas, the more it will expand. Note that in the graph temperature is not one of the axes! That is a little weird. But you can see the volume of the gas getting larger with the pressure staying the same. Something must be making this happen, and that something is the increase in temperature.

Combining these two findings gives an equation that describes ideal gasses.

$$PV = nRT \quad (31.11)$$

where n is the number of moles, and R is called the *universal gas constant*. This is the *ideal gas law*. The law came from researchers that could not access extreme pressures or temperatures, and so for normal conditions, it works just fine. But it will not work at all for extremely low or height pressures and temperatures.

The variables n , V , P , and T , are state variables for ideal gasses. Depending on the units for V , P , and T , the value for R may be different.

$$\begin{aligned} R &= 8.314 \frac{\text{J}}{\text{mol K}} \\ &= 0.08214 \frac{\text{atm l}}{\text{mol K}} \end{aligned} \quad (31.12)$$

31.2.2 Alternate form of the Ideal Gas Law

We know we can describe the number of moles as

$$n = \frac{N}{N_A} \quad (31.13)$$

where N is the number of molecules of gas that we have. Then we can write

$$PV = \frac{N}{N_A} RT \quad (31.14)$$

if we define a new constant

$$k_B = \frac{R}{N_A} \quad (31.15)$$

Then

$$PV = Nk_B RT \quad (31.16)$$

The constant k_B is called *Boltzmann's constant*

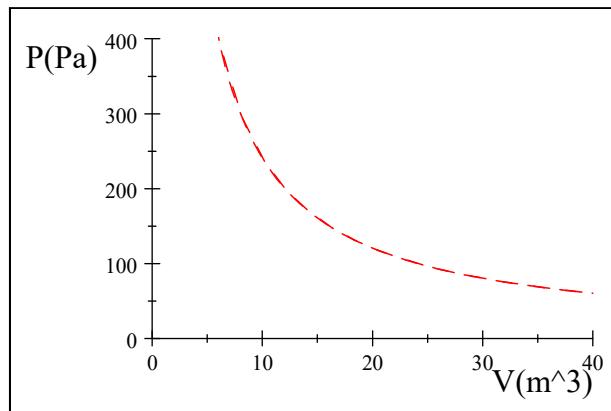
$$k_B = 1.38 \times 10^{-23} \frac{\text{J}}{\text{K}} \quad (31.17)$$

We now have two forms of the ideal gas law!

Syringe
Demo

31.3 PV diagrams

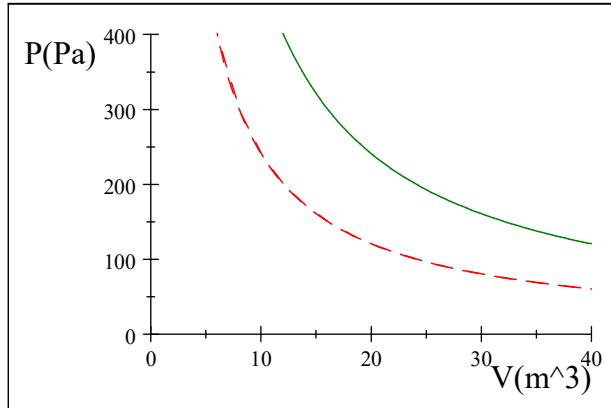
You may have noticed that in the diagrams in the previous sections we put pressure and volume on the axes.



This was awkward because our gas laws have three state variables, P , V , and T

$$P = \frac{nRT}{V}$$

and there wasn't another axis for T . We commonly do this, so we need to know what happens if we change T . In the next figure, you can see the red dashed line is the same



but now there is a green solid line. This is a line for a higher temperature. We expect a $1/V$ line for both, but with a higher temperature the line moves up and a little to the right on the graph. Since the temperature is the same along the whole red dotted line, we call such a line an *isotherm*. The green solid line is a different isotherm with a different temperature.

31.4 Van der Waals Gasses

But of course our ideal gas is missing pieces of the physics of actual gases. For our physics students we have an upper division class in thermodynamics that fills in many of these missing pieces. We also have some quantum mechanics classes that help. Chemists have a thermodynamics class and their physical chemistry class. Mechanical engineers have a thermal class as well. So some of the details will just have to wait. But if you are in one of the disciplines that won't take these fun classes, we should leave you with a slight improvement. We can add in two pieces of physics.

1. As a real gas gets compressed the molecules are more and more attracted to each other. This reduces the pressure in the gas
2. At higher pressures the gas molecules, themselves, constrain where other gas molecules can go (at some point they hit each other!).

A researcher named Van der Waals came up with this new approximation, so the result of adding in these two additional bits of physics is called the Van der Waals gas law equation. Let's see what it looks like.

For the first bit of new physics we can add a term to the pressure that is proportional to the molar density squared.

$$P + a \left(\frac{n}{V} \right)^2$$

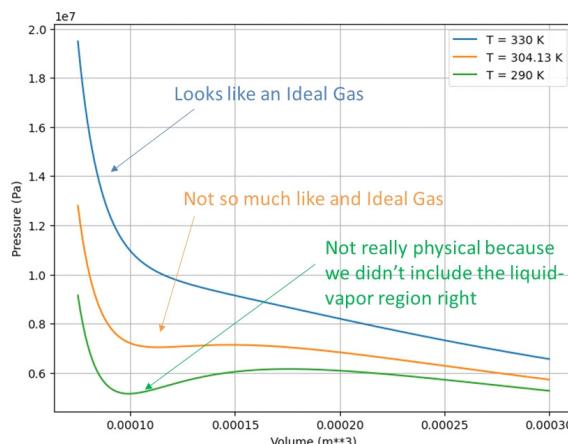
where the constant a depends on the specific gas. You may not find it obvious why this is the right form for our new gas law equation to make the pressure come out right, but it does adjust the pressure under compression and it turns out it kind of works. For the second bit of new physics we can subtract a term proportional to the volume of the molecules, themselves b from the volume of the whole gas.

$$V - nb$$

where b is like a molar volume. All together our gas law becomes

$$\left(P + a \left(\frac{n}{V} \right)^2 \right) (V - nb) = nRT$$

Although it is an improvement, it still doesn't have all the physics that might be needed to describe a real gas. For example, look at the graph below. You see a plot of isotherms for a Van der Waals gas (CO_2).



For high temperatures we see something that looks a lot like our ideal gas graph. but at a certain temperature for each gas the graph flattens out (middle orange line). That is really not behavior that we expect from an ideal gas. This temperature is called a *critical temperature*, T_c , and for our CO_2 it is about 304 K. But look at the lower green line) this line dips down and make a valley. This isn't possible. Our "improved" gas law is giving us unphysical predictions. What is missing is that we can have a change from gas to liquid at this pressure and temperature and we didn't include such changes in our Van der Waals gas law equation. Above T_c our Van der Waals equation isn't too bad. But below T_c

it is not as useful. Of course we could improve on our Van der Waals equation, but that is a subject for an upper division class in thermodynamics. For this class, let's concentrate on the ideal gas equation even though we know it has some limitations.

Chapter 32

32 Kinetic Theory of Gases

2.2.2, 2.2.3

We now have gases with internal energy. But so far, as we have considered the motion of atoms with thermal energy we have been able to treat whole samples of a gas at a time. But you might guess that we are sort of taking averages of the motions of the atoms or molecules. Let's make this more explicit in this lecture.

Fundamental Concepts

- In a gas, the molecules have different speeds
- The distribution of speeds is given by the Boltzmann distribution law
- the root mean square speed ($v_{rms} = \sqrt{\bar{v^2}}$) is a better measure of the speed of gas molecules than the average velocity
- Pressure for an ideal gas is given by $P = \frac{2}{3} \frac{N}{V} \left(\frac{1}{2} m \bar{v^2} \right)$
- Temperature is related directly to the average kinetic energy of the molecules of a gas
- We call a way the molecule can move a *degree of freedom*.
- For an ideal gas $C_V = \frac{3}{2}R$
- For an ideal gas $E_{int} = \frac{3}{2}k_B T$

32.1 Molecular Model of an Ideal Gas

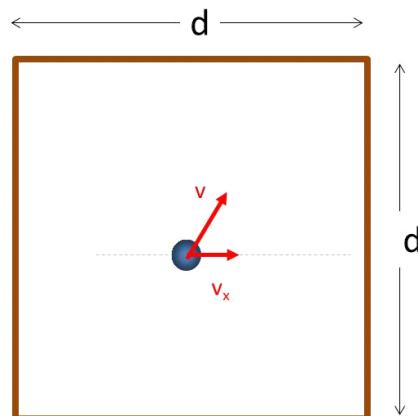
We have hinted that the internal energy of a system must be the energy associated with the atoms and molecules that make the system. By making association in our mental model more concrete, and more mathematical. We will be able to see how the thermodynamic laws we have studied are generated by the basic laws of motion. But we will still limit our study to ideal gases in this class. So let's state some assumptions that follow from the ideal gas approximation.

Ideal Gas Model
The number of molecules in the gas is very large,
The average separation between molecules is large compared to 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.

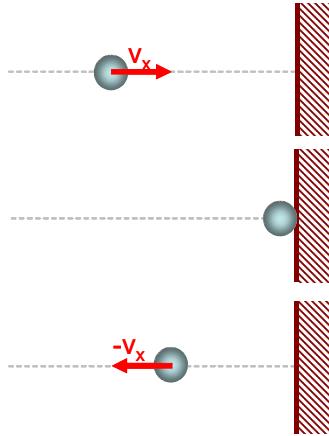
32.2 Microscopic Definition of Pressure

We want to express macroscopic quantities like pressure and volume in terms of microscopic effects. The easiest to start with is pressure



Picture a particle in a cubical box of side length d .

As the particle approaches the wall of the box we may say that it has velocity v_x . The particle impacts the wall and bounces off again. Now it travels with $-v_x$.



We can think of why this would be. We know momentum is conserved. Before the collision we may take the velocity of the wall to be zero, and we may imagine the wall to have a very large mass (which it does, compared to the particle!). If we call the particle mass m and the wall mass M but the wall will not move with any speed after the collision so $v_{wall_f} = 0$

$$\begin{aligned} P_i &= Mv_{wall_i} + mv_{x_i} \\ &= mv_{x_i} \end{aligned}$$

and

$$\begin{aligned} P_f &= Mv_{wall_f} + mv_{x_f} \\ &= -mv_{x_f} \end{aligned}$$

because the wall still isn't going to move. Then for conservation of momentum

$$P_i = P_f$$

We have

$$mv_{x_i} = -mv_{x_f}$$

Then the magnitude of the velocities must be equal and the change in momentum must be

$$\begin{aligned} \Delta p &= mv_{x_i} - (-mv_{x_f}) \\ &= 2mv_{x_i} \end{aligned}$$

Question 123.14.6

Recall that the impulse from the collision of the particle on the wall would be

$$\begin{aligned} \bar{F}_{i,\text{on molecule}} \Delta t_{\text{collision}} &= \Delta p_{x_i} \\ &= -2mv_{x_i} \end{aligned}$$

This is the impulse momentum theorem from Principles of Physics I. The time of the collision is very short. This force is the force of the wall on the particle.

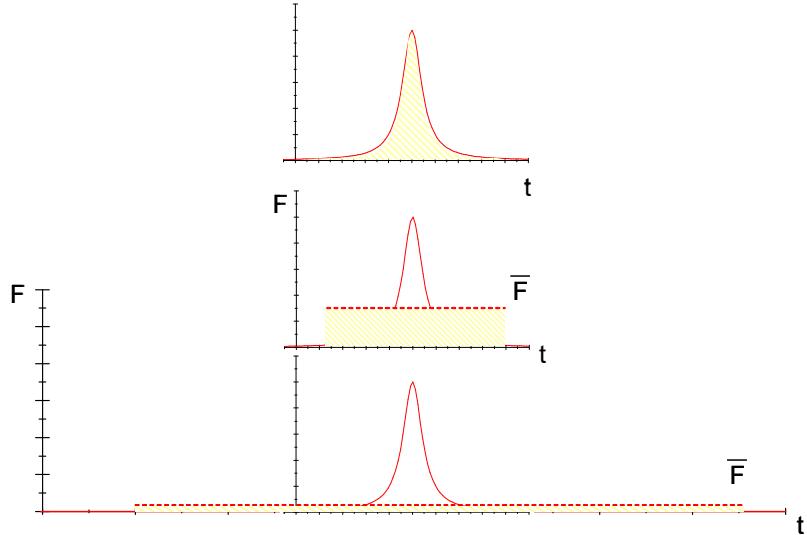
Now consider that if a particle bounces off one wall in the x -direction, it must travel to the opposite wall and back before it can bounce off the same wall again. That is, the time between bounces is roughly

$$\Delta t = \frac{2d}{v_{x_i}} \quad (32.1)$$

Averages are funny things, it is perfectly legal to redefine the time of our average force to be this travel time Δt . What does this mean? In the first graph of the next figure we have the full definition of impulse

$$\Delta p = \int_{t_i}^{t_f} \mathbf{F} dt \quad (32.2)$$

which says the impulse is the area under the F vs. t graph.



The average force \bar{F} is pictured in the second graph. It has the same area as the first graph, but is flat, representing the average force (red dotted line). The final graph has a \bar{F} averaged over a much larger Δt . The area is the same, but the magnitude of the average force is much smaller.

This may seem useless, but at any rate, we can do this. So let's take our average of F over the time

$$\Delta t = \frac{2d}{v_{x_i}}$$

which is the time it takes our particle to travel away from our wall, and bounce back to the wall. We know that sometime within this Δt the collision with our wall actually occurs. The change in momentum is still

$$\bar{F}\Delta t = -2mv_{x_i}$$

so we can write

$$\begin{aligned}\bar{F} \frac{2d}{v_{x_i}} &= -2mv_{x_i} \\ \bar{F} &= -\frac{mv_{x_i}^2}{d}\end{aligned}$$

So far our force has been the force of the wall on the particle, but we know by Newton's third law that the particle force must be equal and opposite.

$$\bar{F}_{i,\text{on molecule}} = -\bar{F}_{i,\text{on the wall}}$$

then if we shorten $\bar{F}_{i,\text{on the wall}}$ to just \bar{F}_{iw}

$$\bar{F}_{iw} = \frac{mv_{x_i}^2}{d}$$

for one particle (molecule or atom).

For many particles

$$\begin{aligned}\bar{F}_w &= \sum_{i=1}^N \frac{mv_{x_i}^2}{d} \\ &= \frac{m}{d} \sum_{i=1}^N v_{x_i}^2\end{aligned}$$

We have found the average force on the container wall!

Question 123.14.7

Now remember how to take an average

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

so if we were to calculate the average velocity, *but square the velocity* before we take the average, it would look like this

$$\overline{v_{x_i}^2} = \frac{1}{N} \sum_{i=1}^N v_{x_i}^2$$

We almost have this in our force equation! Let's put in the missing parts

$$\begin{aligned}\bar{F} &= \frac{m}{d} \frac{N}{N} \sum_{i=1}^N v_{x_i}^2 \\ &= \frac{m}{d} N \overline{v_{x_i}^2}\end{aligned}$$

Going back to a single molecule, we remember that the velocity of that single molecule has components. The velocity squared would be (think of a dot product)

$$v_i^2 = v_{x_i}^2 + v_{y_i}^2 + v_{z_i}^2$$

Then if we take the averages of each piece like we did with v_{xi}^2

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

There is no real reason to prefer the x -direction, we could have chosen the y -direction and had the same value. Then

$$v_x^2 = v_y^2 = v_z^2$$

and

$$\overline{v^2} = 3v_x^2$$

So our force on the wall is

$$\begin{aligned}\bar{F} &= \frac{m}{d} N \overline{v_x^2} \\ &= \frac{m}{d} N \frac{1}{3} \overline{v^2} \\ &= \frac{mN}{3d} \overline{v^2} \\ &= \frac{N}{3d} (m \overline{v^2})\end{aligned}$$

Finally we can find the pressure. We know

$$\begin{aligned}P &= \frac{F}{A} \\ &= \frac{F}{d^2} \\ &= \frac{1}{d^2} \left(\frac{N}{3d} (m \overline{v^2}) \right) \\ &= \frac{1}{d^3} \frac{N}{3} (m \overline{v^2}) \\ &= \frac{1}{3} \frac{N}{V} (m \overline{v^2}) \\ &= \frac{2}{3} \frac{N}{V} \left(\frac{1}{2} m \overline{v^2} \right)\end{aligned}$$

That last step might seem unnecessary, but look how cute the $1/2$ looks with a $m \overline{v^2}$ next to it! We can recognize the part in parenthesis as the average kinetic energy of the molecule! So

$$P = \frac{2}{3} \frac{N}{V} \left(\frac{1}{2} m \overline{v^2} \right) \quad (32.3)$$

and we see that the pressure of the gas is proportional to the number of molecules per unit volume and to the average kinetic energy of the molecules! In fact, this equation shows some of the things we have learned about gas laws. For example, P is inversely proportional to V !

32.3 Temperature

Finally, we are ready to define temperature!

We now know that

$$P = \frac{2}{3} \frac{N}{V} \left(\frac{1}{2} m \overline{v^2} \right)$$

Question 123.15.1

Question 123.15.2

Equation on
Board

and that this has some of the properties of the ideal gas law. We can rearrange this to look a little more like the ideal gas law. Let's start with our pressure, but put the volume on the other side where it is in the ideal gas law.

$$PV = \frac{2}{3} N \left(\frac{1}{2} m \overline{v^2} \right)$$

But temperature does not appear in this equation. Let's recall that we can write the ideal gas law as

$$PV = Nk_B T$$

If we set these two equations equal to each other

$$Nk_B T = \frac{2}{3} N \left(\frac{1}{2} m \overline{v^2} \right)$$

then

$$T = \frac{2}{3} \frac{1}{k_B} \left(\frac{1}{2} m \overline{v^2} \right) \quad (32.4)$$

This is fantastic! We have finally defined temperature. Temperature is related directly to the average kinetic energy of the molecules! The average kinetic energy is

$$\bar{K}_{mol} = \frac{1}{2} m \overline{v^2} = \frac{3}{2} k_B T$$

Note that in each direction we should have

$$\bar{K}_x = \frac{1}{2} m \overline{v_x^2} = \frac{1}{2} k_B T$$

so

$$\bar{K}_y = \frac{1}{2} k_B T$$

and

$$\bar{K}_z = \frac{1}{2} k_B T$$

So in each direction the molecules have

$$\bar{K}_i = \frac{1}{2} k_B T \quad (32.5)$$

We call a way the molecule can move a *degree of freedom*. Our molecules can move in the x , y , and z , direction, so our molecules have three degrees of freedom. Each degree of freedom contributes $\frac{1}{2} k_B T$ worth of energy.

Of course, we only allow translational energy in our formulation for ideal gasses, but for more complex molecules we could have rotational energy (another way to move) and vibrational energy (yet another way to move) etc. And each way the molecule can move is another degree of freedom would contribute $\frac{1}{2}k_B T$. But remember, these more complex molecules would not be ideal gasses.

For the collection of N molecules, we have all together

$$K = N \left(\frac{3}{2}k_B T \right) \quad (32.6)$$

$$= \frac{3}{2}Nk_B T \quad (32.7)$$

$$= \frac{3}{2}nRT$$

and for our ideal gas the only form of internal energy we have is the kinetic energy, so we see that the internal energy depends only on the temperature.

This is only true under our assumption that the molecules or atoms are structureless. So this works well for actual monotonic gases, but not as well for diatomic gasses, etc.

Let's do a problem. Suppose a colleague tells you that in her Statistical Mechanics class they defined a root-mean-squared velocity for atoms in a gas. That is, they took the velocity, squared it, and then took the average. They got

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

Let's show that this works from our kinetic theory of gasses.

We know that each degree of freedom contributes $\frac{1}{2}k_B T$ worth of internal energy for the gas molecules. We just found that

$$\bar{K} = \frac{K}{N} = \frac{3}{2}k_B T$$

and we know that

$$\bar{K} = \frac{1}{2}mv^2$$

Setting these equal gives

$$\begin{aligned} \frac{3}{2}k_B T &= \frac{1}{2}mv^2 \\ 3k_B T &= mv^2 \\ \frac{3k_B T}{m} &= v^2 \end{aligned}$$

so

$$v_{rms} = \sqrt{v^2} = \sqrt{\frac{3k_B T}{m}}$$

which is just what our colleague said. This shows the power of the simple idea of the ideal gas approximation.

The rms speed tells us that if the molecule or atom is tiny (has little mass) it must move very fast. If it is larger, it will move more slowly. Let's look at how molecules travel as they go through a gas.

32.4 Mean Free Path

We started this chapter by saying how molecules interact with the walls of a box. But the molecules really do bounce off each other. The average distance a molecule travels before a molecule-molecule collision is called the *mean free path*. We could give this a symbol, let's re-use the Greek letter λ

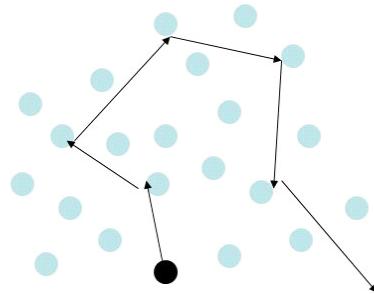


Figure 32.1:

$$\begin{aligned}\lambda &= \frac{\text{Length traveled}}{\text{Number of collisions}} \\ &= \frac{L}{N_{\text{col}}}\end{aligned}$$

This will give the average length before a collision happens.

The average length traveled is easy

$$L = \bar{v} \Delta t$$

But to predict the number of collisions is harder. To do this problem, let's play a geometrical trick. First let's assume the molecules are spheres with diameter d . We see that a collision does not happen unless the distance between the molecules is less than or equal to $2d$. We can model this interaction as a large particle of size $2d$ and many point particles. We let the large particle move in a straight line to create a cylindrical path.

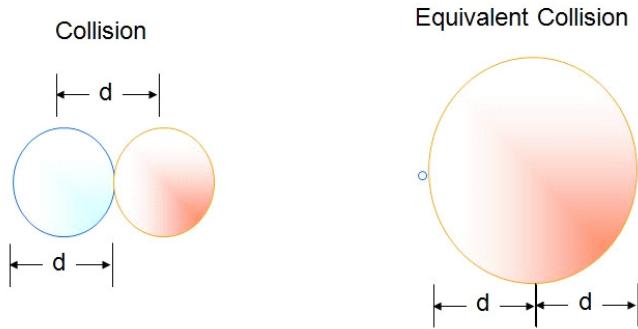
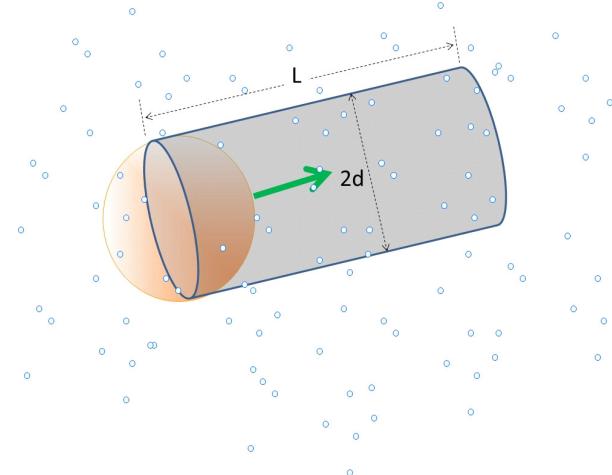


Figure 32.2:



All point particles in the cylinder will collide with our large particle. The volume of such a big cylindrical path is

$$V = (\pi d^2)L$$

but the length L of the cylinder will be

$$L = \bar{v}\Delta t$$

so the volume will be

$$V = \pi d^2 \bar{v} \Delta t \quad (32.8)$$

then, if n_V is the number of molecules per unit volume, our big equivalent particle will hit all of the point particles in the volume.

$$N_{\text{col}} = \pi d^2 \bar{v} \Delta t n_V \quad (32.9)$$

is the number of point molecules in the cylinder. And it is the number of collisions.

The mean free path is the average distance traveled in a time interval Δt divided by the number of collisions that occur in Δt . We now know both of these, so

$$\begin{aligned}\lambda &= \frac{\bar{v}\Delta t}{\pi d^2 \bar{v} \Delta t n_V} \\ &= \frac{1}{\pi d^2 n_V}\end{aligned}\quad (32.10)$$

We can find the the average time between collisions. This is called the mean free time.

$$\begin{aligned}\tau &= \frac{\lambda}{\bar{v}} \\ &= \frac{1}{\pi d^2 n_V \bar{v}}\end{aligned}$$

The frequency of collisions is

$$f = \frac{1}{\tau} = \pi d^2 \bar{v} n_V \quad (32.11)$$

It turns out that our simple model is off by a factor of $\sqrt{2}$ because we assumed the molecules are stationary. We can fix that easily

$$\lambda = \frac{1}{\sqrt{2}\pi d^2 n_V} = \frac{V}{\sqrt{2}\pi d^2 N} \quad (32.12)$$

and

$$f = \sqrt{2}\pi d^2 \bar{v} n_V = \frac{\sqrt{2}\pi d^2 \bar{v} N}{V} \quad (32.13)$$

and finally

$$\tau = \frac{1}{\sqrt{2}\pi d^2 n_V \bar{v}} = \frac{V}{\sqrt{2}\pi d^2 N \bar{v}}$$

We can make a further modification if we assume an ideal gas. We know

$$PV = Nk_B T$$

so

$$\frac{V}{N} = \frac{k_B T}{P}$$

then where there is a V/N we can substitute

$$\lambda = \frac{k_B T}{\sqrt{2}\pi d^2 P} \quad (32.14)$$

and

$$f = \frac{\sqrt{2}\pi d^2 \bar{v} P}{k_B T} \quad (32.15)$$

$$\tau = \frac{k_B T}{\sqrt{2}\pi d^2 \bar{v} P} \quad (32.16)$$

I have used \bar{v} as our average speed for our molecules, but we will find it might be good to use the v_{rms} as our average. But we will get to that soon enough (our colleague from the last problem was right to suggest it!).

Now that we know what temperature is, let's go back to thinking about temperature rise as we transfer energy by heat. We studied heat capacities and specific heats a few lectures ago, but we avoided gasses as we did so. We need to find the temperature rise ΔT in a gas after transferring energy by heat Q for gasses. Let's do that in next.

32.5 Molar Specific Heat

We learned that

$$Q = mc\Delta T$$

or that we could use the molar form of this equation

$$Q = nC\Delta T$$

But we need to find the C values for different gasses. To do this, remember for an ideal gas, we only have translational kinetic energy, so it must be true that

$$E_{int} = K_{tran} = \frac{3}{2}nRT$$

But let's take a special case. If we take the system through a process at constant volume,

$$Q = \Delta E_{int}$$

because no work is done (we didn't change the volume) and for this spacial case we can write C as C_V to remind us that the volume didn't change.

$$Q = nC_V\Delta T$$

so

$$\Delta E_{int} = nC_V\Delta T$$

Notice this is a change in internal energy. We added energy to our gas

$$\begin{aligned} E_{int_f} - E_{int_i} &= nC_V(T_f - T_i) \\ &= nC_V T_f - nC_V T_i \end{aligned}$$

this strongly suggests that if C_V is constant

$$E_{int} = nC_V T$$

we can solve for C_V

$$C_V = \frac{1}{n} \frac{\Delta E_{int}}{\Delta T}$$

and for very small changes in temperature

$$C_V = \frac{1}{n} \frac{dE_{int}}{dT}$$

But for an ideal gas the internal energy is kinetic energy so we know that

$$E_{int} = \frac{3}{2} nRT$$

So we can easily find $\frac{dE_{int}}{dT}$

$$\frac{dE_{int}}{dT} = \frac{3}{2} nR$$

Then our value for C_V is

$$\begin{aligned} C_V &= \frac{1}{n} \frac{dE_{int}}{dT} \\ &= \frac{1}{n} \frac{3}{2} nR \\ &= \frac{3}{2} R \end{aligned}$$

Knowing $R = 8.314 \frac{\text{J}}{\text{mol K}}$, the numerical value is

$$\begin{aligned} C_V &= \frac{3}{2} 8.314 \frac{\text{J}}{\text{mol K}} \\ &= 12.471 \frac{\text{J}}{\text{mol K}} \end{aligned}$$

for all monotonic gasses. It turns out that this is a very good approximation.

You might find the restriction of heating at a constant volume somewhat constrictive. We will lift this restriction in a later lecture.

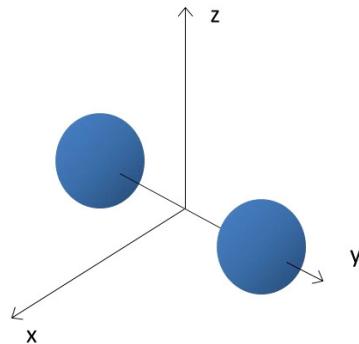
32.6 The Equipartition of Energy

We can 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

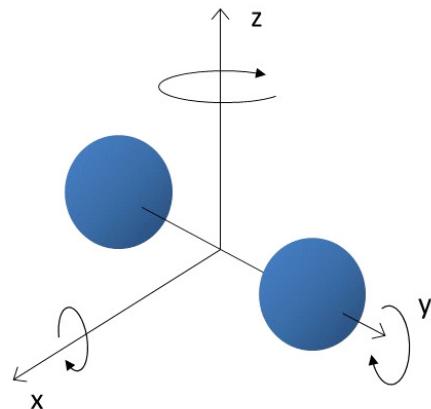
$$E_i = \frac{1}{2} k_B T \tag{32.17}$$

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

$$E_{int} = 3E_i = \frac{3}{2} k_B T \tag{32.18}$$



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

$$\mathbb{I} = \frac{2}{5}mr^2 \quad (32.19)$$

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

$$\mathbb{I} = \sum_i m_i R_i^2 \quad (32.20)$$

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

$$E_{int} = \left(\frac{3}{2}k_B T\right)_{trans} + \left(\frac{2}{2}k_B T\right)_{rot} = \frac{5}{2}k_B T \quad (32.21)$$

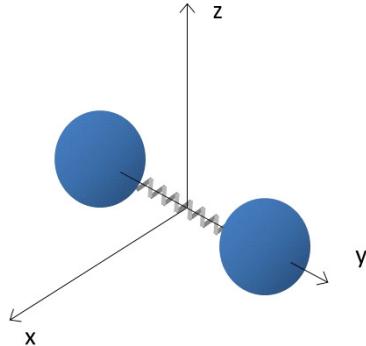
Writing this in molar terms

$$E_{int} = \frac{5}{2}nRT \quad (32.22)$$

then

$$C_V = \frac{1}{n} \frac{dE_{int}}{dT} = \frac{5}{2}R \quad (32.23)$$

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

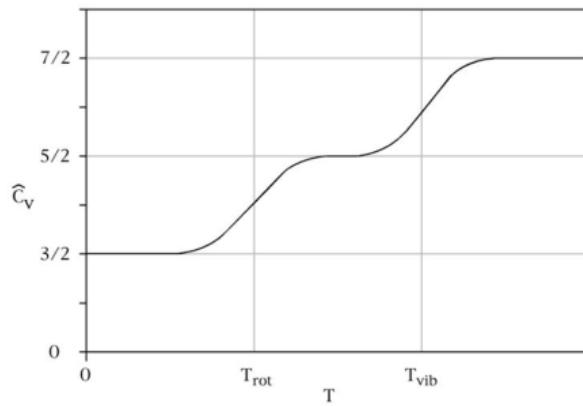
$$E_{int} = \left(\frac{3}{2}k_B T\right)_{trans} + \left(\frac{2}{2}k_B T\right)_{rot} + \left(\frac{2}{2}k_B T\right)_{vib} = \frac{7}{2}k_B T \quad (32.24)$$

which gives

$$C_V = \frac{7}{2}R \quad (32.25)$$

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.

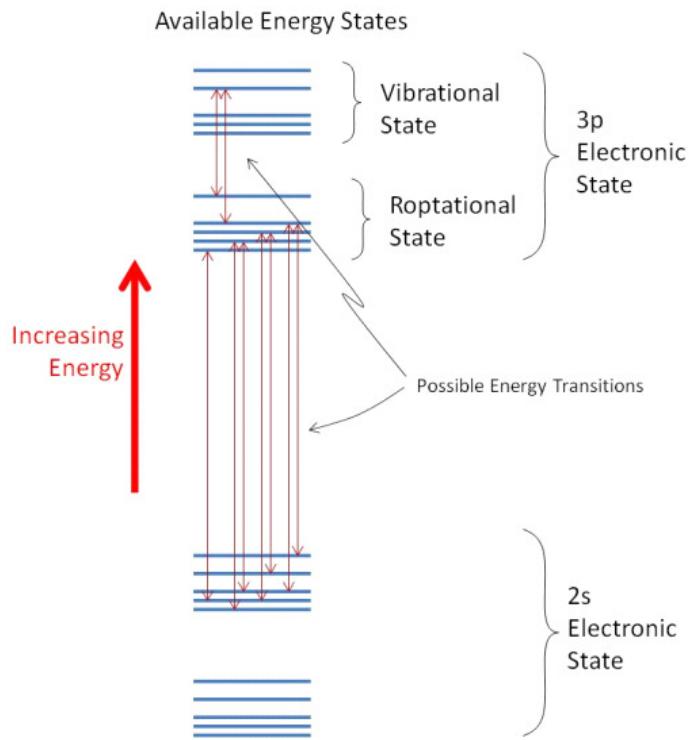


At low temperature, the translational degrees of freedom are preferred. Then, as the temperature rises, the rotational degrees of freedom are filled. Finally the vibrational degrees of freedom are used. The figure shows this relationship for diatomic Hydrogen. Note that there are plateaus at each of our values of C_V that we found ($\frac{3}{2}R$, $\frac{5}{2}R$, $\frac{7}{2}R$).

We have talked about the simplest of molecules. If we had a more complex molecule, there would be more complex rotational and vibrational degrees of freedom. That is why for larger molecules the values of C_V in table in the book don't follow a simple fraction of R .

32.7 Quantization

We have come to the end of what classical theory can explain. We know that different degrees of freedom are filled in groups, but we don't know why. We will only hint at the reason, but it has to do with wave motion that we studied!



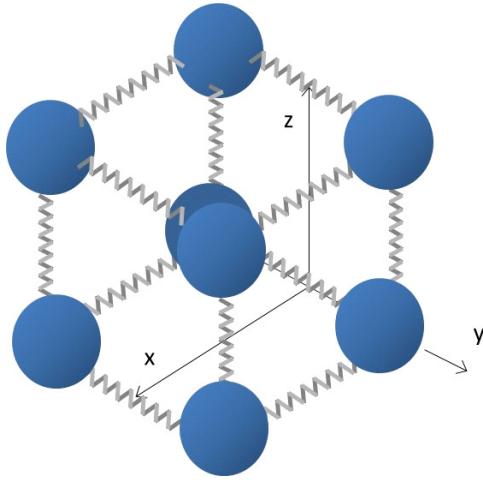
Quantum theory tells us that atoms can be described as waves under boundary conditions. Like a string with fixed ends, these atomic waves have quantized frequencies. The energy of the molecule is proportional to the frequency in quantum theory, so quantum mechanics tells us that the energy of the molecule will be quantized. The figure shows an energy level diagram for a molecule. Higher energy is in the y direction. The vibrational and rotational states are marked. (the maroon arrows show transitions from higher to lower states. These transitions create radiation).

The lowest allowed state is called the *ground state*. It is analogous to the fundamental frequency. At low temperatures the molecule only has enough energy to populate the lower states, but when the temperature rises, there is enough energy transfer due to collisions to push molecules into the rotational states. With higher temperatures, you will see the collisions transfer molecules to the higher vibrational states.

32.8 Molar specific heat of solids

Let's do a problem. Let's find C_V for a solid, and compare it to the specific heat of elemental solids.

We can view the atoms of solids as having a structure of springs and atoms.



We can see that we should have three translational degrees of freedom for each atom. We also have three degrees of freedom associated with the potential energy from the spring-like bonding forces. Recall that

$$E_x = \frac{1}{2}mv_x^2 - \frac{1}{2}kx^2 \quad (32.26)$$

and we have and E_y and E_z as well.

From equipartition of energy we have

$$E_i = 6 \left(\frac{1}{2}k_B T \right) \quad (32.27)$$

per atom! So we have

$$E_{int} = 3Nk_B T \quad (32.28)$$

$$= 3nRT \quad (32.29)$$

and then for solids

$$C_V = \frac{1}{n} \frac{dE_{int}}{dT} = 3R \quad (32.30)$$

numerically this is

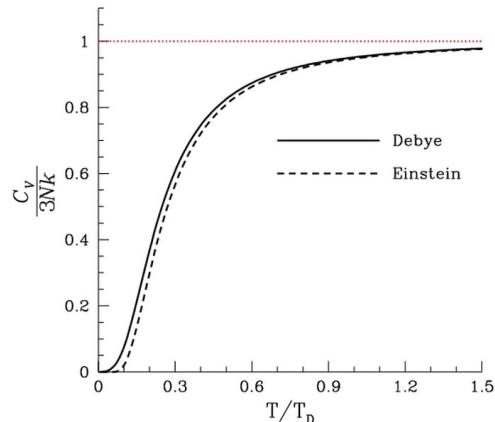
$$\begin{aligned} C_V &= 3 \times 8.314 \frac{\text{J}}{\text{mol K}} \\ &= 24.942 \frac{\text{J}}{\text{mol K}} \end{aligned}$$

This is the *DuLong-Petit law*. Here is our table of specific heat values again.

Substance	$c \left(\frac{\text{J}}{\text{kg K}} \right)$	$C_V \left(\frac{\text{J}}{\text{mol K}} \right)$
Aluminum	900	24.3
Copper	385	24.4
Iron	449	25.1
Gold	129	25.4
Lead	128	26.5
Ice	2090	37.6
Mercury	140	28.1
Water	4190	75.4

we see that our simple analysis did not do too bad! For elemental solids we get about the right number.

This law also breaks down for very cold temperatures where we need even more quantum mechanics to explain what happens. Einstein and Debye both contributed to explaining this departure, but their work is beyond this freshman course. This is a topic for a solid state physics course which is an elective that you might consider if you are interested in seeing a more complete solution.



Debye and Einstein models. (Image in the Public Domain courtesy Frédéric Perez)

32.9 Partial Pressure and Vapor Pressure

So far we have assumed we only have one type of molecule in our gas. But what if we had several kinds of molecules, like in our air that we breath that is mostly N₂ but also has some very important O₂ that we need to live (and a little bit of CO₂ that we worry about)?

We could define a pressure that the gas would make if occupied the entire volume on its own. We call this a *partial pressure*. But think of our kinetic

theory of gasses that we have just developed. If the gasses are in thermal equilibrium, then they have the same temperature, and therefore the same average kinetic energy. They would each make a pressure on the walls of our container.

$$P = \frac{2N}{3V} \left(\frac{1}{2}mv^2 \right)$$

But each gas is doing this at the same time. So the total pressure on the container walls would be the sum of the pressures from the individual gasses in the mix

$$P_{total} = \sum_i P_i$$

This is called Dalton's law. A result of this law is that for any two gasses in equilibrium

$$\frac{P_1}{n_1} = \frac{P_2}{n_2}$$

There is a safety issue hidden in what we have just said. Suppose we have a large container of liquid nitrogen in our lab (because we do). And suppose our container has a leak. The pressure in the room might not change much at all, but the portion of that pressure due to O₂ might change. And we need that O₂ pressure to stay alive. But since the overall pressure didn't change noticeably, we wouldn't have any warning that there was a problem until we felt trouble moving or passed out.

Suppose we have a vapor in a container but also have some of the same material in it's liquid state in the same container. And suppose that the vapor and liquid are in equilibrium. The partial pressure of the vapor under these conditions is called a *vapor pressure*. This is important in many fields including weather prediction. The partial pressure of the water vapor in our atmosphere can't exceed the vapor pressure of water at the same temperature or the water vapor will condense out of the air. This causes dew and the temperature at which the condensation starts to happen is called the *dew point temperature*.

If you listen to a weather report your might hear about the *relative humidity*.

$$R.H. = \frac{\text{Partial pressure of water vapor at } T}{\text{Vapor pressure of water at } T} \times 100$$

Think about what this means. If at a given temperature the partial pressure of water vapor is the same as the vapor pressure of water then the air has as much water as it can hold at that temperature. We say the air is saturated and it feels very muggy.

This almost completes our kinetic theory of ideal gasses. But before we move on we need to look at the speeds of our molecules. We know average speed is related to temperature, but not every molecule is moving at that average speed. What distribution of speeds like?

Chapter 33

33 Speeds and Systems

2.2.4, 2.3.1

We found in our last lecture that v_{rms} is a nice average kind of speed that tells us in bulk how our molecules are moving. But surely if we have an average speed, some molecules are moving more slowly and some are moving more quickly. Can we say how many molecules have a particular speed? We will take on this problem in this lecture.

We will also consider energy conservation in thermodynamic processes. In thermodynamics, conservation of energy is known as the first law of thermodynamics.

Fundamental Concepts

- The number of molecules that have a specific speed in an ideal gas is given by the Maxwell-Boltzmann distribution $N_v = 4\pi N \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}}$
- We characterize a gas by giving the v_{rms} , the average speed \bar{v} , and the most probable speed v_{mp}
- Thermodynamic systems are just the part where we want to know state it's variables P , V , n , and T . All the material and space around the system is called the surroundings.
- We think of quasi-static processes in thermodynamics and would prefer reversible processes.

33.1 Speed of molecules

The internal energy of a system might consist of nuclear internal energy, or chemical internal energy, but let's reserve those internal energies for our Prin-

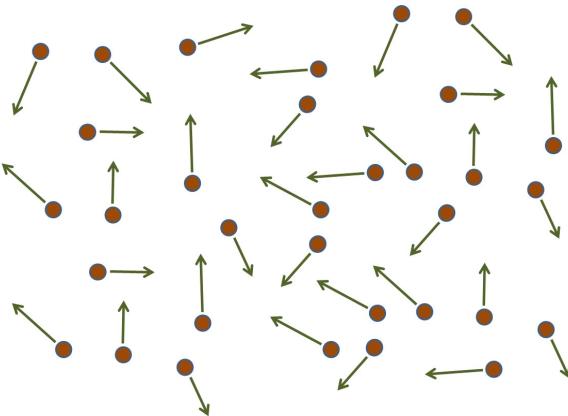
ciples of Physics IV course. For us, internal energy is mostly due to the motion of the molecules in the system. That motion has kinetic energy associated with it. But not a kinetic energy of the whole system moving together. It is a kinetic energy of the parts of the system, the molecules. Think of driving a car. We could describe the kinetic energy of the whole car. That is the kinetic energy we learned about in PH121. But we could also talk about the kinetic energy of one of the pistons in the engine. These two kinetic energies would be different! Just like this, the kinetic energy of the molecules will be different than the kinetic energy of the system as a whole.



To find this kinetic energy of the molecules, we would need to know how fast the molecules are going in an ideal gas at a particular temperate and pressure in a particular volume. Let's review a bit from our last lecture.

The average velocity is not helpful, because with our ideal gas model, the molecules move randomly. So we would guess that on the whole their velocities would cancel. If this is not true, we call the situation wind. But if there is no over-all movement, the average velocity, $\bar{v} = 0$. But we know there must be movement because the molecules have some internal energy. The kinetic energy depends on the speed of the molecules. But not all molecules have the same velocity. So when we say “speed of the molecules” we have to be specific, what speed? The average? The speed the most particles have?

The average velocity won't do.



The molecule motions have random directions. So the average velocity is zero. Some velocities are in negative directions and some are in positive directions, so the velocities cancel out. But if we square the velocities they are all positive. Then we could find the average of the squared velocities and that would be better. But a velocity squared is not a velocity. So to get back to velocity, let's take a square root.

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

This quantity is like an average, and gives a representative value that is near the most probable speeds. We have seen this quantity before, it is the rms speed.

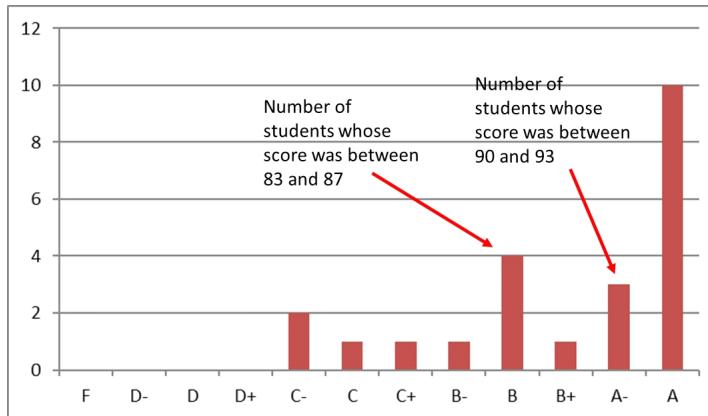
Using this, we can get an expression that relates the temperature of our gas to the speed of the molecules. But it we need to know something about how many molecules have what speed. To find the *distribution* of speeds, let's remember the idea of number density

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

We can express the number of molecules this way assuming we know the volume. But we want the speed of the molecules, and we have learned that concept of energy usually can get us a speed if we don't care to know the direction. Knowing the direction for every air molecule in a room would not be helpful in most cases, so we can be content with just knowing the speed. So lets find 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. If you take quantum mechanics, or statistical mechanics, you will likely see this quantity quite a lot. Then the number of molecules with a particular energy between E_1 and E_2 could be written as

$$n_V(E) dE = \frac{\# \text{ of molecules with energy between } E \text{ and } E + \Delta E}{V} \quad (33.1)$$

This is called a *distribution function*. 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. Think of grades on a test. If we plot the grades, the height of the bar for a “B” grade is the number of students that scored between 83% and 87%.



This graph is a distribution of grades and it tells us for each score segment, say S and $S + \Delta S$ (think 83% and $83\% + 4\% = 87\%$) how many students had scores in that range. We are doing the same for our molecules and their energy. We will make a graph that shows for each energy segment from E to $E + \Delta E$ how many molecules have that energy. This gives us the probability that the molecules will have a particular energy (or speed, since this is kinetic energy). A function that gives the amount of molecules that have a particular amount of energy also is called a *distribution function* just like it was for grades. The distribution function that we will use is written symbolically in this cryptic fashion, $n_V(E)dE$. It is the number of molecules with a particular energy (because our ΔE is so small it is a dE) divided by the total number of molecules. That makes it like a percent.

We won’t derive an equation for this quantity, instead we will borrow a result from our junior level thermal physics class.

$$n_V(E) = n_E e^{-\frac{E}{k_B T}} \quad (33.2)$$

where $n_E dE$ is the number of molecules per unit volume having energy between $E = 0$ and $E = dE$. This distribution function is called the *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 $k_B T$.

33.2 Distribution of Molecular Speeds

Since there is a distribution of energies, we expect our gas molecules to have a distribution of velocities. That is the same as saying the molecules in the gas

do not all go the same speed. Again we won't derive this (at least not in this class!), but the distribution should depend on temperature, T , since we know the internal energy is tied to temperature. The distribution is as follows:

$$N_v(v) = 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}} \quad (33.3)$$

where m is the mass of the molecule. The kinetic energy of the molecules is hiding in the exponent, so we could write this as

$$N_v = 4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} v^2 e^{-\frac{K}{k_B T}} \quad (33.4)$$

Note that this is of the form for our distribution function because for an ideal gas the only internal energy we have is kinetic energy, $E = K$ so we have $E/(k_B T)$ as the exponent. So we used this to turn our energy distribution function into a speed distribution function.

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

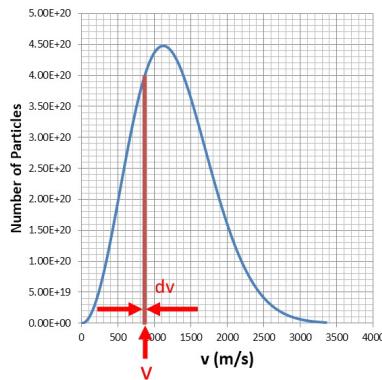
$$dN = N_v dv \quad (33.5)$$

so there should be

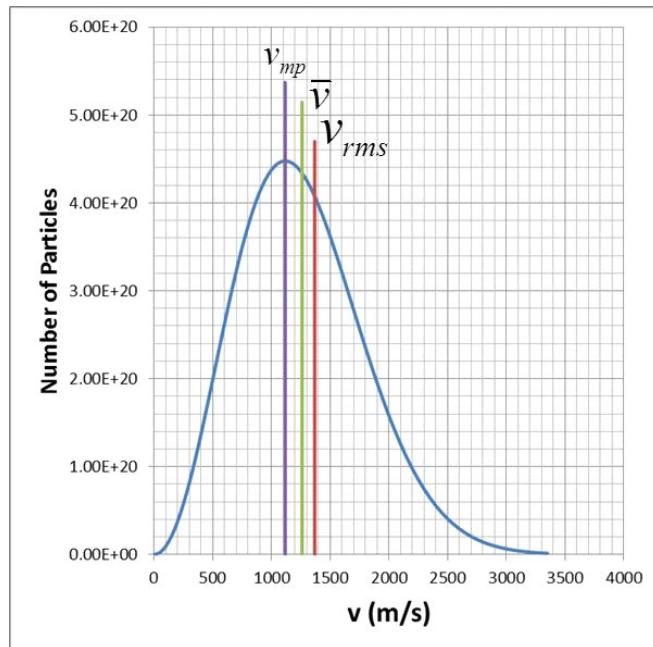
$$N = \int_0^\infty N_v dv \quad (33.6)$$

total molecules. This tells us that the $4\pi N \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}}$ is a *normalization constant* that makes sure our integral of $N_v dv$ gives us N , the total number of particles.

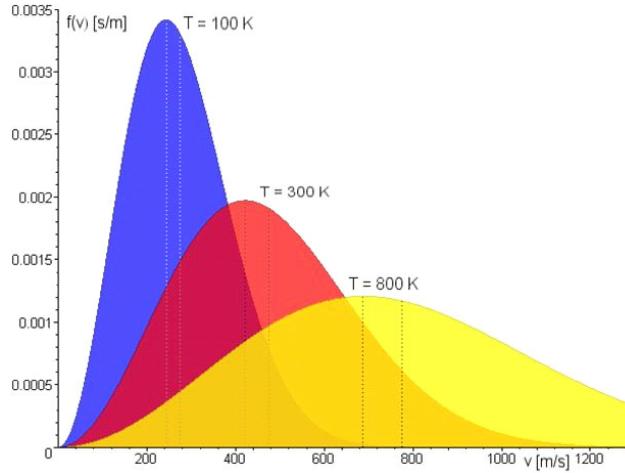
If we plot N_v vs. v we get the figure below. We get a plot that is a little like our grade distribution plot, only for molecular speeds. The number of molecules with speeds between v and $v + dv$ is the area under the blue curve at v . It would be the area of the thin red rectangle right at v with width dv shown in the next figure.



The peak of the curve tells us the most probable speed, that is, the speed the most molecules have, v_{mp} (like in our grade example, the most students got an “A” because that bar was the tallest). 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 . We know from our calculus experience how to find a maximum. We take a derivative

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

and set it equal to zero. I didn't spend a lot of time simplifying, because all the factors out front will go away when we set it equal to zero:

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

$$v_{mp} = \sqrt{\frac{2k_B T}{m}} \quad (33.7)$$

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 expression to be in terms of v_{rms} . And it gives us some insight into using a distribution function.

The average value of v^n is given by

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

Think of this as summing up v^n but using the fact that we already know that for each v there are $N_v dv$ molecules with speed v . So our distribution function makes the integral easier. That 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

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

You might be thinking that this integral looks a little hard to do. You would be right. You could use a symbolic math program to do this. Or you can dust off your second semester calculus knowledge. I prefer the symbolic math program, but let's do it the calc II way here. Looking just at the integral

$$\int_0^\infty v^3 e^{-\frac{mv^2}{2k_B T}} dv = \int_0^\infty v^2 e^{-\frac{mv^2}{2k_B T}} v dv$$

Let's do a substitution. let $x = v^2$, $dx = 2vdv$ and $a = -\frac{m}{2k_B T}$ then we have

$$\int_0^\infty v^3 e^{-\frac{mv^2}{2k_B T}} dv = \int_0^\infty x e^{ax} dx$$

and now we have one of those many integration identities that you memorized (and now have forgotten) from Calc II. So we look in an integral table and find

$$\int_0^\infty x e^{ax} dx = \frac{e^{ax}}{a^2} (ax - 1)$$

This matches our integral that we need. We now know the answer. We just have to undo our substitution.

$$\begin{aligned}
\int_0^\infty v^3 e^{-\frac{mv^2}{2k_B T}} dv &= \left. \frac{e^{\frac{-m}{2k_B T} v^2}}{\left(\frac{m}{2k_B T}\right)^2} \left(\left(-\frac{m}{2k_B T}\right) v^2 - 1 \right) \right|_0^\infty \\
&= \left[\frac{e^{\frac{-m}{2k_B T} (\infty)^2}}{\left(\frac{m}{2k_B T}\right)^2} \left(\left(-\frac{m}{2k_B T}\right) (\infty)^2 - 1 \right) \right] - \left[\frac{e^{\frac{-m}{2k_B T} (0)^2}}{\left(\frac{m}{2k_B T}\right)^2} \left(\left(-\frac{m}{2k_B T}\right) (0)^2 - 1 \right) \right] \\
&= 0 - \left[\frac{1}{\left(\frac{m}{2k_B T}\right)^2} \left(\left(-\frac{m}{2k_B T}\right) (0)^2 - 1 \right) \right] \\
&= \frac{1}{\left(\frac{m}{2k_B T}\right)^2}
\end{aligned}$$

And we can put our result back in our equation for $\overline{v^1}$

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

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

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

Then we have

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

It is time to look up our integral in a table of integrals. My table tells me that

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

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

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

so

$$\overline{v^2} = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \left(-\frac{1}{8 \left(\frac{m}{2k_B T} \right)^{\frac{5}{2}}} (-3\sqrt{\pi}) \right)$$

or

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

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

$$v_{rms} = \sqrt{\frac{3k_B T}{m}} \tag{33.8}$$

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. And you might object that we have already seen this form for v_{rms} and it was much easier to find last time! But it is worth seeing that our distribution of molecular speeds produces this result. It gives us confidence in our distribution function. Notice that, indeed, v_{rms} is close to the average and most probable speeds.

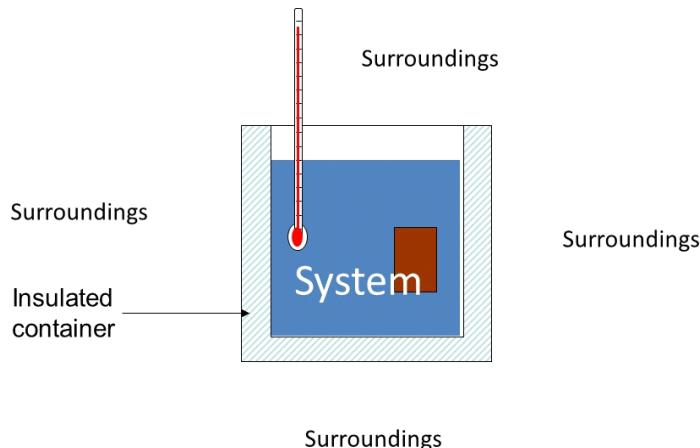
That we have some speeds higher than others explains why things do not evaporate or all boil away at once.

A few rms Speeds for Gasses		
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

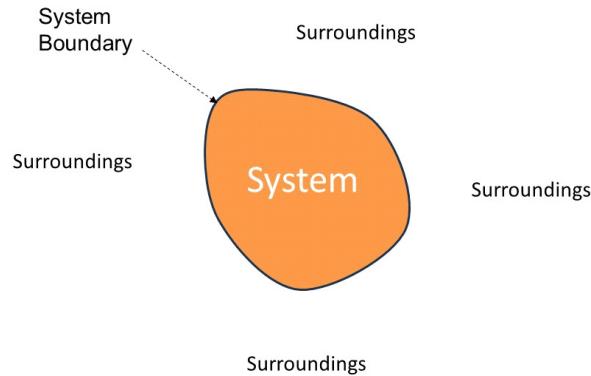
Armed with a molecular model for gasses, let's return to our state variables for a system and see how to find the state variables for different thermodynamic processes. Let's start this by looking at some special ways to change thermodynamic systems.

33.3 Thermodynamic systems

We are going to talk a lot about thermodynamic systems. But what is a thermodynamic system? It is just the group of material that we care about, that we want to know the state variables of in our problem or experiment. In our calorimetry problems we had an isolated container. The stuff in the container was our thermodynamic system and the container and air and table and room and everything else around the container is what we call the surroundings.



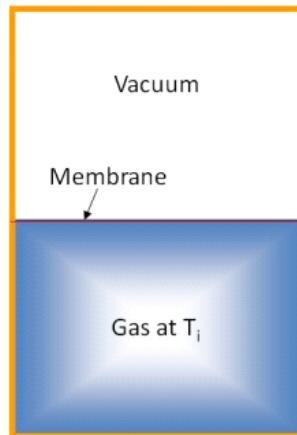
We could draw a more general picture of a thermodynamic system



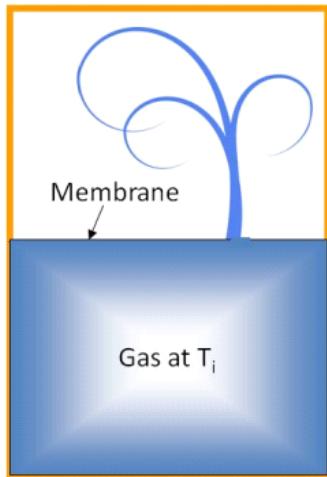
In our first system example we had an insulated container. That insulation prevents energy transfer by heat, and if it is ridged it prevents energy transfer by work. But this is a special case. We call such a system a *closed system*. But most thermodynamic systems do exchange energy with their surroundings. We call them *open systems*.

In our problems, we generally think of the material inside of our system as being in thermodynamic equilibrium. This is not always the case.

Consider the situation shown below



We have a gas in one side of a sealed container and a vacuum on the other side. A membrane separates the two. We have a very definite temperature in the gas, and also a definite temperature in the vacuum. But if I puncture the membrane,



gas goes swirling into the vacuum. The density will not be constant as the gas swirls. Where there is gas, we would say we have a temperature closer to T_i . But where there is no gas, we have a temperature nearer 0 K. It takes a while for the gas to come to equilibrium. During this process we can't say that we have a definite temperature, or definite pressure for the system. Higher physics classes (and engineering classes) deal with the difficult situation that exists before equilibrium is achieved. For now, we will only look at situations that have reached an equilibrium, or are not too far from equilibrium. We call states that change slowly so they are never far from equilibrium, *quasi-static*. The situation we have just described is called a *free expansion* because the gas is allowed to expand with no outside influences. A free expansion is *not* a quasi-static process. Let's think of how we might make a quasi-static process happen.

33.4 Reversible and Irreversible Processes

A reversible process is one in which every point along some path is an equilibrium state and one for which the system can be returned to its initial state along the same path.

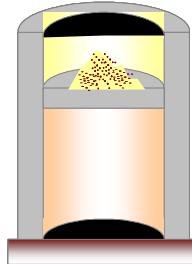
All natural processes are known to be irreversible. Reversible processes are an idealization, like frictionless surfaces or massless strings. But some real processes are good approximations to reversible processes.

Our free expansion, of course, is not a reversible process. Imagine trying to get all the gas atoms back through the hole in the membrane!

We could try to put the gas back into a smaller volume by putting a piston in the system and applying a force. We would be transferring energy by work. This would increase the temperature, so we would have to let energy transfer out of the system (by heat) to keep the temperature the same. This would allow the system to return to its final state, but the surroundings will have changed.

We will have heated up the surroundings. This is not a reversible process!

But suppose we apply our force on our piston in a quasi-static way.



Imagine dropping sand on a frictionless piston to compress a gas. Then slowly removing the sand. The system stays almost in equilibrium as it changes, so that it returns to a new equilibrium right away. After all, you only moved one grain of sand at a time. This process is almost reversible. Of course, you still have to do work in moving the sand, so the outside environment is changed, so even this process is not truly reversible. But it is a good mental model for the sort of processes we will study in this class; slow, almost in equilibrium all the time.

For our quasi-static process we can say we stay in thermodynamic equilibrium. Recall that we have one “law” of thermodynamics. The zeroth law. It tells us that if two objects are in thermodynamic equilibrium they have the same temperature. From what we have done in this and the last lecture we now see this zeroth law is telling us about forces and collisions between molecules. When the molecules of the two objects have the same average kinetic energy, no net energy is transferred between them and this is thermodynamic equilibrium.

Chapter 34

34 First Law and Work

2.3.2, 2.3.3

We have studied the zeroth law of thermodynamics. Zero is an odd place to start numbering (unless you are programing a computer in python or C++). That no energy is transferred between two objects in thermodynamic equilibrium is so obvious that for a while it went unstated. But now we recognize from our kinetic theory of gasses why this must be the case and we can label this fact a law of thermodynamics. But there is more than one law of thermodynamics (other wise, why would we number them at all?). Let's learn another law of thermodynamics, which has to do with conservation of energy in thermodynamic systems.

Fundamental Concepts

- The first law of Thermodynamics is just conservation of energy
- ΔE_{int} is an invariant quantity in thermodynamics
- Work in thermodynamic processes can be described as the integral $w_{int} = \int_{V_i}^{V_f} P dV$
- $\Delta E_{int} = Q + w_{int}$ is the first law of thermodynamics

34.1 Conservation of Energy: First Law of Thermodynamics

Remember when we studied energy in PH 121. We used the work energy theorem

$$\Delta K = w$$

and expanded the work side to show the different types of work explicitly. We have work done by conservative forces w_c and work done by non-conservative forces (dissipative forces like friction) w_{nc} .

$$\Delta K = w_c + w_{nc}$$

and we could further divide up the work by categorizing the source of the work, say, work done by the force of gravity, w_g , or work done by a spring force w_s .

$$\Delta K = w_g + w_s + w_{nc}$$

and we could go on in more complicated systems listing all the different kinds of work.

We relabeled these different types of work as potential energies. For example,

$$\begin{aligned}\Delta U_g &= -w_g \\ \Delta U_s &= -w_s\end{aligned}$$

and rewrote the work energy theorem

$$\Delta K + \Delta U_g + \Delta U_s = w_{nc}$$

This gave us a way to think of mechanical energy in a system. But we realize that we have not covered all types of energy yet. Suppose we consider your car. You put gas in the car. This is a source of energy, a *chemical potential energy*, U_{chem} . So we should really include it.

$$\Delta K + \Delta U_g + \Delta U_s + \Delta U_{chem} = w_{nc}$$

and if your car is a DeLorean that has been modified to use nuclear energy to run, then you might have a nuclear potential energy

$$\Delta K + \Delta U_g + \Delta U_s + \Delta U_{chem} + \Delta U_{nuclear} = w_{nc}$$

and back in PH121 we called the non-conservative work $w_{nc} = \Delta E_{th}$ because friction like forces dissipate mechanical energy by turning them into thermal energy. Then

$$\Delta K + \Delta U_g + \Delta U_s + \Delta U_{chem} + \Delta U_{nuclear} = \Delta E_{th}$$

The quantities ΔU_{chem} , $\Delta U_{nuclear}$, and ΔE_{th} are different than the mechanical energies we studied before. They have to do with the internal workings of the material, itself. We will call an amount of energy that comes from the material, itself *internal energy*. If we think about it, we might expect that temperature has something to do with an internal energy of the object. And we would be right.

34.1.1 Work and Internal energy

Our potential energy terms are all derived from work. Long ago we found that the work done on an object may depend on the path the object takes. We found work using

$$w = \int_{x_i}^{x_f} \vec{F} \cdot d\vec{r} \quad (34.1)$$

and the work was the area under the force vs. displacement curve. But this was work done moving an object as a whole. Now we can do work on an object by compressing it, for example. This is different from moving the whole object around. Let's call this new kind of work *internal work*. This new kind of work should be similar in form to regular mechanical work. We expect it to depend on the details of how the internal work is done.

For mechanical work we defined the potential energy, U , as

$$w = -\Delta U \quad (34.2)$$

Using this concept, we showed that mechanical energy

$$E_{\text{mech}} = K + U \quad (34.3)$$

was often conserved, when $E_{\text{thermal}} = 0$ where E_{mech} was the total mechanic energy. Remember we solved a lot of problems back in Principles of Physics I by knowing that for a closed system, E_{mech} did not change. In other words, E_{mech} was independent of path we take on our F vs. x diagram. This makes E_{mech} much easier to work with. A path independent quantity makes problems easier.

We want another variable to work with that removes the difficulty of worrying about the details of how the internal motion happens (the path). But we can't use the kinetic energy for this. The kinetic energy is part of the mechanical work of the object as a whole, and does not affect the internal workings of the object.

Frictional forces do affect the object. Think of rubbing your hands together. There is mechanical work done by your muscles. There is kinetic energy while your hands move. But your hands heat up. This higher temperature is still there for a while after your hands stop moving. Eventually the temperature of your hands returns to normal. We now know to call this transfer of energy heat, Q .

Now that we remember all of this, we are ready to find our path independent quantity for internal energy. We define the change in internal energy as

$$\Delta E_{\text{int}} = Q + w_{\text{int}}$$

As with the mechanical energy, the change in internal energy does not depend on the details of how the change happens (what we are calling the "path"). The change in energy is used because Q and w are transfer variables, so they *do* depend on the details of the transfer. For small changes we can write

$$\Delta E_{\text{int}} = \Delta Q + \Delta w_{\text{int}}$$

or for very small changes, we could replace the Δ 's with small d 's

$$dE_{int} = dQ + dw_{int}$$

but be careful! these are not true differentials because dw depends on the path! Integration of dw is trick and you would have to think about it carefully. The relationship

$$\Delta E_{int} = Q + w_{int} \quad (34.4)$$

is known as the *first law of thermodynamics*.

Let's look at some systems and think about their internal energies.

34.1.2 An Isolated System

Last lecture we defined isolated systems. An isolated system is a system separated from the rest of the universe. Nothing can act on it. When a system is isolated, we have

$$w_{int} = 0 \quad (34.5)$$

because no work can be done on an isolated system and

$$Q = 0 \quad (34.6)$$

because no heat can flow to or from an isolated system so

$$\Delta E_{int} = w_{int} + Q = 0 \quad (34.7)$$

34.1.3 A Cyclic Process

Suppose we have a process, something that is done to a sample of matter, say a gas, over and over again the say way. We call such a process a *cyclic* process. This is like the pistons in your car going through the same motion over and over again to compress the gas in your engine cylinders. Let's practice with the idea of internal energy, work, heat and PV diagrams for such a process. If a process repeats, it starts and ends in the same state so then

$$\Delta E_{int} = 0 \quad (34.8)$$

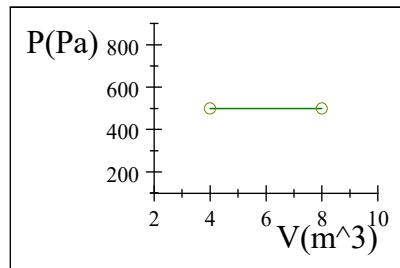
and

$$Q = -w_{int} \quad (34.9)$$

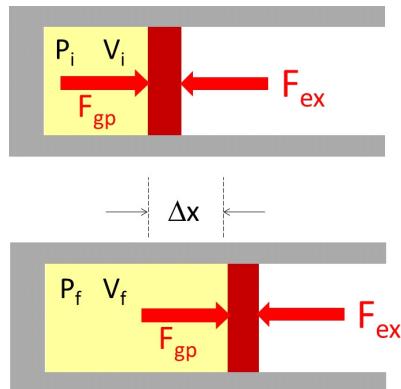
But these examples were too easy. Let's take on a harder problem where we actually find the amount of work done.

34.2 Work

- Consider the work done by a piston as it compresses a gas. It might be good if the amount of work was equal to the amount of energy transfer by heat. We often want this to be true! Otherwise you could melt your engine as your car operates. We want an engine in a car (or train, or airplane) to be a cyclic process. This explains why your car engine surface gets hot. Work is being done in the cylinders in your engine, but to make the process cyclic some energy must be lost by heat or the energy would rise. Unlike your car pistons, let's do this slowly and quasi-statically so that the pressure equals every time we move the piston a little bit.



Then the pressure won't change. Let's call such a process where the pressure doesn't change in the cylinder an *isobaric process*.¹



Lets take a close look at the gas within the piston. At equilibrium we know

$$P = \frac{F_P}{A}$$

or

$$F_P = PA$$

¹There are details of car engines that we won't tackle until later. So not everything in this lecture applies directly to car engines.

This force is the force of the gas pushing on the piston. We could label it

$$\vec{F}_{gp} = F_P$$

Notice that the gas pushes to the right. If we apply an external force to the piston slowly so as to allow the system to remain in thermal equilibrium as we go, then

$$\vec{F}_{ex} = -F_{ex}\hat{x}$$

because we are pushing in the $-x$ -direction. Again, we will move the piston slowly so we don't violate our quasi-static rule. As we move the piston the gas inside the piston adjusts in pressure until we are again in equilibrium. So

$$F_{ex} = F_{gp}$$

but the pressure force F_{gp} pushes the opposite direction

$$\begin{aligned}\vec{F}_{gp} &= +F_{ex}\hat{x} \\ &= PA\hat{x}\end{aligned}$$

We should also note that since neither the piston nor the gas are accelerating, Newton's third law tells us that the force of the gas on the piston and the force of the piston on the gas are equal in magnitude but opposite in direction

$$F_{gp} = F_{pg}$$

$$\vec{F}_{gp} = -F_{pg}\hat{x}$$

When we move the piston to the left, we have a displacement

$$\vec{dr} = -dx\hat{x}$$

so the work done *on the gas by the piston would be*

$$w_{gp} = \int_{x_i}^{x_f} \vec{F}_{ex} \cdot d\vec{r}$$

Remembering how to use a dot product we can write this as

$$w_{gp} = \int_{x_i}^{x_f} F_{ex} dx \cos \theta_{Fx}$$

It is worth remembering that in this formula F_{ex} and dx are magnitudes. Let's write this as

$$w_{gp} = \int_{x_i}^{x_f} |F_{ex}| |dx| \cos \theta_{Fx}$$

to remind us. Substituting in $|F_{gp}|$ for $|F_{ex}|$ (since they have the same magnitude) we have

$$\begin{aligned}w_{gp} &= \int_{x_i}^{x_f} |F_{gp}| |dx| \cos \theta_{Fx} \\ &= \int_{x_i}^{x_f} |PA| |dx| \cos \theta_{Fx}\end{aligned}$$

and the angle between F_{ex} and dx is 180° . If this seems mysterious, look at the last figure, and remember we are really working with F_{ex} . We put in F_{pg} because its magnitude is the same as that of F_{ex} . So 180° is the right angle for our dot product.

$$w_{gp} = - \int_{x_i}^{x_f} |PA| |dx|$$

To get the total work done on the piston we are integrating over the entire path of the piston travels. It is worth noting that P can only be positive, and the same is true for A . Both P and A are constant, since we are using our quasi-static process and allowing the pressure to equalize as we go. So we could write the work done on the piston by the gas as

$$w_{gp} = PA \int_{x_i}^{x_f} |dx|$$

but now we need to consider $|dx|$. We know that dx is a small displacement, and displacements can be negative. And in our case, as the piston goes to the left, dx certainly is negative. But no matter, we don't just have dx , we have $|dx|$. But there is a problem. We got a minus sign from our dot product, but we would like the work done by the piston on the gas to be a positive number. And it looks like our integral will be negative.

But here we can be a little bit clever. If dx really is negative, then the negative hiding inside of dx cancels the negative sign from our dot product. And then

$$w_{gp} = -PA \int_{x_i}^{x_f} dx$$

still gives us positive work (think of what we did with springs in PH121 to deal with $S = -k(x - x_o)$ in work integrals). Now we don't have to worry about integrating with absolute value signs.

Remember that we are using an isobaric process in compressing the gas. That is, we keep the pressure the same but allow the temperature to change, so that the volume and temperature change proportionally $V = (nRT/P)$. So our integral is just

$$\begin{aligned} w_{gp} &= -PA \int_{x_i}^{x_f} dx \\ &= -PA(x_f - x_i) \end{aligned}$$

We can rewrite this result using what we know about the volume of the gas

$$V = A\Delta x$$

so the work done by the gas pushing on the piston is

$$\begin{aligned} w_{gp} &= -P(Ax_f - Ax_i) \\ &= -P(V_f - V_i) \end{aligned}$$

This gives us an idea. We could write the quantity $A dx$ as dV and change the limits of integration from x_f and x_i to V_f and V_i . Then our integral would be

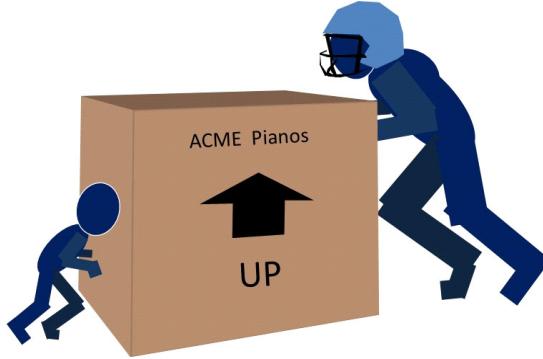
$$\begin{aligned} w_{gp} &= - \int_{x_i}^{x_f} P A dx \\ &= - \int_{V_i}^{V_f} P dV \end{aligned}$$

This gives the same answer for the work done on the piston for our isobaric case

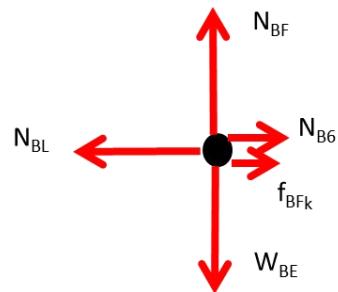
$$w_{gp} = -P(V_f - V_i)$$

But this is the amount of work that the piston does on the gas. Is the gas pushing back on the piston? Of course it is! And that push is also a force, F_{pg} , so it also does work. But let's review what we know about work from PH121.

Suppose we have two people exerting a force on a box. One is a professional American football player, The other is a six-year-old.



They push in opposite directions. How much work is being done? To answer this, let's consider a free-body diagram.



where the subscript B is for the box and the subscript 6 is for the six-year-old

and L is for the football player (In this case, a (L)inebacker). We can see that there will be a net force.

$$F_{net_s} = N_{BL} - N_{B6} - f_{BF}$$

But which way will the box go? We could guess that the football player will push the box *and* the six-year-old to the left. So Δs will be negative. Then the work done on the box by the football player will be

$$\begin{aligned} w_{BL} &= \int_{s_i}^{s_f} N_{BL} ds \\ &= N_{BL} \Delta s \end{aligned}$$

but the work done by the child would be

$$\begin{aligned} w_{B6} &= \int_{s_i}^{s_f} -N_{B6} ds \\ &= -N_{B6} \Delta s \end{aligned}$$

The child's work is negative! What can that mean? Well, for starters, the child's force can't be the force that is making the box move. In fact, the child's force is another obstacle that the linebacker's force must overcome to make the box move. This means that the linebacker must do enough work (push hard enough) to make the box go, *and* to overcome the backward push of the child.

But that is just our case with the gas and piston. The gas pressure force *is* pushing on the piston. But the gas pressure force on the piston is losing ground (the external force is pushing the piston to the left). So we would expect the work done by the gas on the piston to be negative.

Since we assumed a quasi-static process, the two forces must always be nearly the same,

$$|F_{ex}| = |F_{pg}|$$

so the amount of work must be nearly the same, but the sign must be negative for the work on the gas due to the piston's push. We wrote the work done by the piston on the gas as

$$w_{gp} = - \int_{V_i}^{V_f} P dV$$

Then the work done by the gas on the piston must be

$$\begin{aligned} w_{pg} &= - \left(- \int_{V_i}^{V_f} P dV \right) \\ &= \int_{V_i}^{V_f} P dV \end{aligned}$$

that is, the work done on the gas by the piston is the negative of the work done on the piston by the gas.

We are principally interested in what happens to the gas, (we studied objects like the piston in Principles of Physics I) so the first equation is what we are looking for. This is the internal work done on the gas!

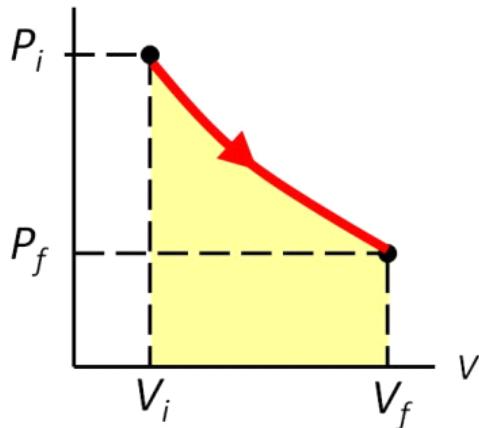
$$w_{gp} = w_{\text{on the gas}} = w_{int} = - \int_{V_i}^{V_f} P dV \quad (34.10)$$

And we could also define

$$w_{pg} = w_{\text{by the gas}} = \int_{V_i}^{V_f} P dV$$

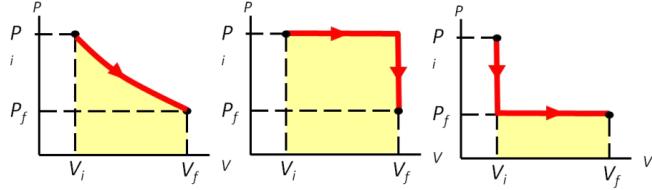
The only difference is the minus sign. We are going to have to be careful!

Still, this looks easy enough, but to integrate we must know how P depends on V during the whole process. It was simple for an isobaric process, but it could be much harder for other processes. We can plot P vs. V to see how it varies for a process. We would expect the integral to be the area under a PV curve. This is very like mechanical work, which was the area under the F - x curve.



We can make a general statement: The work done on a gas in a quasi-static process that takes the gas from an initial state to a final state is the negative of the area under the curve on a PV diagram evaluated between the initial and final states.

Unfortunately, the work done depends on the shape of the curve on the PV diagram. Several are shown below. Each has a different amount of work done (notice the different areas under the curves), but the same initial and final points.



Notice that we have calculated the amount of work done on the gas. For physicists this is normally what we want.

You might be wondering if that is what we want to build car or jet engines. Don't we want the gas to expand to push the piston and make the car go? And of course, you are right in your thinking. Eventually we want to know the work done by the gas on the piston, w_{pg} . But for now, we will look at w_{gp} . But from what we have done we know these two works are very related for a quasi-static process.

$$w_{\text{on the gas}} = -w_{\text{by the gas}}$$

But let's go back to our first law of thermodynamics.

$$\Delta E_{int} = Q + w_{int} \quad (34.11)$$

We know how w_{int} changes on a PV diagram. How do ΔE_{in} and Q change on a PV diagram? For energy transfer by heat, we don't need a pressure or volume change, so a PV diagram isn't a great way to show Q . But we can consider ΔE_{int} on a PV diagram. We already know that for ideal gasses

$$E_{int} = \frac{3}{2}k_B T \quad (34.12)$$

So E_{int} changes with T . But there is no T on a PV diagram! For ideal gasses we can solve this problem. Suppose we have n moles gas in a cylinder. We can solve for P as a function of V

$$P = \frac{nRT}{V}$$

Now we want to let T be constant. Along a line of constant T our E_{int} will be constant.

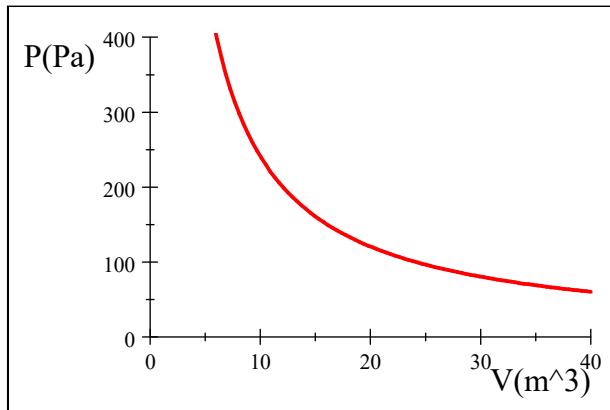
$$P = (nRT) \frac{1}{V}$$

everything in the parenthesis is constant so

$$P \propto \frac{1}{V}$$

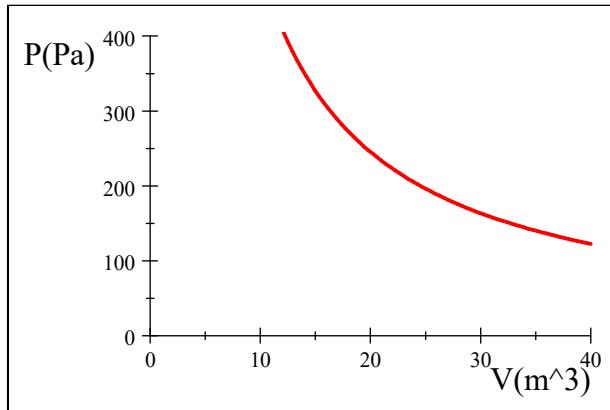
We can plot this for

$n = 1 \text{ mol}$	number of moles
$R = 8.314 \frac{\text{J}}{\text{mol K}}$	Universal gas constant
$T = 290 \text{ K}$	Temperature in Kelvins

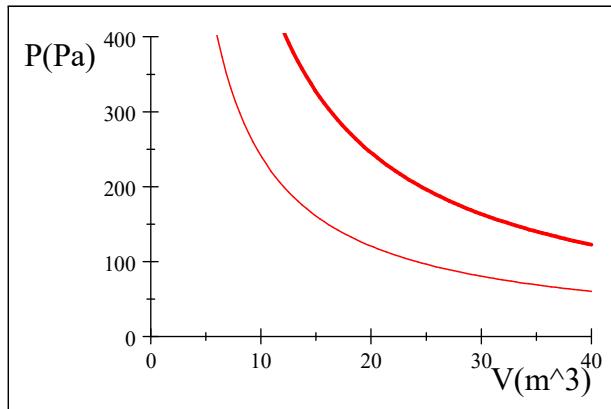


All along this line T is constant so E_{int} is constant. We have already seen such a line on a PV diagram and know it is called an *isotherm*. Now let's plot it again for a different value of T

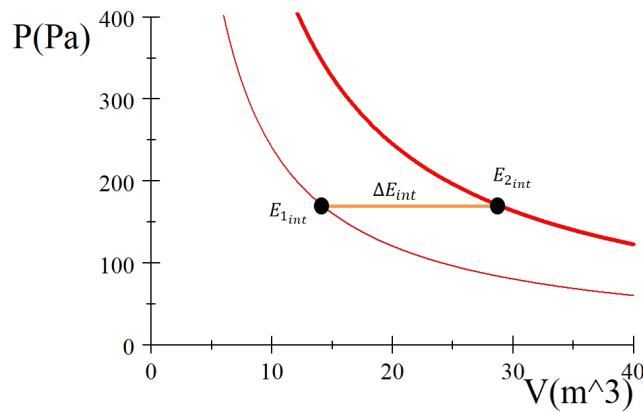
$$\begin{array}{ll} n = 1 \text{ mol} & \text{number of moles} \\ R = 8.314 \frac{\text{J}}{\text{mol K}} & \text{Universal gas constant} \\ T = 590 \text{ K} & \text{Temperature in Kelvins} \end{array}$$



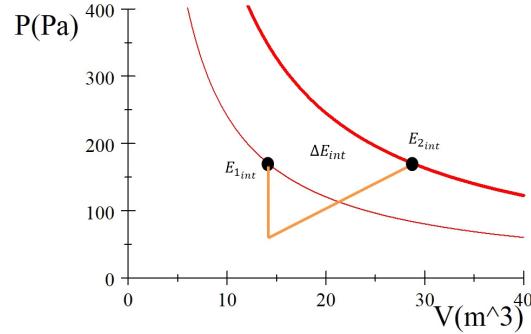
Notice that the curve looks like the same shape, but it has been shifted toward the upper right hand corner of the graph. If we place both the $T = 290$ and $T = 590$ graphs on the same plot, we can better see the shift.



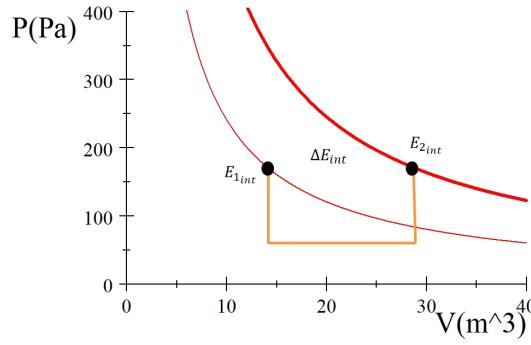
The dark curve is our new pot. The light curve is the old plot for comparison. The curves look a lot alike. They are essentially the same shape. But their location (how far they are from the origin) depends on the temperature, with warmer temperatures moving to the upper right hand part off the graph. So our P vs. V graph can tell us something about temperature. And because $E_{int} = \frac{3}{2}k_B T$ these lines and their positions also tell us about the internal energy. A change in internal energy would be going from one such isothermal line to another.



And because ΔE_{int} is path independent, it doesn't matter how we go from E_{int} . So if we followed this PV path



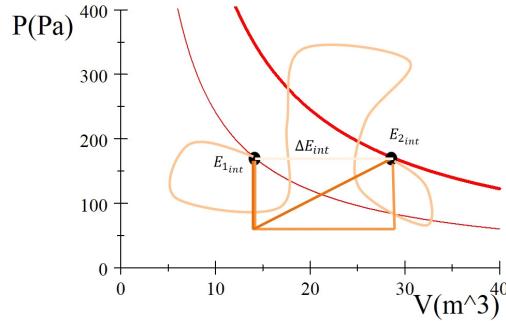
or this PV path



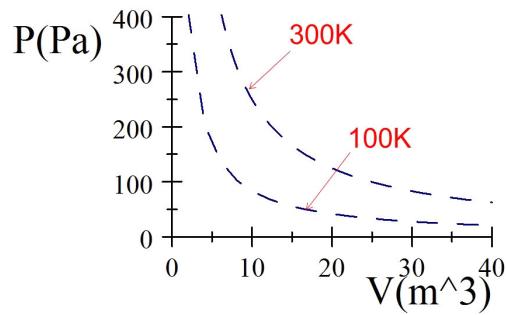
we would get the same ΔE_{int} . A way to think about this is that ΔE_{int} only depends on its starting and end points (think about displacement). So

$$\Delta E_{int} = E_{2,int} - E_{1,int}$$

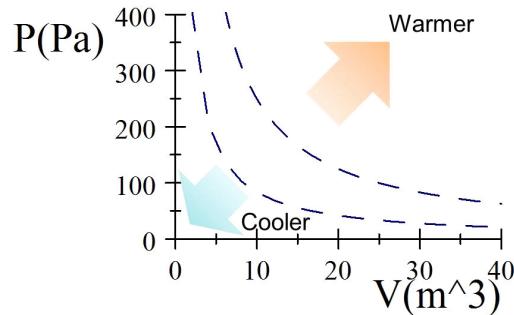
regardless of the path. So any path from $E_{1,int}$ to $E_{2,int}$ gives the same change in internal energy.



Looking at an isothermal process can help us understand our other processes. Suppose we have two different isothermal processes, one at 100 K and one at 300 K. The isotherms look like this on a PV diagram:



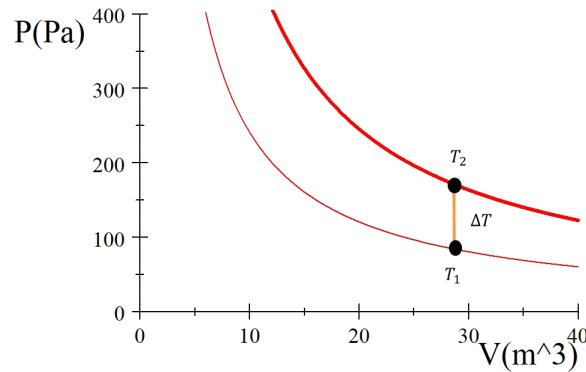
Every point on the 300 K line has the same temperature. Every point on the 100 K line has the same temperature. We can see that points on the PV diagram that are closer to the origin must have lower temperatures. We can see the points farther from the origin have higher temperatures.



Let's go back to energy transfer by heat. We know

$$Q = nC_v\Delta T$$

if the volume doesn't change, we now can see a way to determine Q from a PV diagram.



But this is just for a process where volume doesn't change. Still, our PV diagram does seem to be useful for thinking about thermodynamic process and for filling in our first law equation

$$\Delta E_{int} = Q + w$$

We will work with this some more in our next lecture.

Chapter 35

35 Thermodynamic Processes 2.3.4, 2.3.5

We found in our last lecture that the first law of thermodynamics gives us the change in internal energy

$$\Delta E_{int} = Q + w$$

and that we could illustrate the work, w , process on a PV diagram and we even found a way to show ΔE_{int} on a PV diagram. We did find a way to show Q for the special case of a process that doesn't change volume. But that seems so specific. Aren't there infinite different paths from a specific initial internal energy, $E_{i,int}$ to a final internal energy, $E_{f,int}$? Of course, the answer is yes. But many of these paths from $E_{i,int}$ to $E_{f,int}$ would require a numerical solution (this means use a computer to do the calculation). So instead, we are going to learn a few special processes that are easy to calculate by hand. We will limit our study mostly to these special processes. In industry you might need to go beyond these simple processes, but they give us great insight into how internal energy changes and we can even use them (mostly) to understand complicated systems like gasoline engines. So let's get started and define our special processes.

Fundamental Concepts

- There is a sign convention for energy transfer. If energy leaves it is negative, if it comes into the system it is positive.
- There are two molar specific heats for gasses, C_V and C_P .
- The two molar specific heats are related $C_P = C_V + R$
- There are four special processes for which it is easy to find ΔE_{int} , Q , and w , and three of them are isochoric, isobaric, and isothermal
- For an isochoric process $w_{int} = 0$, $Q = nC_V\Delta T$, $\Delta E_{int} = nC_V\Delta T$

- For an isobaric process $w_{int} = -P\Delta V$, $Q = nC_P\Delta T$, $\Delta E_{int} = nC_P\Delta T - P\Delta V$
- For an isothermal process $w = -nRT \left[\ln \frac{V_f}{V_i} \right]$, $Q = -w$, $\Delta E_{int} = 0$

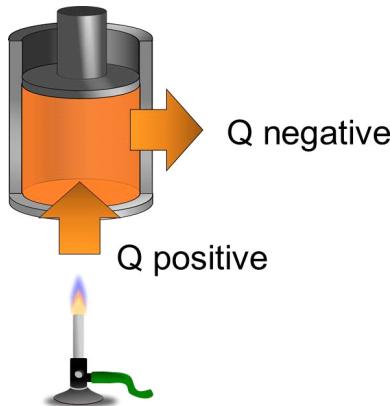
35.1 Heat and systems

Heat is a transfer of energy, so it can't be a property of the system, itself. It is an interaction of the system with its environment. This means that heat is not a state variable. It will depend on process path on a PV-diagram, like work does. Only the combination

$$\Delta E_{int} = Q + w$$

is path dependent.

It is customary to refer to energy gained from an environment as positive heat, and energy lost to the environment as negative heat.



This is a sign convention. We could totally define this the other way. But it is useful for everyone to agree to one system of signs, and the system used in physics today is that energy is negative when it leaves a system, and positive when it enters a system. We will need to remember this for our problems.

It is useful to compare the sign convention for heat with that we built for work (on the gas).

	Work	Heat
Interaction mechanism	Mechanical	Thermal
Process	Macroscopic forces acting on the system	Microscopic collisions between gas particles in the system
Process Requires	External force and displacement	Temperature Difference
Positive when	A gas is compressed: mechanical energy is transferred in to system	The environment is at a higher temperature than the system
Negative when	A gas expands: mechanical energy is transferred out of the system	The environment is at a lower temperature than the system
Equilibrium condition	No net force or torque on the system	System and environment are at the same temperature

Just like with work, the definition of the word “heat” takes some practice to get used to.

Question 123.11.2

It is really important to not use the word “heat” as we use the word “temperature.” Temperature is proportional to the internal energy,

Question 123.11.3

Question 123.11.4

$$E_{int} \propto T$$

as we will show soon. But Q is a transfer of energy. A gas can “have” a temperature, but it “experiences” a transfer of energy. That is why it shows up in the first law as a mechanism for the change in internal energy.

$$\Delta E_{int} = Q + w$$

This is not intuitive because of our left-over language from the caloric theory. So we must continually be on our guard, because poor language leads to poor decisions in experiments and engineering designs.

It would help if we all said things like “to do heat” instead of “heat” when we are warming things. That is what we do with work. We “do work” so we should “do heat,” and this may be a good way to think about it, even though to say “do heat” like this would make our roommates wonder.

Question 123.11.5

We now know that ΔE_{int} is a state variable. It does not depend on path. But it is just one of many state variable for a gas. For ideal gasses, we have

$$P, V, n, T, \Delta E_{int}$$

all as state variables. It takes all of these quantities to describe an ideal gas.

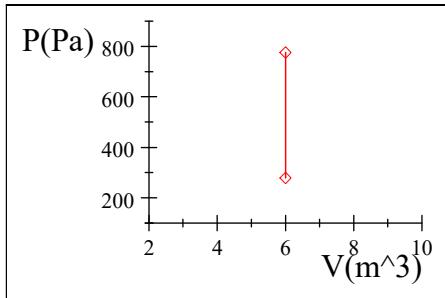
For example, we know that

$$E_{int} \propto T$$

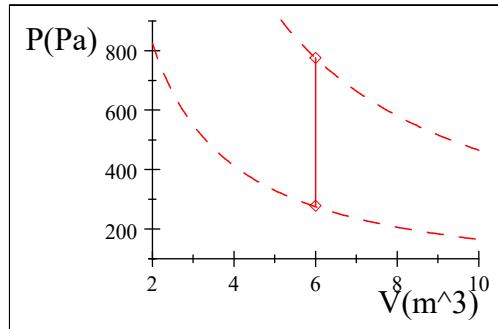
but this does not tell us ΔE_{int} . The quantity ΔE_{int} only tells us about the change in E_{int} . We cannot tell what path was used to accomplish ΔE_{int} . So our first law does not tell us P, V , or n either. We will have to use the first law in cooperation with the ideal gas law.

35.1.1 Constant Volume Process

We ended our last lecture with a constant volume process. But we didn't give it a name. A constant volume process is called an *isochoric process*. Suppose we have a process that takes us between two states on a PV diagram as shown in the next figure.



We start at one pressure, say $P = 275 \text{ Pa}$, and end at another pressure, say, $P = 775 \text{ Pa}$. In this process, notice that the volume does not change. Would we predict that the temperature went up or down in taking this path? Well, we went from the lower to the upper part of the graph, so we expect that the temperature went up. We could check this by plotting the two isotherms that cross through our points.



Sure enough, our temperature went up. Starting again with the ideal gas law

$$P = \left(\frac{nR}{V} \right) T$$

we can see that when V is constant, P goes like T , so if P went up, so did the temperature.

For this isochoric type process no work is done.

$$w_{\text{on the gas}} = w_{\text{int}} = - \int_{V_i}^{V_f} P dV$$

but V doesn't change so $dV = 0$. and therefore

$$w_{\text{int}} = 0$$

But we do have energy transfer by heat

$$Q = nC_v\Delta T$$

so for our first law

$$\begin{aligned}\Delta E_{int} &= Q + w \\ &= nC_v\Delta T + 0 \\ &= nC_v\Delta T\end{aligned}$$

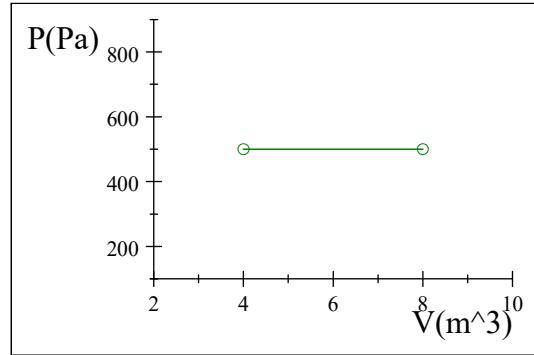
That seems useful! But how would create such a situation? Consider what would happen¹ if you placed an aerosol can in a fire. The aerosol can is rigid, so the volume can't change. But the fire will make the temperature change, and the pressure will change—that is—until the can explodes.

We will also need our differential form for the first law for isochoric processes.

$$dE_{V_{int}} = nC_vdT$$

35.1.2 Constant Pressure Process

But there is more than one way to add energy to a gas. Suppose we envision another process where instead of keeping the volume constant we keep the pressure constant.



We have already considered this process. We called it an *isobaric process*. We found that the work done in an isobaric process is

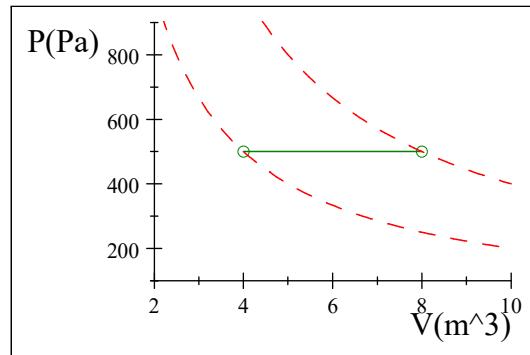
$$w_{\text{on the gas}} = -P(V_f - V_i)$$

¹Only consider this, because it would be irresponsible and dangerous (and illegal in some states) to do this!

Let's review that here

$$\begin{aligned} w_{int} &= - \int_{V_i}^{V_f} P dV \\ &= -P \int_{V_i}^{V_f} dV \\ &= -P(V_f - V_i) \\ &= -P\Delta V \end{aligned}$$

We can see that in this process the pressure is not changing. If we go from left to right along this path, will the temperature rise? We know that going to the right means higher temperature on a PV diagram, so yes it will.



For isochoric processes we defined a molar heat capacity at constant volume C_V that let us find Q

$$Q = nC_V\Delta T$$

In our differential first law equation

$$dE_{int} = dQ + dw$$

and we found that for isochoric processes no work was done so

$$dE_{int} = dQ = nC_v dT$$

but for an isobaric process we have

$$dE_{int} = dQ - PdV$$

But what do we do for dQ if the process is isobaric? It turns out we can play the same trick. Let's define a *molar heat capacity at constant pressure*, C_P . Then

$$Q = nC_P\Delta T$$

and

$$dQ = nC_PdT$$

We found that for ideal gasses

$$C_V = \frac{3}{2}R$$

we need to find a value for C_P . Let's use our ideal gas law

$$PV = nRT$$

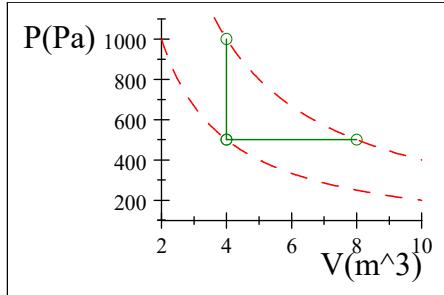
and let the volume and Temperature change, but not P nor n .

$$P\Delta V = nR\Delta T$$

and for an isobaric process, this is our work! Our differential first law would be

$$\begin{aligned}\Delta E_{int} &= Q - w \\ &= nC_P\Delta T - nR\Delta T \\ &= n(C_P - R)\Delta T\end{aligned}$$

But dE_{int} is the same for any two processes that go between the same two temperatures. Suppose we look at a situation where we have the same ΔE_{int} but we make the change two ways, one with an isochoric process and one with an isobaric process.



We already know that for the isochoric process

$$\Delta E_{V_{int}} = nC_v\Delta T$$

Then since

$$\begin{aligned}\Delta E_{V_{int}} &= \Delta E_{P_{int}} \\ nC_v\Delta T &= n(C_P - R)\Delta T\end{aligned}$$

and we can cancel the n 's and the dT 's.

$$C_v = (C_P - R)$$

which gives

$$C_P = C_V + R$$

for an ideal gas. So if we know either C_P or C_V for a gas, we can always find the other. This works pretty well in practice for monotonic gasses (and even for diatomic gasses) at standard temperature and pressure. Not only can we find Q_V and Q_P with molar specific heats, but if we know one molar specific heat, we can easily calculate the other! Since

$$C_V = \frac{3}{2}R$$

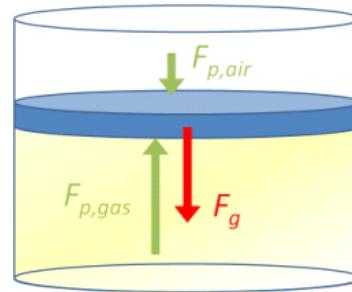
then we know

$$C_P = \frac{5}{2}R$$

So for an isobaric process we have

$$\begin{aligned}\Delta E_{int} &= Q + w \\ &= nC_P\Delta T - P\Delta V\end{aligned}$$

How would we produce such a situation? We saw this in our last lecture, but let's think about it again. Consider a piston in a cylinder. The cylinder has a gas inside, and air on the outside



The piston is free to move up or down. The forces acting on the piston are shown. The weight of the piston pulls it downward, the force due to pressure of the gas, $F_{p,gas}$ pushes up, and the air pressure $F_{p,air}$ pushes down. The net force will be

$$\Sigma F_y = ma_y = F_{p,gas} - F_{p,air} - mg$$

where the mass of the piston is m . We can write this as

$$ma_y = P_{gas}A - P_{air}A - mg$$

If the piston is not accelerating, then

$$P_{gas} = P_{air} + \frac{mg}{A}$$

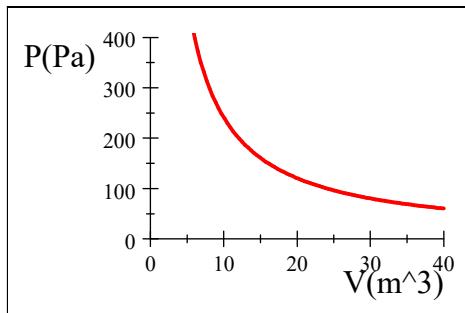
and we find that the pressure of the gas inside is slightly larger than the outside air pressure. What would happen if we heated the gas?

$$PV = NRT$$

would suggest that P might change, but since the piston is free to move, what happens is that the force $F_{p,gas}$ increases when P_{gas} increases, and the piston accelerates upward. It stops when the forces are again balanced. But that means that P_{gas} will only change for a moment, and then we will achieve equilibrium at the same old value of P_{gas} . But something has changed, the volume has increased.

35.2 Constant Temperature Process

We have already done quite a bit of work on this process in our last lecture



We found that $\Delta E_{int} = 0$ for an isothermal process. Then

$$\Delta E_{int} = Q - w$$

gives

$$Q = w$$

but we need to find w . Let's use our ideal gas law for this again.

In this case we know T and n , and that we know P_i and V_i and P_f and V_f . Can we find ΔE_{int} , Q , and w ?

We already know how to find w

$$P = \frac{nRT}{V}$$

so from our work equation

$$w_{int} = - \int_{V_i}^{V_f} P dV \quad (35.1)$$

$$= - \int_{V_i}^{V_f} \frac{nRT}{V} dV \quad (35.2)$$

$$= -nRT \int_{V_i}^{V_f} \frac{1}{V} dV \quad (35.3)$$

$$= -nRT \ln V \Big|_{V_i}^{V_f} \quad (35.4)$$

$$= -nRT [\ln (V_f) - \ln (V_i)] \quad (35.5)$$

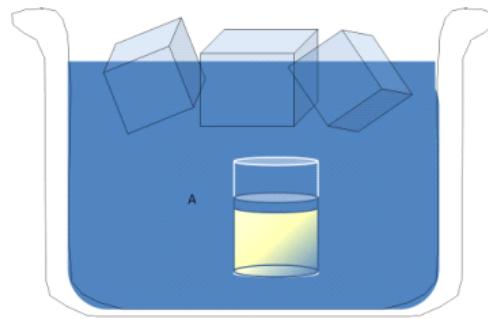
$$w = -nRT \left[\ln \frac{V_f}{V_i} \right]$$

So then

$$\begin{aligned} Q &= -w \\ &= nRT \left[\ln \frac{V_f}{V_i} \right] \end{aligned}$$

and we did not need our molar specific heat equations for this process.

How do we make such a path happen? Consider our piston again.



This time, let's put it in a vat of ice water. We can compress the gas by pushing on the piston. But if the temperature is changed by compressing the gas, it is quickly changed back by the ice water.

We have ΔE_{int} , Q , and w for three of our special processes. This completes three of our four special processes!

Chapter 36

36 Adiabatic Processes and 2nd Law 2.3.6, 2.4.1

Last lecture we found a way to find Q for solids and liquids, $Q = Mc\Delta T$ or $Q = nC\Delta T$. We need to do this for gasses so we can complete our calculations of the terms in the first law of thermodynamics. We will take on this task in this lecture.

Fundamental Concepts

1. There are four special processes and three of them are isochoric, isobaric, and isothermal and the last one is adiabatic
2. The ratio of the molar specific heats is $\gamma = C_P/C_V$
3. The change in internal energy can be expressed as $\Delta E_{int} = nC_V\Delta T$
4. For adiabatic processes we have two new equations $P_iV_i^\gamma = P_fV_f^\gamma$ and $T_fV_f^{\gamma-1} = T_iV_i^{\gamma-1}$
5. For insulation the power transferred through the insulation material is given by $\mathcal{P} = \frac{Q}{\Delta t} = \frac{A(T_h - T_c)}{\sum_i R_i}$
6. The power transferred by radiation is given by $\mathcal{P} = \frac{Q}{\Delta t} = \sigma AeT^4$
 - Doing work on a system can change the system's temperature. This is called the mechanical equivalent of heat.
 - There is a sign convention for energy transfer. If energy leaves it is negative, if it comes into the system it is positive.

- In thermodynamics, we try to find ΔE_{int} , Q , and w for a process. We use the ideal gas law and the first law of thermodynamics to do this.
- We have special processes for which it is easy to find ΔE_{int} , Q , and w
- We are still missing something.

36.1 Constant Q Processes

We found that for isochoric processes there was no work done. And in our first law equation

$$\Delta E_{int} = Q + w$$

the work is half the equation. That made $\Delta E_{int} = Q$. But could we find a process where Q is zero? Then

$$\Delta E_{int} = w$$

We call a process where $Q = 0$ *adiabatic*.

It might see like we are home free, $Q = 0$! We don't need our molar specific heat equations for an adiabatic process either.

But we don't yet know how to find w for an adiabatic process, and therefore we don't know ΔE_{int} . We need to fix this problem. We could do this by finding an expression for the pressure as a function of volume along the adiabatic path, then use our work equation.

$$w = - \int_{V_i}^{V_f} P dV$$

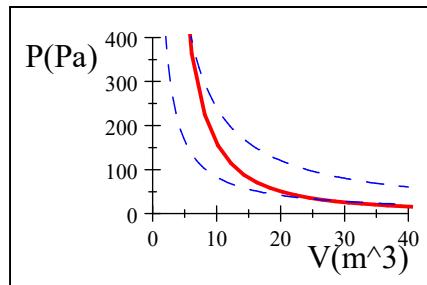
Since

$$\begin{aligned} \Delta E_{int} &= w \\ &= - \int P dV \end{aligned}$$

this would give us both w and ΔE_{int} . But it is not obvious how to find P as a function of V for this path. We need another strategy.

Let's plot an adiabatic process on a PV diagram to see what it looks like. An adiabatic process goes from one temperature to another temperature, and

$$\Delta E_{int} \propto \Delta T$$



And further remember that ΔE_{int} does not depend on path. So we could find ΔE_{int} for an adiabatic path, and we could find ΔE_{int} for a isochoric path, and as long as the ΔT was the same for both paths, the ΔE_{int} would be the same as well! Also remember that for an isochoric path there is no work done. So

$$Q_{isochoric} = \Delta E_{int}$$

And we know an equation for $Q_{isochoric}$!

$$Q_{isochoric} = nC_V\Delta T$$

So if we know T_i and T_f , then we could find ΔE_{int} by finding the equivalent $Q_{isochoric}$ for the same two temperatures.

$$\Delta E_{int} = nC_V\Delta T$$

This is tremendous! We have an equation for ΔE_{int} , and $w = \Delta E_{int}$, so with $Q = 0$ we have a complete set.

But it turns out that we can add two more equations to our adiabatic set that will really help.

The first is that relationship between pressures and volumes along an adiabatic path. We now know for adiabatic processes we can say

$$\Delta E_{int} = nC_V\Delta T$$

so long as we start with the same T_i and end with the same T_f . A very small change in the internal energy would be

$$dE = nC_VdT$$

Now for our adiabatic process we previously noted that there is no energy transfer by heat done, so

$$\begin{aligned}\Delta E_{int} &= 0 + w \\ &= \int PdV\end{aligned}$$

Then a very small amount of internal energy change, dE_{int} would be

$$dE_{int} = -PdV$$

Let's set these two expressions for dE_{int} equal to each other.

$$-PdV = nC_VdT$$

or

$$-\frac{PdV}{nC_V} = dT$$

Now we know the ideal gas law

$$PV = nRT$$

if we take a total differential for a sample of n moles of a gas

$$PdV + VdP = nRdT$$

we can substitute in our expression we have for an adiabatic process for dT

$$PdV + VdP = -nR \frac{PdV}{nC_V}$$

or

$$PdV + VdP = -R \frac{PdV}{C_V}$$

We can use our relationship between C_V and C_P that we found earlier

$$C_V = C_P - R$$

so

$$R = C_P - C_V$$

and let's substitute this in for our R in our total differential equation

$$PdV + VdP = -\frac{C_P - C_V}{C_V} PdV$$

or

$$PdV + VdP = \frac{C_V - C_P}{C_V} PdV$$

Now let's divide through by PV . It's not apparent that this will help, but it does so

$$\frac{PdV}{PV} + \frac{VdP}{PV} = \frac{C_V - C_P}{C_V} \frac{PdV}{PV}$$

and canceling the extra V 's and P 's gives

$$\frac{dV}{V} + \frac{dP}{P} = \frac{C_V - C_P}{C_V} \frac{dV}{V}$$

rearranging terms gives

$$\frac{dP}{P} = \left(\frac{C_V - C_P}{C_V} - 1 \right) \frac{dV}{V}$$

Lets get a common denominator

$$\begin{aligned} \frac{dP}{P} &= \left(\frac{C_V - C_P}{C_V} - \frac{C_V}{C_V} \right) \frac{dV}{V} \\ \frac{dP}{P} &= \left(-\frac{C_P}{C_V} \right) \frac{dV}{V} \end{aligned}$$

Lets define a new quantity

$$\gamma = \frac{C_P}{C_V} \tag{36.1}$$

so we have

$$\frac{dP}{P} = -\gamma \frac{dV}{V}$$

Which is exciting. We are just down to P 's and V 's and a gamma. We may be able to get an expression of P as a function of V yet! We can integrate this last equation

$$\begin{aligned}\int \frac{dP}{P} &= -\gamma \int \frac{dV}{V} \\ \ln(P) + \zeta_P &= -\gamma \ln(V) + \zeta_V\end{aligned}$$

where ζ_P and ζ_V are the constants of integration. We can write this as

$$\begin{aligned}\ln(P) + \gamma \ln(V) &= -\zeta_P + \zeta_V \\ &= \zeta\end{aligned}$$

where $\zeta = \zeta_V - \zeta_P$ is a constant. And now we need to remember a few things about logs.

$$a \ln b = \ln b^a$$

so we can write our equation as

$$\ln(P) + \ln(V^\gamma) = \zeta$$

and

$$\ln a + \ln b = \ln(ab)$$

so our equation becomes just

$$\ln(PV^\gamma) = \zeta$$

or we can exponentiate both sides

$$PV^\gamma = e^\zeta \tag{36.2}$$

where e^ζ is still just a constant.

Note that we have done what we set out to do (sort of). We have an expression of how P varies with V .

$$P = \frac{e^\zeta}{V^\gamma}$$

We could put this into our work equation and integrate to find the work done in an adiabatic process!

$$\begin{aligned}\Delta E_{int} &= w \\ &= - \int \frac{e^\zeta}{V^\gamma} dV\end{aligned}$$

But also note that we have a great conservation equation for an adiabatic process.

$$PV^\gamma = \text{constant}$$

or

$$P_i V_i^\gamma = P_f V_f^\gamma \quad (36.3)$$

and we know that conservation equations come in very handy in solving problems! Let's add this last equation to our set of equations for adiabatic processes.

We can get a second equation for adiabatic processes relating the volume and temperature. If we know an initial state

$$P_i V_i = nRT_i$$

we can use our adiabatic conservation equation from above to write

$$P_f V_f^\gamma = P_i V_i^\gamma$$

we can solve the ideal gas law for P_i and

$$P_i = \frac{nRT_i}{V_i}$$

and use this in our equation

$$\frac{nRT_f}{V_f} V_f^\gamma = \frac{nRT_i}{V_i} V_i^\gamma$$

to yield

$$\frac{T_f}{V_f} V_f^\gamma = \frac{T_i}{V_i} V_i^\gamma$$

or

$$T_f V_f^{\gamma-1} = T_i V_i^{\gamma-1} \quad (36.4)$$

This is another conservation equation for adiabatic processes! We will add this to our equation set for adiabatic processes

These equations are great shortcuts for problem solving *but remember they are only valid for adiabatic processes!*

Notice that we never did complete the integral for work

$$\Delta E_{int} = w = - \int \frac{e^\zeta}{V^\gamma} dV$$

We totally could at this point, but we already know the answer to be

$$\Delta E_{int} = nC_V \Delta T$$

where the ΔT is the same as it would be for our actual adiabatic process. Since we know this and it is easy to calculate, we won't often do the integral

We still need to find γ for our adiabatic processes. We need the ratio of C_P to C_V or γ . We now know that

$$C_V = \frac{3}{2}R$$

$$C_P = \frac{5}{2}R$$

and we may take the ratio of these values

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

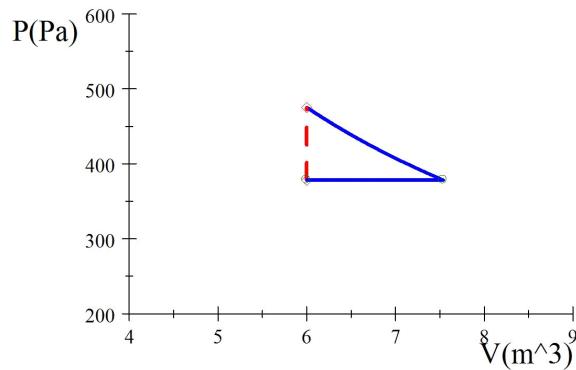
This works well for monotonic gasses, but fails badly for more complex gasses. So we see our ideal gas formulation starts to break down with more complex gasses at this point. To go farther, we would need to include the rotational and vibrational energy of the molecules. Still, we have come a long ways with our simple ideal gas assumptions!

Note that for solids and liquids ΔV is very small so very little work is done. This means that $C_P \approx C_V$ which is why we could get away with only one value for the molar heat capacity in the tables for solids and liquids.

36.1.1 Path Dependence

We should do a problem that shows the path dependence of thermodynamics work. Notice that how we compress the gas (quickly vs. slowly, etc.) makes a difference on our PV diagram. For example, quick actions would not let energy leave by heat, so temperature would rise. Slow actions would allow enough time for energy to leave, allowing the temperature to stay the same. So when we say there is a “path dependence” for thermodynamic work, it is equivalent to saying we have different amounts of work for “different paths” on a PV-diagram. The PV-diagram shows the different physical paths as different lines on the graph. Here is an example.

Suppose we change 1 mol of an ideal gas from one state to another by two different paths. The two paths are shown in the next figure.



We recognize the blue path as a combination of an isobaric and an isothermal process. So for the blue path, the work done will be a combination of

$$\begin{aligned} w_{isobaric} &= -P(V_f - V_i) \\ w_{isothermal} &= nRT \left[\ln \frac{V_i}{V_f} \right] \end{aligned}$$

Then

$$\begin{aligned} w_{blue} &= -P(V_f - V_i) + nRT \left[\ln \frac{V_i}{V_f} \right] \\ &= -378.29 \text{ Pa} (7.5384 \text{ m}^3 - 6 \text{ m}^3) \\ &\quad + (1) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (343 \text{ K}) \left[\ln \frac{7.5384 \text{ m}^3}{6 \text{ m}^3} \right] \\ &= 650.9 \text{ J} \end{aligned}$$

But the red path is an isovolumetric path. We know that for isovolumetric paths the work is zero!

$$\begin{aligned} w_{red} &= - \int_{V_1}^{V_2} P dV \\ &= 0 \end{aligned}$$

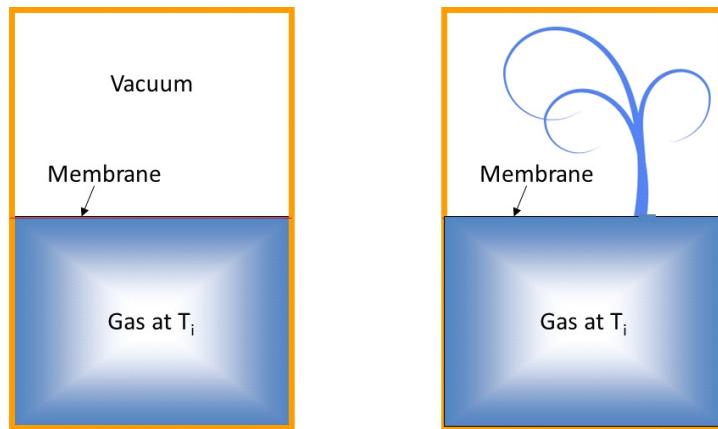
and we can see that zero is very different than 651 J! Indeed the work for each PV-diagram path is different. But why?

The red path is what we would get if we put a sealed strong container in a fire. The fire would provide energy to the gas in the container by heat. But the strong walls of the container would keep the gas from expanding. The pressure would rise and the temperature as well. But for the blue path we envision a movable piston in a cylinder of gas. When the fire adds energy by heat, the piston rises, keeping the pressure the same. The temperature does go up. Then we remove the fire, and slowly compress the gas back to its original volume. By doing this slowly the work we do adds a little bit of energy to the gas, but the energy can leave by heat, so the temperature of the gas stays the same.

Note that these are really different processes! So it is no wonder that the internal work done is different. And we see all of this with our PV-diagram!

36.2 Irreversibility

We should ask, are our special processes reversible? We found that for a free expansion the process was not reversible.



But how about for our special processes: It would make our math easier if they were reversible.

For an isochoric process we know $w_{int} = 0$, $Q = nC_V\Delta T$, $\Delta E_{int} = nC_V\Delta T$. It seems like we can change the pressure of a gas, say by transferring energy by heat, and let the temperature rise, but then cool the cylinder of gas back to the same pressure all without changing the volume. It seems like this would be doable.

For an isobaric process we know $w_{int} = -P\Delta V$, $Q = nC_P\Delta T$, $\Delta E_{int} = nC_P\Delta T - P\Delta V$. It seems like we could make a cylinder with a free floating position that could go up or down. We could transfer energy by heat and let the volume increase. But then transfer energy back out of the cylinder to restore it to the same pressure. This seems reversible.

For an isothermal process $w = -nRT \left[\ln \frac{V_f}{V_i} \right]$, $Q = -w$, $\Delta E_{int} = 0$. Could we do work on a cylinder of gas that is in an ice water bath and transfer energy out by heat such that the temperature stays the same, then pull the piston back out allowing energy to transfer back in by heat, still keeping the temperature the same? This also seems possible.

Finally, for an adiabatic process we now know that $w = \Delta E_{int}$, $Q = 0$ and $\Delta E_{int} = nC_v\Delta t$ for an equivalent process. So we insolate our cylinder and do some work. That increases the internal energy (and the temperature) but we pull the piston back out and the internal energy reverts back. These all seem like reversible processes. But notice that in each case, though the cylinder contents came back to the same state, the surroundings would not be the same. In each case energy was transferred either work or by heat. And the surroundings had to supply that work or energy transfer by heat. And we expect that in energy transfer by heat that not only was our cylinder affected by a stove or hotplate and the air in the room would also have a larger temperature. In the case of work our diligent coresearcher that kept pushing and pulling pistons is now tired. The surroundings are different. Causing a change in the surroundings is something we need to think about, and it gives us our first version of the *Second*

Law of Thermodynamics.

36.3 The Second Law of Thermodynamics

We did a lot with our first law of thermodynamics (using our ideal gas law). But there is a second law of thermodynamics. It is a little harder to state in a nice equation. So we are going to build up to it a little at a time. Our first version is due to a researcher called Clausius, so it is called the Claussius statement of the Second Law of Thermodynamics.

It is impossible to make a cyclical machine that transfers energy spontaneously
from a colder object to a hotter object

This seems obvious. Refrigerators have to be plugged in, so do air conditioners. They require an input of energy to run. But this is just what we have been discussing. for a refrigerator or an air conditioner to work the surroundings have to change. Energy in the form of electricity has to be supplied. And that not only changes the surroundings it also changes our budget. In our next lectures, we will form more mathematical statements of the second law.

Chapter 37

37 Producing Useful Work

2.4.2, 2.4.3

So far we have talked about doing work on gas in cylinders. That may not seem all that useful. But if we think that car engines have gas in cylinders with pistons, then it may seem like this could be important.

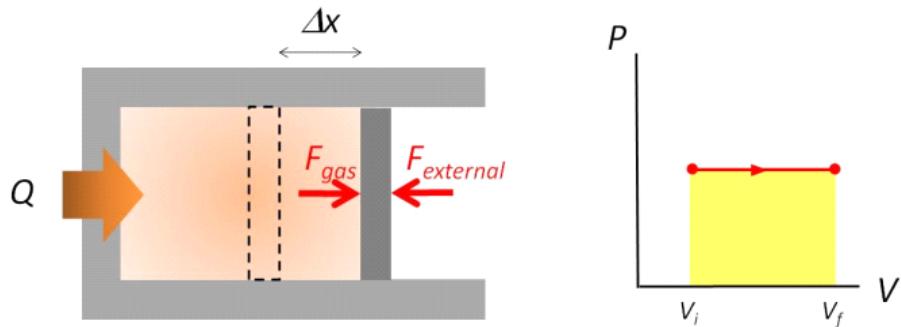
Fundamental Concepts

- $W_{\text{useful}} = -W_{\text{on the gas}}$

37.1 Useful work.

Question 123.17.1

Suppose that we want to produce useful work from a thermodynamic device. For building such a device we have a nomenclature issue.



Consider our piston. If we add energy by heat, and the piston is free to move, then we do an amount of work on the gas

$$w = -P(V_f - V_i)$$

This is a negative amount of work, since V_f is larger than V_i . But think of a steam engine. We want the force by the gas on the piston to win, because this pushing of the piston is what we can use to push the wheel of a steam engine, or a turbine, etc. That is negative work on the gas. And it is inconvenient to have the useful work be negative. Let's define the useful work done by a thermodynamic device as the negative of work done on the gas.

$$w_{\text{useful}} = -w$$

But, w_{useful} is kind of a awkward name. Usually a machine that does useful work is called an engine, so let's call our useful work W_{eng} for "work done by an engine."

$$w_{\text{eng}} = -w$$

37.2 Reservoirs

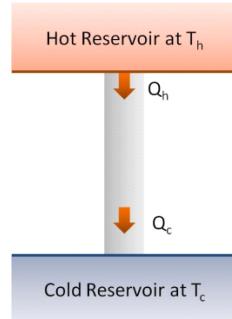
Question 123.17.2

Think of a water reservoir. In the western US we know what these are. It is a large man-made body of water that is designed to provide irrigation water for farms and homes.

Think of what happens when you water your lawn with water from the reservoir. Some water does leave from the reservoir, but compared to the whole reservoir of water, one lawn watering makes very little difference. Often there is more water coming into the reservoir, so really there is negligible water loss from your individual lawn.

A heat reservoir is like this. A hot reservoir is a source of thermal energy that can be tapped without causing any measurable change in temperature. We can also have a cold reservoir that has a large lack of thermal energy. A small amount of thermal energy transferred to this cold reservoir won't measurably change its temperature.

The next figure is called an *Energy Transfer Diagram*.



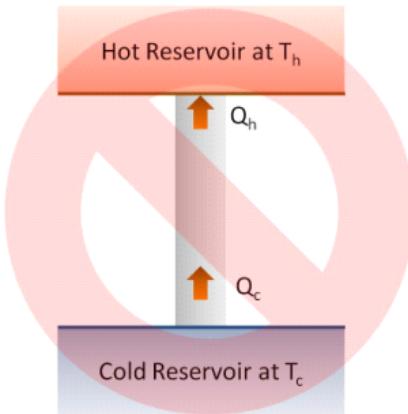
Question 123.17.3

It shows energy moving by heat from a hot reservoir to a cold reservoir.

Using our understanding of the second law of thermodynamics, we can see that an amount of energy, Q_h , can flow from the hot reservoir and an amount

of energy, Q_c , can flow by heat to the cold reservoir. We expect that if there is no other mechanism for dissipating the energy, then $Q_c = Q_h$. We will define both Q_h and Q_c as positive quantities. If they are negative, we will explicitly write a negative sign.

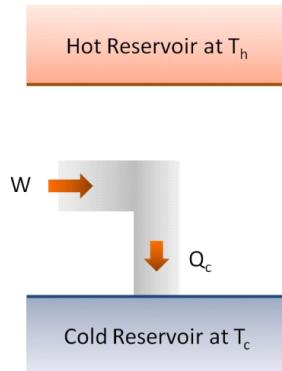
Look at this energy transfer diagram.



This diagram says that energy flows from the cold reservoir to the hot reservoir. From the second law of thermodynamics we would say that this won't happen. This would increase the order of the cold reservoir spontaneously. This does not happen.

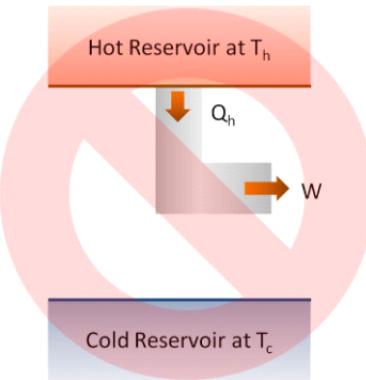
Question 123.17.4

Here is another energy transfer diagram



This diagram says that we do work that is directly converted into heat. This is a common experience. Rub your hands together. Your hands warm up. The work has created more internal energy, But that energy quickly dissipates into the room. The room temperature is not affected by this increase in internal energy of the air. The air in the room is a cold reservoir.

Here is another diagram.



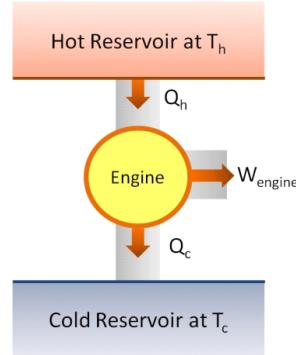
This diagram says that energy from the hot reservoir is converted directly into useful work. This is not something we expect to happen. And our second law of thermodynamics again would prohibit this situation.

This would be a perfect engine, converting all the energy transferred from the hot reservoir into useful work. In fact, this idea could give us a perfectly good restatement of the second law of thermodynamics.

Perfect engines are not possible

37.3 Heat Engines

Using our energy transfer diagram, let's introduce a more practical engine. Here is the diagram. We have added the engine explicitly on the diagram.



The engine takes energy from the heat reservoir and turns it into useful work. But there is some energy that is wasted and this energy ends up in the cold reservoir. The amount of energy that becomes work must be

$$W_{\text{engine}} = Q_h - Q_c$$

We call $Q_h - Q_c$ the net heat

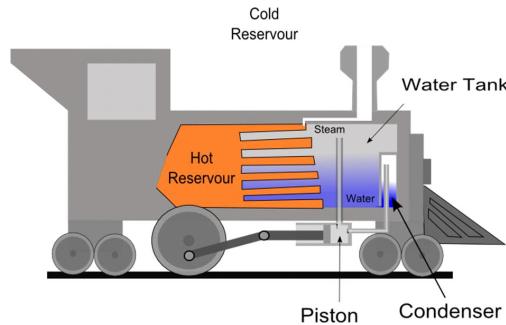
$$Q_{net} = Q_h - Q_c$$

Notice that Q_h is positive *for the engine* and Q_c is negative *for the engine*.

A useful heat engine is a device that takes in energy by heat and operates in a cycle. If it did not periodically return to its original state, then it could not keep going. The basics of the cycle are as follows

1. The working substance absorbs energy by heat from the high temperature reservoir
2. Work is done by the engine
3. Energy is expelled by heat to a low temperature reservoir

37.3.1 Example: Steam engine



In a steam engine

1. Water in the boiler absorbs energy from burning fuel (hot reservoir) and evaporates to steam
2. The steam does work by expanding against a piston
3. The steam cools and condenses releasing heat to the outside air (cold reservoir), and the liquid water returns to the boiler.

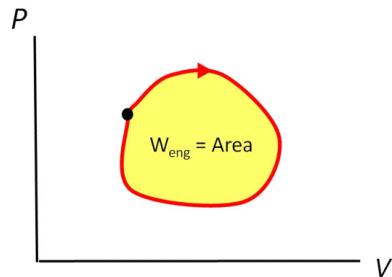
Since the engine goes through a cyclical process,

$$\Delta E_{int} = 0$$

Its initial and final internal energies are the same, so

$$Q_{net} = W_{eng} \quad (37.1)$$

The work done by the engine equals the *net* energy absorbed by the engine. The work is equal to the area enclosed by the curve of the PV diagram.



Thermal efficiency is defined as the ratio of the net work done by the engine during one cycle to the energy input at the higher temperature

$$\eta = \frac{W_{eng}}{|Q_h|} \quad (37.2)$$

$$= \frac{Q_{net}}{Q_h} \quad (37.3)$$

$$= \frac{|Q_h| - |Q_c|}{Q_h} \quad (37.4)$$

$$= 1 - \frac{|Q_c|}{|Q_h|} \quad (37.5)$$

We can think of this as

$$\eta = \frac{\text{what you gain in work}}{\text{what you gave in energy by heat}} \quad (37.6)$$

For a car engine

$$\eta = 20\% \quad (37.7)$$

is a very good number (most student and faculty cars are far less efficient than this).

Suppose that we have a 100% efficient engine. We know that this is not possible, but consider what this would mean.

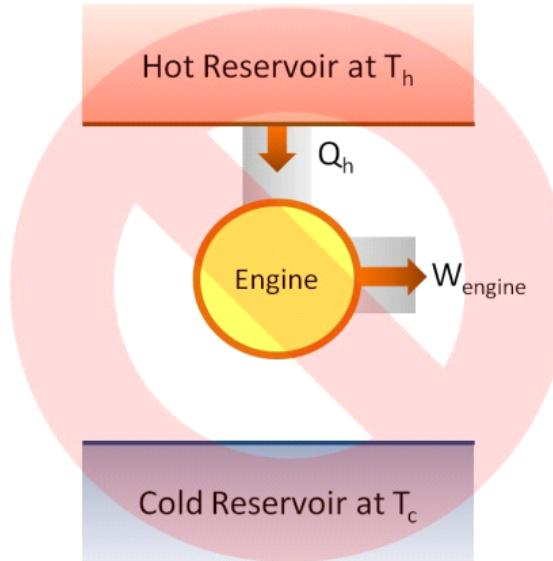
$$\eta = 1 - \frac{|Q_c|}{|Q_h|} \quad (37.8)$$

$$= 1 \quad (37.9)$$

this can only happen when $|Q_c| = 0$ which means that no energy is transferred out of the engine by heat – which is not really possible. The fact that real engines have efficiencies much lower than 1 leads to another statement of the second law of thermodynamics.

Question 123.17.5

Question 123.17.6



It is impossible to construct a heat engine that, operating in a cycle, inputs energy by heat from a hot reservoir and converts the energy entirely into useful work.

This is called the Kelvin-Plank form of the second law, and it means that

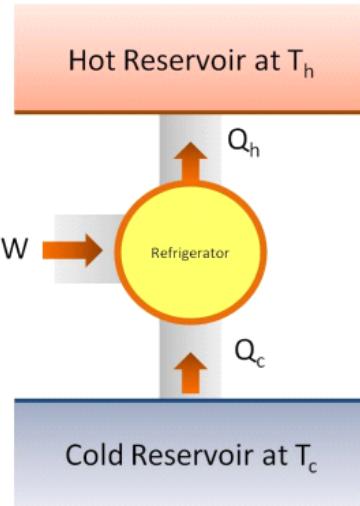
$$W_{eng} < |Q_h| \quad (37.10)$$

so the device represented by our drawing above is impossible to achieve.

37.4 Heat Pumps and Refrigerators

Let's look at a special form of a heat engine, a head pump..

Question 123.17.7



A heat pump is an engine run in reverse. We wish to pump heat from a cold reservoir to the heat reservoir. To do this we must do work *on* the heat pump. This is not a natural process, so by adding in work we are adding energy to make it happen.

A window air conditioner unit is a good example of a heat pump. It removes energy by heat from the inside of your apartment and transfers that energy to the outside of your apparent by heat process. To make this happen, you need mechanical device that does work

A refrigerator is another example of a heat pump. It transfers energy by heat from the cold interior to the warmer environment of your apartment kitchen. It would be great if we could transfer heat from our cold reservoir (room of a house or freezer compartment of a fridge) to a hot reservoir without doing work, but this is not possible. This fact was discovered by Clausius and lead to the Clausius statement of the second law—*It is impossible to construct a cyclical machine whose sole effect is to transfer energy continuously by heat from one object to another object at a higher temperature without the input of energy by work.*

Or more simply, as we have already said: Energy does not transfer spontaneously by heat from a cold object to a hot object.

37.4.1 Coefficient of Performance

How well a heat pump works can be described by the *coefficient of performance* (COP).

$$COP_h = \frac{\text{energy transferred at high temperature}}{\text{work done on heat pump}} \quad (37.11)$$

$$= \frac{|Q_h|}{W} \quad (37.12)$$

COP is similar to efficiency. Think of our air conditioner. Q_h is typically higher than W . That is, we usually transfer more energy by heat to the outside world than we provide in work. The outside part of the airconditioner does indeed get hotter than the outside air temperature. That is why energy will transfer by heat. Values of *COP* are generally greater than 1, though it is possible for them to be less than 1. Of course we would like the *COP* to be as high as possible.

For refrigeration, the important thing is how much energy we transfer out of the cool area, Q_c , so for cooling the *COP* is

$$COP_c = \frac{|Q_c|}{W} \quad (37.13)$$

A good refrigerator should have a high COP. Typical values are 5 or 6.

Chapter 38

38 Ideal Gas Heat Engines

2.4.4, 2.4.5

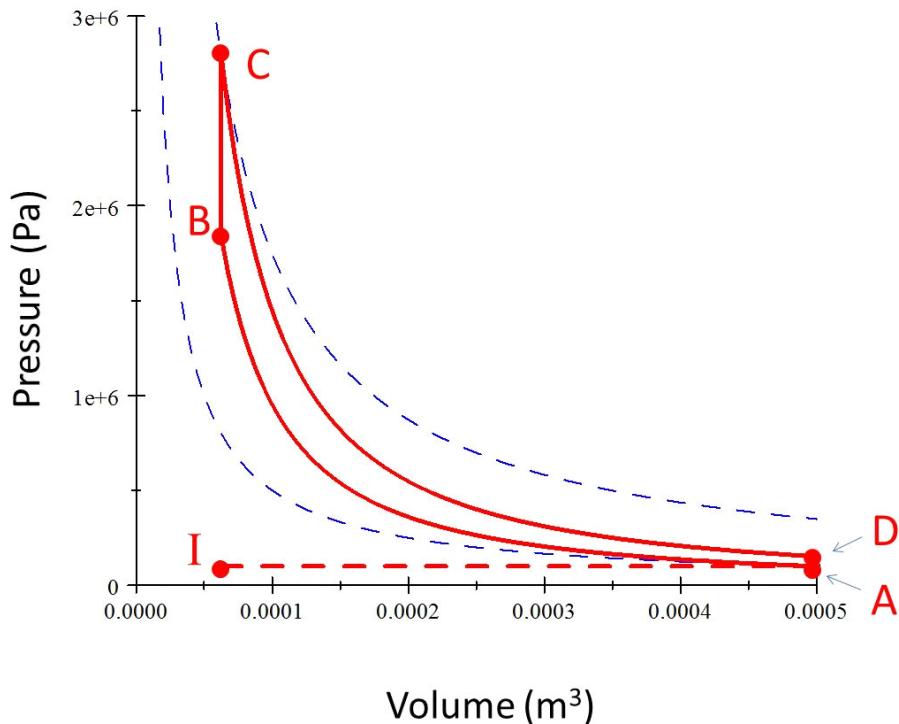
So far we have talked about idealized heat engines and refrigerators. We should try a somewhat realistic engine. Our goal in such a problem is to find for each process the values of ΔE_{int} , Q , W , and W_{eng} , Q_h , Q_c . We also want to find the ideal gas state variables P , V , n , and T at the transition points between processes.

Fundamental Concepts

- To describe a heat engine cycle you need to find P , V , n , T , ΔE_{int} , Q , W , for each part of the cycle and W_{eng} , Q_h , Q_c for the entire cycle.
- The Otto Cycle
- The Carnot cycle represents the maximum theoretical limit of efficiency for a heat engine.
- The efficiency of the Carnot cycle depends only on the two extreme temperatures $\eta_C = 1 - \frac{T_c}{T_h}$.
- An engine design that is more efficient than a Carnot engine must have a fundamental flaw in its calculations.

Question 123.19.1

38.1 An example: The Otto Cycle



The *PV* diagram above shows the Otto cycle. This cycle is approximately what happens in a gasoline engine, like those in cars or, more like what we find in lawn mowers.

For our engine, suppose the compression ratio is

$$\frac{V_A}{V_B} = 8.00$$

Let's take

$$\begin{aligned} V_A &= 500 \text{ cm}^3 = 0.0005 \text{ m}^3 \\ P_A &= 100 \text{ kPa} \\ T_A &= 293 \text{ K} \end{aligned}$$

then we can find V_B

$$V_B = \frac{V_A}{8} = 62.5 \text{ cm}^3 = 6.25 \times 10^{-5} \text{ m}^3$$

and we will need to measure the temperature at C

$$T_C = 1023 \text{ K}$$

and the intake temperature

$$T_I = 273 \text{ K}$$

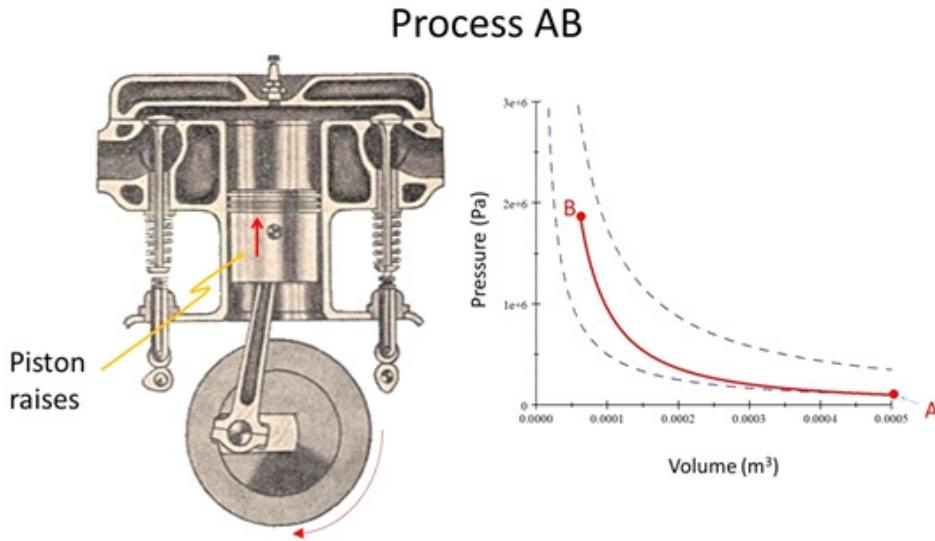
(which probably tells us that this must be a road trip in Rexburg, not Phoenix). Further suppose our engine uses a diatomic ideal gas with

$$\begin{aligned} C_V &= \frac{5}{2}R \\ R &= 8.314 \frac{\text{J}}{\text{mol K}} \end{aligned}$$

so

$$\gamma = 1.4$$

Let's follow the process through to see what the state variables are at each point on the PV diagram, and what Q_h and Q_c are for each process, and what the work, W , is and what the change in internal energy will be, ΔE_{int} . Finally, we will find the efficiency of the engine.



Process $A \rightarrow B$ is an adiabatic compression. The piston moves upward, compressing the fuel-air mixture. This happens rapidly, so there is no time for energy to leave by heat, $Q = 0$. In this process $\Delta E_{int} = W$.

we can use

$$P_i V_i^\gamma = P_f V_f^\gamma$$

to find the final pressure for this process at B

$$P_B = P_A \frac{V_A^\gamma}{V_B^\gamma}$$

We can find the work from

$$\begin{aligned}
 W_{AB} &= - \int_{V_A}^{V_B} P dV \\
 &= - \int_{V_A}^{V_B} P_A \frac{V_A^\gamma}{V^\gamma} dV \\
 &= -P_A V_A^\gamma \int_{V_A}^{V_B} \frac{1}{V^\gamma} dV \\
 &= -P_A^\gamma V_A^\gamma \left(\frac{V^{1-\gamma}}{1-\gamma} \Big|_{V_A}^{V_B} \right) \\
 &= -\frac{P_A V_A^\gamma}{1-\gamma} (V_B^{1-\gamma} - V_A^{1-\gamma})
 \end{aligned}$$

We can rearrange this into a more convenient formula

$$\begin{aligned}
 W_{AB} &= \frac{P_A V_A^\gamma}{\gamma - 1} \left(\frac{1}{V_B^{\gamma-1}} - \frac{1}{V_A^{\gamma-1}} \right) \\
 &= \frac{1}{\gamma - 1} \left(\frac{P_A V_A^\gamma}{V_B^{\gamma-1}} - \frac{P_A V_A^\gamma}{V_A^{\gamma-1}} \right) \\
 &= \frac{1}{\gamma - 1} \left(\frac{P_A V_A^\gamma}{V_B^\gamma V_B^{-1}} - \frac{P_A V_A^\gamma}{V_A^\gamma V_A^{-1}} \right) \\
 &= \frac{1}{\gamma - 1} (P_B V_B - P_A V_A)
 \end{aligned}$$

which we can generalize into another basic equation for adiabatic processes.

$$W_{AB} = \left(\frac{1}{\gamma - 1} \right) (P_f V_f - P_i V_i)$$

then for process $A \rightarrow B$ we have

$$\begin{aligned}
W_{AB} &= \left(\frac{1}{\gamma - 1} \right) (P_B V_B - P_A V_A) \\
&= \left(\frac{1}{\gamma - 1} \right) \left(P_A \frac{V_A^\gamma}{V_B^\gamma} V_B - P_A V_A \right) \\
&= \left(\frac{1}{\gamma - 1} \right) \left(P_A \frac{V_A^\gamma}{\left(\frac{V_A}{8}\right)^\gamma} V_B - P_A V_A \right) \\
&= \left(\frac{1}{\gamma - 1} \right) \left(8^\gamma P_A \left(\frac{V_A}{8} \right) - P_A V_A \right) \\
&= \left(\frac{V_A}{\gamma - 1} \right) (8^{\gamma-1} P_A - P_A) \\
&= \left(\frac{V_A P_A}{\gamma - 1} \right) (8^{\gamma-1} - 1)
\end{aligned}$$

so the work is

$$\begin{aligned}
W_{AB} &= \left(\frac{V_A P_A}{\gamma - 1} \right) (8^{\gamma-1} - 1) \\
&= \left(\frac{(0.0005 \text{ m}^3)(100 \text{ kPa})}{1.4 - 1} \right) (8^{1.4-1} - 1) \\
&= 162.17 \text{ J}
\end{aligned}$$

But this is the work done on the gas, the useful work is the negative of this

$$W_{engAB} = -162.17 \text{ J}$$

and the change in internal energy is

$$\Delta E_{AB} = 162.17 \text{ J}$$

and we can find

$$E_{int} = \frac{5}{2} nRT$$

but we need n . We can get it from what we know at A

$$\begin{aligned}
n &= \frac{P_A V_A}{R T_A} \\
&= \frac{(100 \text{ kPa})(0.0005 \text{ m}^3)}{(8.314 \frac{\text{J}}{\text{mol K}})(293 \text{ K})} \\
&= 2.0525 \times 10^{-2} \text{ mol}
\end{aligned}$$

then

$$\begin{aligned} E_A &= \frac{5}{2}nRT_A \\ &= \frac{5}{2}(2.0525 \times 10^{-2} \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}}\right) (293 \text{ K}) \\ &= 125.00 \text{ J} \end{aligned}$$

We used

$$P_B = P_A \frac{V_A^\gamma}{V_B^\gamma}$$

but we should find a numeric answer for P_B

$$\begin{aligned} P_B &= P_A \frac{V_A^\gamma}{\left(\frac{V_A}{8}\right)^\gamma} \\ &= P_A \frac{V_A^\gamma}{\frac{V_A^\gamma}{8^\gamma}} \\ &= 8^\gamma P_A \\ &= 8^{1.4} (100 \text{ kPa}) \\ &= 1837.9 \text{ kPa} \end{aligned}$$

and finally, from the ideal gas law

$$\begin{aligned} T_B &= \frac{P_B V_B}{nR} \\ &= \frac{(1837.9 \text{ kPa})(6.25 \times 10^{-5} \text{ m}^3)}{(2.0525 \times 10^{-2} \text{ mol})(8.314 \frac{\text{J}}{\text{mol K}})} \\ &= 673.15 \text{ K} \end{aligned}$$

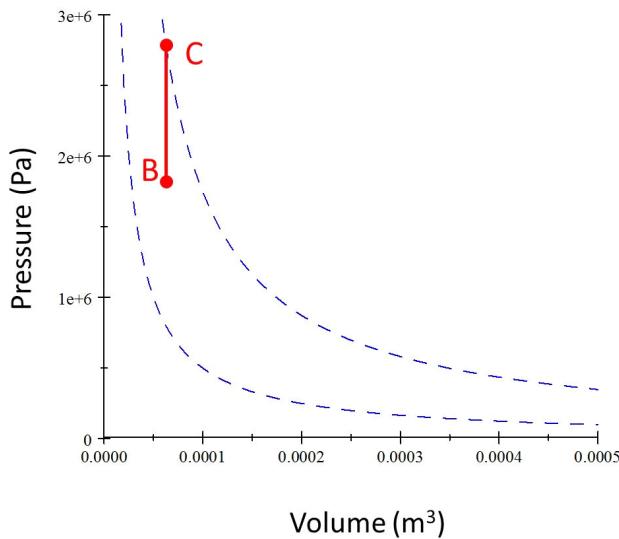
at last we can find E_B

$$\begin{aligned} E_B &= \frac{5}{2}nRT_B \\ &= \frac{5}{2}(2.0525 \times 10^{-2} \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}}\right) (673.15 \text{ K}) \\ &= 287.17 \text{ J} \end{aligned}$$

what we have learned is summarized in this table

State	T (K)	P (kPa)	V (m ³)	n (mol)	E _{int} (J)	-
A	293	100	0.0005	2.0525 × 10 ⁻²	125.00	-
B	673.15	1837.9	6.25 × 10 ⁻⁵	2.0525 × 10 ⁻²	287.17	-
Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
AB	0	162.17 J	162.17 J	-162.17 J	0	0

That was just one process!



Let's take on the next process $B \rightarrow C$. This is where the spark plug ignites the air-fuel mixture created by the fuel injectors. A great deal of energy is released in a hurry. The temperature jumps up and the pressure does too, but the volume is the same because this process happens in a flash—literally!

So this must be an isochoric process. We expect $W_{BC} = 0$, so $\Delta E_{BC} = Q_{BC}$. The pressure increases, so we expect Q_{BC} to be positive. This must be part of Q_h ! We know the temperature at C and we know the volume so we can find the final pressure

$$\begin{aligned} P_C &= \frac{nRT_C}{V_C} \\ &= \frac{(2.0525 \times 10^{-2}) (8.314 \frac{\text{J}}{\text{mol K}}) (1023)}{6.25 \times 10^{-5} \text{ m}^3} \\ &= 2.7931 \times 10^6 \text{ Pa} \\ &= 2793.1 \text{ kPa} \end{aligned}$$

We can find the internal energy at C now

$$\begin{aligned} E_C &= \frac{5}{2} nRT_C \\ &= \frac{5}{2} (2.0525 \times 10^{-2} \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (1023 \text{ K}) \\ &= 436.42 \text{ J} \end{aligned}$$

We can use

$$\begin{aligned} Q &= nC_V\Delta T \\ &= n\frac{5}{2}R(T_C - T_B) \end{aligned}$$

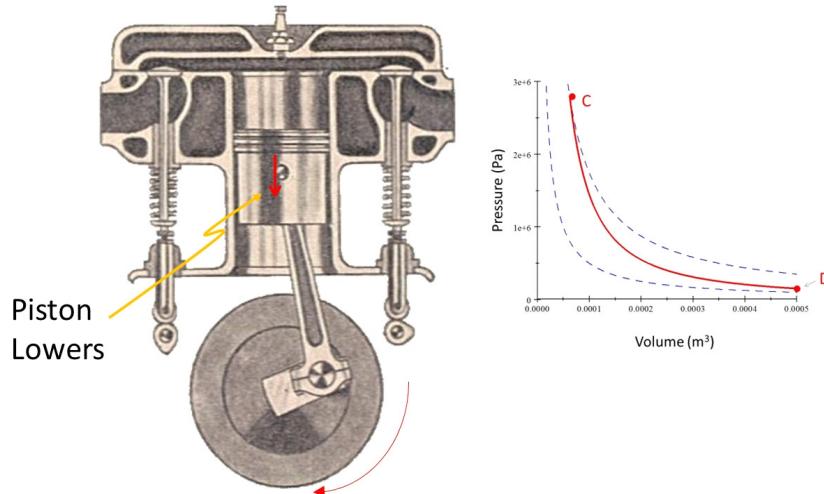
For the process BC to find the energy transfer by heat

$$\begin{aligned} Q_{BC} &= (2.0525 \times 10^{-2} \text{ mol}) \frac{5}{2} \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (1023 \text{ K} - 673.15 \text{ K}) \\ &= 149.25 \text{ J} \end{aligned}$$

We already mentioned that $Q_{BC} = \Delta E_{int,BC}$ and $W_{BC} = 0$, so we can complete our table of state variables and process variables for this process.

State	T (K)	P (kPa)	V (m ³)	n (mol)	E _{int} (J)	-
B	673.15	1837.9	6.25×10^{-5}	2.0525×10^{-2}	287.17	-
Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
BC	149.25	0	149.25	0	149.25	0

Process CD (Power Stroke)



The next process is $C \rightarrow D$. This is another adiabatic process, this time an expansion. This is sometimes called the “power stroke” because the hot, high pressure gas pushes the piston down. This push is what makes the car go. We know that $Q_{CD} = 0$ so $\Delta E_{CD} = W_{CD}$

To find the ideal gas state variables at D let's start with

$$P_i V_i^\gamma = P_f V_f^\gamma$$

again, then

$$P_D = P_C \frac{V_C^\gamma}{V_D^\gamma}$$

but from our PV diagram we see $V_C = V_B$ and $V_D = V_A$. This makes sense because the cylinder and piston system maximum and minimum volumes don't change. So we can write

$$P_D = P_C \frac{V_B^\gamma}{V_A^\gamma} = 8^{-\gamma} P_C$$

and knowing this we can find the temperature at D using the ideal gas law

$$T_D = \frac{P_D V_D}{nR}$$

let's numerically calculate P_D and T_D at this point.

$$\begin{aligned} P_D &= 8^{-1.4} (2793.1) \\ &= 151.97 \text{ kPa} \end{aligned}$$

and

$$\begin{aligned} T_D &= \frac{(151.97 \text{ kPa}) (0.0005 \text{ m}^3)}{(2.0525 \times 10^{-2} \text{ mol}) (8.314 \frac{\text{J}}{\text{mol K}})} \\ &= 445.28 \text{ K} \end{aligned}$$

and while we are at it we can complete our state D by calculating E_D

$$\begin{aligned} E_D &= \frac{5}{2} n R T_D \\ &= \frac{5}{2} (2.0525 \times 10^{-2} \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (445.28 \text{ K}) \\ &= 189.96 \text{ J} \end{aligned}$$

Now for the work done, we again use our new adiabatic equation

$$\begin{aligned} W_{CD} &= \left(\frac{1}{\gamma - 1} \right) (P_D V_D - P_C V_C) \\ &= \left(\frac{1}{\gamma - 1} \right) (8^{-\gamma} P_C V_A - P_C V_B) \\ &= \left(\frac{P_C}{\gamma - 1} \right) (8^{-\gamma} V_A - V_B) \end{aligned}$$

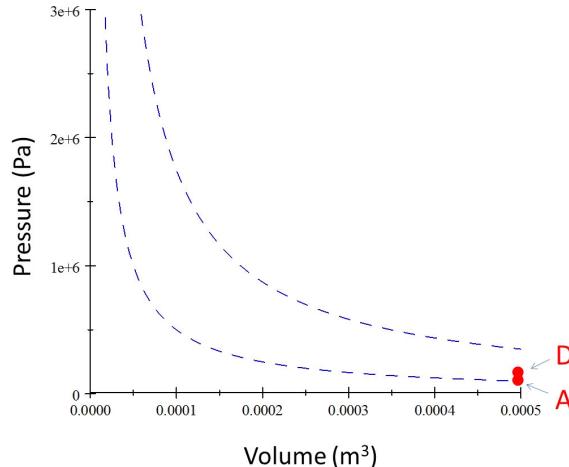
all of which we know, so

$$\begin{aligned} W_{CD} &= \left(\frac{2793.1 \text{ kPa}}{1.4 - 1} \right) (8^{-1.4} (0.0005 \text{ m}^3) - 6.25 \times 10^{-5} \text{ m}^3) \\ &= -246.46 \text{ J} \end{aligned}$$

which completes our set for process *CD*

State	T (K)	P (kPa)	V (m ³)	n (mol)	E _{int} (J)	-
C	1023	2793.1	6.25 × 10 ⁻⁵	2.0525 × 10 ⁻²	436.42	-
D	445.28	151.97	0.0005	2.0525 × 10 ⁻²	189.96 J	-
Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
CD	0	-246.46	-246.46	246.46	0	0

We are nearing the end of the cycle, In process *D* → *A* a valve opens letting the pressure drop quickly.



Since this happens quickly, the piston does not have time to change position, so the volume does not change. Process *DA* is (almost) isochoric. We already know *P_A*, *V_A*, *n*, and *T_A*. But we need to find the process work, heat and change in internal energy. Since the volume does not change, *W_{DA}* = 0 and *ΔE_{DA}* = *Q_{DA}* which we can find using

$$\begin{aligned} Q_{DA} &= nC_V\Delta T \\ &= n\frac{5}{2}R(T_A - T_D) \end{aligned}$$

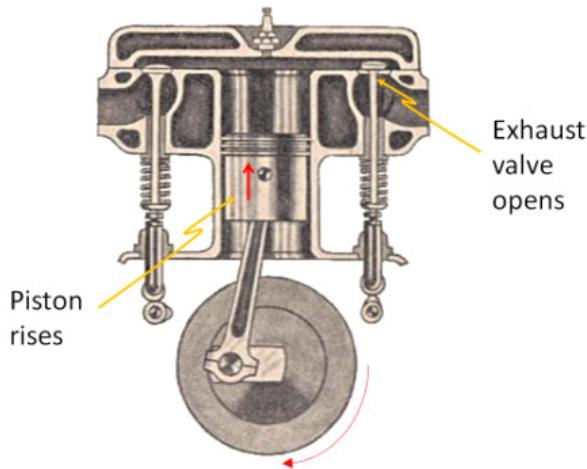
which gives

$$\begin{aligned} Q_{DA} &= (2.0525 \times 10^{-2} \text{ mol}) \frac{5}{2} \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (293 \text{ K} - 445.28 \text{ K}) \\ &= -64.964 \text{ J} \end{aligned}$$

which tells us that we have lost energy by heat. So Q_{DA} must contribute to Q_c .

State	T (K)	P (kPa)	V (m^3)	n (mol)	E_{int} (J)	-
D	445.28	151.97	0.0005	2.0525×10^{-2}	189.96 J	-
A	293	100	0.0005	2.0525×10^{-2}	125.00	-
Process	Q (J)	W_{int} (J)	ΔE_{int} (J)	W_{eng} (J)	Q_h	Q_c
DA	-64.964	0	-64.964	0	0	64.964

Process AI



But we are not done. The opening of the valve let the spent fuel leave as exhaust. But we can help remove the exhaust by pushing the piston up. This is done at constant pressure, because the valve is open to the exhaust system air pressure. We will lose gas by venting it. We can find the new number of moles using the ideal gas law

$$P_I V_I = nRT$$

since $P_I = P_A$ and

$$\frac{V_A}{V_I} = 8.00$$

$V_I = 8/V_A$ so we can guess that we should have

$$\frac{P_A V_A}{P_I V_I} = \frac{n_A R T_A}{n_I R T_I}$$

$$8 = \frac{n_A T_A}{n_I T_I}$$

so

$$n_I = \frac{n_A T_A}{8 T_I}$$

or

$$\begin{aligned} n_I &= \frac{(2.0525 \times 10^{-2} \text{ mol})(293 \text{ K})}{8(273 \text{ K})} \\ &= 2.7536 \times 10^{-3} \text{ mol} \end{aligned}$$

and then E_I will be

$$\begin{aligned} E_D &= \frac{5}{2} n_I R T_I \\ &= \frac{5}{2} (2.7536 \times 10^{-3} \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (273 \text{ K}) \\ &= 15.625 \text{ J} \end{aligned}$$

and the work done for an isobaric process is

$$W_{AI} = P_A (V_I - V_A)$$

but this assumed that n did not change. Really we are just moving the gas, so our work is

$$\begin{aligned} W &= F \Delta x \\ &\approx 0 \end{aligned}$$

because it takes very little force to push out the gas, so we can ignore W_{AI} .

Finding Q_{AI} we would like to use

$$Q = n C_P \Delta T$$

but we quickly run into trouble again because n changed! So what has gone wrong?

Well we are using convection to remove heat instead of conduction. We can get around this by finding ΔE_{int} directly.

$$\begin{aligned} \Delta E_{AI} &= E_I - E_A \\ &= 15.625 \text{ J} - 125.00 \text{ J} \\ &= -109.38 \text{ J} \end{aligned}$$

From this and the first law

$$\Delta E_{int} = Q + W$$

we can find Q

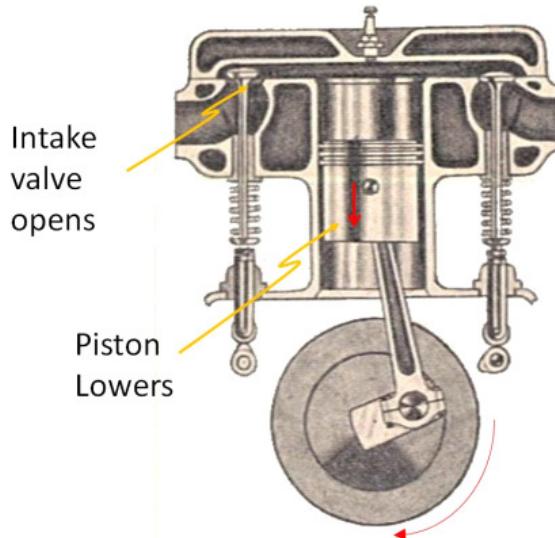
$$\begin{aligned} Q &\approx \Delta E_{int} \\ &\approx -109.38 \text{ J} \end{aligned}$$

Energy is leaving with the gas, so we seem to pick up a contribution to Q_c . But really this is not part of our thermodynamic process, because we mechanically removed the energy by removing the gas. Let's denote this by putting our

energy in parenthesis to remind us that this is a more mechanical than thermal process.

State	T (K)	P (kPa)	V (m ³)	n (mol)	E _{int} (J)	-
A	293	100	0.0005	2.0525 × 10 ⁻²	125.00	-
I	273	100	6.25 × 10 ⁻⁵	2.7536 × 10 ⁻³	15.625	-
Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
AI	(-109.38)	~ 0	(-109.38)	~ 0	0	(109.38)

Process IA



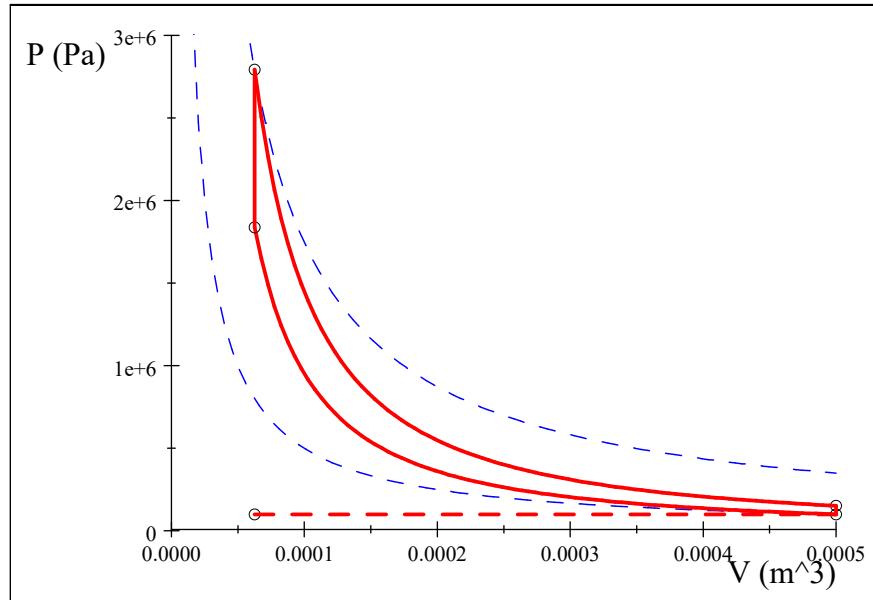
Now the exhaust valve closes, and the intake valve opens. The new fuel mixture is brought in. We essentially reverse process *AI* in process *IA*.

State	T (K)	P (kPa)	V (m ³)	n (mol)	E _{int} (J)	-
I	293	100	6.25 × 10 ⁻⁵	2.5657 × 10 ⁻³	15.625	-
A	293	100	0.0005	2.0525 × 10 ⁻²	125.00	-
Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
IA	(109.38)	~ 0	(109.38)	~ 0	(109.38)	0

But there is a difference. The new gas has useful fuel. so we are ready to complete the cycle again by compressing the gas, igniting it, and receiving the power stroke and exhaust. Note that Processes *AI* and *IA* have a net zero effect thermodynamically, but are very important to making your car or lawn mower go!

We can plot what we have found. Note that since *n* changed for the *AI*–*IA* processes, it is not really possible to include it on the PV diagram. That is why

it has a dashed line. We know the volume and pressure along the path, but the predicted temperature from the graph may not be right because n changed too.



Let's summarize what we found in large tables. First the ideal gas state variables and E_{int}

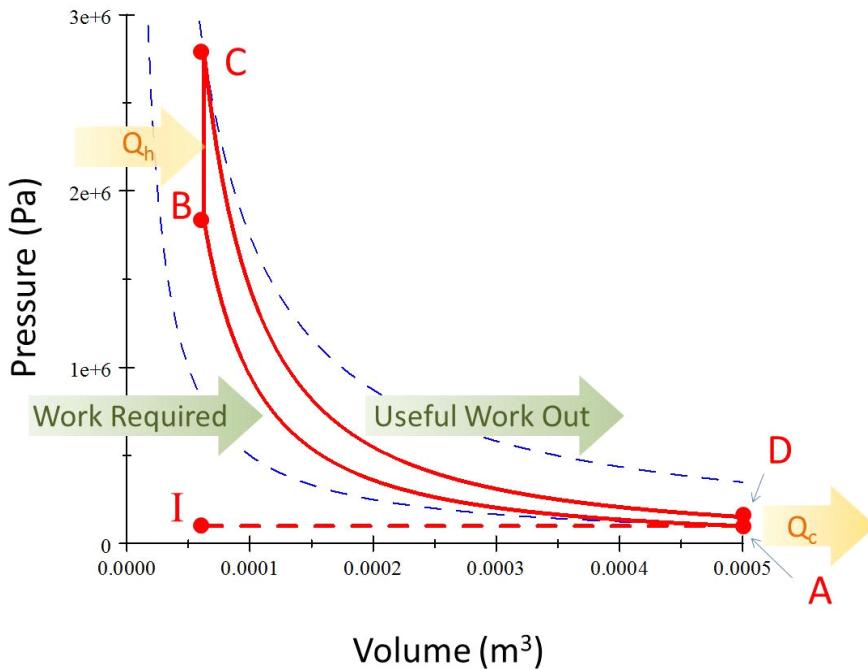
State	T (K)	P (kPa)	V (m ³)	n (mol)	E _{int} (J)
A	293	100	0.0005	2.0525×10^{-2}	125.00
B	673.15	1837.9	6.25×10^{-5}	2.0525×10^{-2}	287.17
C	1023	2793.1	6.25×10^{-5}	2.0525×10^{-2}	436.42
D	445.28	151.97	0.0005	2.0525×10^{-2}	189.96 J
I	293	100	6.25×10^{-5}	2.5657×10^{-3}	15.625

and now the values of Q , W , and E_{int} for the processes. We have included W_{eng} , Q_h and Q_c as well

Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
AB	0	162.17	162.17	-162.17	0	0
BC	149.25	0	149.25	0	149.25	0
CD	0	-246.46	-246.46	246.46	0	0
DA	-64.964	0	-64.964	0	0	64.964
AI	(-109.38)	(~ 0)	(-109.38)	(~ 0)	(0)	(109.38)
IA	(109.38)	(~ 0)	(109.38)	(~ 0)	(109.38)	(0)
Whole Cycle	84.286	-84.29	0	84.29	149.25	64.964

Note that the last row of the table is a total for the entire cycle. We see that we have useful work of 84.3 J and that $Q_h = 258.63$ J and $Q_c = 174.34$. We can

identify Q_h and Q_c on our graph. The useful work is the difference between the work done in the two adiabatic processes.



We can see that it takes some work to make process AB happen. That is why we need a starter motor to get our engine started. But once the engine gets going the momentum from the piston and crank shaft keep the engine running. But still process AB requires energy input by work, so it is like a resistance for the cycle.

What is the efficiency of this cycle? We don't have a really good way with our simple view of thermodynamics to accurately account for the change in chemical potential energy involved with state I . The energy transfer in processes AI and IA are, in a sense, mechanical energy processes. So for this class it might be more fair to say that processes AI and IA have no net effect (except for refueling) and exclude their effect from Q_c and Q_h then ignoring the I processes

$$\begin{aligned}\eta &= \frac{W_{eng}}{Q_h} \\ &= \frac{84.29}{149.25} \\ &= 0.56476\end{aligned}$$

For a gasoline engine, it is often quoted that the efficiency is

$$\eta_{otto} = 1 - \frac{1}{\left(\frac{V_A}{V_B}\right)^{\gamma-1}}$$

Let's test this.

$$\begin{aligned}\eta_{otto} &= 1 - \frac{1}{\left(\frac{0.0005}{6.25 \times 10^{-5}}\right)^{1.4-1}} \\ &= 0.56472\end{aligned}$$

But you might guess that this is very optimistic. Efficiencies for actual engines come in at values around 20%.

38.2 Carnot, the Limit of Efficiency for Heat Engines

We now know that we can design actual engines that power cars and useful machines by combining several of our special thermodynamic processes into larger, more complex processes that form a cycle. You might guess that there are more than just isochoric, isobaric, isothermal, and adiabatic processes that exist. A good question is what processes should we combine to form our engine? Past researchers have asked this question over and over and many new engines have been developed. Of course we want our engines to be as efficient as possible, and the second law of thermodynamics says the engines can't be 100% efficient. But is there a "most efficient" engine, one that no other engine can beat in efficiency? If there is, we could use this to judge new engine designs. Any engine with a better efficiency would be a mistake that won't work, and that approach this maximum efficiency would be a possible improvement in technology. There is such a maximally efficient engine.

Carnot was a French Engineer. He thought of an engine that would operate on an ideal reversible cycle. He hit upon the most efficient cycle possible. Sadly, it is not possible to create his engine in practice. Carnot's engine assumed no friction and ideal thermodynamic processes. So why is it useful to know about Carnot's engine design? The Carnot cycle provides us with a useful upper limit for all practical engines. To tell if a design is practical, we can compare it to the Carnot cycle. If our design looks like it will outperform the Carnot engine, we know we have done something wrong. Or conversely, if our system design requires an engine to be better than a Carnot engine, we need to go back to the drawing board.

Knowing just this much we can state Carnot's theorem.

No real heat engine operating between two energy reservoirs can be more efficient than a Carnot engine operating between the same two reservoirs

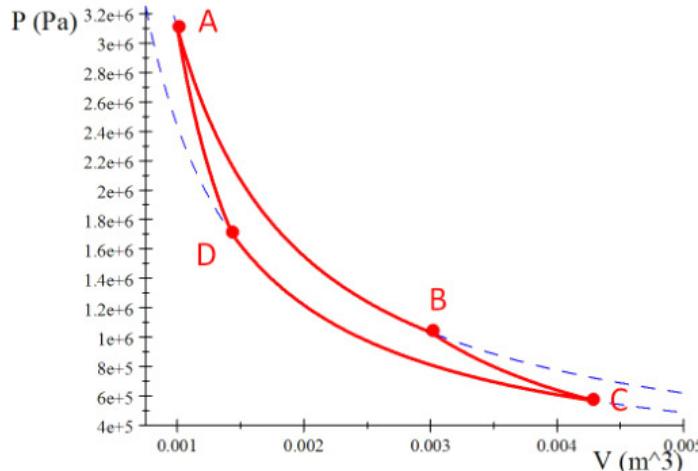
All real engines are less efficient than a Carnot engine because they do not operate through a reversible cycle. The efficiency of a real engine is reduced by friction, energy losses through conduction, etc.

38.2.1 The Carnot Cycle

Carnot found a set of special processes that when combined made a theoretical heat engine. And he proved mathematically that this particular combination is theoretically the most efficient possible. Let's take a look at this engine design.

The Carnot cycle is shown in the figure. It consists of the following parts:

1. Process $A \rightarrow B$ is an isothermal expansion
2. Process $B \rightarrow C$ is an adiabatic expansion
3. Process $C \rightarrow D$ is an isothermal contraction
4. Process $D \rightarrow A$ is an adiabatic compression



Let's take an example. Let's choose a Carnot engine that operates between the same two extreme temperatures as our Otto cycle example.

$$T_c = 273 \text{ K}$$

$$T_h = 1023 \text{ K}$$

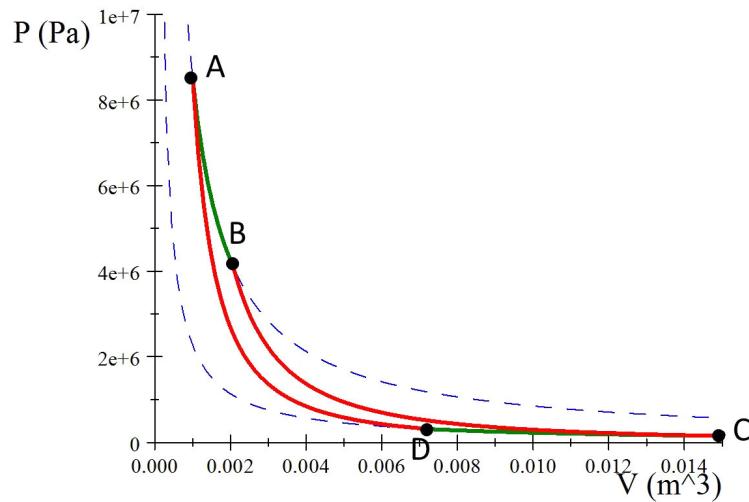
Let's suppose the carnot engine is made from a piston with a monotonic ideal gas. Let's choose the minimum and maximum volumes to be

$$\begin{aligned} V_A &= 0.001 \text{ m}^3 \\ V_C &= 0.015 \text{ m}^3 \end{aligned}$$

and let's have

$$n = 1 \text{ mol}$$

of gas. For these numbers, our graph looks like this:

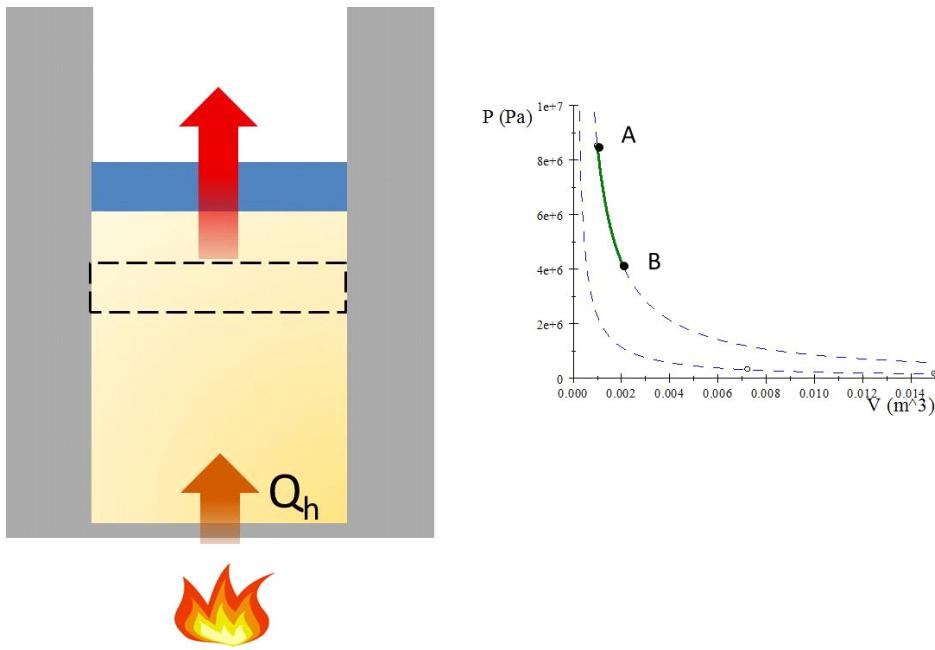


Our job is to find P , V , T , W , Q , and ΔE_{int} for each part of the cycle and Q_h , Q_c , and W_{eng} for the whole cycle.

Let's take each process separately like we did for the Otto cycle.

38.3 Process $A \rightarrow B$ isothermal expansion

The gas is placed in contact with the high temperature reservoir, T_h . The gas absorbs heat $|Q_h|$. The gas does work W_{AB} in raising the piston



Remember for isothermal processes we can write

$$\frac{P_i V_i}{P_f V_f} = \frac{nRT_i}{nRT_i}$$

because the temperature does not change, so

$$P_A V_A = P_B V_B$$

and we know for an isothermal process that

$$W_{AB} = nRT_A \ln \left(\frac{V_A}{V_B} \right)$$

and that

$$\Delta E_{int} = 0$$

so

$$|Q_{AB}| = |-W_{AB}|$$

and we can identify this as $|Q_h|$ and $T_A = T_h$

We can complete state A using the ideal gas law

$$\begin{aligned} P_A &= \frac{nRT_h}{V_A} = \frac{(1 \text{ mol}) (8.314 \frac{\text{J}}{\text{mol K}}) (1023 \text{ K})}{0.001 \text{ m}^3} \\ &= 8.5052 \times 10^6 \text{ Pa} \end{aligned}$$

To complete state *B* we need the volume and the pressure. At point *B* both the adiabat and the isotherm have the same pressure and volume. So we know

$$P_B = \frac{nRT_h}{V_B}$$

and

$$P_B = \frac{K_{BC}}{V_B^\gamma}$$

where the constant $K_{BC} = P_B V_B^\gamma$ from our adiabatic equations. But we don't know K_{BC} , P_B , or V_B

We do know that at *C*

$$\begin{aligned} P_C &= \frac{nRT_c}{V_C} \\ P_C &= \frac{K_{BC}}{V_c^\gamma} \end{aligned}$$

and at point *C* and we do know V_C , so we can find K_{BC}

$$\frac{nRT_c}{V_C} = \frac{K_{BC}}{V_C^\gamma}$$

$$\begin{aligned} K_{BC} &= \frac{nRT_c}{V_C} V_C^\gamma \\ &= nRT_c V_C^{\gamma-1} \end{aligned}$$

so

$$\begin{aligned} K_{BC} &= nRT_c V_C^{\gamma-1} = (1 \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (273 \text{ K}) (0.015 \text{ m}^3)^{\frac{2}{3}} \\ &= 138.05 \text{ J m}^2 \end{aligned}$$

then moving back to point *B* we can set the two equations for pressure equal

$$\begin{aligned} P_B &= \frac{nRT_h}{V_B} \\ P_B &= \frac{K_{BC}}{V_B^\gamma} \end{aligned}$$

to get

$$\frac{nRT_h}{V_B} = \frac{K_{BC}}{V_B^\gamma}$$

then

$$\begin{aligned} \frac{V_B^\gamma}{V_B} &= \frac{K_{BC}}{nRT_h} \\ V_B^{\gamma-1} &= \frac{nRT_h}{K_{BC}} \end{aligned}$$

which gives us an awkward

$$V_B = \sqrt[{\gamma-1}]{\frac{K_{BC}}{nRT_h}}$$

but it works.

$$\begin{aligned} V_B &= \sqrt[{\gamma-1}]{\frac{138.05 \text{ J m}^2}{(1 \text{ mol}) (8.314 \frac{\text{J}}{\text{mol K}}) (1023 \text{ K})}} \\ &= 2.0679 \times 10^{-3} \text{ m}^3 \end{aligned}$$

and we are back to the ideal gas law to find the pressure

$$\begin{aligned} P_B &= \frac{nRT_B}{V_B} \\ &= \frac{(1 \text{ mol}) (8.314 \frac{\text{J}}{\text{mol K}}) (1023 \text{ K})}{2.0679 \times 10^{-3} \text{ m}^3} \\ &= 4.1130 \times 10^6 \text{ Pa} \end{aligned}$$

Armed with all of this, we can compute the work and heat.

$$W_{AB} = nRT_h \ln \left(\frac{V_A}{V_B} \right)$$

$$\begin{aligned} W_{AB} &= (1 \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (1023 \text{ K}) \ln \left(\frac{0.001 \text{ m}^3}{2.0679 \times 10^{-3} \text{ m}^3} \right) \\ &= -6179.3 \text{ J} \end{aligned}$$

We can find

State	T (K)	P (Pa)	V (m ³)	n (mol)	—	—
A	1023	8.5052×10^6	0.001	1	—	—
B	1023	4.1130×10^6	2.0679×10^{-3}	1	—	—
Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
AB	6179.3	-6179.3	0	6179.3	6179.3	0

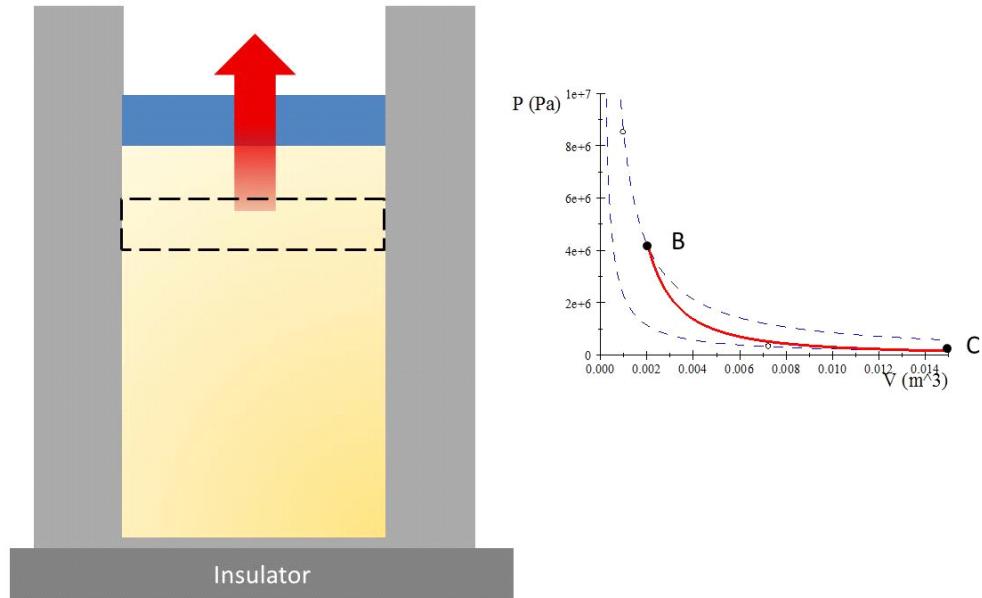
38.4 Process $B \rightarrow C$ adiabatic expansion

The base of the cylinder is replaced by a thermally nonconducting wall. No energy enters or leaves the system by heat.

$$Q = 0$$

The temperature falls from T_h to T_c . The gas does work W_{BC} .

Process B→C



Remember that for an adiabatic process

$$P_A V_A^\gamma = P_B V_B^\gamma$$

and we usually find a more convenient path for calculating ΔE_{int} .

$$\Delta E_{int} = nC_V \Delta T$$

We can find the pressure at C using the ideal gas law

$$\begin{aligned} P_C &= \frac{nRT_c}{V_C} = \frac{(1 \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}}\right) (273 \text{ K})}{0.015 \text{ m}^3} \\ &= 1.5131 \times 10^5 \text{ Pa} \end{aligned}$$

and

$$\begin{aligned} \Delta E_{int} &= (1 \text{ mol}) \frac{3}{2} \left(8.314 \frac{\text{J}}{\text{mol K}}\right) (273 \text{ K} - 1023 \text{ K}) \\ &= -9353.3 \text{ J} \end{aligned}$$

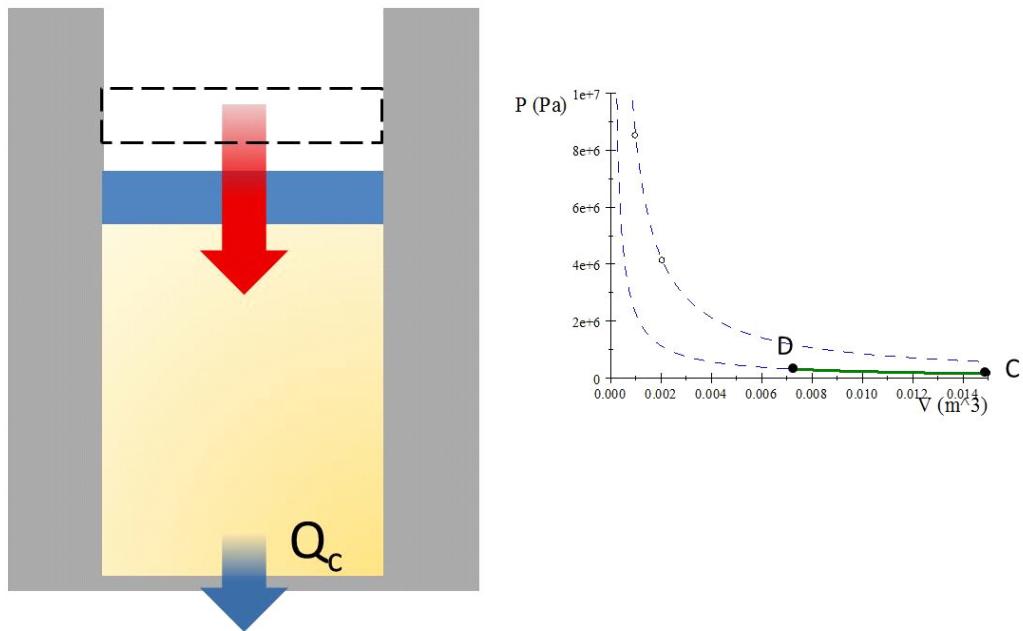
we can summarize all this.

State	T (K)	P (Pa)	V (m^3)	n (mol)	-	-
B	1023	4.1130×10^6	2.0679×10^{-3}	1	-	-
C	273	1.5131×10^5	0.015 m^3	1	-	-
Process	Q (J)	W_{int} (J)	ΔE_{int} (J)	W_{eng} (J)	Q_h	Q_c
BC	0	-9353.3	-9353.3	9353.3	0	0

38.5 Process $C \rightarrow D$ isothermal compression

The gas is placed in contact with the cold temperature reservoir. The gas expels energy $|Q_c|$ and work W_{CD} is done on the gas.

Process C → D



The process is isothermal, so $T_C = T_D$ and $nRT_C = nRT_D$ so we have

$$P_C V_C = P_D V_D \quad (38.1)$$

and

$$W_{CD} = nRT_C \ln \left(\frac{V_C}{V_D} \right) \quad (38.2)$$

again

$$\Delta E_{int} = 0 \quad (38.3)$$

so

$$|Q_{CD}| = |-W_{CD}| \quad (38.4)$$

and we can identify this as $|Q_c|$

We again need to find V_D and we can use the same method as before, finding V_B . At position D

$$P_D = \frac{nRT_c}{V_D}$$

and

$$P_D = \frac{K_{DA}}{V_D^\gamma}$$

but we don't know K_{DA} , P_D , or V_D

But we also know that at A

$$\begin{aligned} P_A &= \frac{nRT_h}{V_A} \\ P_A &= \frac{K_{DA}}{V_A^\gamma} \end{aligned}$$

and at point A and we do know V_A , so we can find K_{DA}

$$\frac{nRT_h}{V_A} = \frac{K_{DA}}{V_A^\gamma}$$

so

$$K_{DA} = nRT_h V_A^{\gamma-1}$$

or

$$\begin{aligned} K_{DA} &= (1 \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (1023 \text{ K}) (0.001 \text{ m}^3)^{\frac{2}{3}} \\ &= 85.052 \text{ J m}^2 \end{aligned}$$

then moving back to point D we have

$$\frac{nRT_c}{V_D} = \frac{K_{DA}}{V_D^\gamma}$$

and

$$V_D = \sqrt[{\gamma-1}]{\frac{K_{DA}}{nRT_c}}$$

numerically we get

$$\begin{aligned} V_D &= \sqrt[{\frac{2}{3}}]{\frac{85.052 \text{ J m}^2}{(1 \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (273 \text{ K})}} \\ &= 7.2537 \times 10^{-3} \text{ m}^3 \end{aligned}$$

And we are back to the ideal gas law to find the pressure

$$\begin{aligned} P_D &= \frac{(1 \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}}\right) (273 \text{ K})}{7.2537 \times 10^{-3} \text{ m}^3} \\ &= 3.1291 \times 10^5 \text{ Pa} \end{aligned}$$

And so we can find the work as

$$\begin{aligned} W_{CD} &= (1 \text{ mol}) \left(8.314 \frac{\text{J}}{\text{mol K}}\right) (273 \text{ K}) \ln \left(\frac{0.015 \text{ m}^3}{7.2537 \times 10^{-3} \text{ m}^3} \right) \\ &= 1649.0 \text{ J} \end{aligned}$$

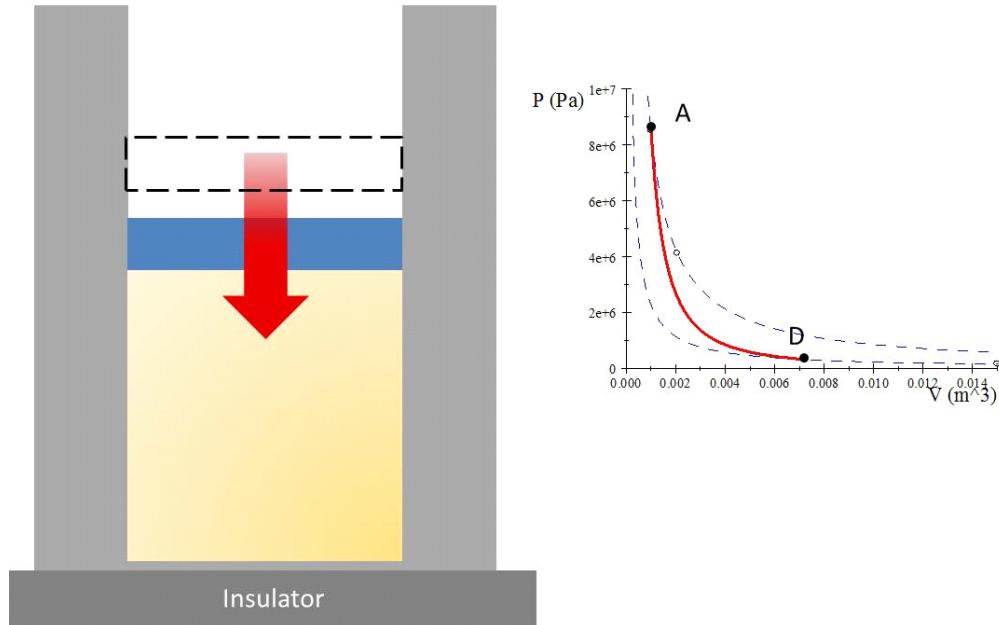
and our summary is as follows

State	T (K)	P (Pa)	V (m ³)	n (mol)	—	—
C	273	1.5131×10^5	0.015 m^3	1	—	—
D	273	3.1291×10^5	7.2537×10^{-3}	1	—	—
Process	Q (J)	W_{int} (J)	ΔE_{int} (J)	W_{eng} (J)	Q_h	Q_c
CD	-1649.0	1649.0	0	0	0	1649.0

38.6 Process $D \rightarrow A$ adiabatic compression

The gas is again placed against a thermally nonconducting wall so no heat is exchanged with the surroundings. The temperature of the gas increases from T_c to T_h . The work done on the gas is W_{DA}

Process D→A



We know $Q = 0$ and that $\Delta E_{int} = W$. Again we can use

$$\Delta E_{int} = nC_V\Delta T$$

$$\begin{aligned}\Delta E_{int} &= (1 \text{ mol}) \frac{3}{2} \left(8.314 \frac{\text{J}}{\text{mol K}} \right) (1023 \text{ K} - 273 \text{ K}) \\ &= 9353.3 \text{ J}\end{aligned}$$

so then

State	T (K)	P (Pa)	V (m³)	n (mol)	-	-
D	273	3.1291×10^5	7.2537×10^{-3}	1	-	-
A	1023	3.1×10^6	0.001	1	-	-
Process	Q (J)	W _{int} (J)	ΔE _{int} (J)	W _{eng} (J)	Q _h	Q _c
DA	0	9353.3	9353.3	-9353.3	0	0

All together we have

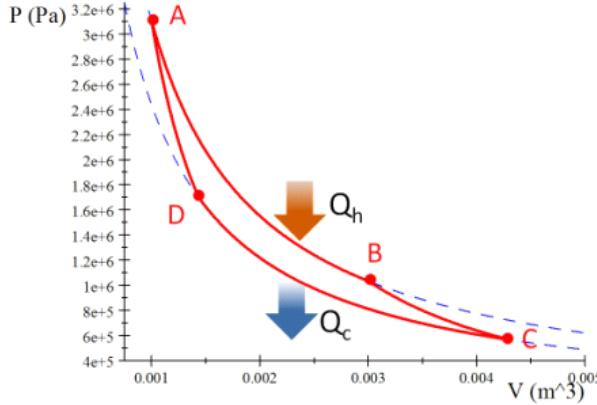
State	T (K)	P (Pa)	V (m³)	n (mol)
A	1023	8.5052×10^6	0.001	1
B	1023	4.1130×10^6	2.0679×10^{-3}	1
C	273	1.5131×10^5	0.015	1
D	273	3.1291×10^5	7.2537×10^{-3}	1

and our internal energy, heat, and work are

Process	Q (J)	W_{int} (J)	ΔE_{int} (J)	W_{eng} (J)	Q_h	Q_c
AB	6179.3	-6179.3	0	6179.3	6179.3	0
BC	0	-9353.3	-9353.3	9353.3	0	0
CD	-1649.0	1649.0	0	-1649.0	0	1649.0
DA	0	9353.3	9353.3	-9353.3	0	0
Total	4530.3	-4530.3	0	4530.3	6179.3	1649.0

Question 123.19.2

We can see that process $A \rightarrow B$ provides Q_h and process $C \rightarrow D$ provides Q_c . Let's label this in our PV diagram.



We can also find the efficiency of the Carnot cycle. The net work is W_{ABCDA} and the efficiency is

$$\eta = \frac{W_{eng}}{|Q_h|} = 1 - \frac{|Q_c|}{|Q_h|}$$

No heat is transferred during the adiabatic processes, so the only heat transfer is only during the isothermal processes, we have

$$\begin{aligned} \eta &= 1 - \frac{\left| nRT_c \ln \left(\frac{V_c}{V_d} \right) \right|}{\left| nRT_h \ln \left(\frac{V_A}{V_B} \right) \right|} \\ &= 1 - \frac{\left| T_c \ln \left(\frac{V_c}{V_d} \right) \right|}{\left| T_h \ln \left(\frac{V_A}{V_B} \right) \right|} \end{aligned}$$

and using the adiabatic temperature-volume relationships

$$T_h V_A^{\gamma-1} = T_c V_D^{\gamma-1}$$

and

$$T_h V_B^{\gamma-1} = T_c V_C^{\gamma-1}$$

$$\frac{T_h V_A^{\gamma-1}}{T_h V_B^{\gamma-1}} = \frac{T_c V_D^{\gamma-1}}{T_c V_C^{\gamma-1}}$$

or

$$\frac{V_A^{\gamma-1}}{V_B^{\gamma-1}} = \frac{V_C^{\gamma-1}}{V_D^{\gamma-1}}$$

which gives us

$$\sqrt[\gamma-1]{\frac{V_A^{\gamma-1}}{V_B^{\gamma-1}}} = \sqrt[\gamma-1]{\frac{V_C^{\gamma-1}}{V_D^{\gamma-1}}}$$

then

$$\frac{V_A}{V_B} = \frac{V_D}{V_C}$$

using this we find

$$\begin{aligned}\eta &= 1 - \frac{\left| T_c \ln \left(\frac{V_C}{V_D} \right) \right|}{\left| T_h \ln \left(\frac{V_D}{V_C} \right) \right|} \\ &= 1 - \frac{\left| T_c \ln \left(\frac{V_C}{V_D} \right) \right|}{\left| -T_h \ln \left(\frac{V_C}{V_D} \right) \right|}\end{aligned}$$

or, removing the minus sign because of the absolute values

$$\eta = 1 - \frac{T_c}{T_h}$$

where the temperatures, of course, is in Kelvin. This means that for this process

$$\frac{|Q_c|}{|Q_h|} = \frac{T_c}{T_h}$$

and

$$\eta_C = 1 - \frac{T_c}{T_h} \quad (38.5)$$

This is the general form for a Carnot cycle. Then for our example

$$\begin{aligned}\eta_C &= 1 - \frac{273 \text{ K}}{1023 \text{ K}} \\ &= 0.73314\end{aligned}$$

This is a great review of how to use our thermodynamic equations for special process, even though we can't build a Carnot engine. But more importantly, we can judge the efficacy of engine designs. For example, let's take our Otto cycle from last time. Our design gave

$$\begin{aligned}\eta_{otto} &= 1 - \frac{1}{\left(\frac{0.0005}{6.25 \times 10^{-5}} \right)^{1.4-1}} \\ &= 0.56472\end{aligned}$$

So the Otto cycle example might be achievable. Of course, we can tell that our Otto cycle is still an idealization. We recognize that we neglected all friction, and assumed perfect adiabatic processes. But we can see that there is some hope of making our Otto cycle-based engines more efficient. But no Otto cycle that operates between the same $T_c = 273\text{ K}$ and $T_h = 1023\text{ K}$ can never be better than 73% effective because that is the maximum efficiency for a heat engine operating between these two temperatures.

We are done with our introductory study of thermodynamics. Many of you will have junior level thermodynamics classes, so you are likely to see more of this. But we can stop for now. For physics majors the next physics class is likely to be PH220, electricity and magnetism. I would prefer to call this class “introduction to field theory” because we really study electric and magnetic fields for the first time. We will take up the topic of wave motion at the end of PH220 to mathematically describe light as electromagnetic waves. And then we will describe everything using waves in PH279, “modern physics,” as our first introduction to quantum mechanics.

I hope you have enjoyed learning about waves, optics, and thermodynamics. I hope you will see the world a little differently (and think about the processing of seeing the world!). For physics majors, there is still so many cool things to learn. If you are leaving physics at this point, you can continue to learn about the world and the universe by reading and watching physics related material. If you are innovating new techniques or products, you might even want to hire the services of a physicist to help with the underlying principles for your new venture!

Chapter 39

39 Microscopic Energy Transfer Considerations

2.4.6, 2.4.7

Fundamental Concepts

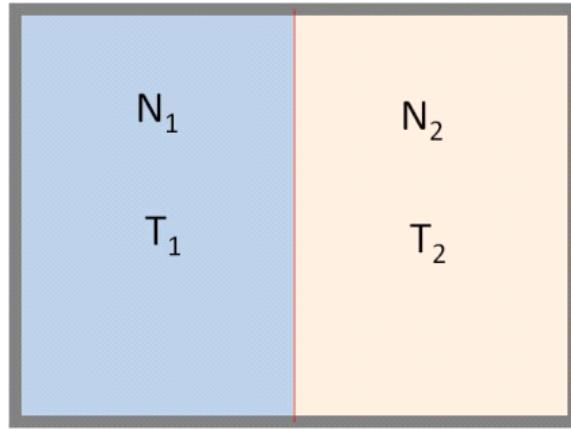
1. Microscopic idea of Energy Transfer
2. Microscopic idea of disorder
3. Entropy

39.1 Microscopic View of Energy Transfer

Question 123.16.1

Suppose we have a chamber divided into two volumes. The volumes are separated by a membrane. If they have different temperatures, we expect that energy will transfer, but how?

Think of a gym. In our gyms, we have large nets that divide courts to keep balls from passing from one court to another. Now think of millions of basket balls in each side of the gym. The basket balls all are moving. Some of the basket balls will strike the net. If a ball on the other side strikes the net at the same place at the same time, the collision will transfer energy. This would be a conservation of momentum and energy problem. The slower ball would gain energy, the faster ball would lose energy.



This is analogous to our gasses in chambers. The molecules, like the basketballs, transfer energy by collision.

The internal energy for each side of the chamber depends on both N and T

$$\begin{aligned} E_{1i} &= \frac{3}{2}N_1k_B T_{1i} \\ E_{2i} &= \frac{3}{2}N_2k_B T_{2i} \end{aligned}$$

The total energy is

$$E_{tot} = E_{1i} + E_{2i}$$

if our container does not change volume (no work done) and is insulated ($Q = 0$) then E_{tot} will not change. But eventually the two sides come to the same temperature

$$T_f = T_{1f} = T_{2f}$$

this happens when the average energy of the molecules (or basketballs) are the same. On average, neither side gains energy due to the collisions on the membrane (net). We can rewrite our energy equation for the final condition as

$$\bar{E}_{tot} = \bar{E}_{1i} = \bar{E}_{2f}$$

because the average energies will all be the same. Using

$$\begin{aligned} E_{1f} &= N_1 \frac{3}{2}k_B T_{1f} = N_1 \bar{E}_1 \\ E_{2f} &= N_2 \frac{3}{2}k_B T_{2f} = N_2 \bar{E}_2 \end{aligned}$$

we can rewrite this as an energy per molecule

$$\frac{E_{tot}}{N_1 + N_2} = \frac{E_{1f}}{N_1} = \frac{E_{2f}}{N_2}$$

which gives a set of expressions for the final energy of each side

$$\begin{aligned} E_{1f} &= \frac{N_1 E_{tot}}{N_1 + N_2} \\ E_{2f} &= \frac{N_2 E_{tot}}{N_1 + N_2} \end{aligned}$$

and we can verify that

$$\begin{aligned} E_{tot} &= E_{1f} + E_{2f} \\ &= \frac{N_1 E_{tot}}{N_1 + N_2} + \frac{N_2 E_{tot}}{N_1 + N_2} \\ &= E_{tot} \end{aligned}$$

Let's take what we have learned and mathematically expressed as macroscopic formula and show it at a microscopic level as a problem example.

Let's show that if $Q_{tot} = 0$, and no work is done, then

$$Q_1 = -Q_2$$

This is a calorimetry problem.

Start with the first law,

$$\begin{aligned} Q_1 &= \Delta E_1 = E_{1f} - E_{1i} \\ Q_2 &= \Delta E_2 = E_{2f} - E_{2i} \end{aligned}$$

since no work is done. Let's take Q_1 first

$$Q_1 = E_{1f} - E_{1i}$$

since

$$E_{1f} = \frac{N_1 E_{tot}}{N_1 + N_2}$$

and

$$E_{tot} = E_{1i} + E_{2i}$$

then

$$Q_1 = \frac{N_1 (E_{1i} + E_{2i})}{N_1 + N_2} - E_{1i}$$

Then, applying some algebra,

$$\begin{aligned} Q_1 &= \frac{N_1 (E_{1i} + E_{2i})}{N_1 + N_2} - E_{1i} \frac{N_1 + N_2}{N_1 + N_2} \\ &= \frac{N_1 E_{1i} + N_1 E_{2i} - E_{1i} N_1 - E_{1i} N_2}{N_1 + N_2} \\ &= \frac{N_1 E_{2i} - E_{1i} N_2}{N_1 + N_2} \end{aligned}$$

Expanding this out gives

$$\begin{aligned}
 Q_1 &= \frac{N_1 E_{2i}}{N_1 + N_2} - \frac{E_{1i} N_2}{N_1 + N_2} \\
 &= \frac{N_1 E_{2i}}{N_1 + N_2} - \frac{E_{1i} N_2}{N_1 + N_2} + \frac{N_2 E_{2i}}{N_1 + N_2} - \frac{N_2 E_{2i}}{N_1 + N_2} \\
 &= \frac{N_1 E_{2i}}{N_1 + N_2} + \frac{N_2 E_{2i}}{N_1 + N_2} - \frac{E_{1i} N_2}{N_1 + N_2} - \frac{N_2 E_{2i}}{N_1 + N_2} \\
 &= \frac{N_1 E_{2i} + N_2 E_{2i}}{N_1 + N_2} - \frac{E_{1i} N_2}{N_1 + N_2} - \frac{N_2 E_{2i}}{N_1 + N_2}
 \end{aligned}$$

The first term is just E_{2i} , and using $E_{tot} = E_{1i} + E_{2i}$ again we see

$$\begin{aligned}
 Q_1 &= E_{2i} - \left(\frac{(E_{1i} N_2 + N_2 E_{2i})}{N_1 + N_2} \right) \\
 &= E_{2i} - \left(\frac{N_2 (E_{1i} + E_{2i})}{N_1 + N_2} \right) \\
 &= E_{2i} - \left(\frac{N_2 (E_{tot})}{N_1 + N_2} \right) \\
 &= E_{2i} - E_{2f} \\
 &= -Q_2
 \end{aligned}$$

which is what we expect. We have shown our calorimetry equation using the ideas of microscopic energy transport for ideal gasses.

39.2 The arrow of time.

Why does time go only one way? The answer to this question may surprise you. *We don't know.* There is nothing in our laws of motion that tells us that time should only go one way. Take motion under constant acceleration.

$$y_f = y_i + v_o \Delta t + \frac{1}{2} a \Delta t^2$$

and let's start with $t_i = 0$ so

$$y_f = y_i + v_o t_f + \frac{1}{2} a t_f^2$$

this equation describes the motion of an object under constant acceleration, like that of gravity near the Earth's surface. But nothing in the equation tells us that time must take on ever greater more positive values.

The truth is that there is only one physical law that tells us that time goes only one way, and that law is an empirical law that says essentially "time only goes one way." We experience time going one way, so we developed a law that says so. This is not a terribly convincing argument. But this is the state of physics.

Some processes (really all processes) can't actually run forward or backward. Idealized processes can be run in reverse, but not real process. We call processes that can only work one direction *irreversible processes*. If we shot a movie of an irreversible process and ran the movie backwards, it would be funny, because we would know that the reverse process can't happen.

Let's take an example. Suppose our process is dropping a marker on the floor. Because the marker bounces on the floor are not perfectly elastic, the potential energy of the marker is converted into heat energy. If we added the lost energy back to the marker on the floor, it won't bounce up to a height for us to catch. This process of dropping the marker is irreversible.

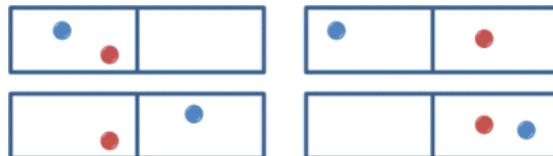
To make this idea of irreversible processes and the arrow of time sound a little better, let's study thermodynamic situations and see that at least the arrow of time make sense.

39.3 Disorder

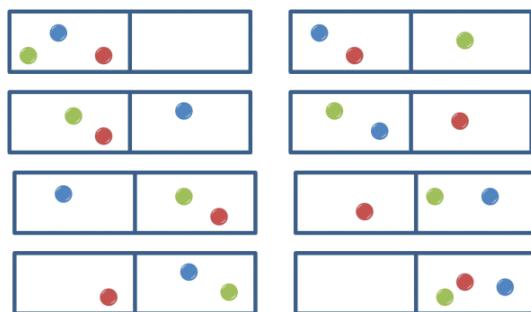
Suppose we have a system that has particles and places for those particles to be placed. We can start with a small space (see figure below). If I have one particle moving randomly, and divide my space into two compartments, then I have a probability of $\frac{1}{2}$ that the particle will be in the left compartment.



Now if I have two particles, the probability that two particles will be in the left compartment is $(\frac{1}{2})^2$ or $\frac{1}{4}$.



I can keep adding particles. For three particles the probability of being in the left side is $(\frac{1}{2})^3$ or $\frac{1}{8}$.



For 50 particles it is

$$\left(\frac{1}{2}\right)^{50} = \frac{1}{1125\,899\,906\,842\,624}$$

We can extend this thinking to divisions of particles. Suppose I have 100 particles. We know that they will have a distribution of speeds (Maxwell-Boltzman distribution). What is the probability that the 50 fastest particles will be in the left side and the 50 slowest will be on the right side? This is the product of the last result with itself

$$\begin{aligned} \left(\frac{1}{2}\right)^{50} \left(\frac{1}{2}\right)^{50} &= \\ &= \frac{1}{1267\,650\,600\,228\,229\,401\,496\,703\,205\,376} \\ &= 7.888\,6 \times 10^{-31} \end{aligned}$$

BYU 50 particle demonstrationn

This is not very probable. And we are only working with 100 particles and two sides of a box!

The point to this exercise is to realize there are very many *states* or ways to place particles in a divided box. Although each configuration is equally probable, suppose we want a particular outcome, say, the fastest 50 molecules on one side and the slowest 50 molecules on the other. There are many many more ways to *not* have this configuration than there are to have this configuration.

Each configuration of molecules and locations is called a *microstate*. There are a huge number of microstates in our example.

The collection of microstates is called a *macrostate*. There is more than one configuration of particles in our fast-slow division that places all the fast on one side and all the slow on another side. But the collection of all microstates that make up this division would be a macrostate. For example, we could have an even distribution of particles (within their fast/slow groups) on each side, or we could have all the fastest particles in each group on the top, and the slowest on the bottom within each major fast or slow side of the room. These two microstates still preserve the division between the 50 fastest on the left and the slowest 50 on the right. They just rearrange the molecules within their groups.

39.4 Extension to large spaces and many particles

Now let's take a room full of molecules. We count the number of microstates by considering the number of places we can put a molecule. Each molecule has a volume V_m . So there are

$$w = \frac{V}{V_m} \tag{39.1}$$

places to put the molecule that are distinct. Now suppose we have N molecules. If we ignore the probability that we might have two molecules in one space (which is different than the case above!) we have

$$(w)^N = \left(\frac{V}{V_m} \right)^N \quad (39.2)$$

ways to place these molecules among our w states. I'll call this

$$W_i = \left(\frac{V_i}{V_m} \right)^N \quad (39.3)$$

where V_i is a particular volume (we will let it change to V_f in just a minute).

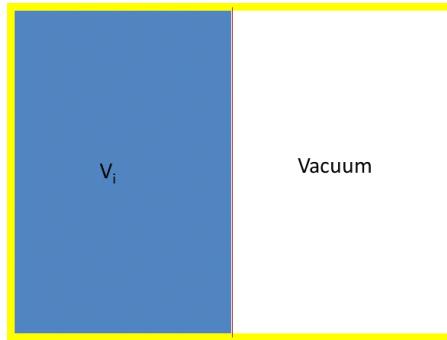
39.5 A Measure of Disorder

It would be convenient to find a way to measure the disorder. Then we could see how a thermodynamic process affects the amount of disorder. The traditional way to do this is to define a quantity

$$S \equiv k_B \ln(W) \quad (39.4)$$

This has the property that S increases as the number of states, W , increases.

Let's take an example, our old friend, the adiabatic free expansion.



For this case we start with all the atoms on one side of a membrane. No energy transfers out by heat. No work is done, we just puncture the membrane in the center and let the gas flow from one side to the other. How does the disorder change in this process? Consider the number of states.

$$W_i = \left(\frac{V_i}{V_m} \right)^N$$

$$W_f = \left(\frac{V_f}{V_m} \right)^N$$

We can write our disorder measure as

$$S_i \equiv k_B \ln(W_i)$$

and

$$S_f \equiv k_B \ln(W_f)$$

then the change is order is

$$\begin{aligned}\Delta S &= S_f - S_i \\ &= k_B \ln(W_f) - k_B \ln(W_i) \\ &= k_B \ln\left(\frac{W_f}{W_i}\right)\end{aligned}$$

If we look at the ratio of the number of states

$$\begin{aligned}\frac{W_f}{W_i} &= \frac{\left(\frac{V_f}{V_m}\right)^N}{\left(\frac{V_i}{V_m}\right)^N} \\ &= \left(\frac{V_f}{V_i}\right)^N\end{aligned}$$

so

$$\begin{aligned}\Delta S &= k_B \ln\left(\left(\frac{V_f}{V_i}\right)^N\right) \\ &= N k_B \ln\left(\left(\frac{V_f}{V_i}\right)\right)\end{aligned}\tag{39.5}$$

We should ask if the order increased or decreased. Since V_f is larger for our case than V_i , then ΔS will be a positive number. This means the disorder has increased in this case.

It is a good question to ask, will ΔS ever be negative? In other words, will the gas ever line up on only one side every again? If the answer is no, then this is an irreversible process. We can guess that this process is irreversible. We would have to do work to pump the gas back into one side.

39.6 Definition of entropy

Question 123.16.1

We have learned to use one state variable, ΔE_{int} . The trick is to realize that ΔE_{int} only depends on ΔT and use any easy path to calculate it's value. That is why we can use

$$\Delta E_{int} = nC_V\Delta T$$

no matter what process we have. State variables, like ΔE_{int} give us powerful ways to calculate what will happen in a system.

Question 123.16.2

There is another state variable in Thermodynamics, *change in entropy*, ΔS . We will again find that we can calculate ΔS by any path because it depends only on the starting and ending points. And from our previous discussion we recognize that this new state variable, it is just a change in our disorder parameter!

$$\Delta S = k_B \ln \left(\frac{W_f}{W_i} \right)$$

Note that this is a difference between two disorder parameters. We will call our disorder parameter the *entropy* of the system.

$$S \equiv k_B \ln (W) \quad (39.6)$$

Entropy was originally defined at the macroscopic level. Statistical Mechanics (the microscopic study of thermodynamics) gives us our modern definition we have followed. We can state this definition as

Definition 1 *Entropy is a measure of the amount of disorder in a system and its environment.*

39.6.1 Entropy and the Second Law

In our example, because, $V_f > V_i$ we found that ΔS will be positive. This implies that the disorder has increased. We can see that this is true, because there are now molecules distributed among many more states (position, occupation combinations) than before.

We will find this to be true in general, for irreversible processes (all natural process that are recognized). The total entropy of an isolated system that undergoes a change cannot decrease.

What about non-isolated systems? We must then consider the entropy of the system *and* its surroundings. We again find that entropy of a system *and* its surroundings cannot decrease. But the entropy of a system can decrease if the entropy of the surroundings increases more so the total effect is an increase in entropy.

These statements constitute the most commonly seen form of the second law of thermodynamics.

39.6.2 Macroscopic view

We don't really want to count the microstates of a process and look for changes in their number to do calculations. It would get tedious very fast! And so far our definition of ΔS was done only for a free expansion. We expect that in some way disorder would increase if we transferred in energy by heat, or did some kind of work on the system. We need a macroscopic version of our definition of entropy.

We will give a definition and then show that it works. The change in entropy is given by

$$\Delta S = \frac{\Delta Q_r}{T} \quad (39.7)$$

where Q_r is the energy transferred by heat through a reversible path and T is the absolute temperature (K) of the system. The assumption of a single temperature is because we have assumed ΔQ_r is small. But why do we use a reversible path?

Suppose we have an irreversible process. It might be hard to compute Q . But remember that, like ΔE_{int} , ΔS does not depend on path,! it depends only on the initial and final states. So we may calculate it using any convenient path, and a reversible path is more convenient.

Note that macroscopically we have not defined entropy but the change in entropy, ΔS . This is the meaningful quantity (like ΔE_{int}) for thermodynamic calculations.

Let's do an example: What is the change in entropy when a block of ice melts? We will use

$$\Delta S = \int \frac{dQ_r}{T}$$

Note that while the ice is melting, we will have ice water at a constant $T_{melt} = 273\text{ K}$. Then our integral is

$$\Delta S = \frac{1}{T_{melt}} \int dQ_r$$

and the integral is not too hard

$$\Delta S = \frac{Q_r}{T_{melt}}$$

and for melting we know what Q_r could be. We can freeze water as well as melt ice. So we could model this as a reversible process. The amount of energy transfer needed to melt the ice is just

$$Q = mL$$

where m is the mass of the ice and L is the latent heat of fusion for water. So

$$\Delta S = \frac{mL}{T_{melt}}$$

This wasn't too hard!

Let's do another example to tie our macroscopic ΔS to our microscopic ΔS . Let's find the change in entropy for a gas brought through an isothermal expansion. Again we have

$$\Delta S = \int_{V_i}^{V_f} \frac{dQ_r}{T}$$

and since the process is isothermal, T is constant

$$\Delta S = \frac{1}{T} \int_{V_i}^{V_f} dQ_r$$

and the integral would seem easy. It is just Q , but it is a particular Q for an isothermal process. For an isothermal process we know

$$\Delta E_{int} = 0$$

and

$$\Delta E_{int} = Q + w$$

so

$$Q = -w$$

and further we know that for an isothermal process

$$\begin{aligned} w &= - \int P dV \\ &= -nRT \ln\left(\frac{V_f}{V_i}\right) \end{aligned}$$

so the Q_r for our process could be just

$$Q = nRT \ln\left(\frac{V_f}{V_i}\right)$$

Then

$$\begin{aligned} \Delta S &= \frac{1}{T} nRT \ln\left(\frac{V_f}{V_i}\right) \\ &= nR \ln\left(\frac{V_f}{V_i}\right) \end{aligned}$$

and remember that $nR = Nk_B$ so

$$\begin{aligned} \Delta S &= \\ &= Nk_B \ln\left(\frac{V_f}{V_i}\right) \end{aligned}$$

which is just what we found from our microscopic view of entropy change! Now we realize that when we let the gas flow from one side of our divide box to the other, we did this isothermally. And that is true, the gas did not change the average kinetic energy per molecule as it flowed from one side of the box to the other. Since our system was isolated, $Q = 0$ and $W = 0$ for the whole system, so $\Delta E_{int} = 0$ and the system did not change temperature. But surely something changed! And that something was ΔS !

39.6.3 Heat death of the Universe

We should ask the question, is the universe an isolated system? We don't know the answer to this for sure¹, but we believe this is true that the universe is

¹ After all, the "Flash" and his colleagues seem to be able to come and go from different "universes all the time."

isolated.² That means that the change in entropy of the universe will always be positive. The universe is “running down.” The ultimate end of this increase in energy would mean that all matter in the universe will be spread out and cold. This is called the *heat death of the universe*.

Note that this is a little different than the eschatological view that we have in the Church of Jesus Christ of Latter Day Saints.

Throughout my career, many colleagues and acquaintances have said it is foolish to believe in a church and a resurrection because the second law of thermodynamics tells us that the universe will die in a cold blur. How do we respond?

I can only give my opinion. But I think it is worth discussing, so here is a disclaimer for what follows.

WARNING: This is not an official statement of doctrine. This is an answer to a question that many scientists must face from colleagues who will question their faith using the principals of thermodynamics. Official statements by the General Authorities supersede any opinions expressed here

39.6.4 The First Law and LDS Thought³

I suppose we should ask, do we believe in the first law of thermodynamics. If this law does not work for us, we have no reason to expect the second law to work. The first law is basically conservation of energy. Here are some quotes to consider.

And there stood one among them that was like unto God, and he said unto those who were with him: We will go down, for there is space there, and we will take of these materials, and we will make an earth whereon these may dwell; (Abraham 3:24)

You ask the learned doctors why they say the world was made out of nothing; and they will answer, ‘Doesn’t the Bible say He created the world?’ And they infer, from the word create, that it must have been made out of nothing. Now, the word create came from the word baurau which does not mean create out of nothing; it means to organize; the same as a man would organize materials and build a ship. (Joseph Smith, King Follett Discourse)

The idea of the first law of thermodynamics seems in harmony with LDS thought.

So let’s take on the Second Law.

²Dispite what the television shows seem to say.

³Comparing physical theory to LDS thought has its dangers, I am aware. But as scientists, we can’t escape thinking about this. Just as a caution, John A. Widsoe wrote a book called *Joseph Smith as Scientist* in which Widsoe tried very hard to show that LDS thought is in harmony with Universal Ether Theory. But we now know Universal Ether Theory is not correct, so being in harmony with it is meaningless. So we should be cautious!

39.6.5 The Second Law and LDS Thought

Again here are some quotes

Wherefore, the first judgment which came upon man must needs have remained to an endless duration. And if so, this flesh must have laid down to rot and to crumble to its mother earth, to rise no more. (2 Nephi 9:7)

And our spirits must have become like unto [Satan], and we become devils, angels to a devil, to be shut out from the presence of our God, and to remain with the father of lies, in misery, like unto himself. (2 Nephi 9:9)

We seem to accept the concept of the second law as well! –This will be a surprise to our non-member friends.

But then how can we expect eternal life and exaltation? Does something counter the second law (a Third or Fourth Law of Thermodynamics)?

A Third Law of Thermodynamics?

Again let's go to the scriptures. Jacob describes another force not recognized by current scientific theory that counteracts this tendency to decay

Wherefore, it must needs be an infinite atonement—save it should be an infinite atonement this corruption could not put on incorruption. Wherefore, the first judgment which came upon man must needs have remained to an endless duration. And if so, this flesh must have laid down to rot and to crumble to its mother earth, to rise no more. (2 Nephi 9:7-8)

O how great the goodness of our God, who prepareth a way for our escape from the grasp of this awful monster; yea, that monster, death and hell, which I call the death of the body, and also the death of the spirit. (2 Nephi 9:10)

Atonement means to bring together. This is an apt description of a process or law that would counter or balance the second law⁴. Since the atonement can only be accessed through Christ, it is not likely to be discovered in a laboratory, but is a very real, physical, process. (see for example Luke 24:39)

This is not doctrine, but is a way to think about the second law until we get direct revelation on the subject. We really should not be worried about a conflict between thermodynamics and our doctrine. I don't believe one exists as you can see.

⁴Nibley, The meaning of the temple, reprinted in the Collected Works of Hugh Nibley: volume 12, Deseret Book, Salt Lake City, UT, 1992

Appendix A

Experiment 1

A.1 Introduction

We have learned that

$$\omega^2 = \frac{k}{m}$$

and we know that

$$\omega = 2\pi f$$

and

$$T = \frac{1}{f}$$

so we would predict that

$$\begin{aligned} T &= \frac{1}{\frac{\omega}{2\pi}} = \frac{2\pi}{\omega} \\ &= 2\pi \sqrt{\frac{m}{k}} \end{aligned} \tag{A.1}$$

We would like to verify that this is true experimentally.

Unfortunately, springs don't come marked with their spring constants. You will need to measure your spring constant. A way to do this is to hang your spring motionless. When the mass is not moving Newton's second law tells us

$$\sum_y F = ma_y = F_S - F_g$$

With no change in velocity, the acceleration must be zero

$$0 = F_S - F_g$$

so

$$\begin{aligned} F_S &= F_g \\ k\Delta y &= mg \end{aligned}$$

or

$$k = \frac{mg}{\Delta y} \quad (\text{A.2})$$

We believe that the equation of motion for our mass-spring system would be

$$\begin{aligned} y(t) &= y_{\max} \cos(\omega t + \phi_o) \\ &= y_{\max} \cos\left(\frac{2\pi}{T}t + \phi_o\right) \end{aligned} \quad (\text{A.3})$$

A.2 Equipment

We have springs, masses, and motion detectors. Your motion detector should be placed below the mass as was demonstrated in class. Be careful to not drop the mass on the motion detector.

The motion detectors connect to our laptop computers. Use the LoggerPro program to collect data. Your lab instructor will help you set up the program and show you how to collect data.

A.2.1 Assignment

1. Find your spring constant
2. Using equation (A.1) predict the period of oscillation of your mass and spring.
3. Suspend your mass from the spring and capture several cycles of the motion using the motion detector.
4. Verify that equation (A.3) is a good model for our physical situation by fitting a sinusoidal curve to the data (ask instructor for help).
5. Using your curve fit equation, compare your predicted period to your measured period. Do this by matching the fit equation to our theoretical equation of motion

$$\begin{aligned} Y &= A * \cos(Bx + C) + D \\ y(t) &= y_{\max} \cos\left(\frac{2\pi}{T}t + \phi_o\right) \end{aligned}$$

we can see that

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

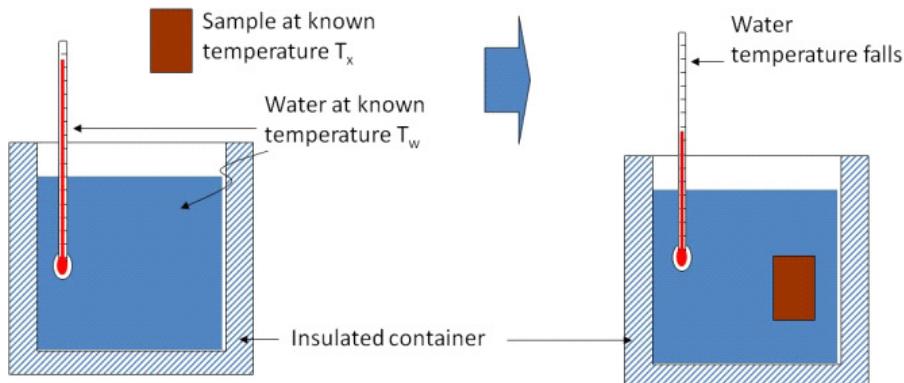
and solve for T .

6. Do your predicted and measured periods match?
7. If they don't match, consider your inputs in equation (A.1) and try again.

Appendix B

Experiment 2

B.1 Introduction



We studied calorimetry earlier in the semester. The idea behind calorimetry is just conservation of energy. A calorimeter is supposed to be an insulated system, so energy can't leave by heat.

Your job today is to measure the latent heat of fusion of ice/water.

Some helpful reminders are the specific heat equation

$$Q = m_i c_i \Delta T$$

and the latent heat equation.

$$\begin{aligned} Q &= \pm L_f M && \text{melt/freeze} \\ Q &= \pm L_v M && \text{boil/condense} \end{aligned}$$

From our table we used in class, we know what the value should be.

Substance	T_m (°C)	L_f ($\frac{\text{J}}{\text{kg}}$)	T_m (°C)	L_v ($\frac{\text{J}}{\text{kg}}$)
Nitrogen (N_2)	-210	0.26×10^5	-196	1.99×10^5
Ethyl alcohol	-114	1.09×10^5	78	8.79×10^5
Mercury	-39	0.11×10^5	357	2.96×10^5
Water	0	3.33×10^5	100	22.6×10^5
Lead	328	0.25×10^5	1750	8.58×10^5

Find the latent heat of fusion for water/ice and calculate a percent error to see how well you did

$$\%Error = \frac{measured - actual}{actual}$$

B.2 Equipment

I will have ice water available in a tub. I will also have calorimeter cups, thermometers, and triple beam balances for finding mass. Water will be available through our store room. Let me know if you need something else.

B.3 Assignment

Measure the latent heat of fusion of water.

1. Place some ice in water to form a slush (this will give you ice at 0 °C).
2. Place some water into the calorimeter cup, and measure its temperature and mass.
3. Place a piece of ice into the water in the calorimeter. You will need to find a way to measure its mass.
4. Measure the temperature change of the water after the ice has melted.
5. Calculate the latent heat, and compare to the accepted value.
6. Turn in your answer, $\%Error$, and all your group names on a single sheet of paper (the lab instructions sheet is fine)

Appendix C

Experiment 3

C.1 Introduction

A converging lens is one which is thicker at the center than at the edge, and which converges incident parallel rays to a real focus on the opposite side of the lens from the object. A diverging lens is thinner at the center than at the edge and diverges the light from a virtual focus on the same side of the lens as the object. The optic axis of a lens is a line drawn through the center of the lens perpendicular to the face of the lens. The principal focal point is a point on the principal axis through which incident rays parallel to the principal axis pass, or appear to pass, after refraction by the lens. The focal length, f , of the lens is the distance from the optical center of the lens to the principal focal point. The reciprocal of the focal length in meters is called the power of the lens and is expressed in diopters (d).

$$P = \frac{1}{f \text{ (m)}}$$

If s is the distance of the object from the optical center of the lens, s' the distance of the image from the lens, and f the focal length, the thin lens equation is

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \quad (\text{C.1})$$

The magnification is given by

$$m = -\frac{s'}{s} \quad (\text{C.2})$$

For the case where more than one lens is present, Equation (C.2) is applied successively to each lens, treating the image formed by one lens as the object for the next lens, keeping in mind that the distances in Equation (C.2) are measured with respect to each individual lens, and not a single fixed point.

C.2 Equipment

We have 2 m sticks, lens and holders, paper screens, and light sources.

C.3 Assignment

Verify the lens equation (C.1), and the image magnification equation (C.2), for the lens with the longer focal length. I suggest you use at least three object distances greater than two focal lengths. I suggest you measure the image distance and the image height. I suggest you also measure at least three object distances between one and two focal lengths. Measure the image distance and the image height.

Plot your magnification results. Interpret the results, does the lens equation work?

Compute the magnification in each case and verify that equation (C.2) holds.

The lens equation is supposed to be good for virtual images as well as real images. We need to verify the equation for this case as well. I suggest you do this by measuring at least one object distance closer to the lens than one focal length, locate and measure the image distance. Measure the image height. Verify that equations (C.1) and (C.2) are valid for this case. Include this case on your graph.

Problem Types

Simple Harmonic Motion (SHM)

$$\begin{aligned}x(t) &= x_{\max} \cos(\omega t + \phi_o) \\v(t) &= -\omega x_{\max} \sin(\omega t + \phi_o) \\a(t) &= -\omega^2 x_{\max} \cos(\omega t + \phi_o)\end{aligned}$$

$$\omega = 2\pi f$$

$$f = \frac{1}{T}$$

$$v_{\max} = \omega x_{\max}$$

$$a_{\max} = \omega^2 x_{\max}$$

Energy in SHM

$$\begin{aligned}K &= \frac{1}{2}mv^2 \\&= \frac{1}{2}kx_{\max}^2 \sin^2(\omega t + \phi)\end{aligned}$$

$$\begin{aligned}U_s &= \frac{1}{2}kx^2 \\&= \frac{1}{2}kx_{\max}^2 \cos^2(\omega t + \phi)\end{aligned}$$

Angular Frequency and oscillation

For a mass sprig system

$$\omega = \sqrt{\frac{k}{m}}$$

for a pendulum

$$\omega = \sqrt{\frac{g}{L}}$$

for a physical pendulum

$$\omega = \sqrt{\frac{mgd}{I}}$$

Damped oscillation

$$x(t) = x_{\max} e^{-\frac{b}{2m}t} \cos(\omega t + \phi_o)$$

$$\omega = \sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2}$$

$$x(t) = x_{\max} e^{-\frac{b}{2m}t} \cos \left(\left(\sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2} \right) t + \phi_o \right)$$

Forced oscillation

$$\mathbf{F}_d = -b\mathbf{v}$$

$$F(t) = F_o \sin(\omega_f t)$$

$$\Sigma F = F_o \sin(\omega_f t) - kx - bv_x = ma$$

$$x(t) = A \cos(\omega_f t + \phi)$$

$$A = \frac{\frac{F_o}{m}}{\sqrt{\left(\omega_f^2 - \omega_o^2\right)^2 + \left(\frac{b\omega_f}{m}\right)^2}}$$

$$\omega_o = \sqrt{\frac{k}{m}}$$

Waves

$$y(t) = A \cos(ax - \omega t + \phi)$$

$$k = \frac{2\pi}{\lambda}$$

$$v = \frac{\lambda}{T}$$

$$v = \lambda f$$

Waves on strings

$$v = \sqrt{\frac{T_s}{\mu}}$$

$$\mu = \frac{m}{L}$$

Sound waves

$$v_{sound} = v_o \sqrt{1 + \frac{T_c}{T_o}}$$

$$v_o = 331 \frac{\text{m}}{\text{s}}$$

$$T_o = 273$$

$$P \equiv \frac{F}{A}$$

$$s(x, t) = s_{\max} \cos(kx - \omega t + \phi_o)$$

Spherical waves

$$y = A(r) \sin(kr - \omega t + \phi_o)$$

Light waves

$$n = \frac{c}{v}$$

$$c = 299792458 \frac{\text{m}}{\text{s}}$$

Power and Intensity

$$\mathcal{P} = \frac{\Delta E}{\Delta t}$$

$$\mathcal{I} \equiv \frac{\mathcal{P}}{A}$$

$$A = 4\pi r^2$$

Sound Intensity Level

$$\beta = 10 \log_{10} \left(\frac{I}{I_o} \right)$$

$$I_o = 1 \times 10^{-12} \frac{\text{W}}{\text{m}^2}$$

Doppler Effect

$$f' = \frac{v + v_d}{v} f \quad \text{observer moving toward the source}$$

$$\begin{aligned}
 f' &= \frac{v - v_d}{v} f && \text{observer moving away from the source} \\
 f' &= \frac{v}{v - v_s} f && \text{Source moving toward observer} \\
 f' &= \frac{v}{v + v_s} f && \text{Source moving away from observer} \\
 f' &= \frac{v \pm v_d}{v \mp v_s} f
 \end{aligned}$$

Superposition

$$\begin{aligned}
 \sin a + \sin b &= 2 \cos\left(\frac{a-b}{2}\right) \sin\left(\frac{a+b}{2}\right) \\
 y_r &= y_{\max} \sin(kx - \omega t) + A \sin(kx - \omega t + \phi_o) \\
 &= 2y_{\max} \cos\left(\frac{\phi_o}{2}\right) \sin\left(kx - \omega t + \frac{\phi_o}{2}\right)
 \end{aligned}$$

Standing waves

$$\begin{aligned}
 \sin(a \pm b) &= \sin(a) \cos(b) \pm \cos(a) \sin(b) \\
 y_r &= y_{\max} \sin(kx - \omega t) + y_{\max} \sin(kx + \omega t) \\
 &= (2y_{\max} \sin(kx)) \cos(\omega t) \\
 x &= n \frac{\lambda}{4} \\
 f_n &= \frac{v}{\lambda_n} = \\
 &= \frac{n}{2L} v \\
 &= \frac{n}{2L} \sqrt{\frac{T}{\mu}}
 \end{aligned}$$

Standing waves in pipes
Open on both ends

$$\begin{aligned}
 \lambda_n &= \frac{2}{n} L \\
 f_n &= n \frac{v}{2L} \quad n = 1, 2, 3, 4, \dots
 \end{aligned}$$

Closed on one end

$$\begin{aligned}
 \lambda_n &= \frac{4}{n} L \\
 f_n &= n \frac{v}{4L} \quad n = 1, 3, 5, \dots
 \end{aligned}$$

Afterword

The back matter often includes one or more of an index, an afterword, acknowledgements, a bibliography, a colophon, or any other similar item. In the back matter, chapters do not produce a chapter number, but they are entered in the table of contents. If you are not using anything in the back matter, you can delete the back matter `TEX` field and everything that follows it.