Physics 135 Coursepack
Physics for the Life Sciences I

Author:
Tim McKay

Table of Contents

Syllabus

Important Dates to Remember

Lecture Notes
Physics 135 Fall 2009: Lecture Notes

Physics 135: Physics for Life Sciences I
Syllabus for Fall 2009

Lecture: MW: 11–12
Location: Room 170 Dennison

Discussion: TuTh: 10-11, 11-12, or 12-1
Location: 1250 Undergraduate Science Building

Primary Lecture: Professor Michal Zochowski
• Office Hours: Wednesday 1:30-3:30 in the Physics Help Room
• Email: michalz@umich.edu

Discussion Instructor: Dr. Toby Eckhause
• Office Hours: Mondays 10:00-12:00 and Tuesdays 2:00-4:00 in the Physics Help Room and by appointment
• Email: eckhause@umich.edu

<table>
<thead>
<tr>
<th>Lecture #</th>
<th>Date</th>
<th>Lecture Title</th>
<th>Pre-Lecture Reading</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wednesday, Sept. 9</td>
<td>Physics and Life</td>
<td>Lecture notes #1</td>
<td>Physics and life</td>
</tr>
<tr>
<td>3</td>
<td>Wednesday, Sept. 16</td>
<td>Not Changing: Newton’s First and Third Laws</td>
<td>Lecture notes #3</td>
<td>Newton’s Laws, forces, friction, Free Body Diagrams</td>
</tr>
<tr>
<td>4</td>
<td>Monday, Sept. 21</td>
<td>Standing Up and Staying Still</td>
<td>Lecture notes #4</td>
<td>Tension, force transmission, and simple machines</td>
</tr>
<tr>
<td>5</td>
<td>Wednesday, Sept. 23</td>
<td>How Do Forces Happen I: Stretching and Squashing</td>
<td>Lecture notes #5</td>
<td>Rotational equilibrium, Stress and Strain</td>
</tr>
<tr>
<td>6</td>
<td>Monday, Sept. 28</td>
<td>How Do Forces Happen II: Friction</td>
<td>Lecture notes #6</td>
<td>Static friction, kinetic friction</td>
</tr>
<tr>
<td>7</td>
<td>Wednesday, Sept. 30</td>
<td>Describing Motion: Velocity and Acceleration</td>
<td>Lecture notes #7</td>
<td>Position, speed, acceleration, PT, VT, AT graphs. Do this for both linear and rotational</td>
</tr>
<tr>
<td>Date</td>
<td>Day</td>
<td>Topic</td>
<td>Lecture Notes</td>
<td>Notes</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------</td>
<td>---------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Monday, Oct. 5</td>
<td>8</td>
<td>Starting and Stopping: Newton’s Second Law</td>
<td>Lecture notes #8</td>
<td>Movement in a line, momentum, fluid friction</td>
</tr>
<tr>
<td>Wednesday, Oct. 7</td>
<td>9</td>
<td>Changing Directions</td>
<td>Lecture notes #9</td>
<td>General curved paths, circular motion, parabolic motion</td>
</tr>
<tr>
<td>Thursday, Oct. 8</td>
<td></td>
<td>Exam #1 Today</td>
<td>8-10 pm</td>
<td>Location TBA</td>
</tr>
<tr>
<td>Monday, Oct. 12</td>
<td>10</td>
<td>How Much Motion Do You Have: Kinetic Energy</td>
<td>Lecture notes #10</td>
<td>Work and Kinetic Energy</td>
</tr>
<tr>
<td>Monday, Oct. 19</td>
<td></td>
<td>No Lecture Today: Fall Study Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday, Oct. 20</td>
<td></td>
<td>No Discussion Today: Fall Study Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday, Oct. 21</td>
<td>12</td>
<td>Collisions: Impulse and Momentum</td>
<td>Lecture notes #12</td>
<td>Impulse and change in momentum, collisions and the third law</td>
</tr>
<tr>
<td>Monday, Oct. 26</td>
<td>13</td>
<td>Small Disturbances: Oscillation Around Equilibrium</td>
<td>Lecture notes #13</td>
<td>Simple Harmonic Motion, stable equilibrium, oscillations, and restoring forces</td>
</tr>
<tr>
<td>Wednesday, Oct. 28</td>
<td>14</td>
<td>Resonance: Damped and Driven Oscillations</td>
<td>Lecture notes #14</td>
<td>Energy in SHM, damping/driving forces, and resonance</td>
</tr>
<tr>
<td>Monday, Nov. 2</td>
<td>15</td>
<td>Energy and the Atomic Picture of Matter: Temperature and Ideal Gases</td>
<td>Lecture notes #15</td>
<td>Temperature, kinetic energy, and ideal gases</td>
</tr>
<tr>
<td>Wednesday, Nov. 4</td>
<td>16</td>
<td>Random Processes and Inevitable Conclusions</td>
<td>Lecture notes #16</td>
<td>Fundamental ideals of statistical physics</td>
</tr>
<tr>
<td>Thursday, Nov. 5</td>
<td></td>
<td>Exam #2 Today</td>
<td>8-10 pm</td>
<td>Location TBA</td>
</tr>
<tr>
<td>Week</td>
<td>Date</td>
<td>Topic</td>
<td>Notes</td>
<td>Additional Info</td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>--------------------------------------------</td>
<td>-------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>17</td>
<td>Monday, Nov. 9</td>
<td>Microscopic Transport in Life</td>
<td>Lecture notes #17</td>
<td>Diffusion and osmosis</td>
</tr>
<tr>
<td>18</td>
<td>Wednesday, Nov. 11</td>
<td>Thermal Properties of Liquids and Solids</td>
<td>Lecture notes #18</td>
<td>Heat capacity, thermal expansion</td>
</tr>
<tr>
<td>19</td>
<td>Monday, Nov. 16</td>
<td>Heat Transfer: Staying Warm</td>
<td>Lecture notes #19</td>
<td>Conduction, convection, and radiation</td>
</tr>
<tr>
<td>20</td>
<td>Wednesday, Nov. 18</td>
<td>Life and Random Processes: Something from Nothing?</td>
<td>Lecture notes #20</td>
<td>Formation of structure, processes and cycles and the flow of energy</td>
</tr>
<tr>
<td>21</td>
<td>Monday, Nov. 23</td>
<td>Keeping it going: thermodynamic cycles and engines</td>
<td>Lecture notes #21</td>
<td>Thermal cycles, heat engines, and entropy</td>
</tr>
<tr>
<td></td>
<td>Wednesday, Nov. 25</td>
<td><strong>No Lecture: Thanksgiving</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thursday, Nov. 26</td>
<td><strong>No Discussion: Thanksgiving</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Monday, Nov. 30</td>
<td>Life in a Fluid: dealing with pressure</td>
<td>Lecture notes #22</td>
<td>Hydrostatic pressure, Pascal’s Principle, and atmospheric pressure</td>
</tr>
<tr>
<td>23</td>
<td>Wednesday, Dec. 2</td>
<td>At the Edge: Surface Tension</td>
<td>Lecture notes #23</td>
<td>Surface tension, surfactants, bubbles and drops, capillary action</td>
</tr>
<tr>
<td></td>
<td>Thursday, Dec. 3</td>
<td><strong>Exam #3 Today</strong></td>
<td>8-10 pm</td>
<td>Location TBA</td>
</tr>
<tr>
<td>24</td>
<td>Monday, Dec. 7</td>
<td>Flowing Fluids: If Only They Were Perfect</td>
<td>Lecture notes #24</td>
<td>Fluids in motion, Bernoulli’s Equation</td>
</tr>
<tr>
<td>25</td>
<td>Wednesday, Dec. 9</td>
<td>Real Flowing Fluids: Compressibility and Viscosity</td>
<td>Lecture notes #25</td>
<td>Real fluids, viscosity, and flow in a pipe</td>
</tr>
<tr>
<td>26</td>
<td>Classes End Monday, Dec. 14</td>
<td>Real Flowing Fluids: turbulence and mixing</td>
<td>Lecture notes #26</td>
<td>Reynolds Number, viscosity</td>
</tr>
<tr>
<td></td>
<td>TBA</td>
<td>Monday, Dec. 21, 4:00-6:00 pm</td>
<td>Study, study, study!</td>
<td>Location TBA</td>
</tr>
</tbody>
</table>

Copyright: Regents of the University of Michigan, 2009
**Important Dates to Remember**

*Physics 135: Physics for Life Sciences I*

**Exams:**

<table>
<thead>
<tr>
<th>Exam</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam 1</td>
<td>Thursday, Oct. 8: 8-10 pm</td>
</tr>
<tr>
<td>Exam 2</td>
<td>Thursday, Nov. 5: 8-10 pm</td>
</tr>
<tr>
<td>Exam 3</td>
<td>Thursday, Dec. 3: 8-10 pm</td>
</tr>
<tr>
<td>Final Exam</td>
<td>Monday, Dec. 21, 4:00-6:00 pm</td>
</tr>
</tbody>
</table>
Physics for the Life Sciences: Fall 2008 Lecture #1

Physics 135: Physics for the Life Sciences

This is a course for aspiring scientists; including biologists, chemists, medical doctors, bioengineers, biochemists, and physicists. The goal is to help you learn some very fundamental aspects of physics, especially those which most important for understanding life. In this course, we’ll teach you not only what the laws of physics are, but how to use them to analyze how life works. By covering many of the basics, we’ll also provide you with the physics framework you need to build a more detailed understanding of life later.

This course is the first in a sequence of two courses especially designed for life scientists. During this first term, we will focus on mechanical and thermal aspects of life, including fluids. Here are some of the many questions we will address.

- How does the inexorable pull of gravity affect the sizes and shapes of organisms? What must they do to move around?
- How does inanimate matter apply forces? How do organisms use membranes, muscle, tendon, and bone to support themselves, get up and move about?
- Why does it take so much effort to jog along at a constant speed? Why haven’t organisms evolved wheels to make this easier?
- What is energy, and how do organisms take energy from or give energy to an object? What forms can energy take?
- What is temperature and thermal energy? How does life harness purely random thermal motion to get things done?
- How do organisms manage to survive winters and live in cold oceans?
- How do living things get along within life’s two great fluids: air and water?

In the second semester course, Physics 235, we’re going to learn about several new aspects of the physics important for life:

- How do living things sense the world around them? Sound and light, imaging and detection.
- How can we extend our senses? Instrumental imaging.
- How does life send signals within an organism? Electric fields and potentials, electric currents and circuits, electricity and magnetism.
- What is life built of? The elements, nuclei, radiation, and the origins of these.

Each course will include a number of components:

- Textbook: There will be no traditional textbook for this course. Instead we will provide detailed lecture notes for each class. These are available as a coursepack and also as a PDF file on the Ctools site. No text on the market approaches
physics in quite the way we will, so it doesn’t make sense to use one. That said, every introductory physics text provides a useful overview of many of the topics we will study. If you have one, feel free to use it. You may find it helpful in studying.

- Supplementary Problems: We will ask you to purchase the Schaum’s *Outline of College Physics*. This is a very cheap, basic book which will give you another look at many of the topics we’re covering, and includes a lot of example problems which are likely to be useful for exam reviews. It should cost ~$15 purchased online, and you can also get it at the local bookstores. Here are some details:
  - Publisher: McGraw-Hill; 10 edition (November 15, 2005)
  - ISBN-10: 0071448144

- Readings: The time we spend in class will be focused on trying to understand the most difficult aspects of the material, rather than on providing a first-look introduction to each new topic. To make this work, you have to come to class prepared. This means you will have material to read and think about before every class. This will typically be 5-10 pages from lecture notes, occasionally accompanied by an additional reading.

- Daily lecture preparation homework: To help you to prepare for class, you will have to answer a few simple questions and solve a few straightforward problems for each lecture meeting of the class. These are to be written out and turned in at front of the class *before* the start of lecture. Working through these will help make sure you’re ready for what we’ll do during the lecture period. These assignments will be available on the course Ctools site. Print each assignment, do the work in the space provided, and turn it in at the start of class.

- Lecture/Demonstration: All students in this course will meet together twice a week (on Mondays and Wednesdays) in a lecture/demonstration setting in the Dennison building. During these sessions we will go over some details of the material, view and analyze demonstrations of the phenomena in question, and work through questions designed to challenge your understanding of the material. During lecture we will “Qwizdom” electronic response units to test your understanding of lectures in real time. You will need to purchase a Qwizdom unit from the Computer Showcase for this purpose. The details are available at: [http://showcase.itcs.umich.edu/pages/remotes/](http://showcase.itcs.umich.edu/pages/remotes/)

- Discussion: On Tuesdays and Thursdays you will be split out into smaller discussion sections, which will meet in the Undergraduate Science Building. During these sessions you will work in small groups solving problems, constructing models for biological problems, and occasionally working with auxiliary readings and the primary scientific literature.
Weekly online exercises: You will have online homework due once a week. These assignments will be done using the online homework system called “Mastering Physics”. You can purchase access to this system online following instructions which will be posted on the Ctools site.

Literature summaries and readings: Several times during the term we will do assignments which require you to read and respond to additional readings or to do work with the primary scientific literature. These assignments have two goals; to show you that the physics you’re learning is at the core of much modern biological research, and to give you some introduction to what the scientific literature is like.

Exams: We will have three exams and a final during the term. Like the homework, these will include a mix of quantitative problems and written explanations. Practice exams will be provided for the midterms.

Grades will include contributions from all of the above components:

- Daily homework & Classroom responses: 10%
- Weekly online homework: 10%
- Auxiliary readings and literature reviews: 5%
- Discussion group work: 10%
- Midterm exams 15% each and Final exam 20%

Grading: grades will be assigned on a fixed scale. There’s no curve, so it’s perfectly possible for everyone to get A’s. As a rule, students who put in all the effort we expect rarely fail to get A’s or B’s. Here’s the scale we’ll use:

87-100%: A
77-87%:   B
62-77%:   C
45-62%:   D

If the median score in the course ends up significantly below 77%, we will lower the grade scale to ensure that half of the students receive A and B grades.

Introduction to the course

Science is a great human endeavor. For centuries, millions of people have worked together to expand our collective understanding of what is in the world and how those things interact. To explain, scientists seek consistent patterns in the way things work; laws of nature. There is an essential assumption made in all this; that everything we see comes about as a consequence of these laws of nature playing out in every way possible. The laws govern what can happen, and everything that can happen, does. If something that could happen didn’t, we should recognize the presence of some other, previously unseen law preventing it.
These days, nearly 400 years after Galileo’s death, science has become a broad, complicated endeavor. People speak of “the sciences”, and include among them things like biology, chemistry, geology, medicine, astronomy, and physics. There is an exaggerated impression of segregation among these fields. It is true that the focus of study, and often the methods, vary from field to field. But much more unites the sciences than separates them. All rely on the same basic notion; that everything comes about as the consequence of a limited set of laws which govern what can happen. These are not laws of physics, or of chemistry, or of biology. They are not principles which you study in one class just for that class, or just to please your professor. They are laws of nature: unchanging, universal, and inescapable.

The particular focus in this course will be “physics”. In the past, this word had the broadest meaning; it meant simply “natural science”. The subject was everything in nature. At the time, most people believed there was more to the world than nature. This imagined super-natural world was the topic of “meta-physics”. From a very early time, astronomy was considered separate. The precision and timelessness observed in the stars and planets seemed unearthly enough to suggest an utterly different reality, and astronomy was held apart. Newton was the first to suspect that heavens and the Earth might all obey the same set of laws. The last century of research has confirmed this suspicion to an astonishing degree, and now we speak of a united “astrophysics”.

As time went on and the successes of science grew, other disciplines were defined, carving off corners of the natural world for specialization. Chemistry first emerged around 1700, focusing on the elements of matter, their states, and how they come together to form the compound substances. Still more recent is biology; a term invented in 1802 by Gottfried Reinhold Treviranus, a German naturalist. Biology is defined today as “the division of physical science which deals with organized beings or animals and plants, their morphology, physiology, origin, and distribution”\textsuperscript{1}. Notice that biology is described as a division of the physical sciences, as if to emphasize that there is no fundamental distinction between physical and life sciences.

In this course, we will treat the suggestion that biology is “a division of physical science” in a deadly serious way. The most important idea in this course is that life is the outcome of physical laws, and nothing else. There is no “spark of life” which separates the animate and inanimate. Biology is not a subject apart from chemistry and physics, but just a complex and interesting application of them.

This assertion may surprise you. It may conflict with beliefs you have long held, with what you have been taught in the past, or with your gut feeling about things. This is good. Intellectual dissonance is the surest sign that you’re in a position to really learn.

We are not going to insist, or even ask, that you believe this assertion. Instead we will take it as a starting point, a possibility worth thinking about and testing. During this class we will be learning some of the laws of nature. In each case, we will examine this

\textsuperscript{1} Oxford English Dictionary
assumption; that life is shaped by and exists within the constraints of purely physical laws. We will ask whether living things ever evade the limits these laws place, as they might if life involved something beyond the physical, something metaphysical or supernatural. We will also ask how the extraordinary diversity we see in life could come about as a result of these physical laws.

There is one extremely important tool needed to understand how the interplay between physical laws and life takes place: evolution. The idea that life evolves through natural selection of random variations provides our only tool for understanding the diversity of life. Evolution has allowed life to find incredibly various and seemingly ingenious ways to function. Evolution, working within the limits provided by physical laws, will allow us to understand why animals never evolved wheels, why cells are the size they are, why hummingbirds eat several times their weight in food each day, and why the largest animals which have ever lived all swim in the sea.

What we’ll accomplish in this term and next will only scratch the surface of this profound and important topic. But I think you’ll find even a superficial look can teach you a lot about the inescapable unity between the physical and biological worlds. At a minimum you should learn how physical laws constrain organisms, and with luck this approach will change the way you think about life.

**Three examples to set the stage**

Let’s start with a few examples to illustrate how life has evolved to work around the limits placed on it by physical laws. Let’s start with one of the most obvious connections: size and shape.

**1: The Spherical Cow and Modeling**

There is a famous joke about cows. A dairy farmer is having trouble making ends meet, and hopes to find a way to enhance the productivity of his farm. For reasons lost in the mists of time, he calls in a psychologist to help.

The psychologist conducts a series of interviews with the farmer, his family, and the cows, and collates their reactions to a variety of visual stimuli. In the end, he tells the farmer “Your cows are suffering from Ruthvenian post-lactic stress disorder. To enhance your productivity you need to provide them with a more nurturing climate. Paint the walls of the barn a cool, neutral color, provide them with quietly energetic music, and be sure to reinforce their sense of self-worth daily.” This doesn’t work. Indeed the cows become very relaxed, but this only makes them harder to milk.

Next he turns to a biologist, who sends a graduate student to take cell cultures from inside each cow’s mouth, extracts the DNA, and sequences it. Her report to the farmer suggests supporting a new research program to splice the DNA of a Minke Whale into the cow, allowing them to grow much larger, produce more and richer milk, and not wander about the farm so much. The Minke Whale, after all, has the richest milk of all mammals. The
farmer, imagining the bad press he would receive for this Frankenstein milk, politely thanks the biologist and sends her on her way.

Finally, the farmer calls on a physicist. Unlike the others, the physicist doesn’t examine or consult the cows at all. Instead, she goes to the chalkboard, draws a large circle, and says: “Assume each cow is a sphere” and begins to write long equations on the board…

The point of this joke, like so many, is to illustrate some partial truths. Psychologists do look for answers in the minds of their subjects. Biology has a strong focus on genetic research. And physicists achieve much of their success by building mathematical models. Let’s see where we can go with this “spherical cow approximation” (SCA)…

If a cow were a sphere, it would be easy to calculate things about it. Tell me it’s radius, and I can quickly calculate both its volume \((4/3\pi r^3)\) and its surface area \((4\pi r^2)\). For a real cow, with its complicated shape, this is hard to do. What’s the formula for the volume of a cow given it’s height at the shoulder \(h\)? There isn’t one. But although these spherical answers are easy to calculate, they’re also obviously wrong. I don’t expect to precisely predict the mass of a cow from the SCA. But there are things I can accurately predict.

What would happen if I changed the size of a cow? How would its volume and surface area change? Without a model, we could only do the experiment: grow a big cow and measure it. But given the SCA, we can predict what will happen. For our sphere, volume is related to radius through the equation \(V = 4/3\pi r^3\). If we double the size of \(r\), the volume increases by a factor of \(2^3\), or 8. We can also predict how the surface area \((S = 4\pi r^2)\) changes: it should increase by a factor of \(2^2\), or 4.

Why might the farmer care? Suppose he is raising beef cattle. Doubling the size of the cow would yield 8 times as much meat. What if he’s out to make leather? Doubling the size of the cow would yield 4 times as much hide. But of course there are other implications of size. The amount of food and oxygen an organism needs is largely governed by mass. Each cell must be kept alive, and more mass means more cells. So doubling the size of a cow increases its food needs by a factor of eight! A farmer out to make leather would, as a result, generally prefer a lot of smaller cows to a few big ones!

We will make use of the SCA, along with many other simple models, quite often in this class. They will allow us to extract, from all the specificity of real biological circumstances, a few important facts. Some examples follow.

2: Convergent Evolution

I have said that evolution, the random change and selection of subsequent generations of life, allows life to find strikingly effective ways to work within the limits placed by physical laws. A beautiful example of this is “convergent evolution”. Here’s the idea: if physical laws present a difficult problem for life, there may be very few workable solutions. When this is so, the same solutions will often be found, completely
independently, over and over during the course of evolution. While there are many examples of this, perhaps the most visible, well documented, and delightful, is flight.

We all know what will happen when we release an object in the air: gravity pulls it to the ground. Sometimes the friction an object feels when moving through the air slows its fall, but in the end “what goes up must come down”. To fly, to remain in the air at will, climbing and diving when you want, you need to be able to generate forces large enough to overcome the gravity which pulls you down. As we will see in a bit, if you want something to push up on you (to lift you up more strongly than gravity pulls you down) you have to push down on it. Up there in the air, the only thing to push on is the air. If you want to push hard on the air, you need something big; you need a wing. Now it would also help to minimize how hard gravity pulls you down by being as light as possible.

Life, through evolution, has found this solution at least four quite independent times, among the insects, reptiles, birds, and mammals. In each case the solutions are strikingly similar: large, thin, flexible wings are attached to bodies with many adaptations designed to reduce weight. Insects might seem the exception to this focus on slenderness. After all, a big fat June Bug flies along just fine. Why is that?

Recall the SCA. The volume of a creature increases like size$^3$, while the surface area increases like size$^2$. So if shapes remain the same, the ratio of mass to wing area (volume to surface area) changes like size$^3$/size$^2$, or like size. The bigger a flier is, the more important it will be for it to limit its weight and increase the relative size of its wings. The largest fliers, birds like condors and cranes, have quite enormous wingspans and surprisingly tiny masses. Insects live at the low end of this tradeoff, where the benefits of reducing weight are really not important. So there are lots of chubby shaped insects which can still fly. Why are insects always pretty small? It turns out their sizes are not limited by flight. We’ll see what does limit them in a bit too.

3: Diffusion

Most of the motion we associate with life seems willful: I throw a ball, a bird flaps its wing, a snail crawls across the floor. But one kind of motion, incredibly important for life, clearly doesn’t require will: it just happens. This is transport on the molecular scale: what we would generally call “diffusion”.

Imagine a rectangular box divided in two. The left hand side is filled with Nitrogen gas. In it huge numbers of N$_2$ molecules fly freely through space, colliding occasionally with one another or the walls. The right hand side is completely empty; a vacuum. Now suppose we remove the dividing wall. What will happen? Some of the molecules which would have hit the divider and bounced back will now just continue into the empty side, eventually reaching the far wall and bouncing back. After a while (a very short while indeed in this case) there will be essentially equal numbers of molecules on both sides of the box.
How did this happen? Did anyone decide to push those molecules across and make them spread them out evenly? This kind of motion, which just happens as a result of random thermal motion, is called “diffusion”. It happens because atoms and molecules are always moving around (with speeds which depend on their temperature). If you give these things which are just rattling around a chance to spread out, they will. Not because they want to or because anything is pushing them, but just by chance.

How does life use this? When it comes time to deliver not a whole mess of stuff (like a mouthful of food or a big gulp of air) but individual molecules to where they’re needed, our familiar pushes and pulls don’t work, and diffusion has to take over. Let’s take a basic example: delivering oxygen molecules into a cell. Imagine you put some air next to a cell. There will be oxygen in the air, and let’s assume no oxygen inside the cell. If the cell membrane is “permeable” to oxygen (if it allows oxygen to pass through), then when oxygen outside the cell hits the wall it will pass through. It will continue to build up inside the cell until there is, on average, just as much flowing out as flowing in.

How fast does this diffusion process happen? The rate depends on a number of things, including at least how much oxygen is on one side and the other, the temperature of the stuff on both sides, how much surface area there is to diffuse through, and how permeable the membrane is. A key factor here is surface area. Remember, the amount of oxygen something needs depends on its total volume, which changes like size$^3$. The amount of oxygen it can absorb through diffusion depends on surface area, which changes like size$^2$. So in our SCA, the amount of oxygen available for each little bit of an animal should change like surface area / volume, or like size$^{-1}$.

This sounds like a losing game. The bigger you get, the less able you are to supply oxygen to your cells. As organisms get larger, it should become harder and harder for them to deliver oxygen to their cells. How has evolution gotten around this limitation? The answer is to break the spherical cow approximation. A key part of the SCA is the assumption that when things change size, they stay the same shape. Our spherical cow is always a sphere. We call this kind of change in size “isomorphic”, meaning keeping the same shape. Evolution has overcome this problem by changing shape while changing size, and doing this in ways which make the surface area for oxygen exchange increase more rapidly than it would in the SCA.

For most “big” organisms, this is done by developing lungs. Your own lungs, for example, are a spongy mass of more than a half-billion tiny sacks called alveoli. Their total surface area is around 100 square meters. Is this big? You’re about 2 m tall and 0.3 m wide, so your surface area (in skin) is perhaps 1.2 m$^2$. By growing complex, spongy lungs, your body increases the available surface area for oxygen diffusion by a factor of 100, easily enough to let you grow big. Fish do the same trick by growing gills; huge numbers of very thin feathery sheets they expose to oxygenated water. If you wanted to deliver oxygen to your cells more directly, without the complication of squishy, disease-prone lungs, you’d have to be about 100x smaller. That would make you 2 cm tall; about the size of a cricket.
In the end, this is why insects have never become large. They mostly deliver oxygen directly to their cells through a system of narrow tracheal tubes. This system acts almost entirely through diffusion, and doesn’t allow them to really change shape while changing size. As a result, it provides a fundamental limit to the size of insects.

**Some differences among the sciences and how you study them**

Many of you will take courses in many different sciences while you’re here at Michigan. There are superficial aspects of practice in each discipline which combine to make Physics, Chemistry, and Biology seem very different, obscuring the connections among them. There is one in particular you might want to watch out for from the very beginning.

Biology and Chemistry make extensive use unique, specially invented terminology, often derived from Latin. This choice of Latin, an ancient language preserved in part to support this penchant for specific naming, makes clear the desire to avoid using “everyday” language for the naming of things. These specific names are adopted to emphasize as clearly as possible the differences among things, and to stress for example the distinction between deoxyribonucleic acid and ribonucleic acid.

Physics, by contrast, tends to focus on describing things in the simplest possible way. In so doing, it often picks up terminology from everyday life for use in its theoretical framework. Some examples of physics terms include the normal force, friction, work, energy, force, action, pressure, tension, stress and strain, the big bang, the standard model, and string theory. Using this kind of language is nice because it tends to avoid the possible obscuring effects of unfamiliar terms. It can also be treacherous, because these familiar words bring with them everyday meanings different from those used in physics. To overcome this pitfall, you have to be particularly careful to notice when these familiar words are adopted with much more particular meanings. We will try to point this out, but you have to heighten your sensitivity.

**Tools and skills for this course**

Here are some things I’m sure you know, but which it won’t hurt much to be reminded of. These are all things we’ll use extensively, and rather than remind you of them every time they come up we’re going to put them all here.

**Volume and Surface Area**

There are many simple shapes for this volume and surface area can be simply expressed:

<table>
<thead>
<tr>
<th>Shape</th>
<th>Surface Area</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cube of edge length L</td>
<td>$6L^2$</td>
<td>$L^3$</td>
</tr>
<tr>
<td>Sphere of radius R</td>
<td>$4\pi R^2$</td>
<td>$4\pi R^3/3$</td>
</tr>
<tr>
<td>Cylinder with radius R and length L</td>
<td>$2\pi R^2 + 2\pi RL$</td>
<td>$\pi R^2 L$</td>
</tr>
</tbody>
</table>
For other shapes, like a wombat or an automobile, we don’t have simple formulas for surface area and volume. There are some simple scaling rules though. If you change the size of an object, keeping the shape exactly the same, the surface area will increase like the size$^2$ and the volume will increase like the size$^3$. You can see this is true in the above trivial examples, and you can extend it to any shape you like by building that shape up out of tiny cubes.

**Density and Mass**

The average density of an object can be found by dividing its total mass by its total volume:

$$\rho_{\text{av}} = \frac{M}{V}$$

The density of an object is determined by what it’s made of. Most living things (like you) are made of a mix of things (muscle, bone, brain, etc.). In most animals the main ingredient is water, so often you’re not too far wrong if you use the density of water to estimate the density of an animal. Conveniently, water has a memorable density of about 1000 kg/m$^3$.

How could you use this to estimate your mass? There’s no formula to get your volume from your height. So let’s estimate it by imagining you’re a cylinder, say 1.8 m high and with a radius of 10 cm (about 4 inches). This would give us a volume of 0.23 m$^3$ and a mass of 56 kilograms.

**Trigonometry and the Pythagorean Theorem**

As we discuss various geometric properties such as size, shape, motion, etc., you will have to use trigonometry in many basic ways. Here are some basics which would be good to recall. Given a right triangle with sides that have length A, B, and C, we can write the following:

- $A^2 + B^2 = C^2$
- $\cos \theta = A/C$
- $\sin \theta = B/C$
- $\tan \theta = B/A$

Here’s another useful thing to remember: angles can be measured either in radians (which run from 0 to $2\pi$) or in degrees (which run from 0 to 360). You’ll need to be careful about which you’re using with your calculator.

**Some basic calculus**

Calculus is a branch of mathematics dedicated to describing change. Physics is all about change; not about how things are but about how they change. Calculus was invented, in large part by Newton, as the central tool of physics. As a result, any serious understanding of physics requires reference to calculus. It does not, fortunately, require a
very elaborate application of calculus. So while we will very often incorporate the ideas
of calculus in what we discuss, you won’t have to deploy the many methods of calculus
particularly often in this class.

You will need to understand that derivatives of functions describe their slopes, their rates
of change, and that integrals of functions describe areas under them. We will do some
simple calculus derivations occasionally, and you should be comfortable with this.

**Estimation**

This provides a nice introduction to the topic of estimation. We will often estimate things
in this course. Why not just be precise, use equations, and calculate exact answers? There
are at least two different reasons.

First, most of what happens in the world is incredibly complicated. This makes precise
description, in the form of a perfect, tidy equation, impossible. Fortunately, this
complexity doesn’t leave us helpless to describe or predict what will happen. It simply
means we will have to approximate; to construct models which capture some of the most
important features of the situation, while glossing over less significant details. Our
spherical cow is a great example of this. It doesn’t tell us the volume of a cow. But it
does give us an idea of how that volume changes as a cow grows.

There is another important reason to estimate. Even if we had a perfectly precise, tidy
analytic theory, we still may not perfectly know the parameters involved. For example, if
we want to know the mass of a cow, we may not know its precise height or length. We
may not know its density or detailed shape. What’s the density of a cow? How could you
estimate this? Well, like most land animals, cows can swim, a little at least. This means
they can nearly float. This means their density must be close to the density of water,
which is one of those nice, useful, numbers you should just know…

It is very useful, and important, to be able to make quantitative estimates in situations
where perfect knowledge is absent. This appears in all arenas of life. If, for example,
someone told you there were 5,000 piano tuners in Ann Arbor, should you believe them?
Most of you have probably heard the five second rule: “if you drop a piece of food on the
ground and pick it up in less than five seconds, it’s OK to eat it”. Is this nonsense or true?
Why five seconds and not 2.5 or 10? What happens in five?
Before we embark on any really serious analysis we need to make sure a few basic tools are in place, things we’ll need to use.

Units of Time, Space, and Mass, and Questions of Scale

Quantitative description of anything implies measurement: comparison to some frame of reference. All measurements in this class will involve comparisons to just three fundamental standards, measuring time, distance, and mass:

Time: seconds (defined as time required for a Cesium atom to vibrate 9,192,631,170 times)

Distance: meters (defined as distance traveled by light in 1/299,752,458 second, about 3.28 feet)

Mass: Kilograms (defined as mass of a little cylinder kept in Paris, about 2.2 pounds in more familiar units)

Now each thing you might measure, like distance, or time, might be measured in a variety of different units. Time, for example, might be measured in seconds, or hours, or days. Now some particular period of time, 45 seconds say, actually has only one duration. We might measure it in many different units, but it’s always really the same thing. To convert this one period from one set of units to another we can take advantage of conversion factors:

\[
45 \text{ seconds} \times \left( \frac{1 \text{ minute}}{60 \text{ seconds}} \right) = \left( \frac{45}{60} \right) \text{ minutes} = 0.75 \text{ minutes}
\]

\[
45 \text{ seconds} \times \left( \frac{1 \text{ day}}{86400 \text{ seconds}} \right) = \left( \frac{45}{86400} \right) \text{ days} = 5.2 \times 10^{-4} \text{ days}
\]

Notice what we do in each case. Start with what you are given (45 seconds), then multiply by a “conversion factor”; a ratio of two periods that are equal to one another, but measured in different units. Since the two are equal, the ratio is actually equal to one, and when you multiply by it, you leave the original period unchanged. What the conversion factor does, then, is to change the units without changing the value of the measured quantity. Here are a few more examples:

\[
1.8 \text{ meters} \times \left( \frac{100 \text{ centimeters}}{1 \text{ meter}} \right) = 180 \text{ centimeters}
\]

\[
56 \text{ kilogram} \times \left( \frac{1 \text{ pound}}{0.454 \text{ kilogram}} \right) = 123.4 \text{ pounds}
\]

Since we will work with a variety of different units, you will need to develop some facility with doing these conversions. Sometimes they will be more complicated. Let’s convert speed in meters per second to miles per hour:
1 meter/second * (1 mile / 1609 meter) * (3600 second / 1 hour) = 2.24 mph

So that’s what “units” are. What about “dimensions”?

When we ask what dimensions something has, we’re asking about what it is we’re measuring, not how we’re measuring it. Is it a distance, a time, a mass? Notice that this is different from asking about units. When we talk about units, we’re asking something more specific. How are we measuring this? What is it we’re comparing this thing to? When we talk of dimensions, we’re only asking about the nature of the thing we’re measuring, not the particular system of comparison we use to measure it.

As it happens, there aren’t so many fundamentally different dimensions of measurement in physics. For much of this course, we’ll need just three:

- Length: L
- Time: T
- Mass: M

Other quantities of interest are measured in terms of these. For example, speed has dimensions of length divided by time (L/T), and density has dimensions of mass divided by volume (M/L^3). Paying attention to the dimensions of things is important. The dimensions of a thing tell you what it really is. While the units can be changed by conversion factors, the dimensions cannot. There’s no way to turn a length into a mass.

**Orders of magnitude and scientific notation**

In discussing the physics of living things we will need to talk about things both very small and very large. In the spatial dimension we will talk about things ranging from atoms (0.000,000,000,1 m) to the Earth (6,400,000,000 m). In time we will consider the time it takes two atoms to bond (0.000,000,000,001 s) to the age of the universe (4,400,000,000,000,000 s). In mass, we will ponder both electrons (0.000,000,000,000,000,000,000,000,000,001 kg) to, again, the Earth (6,000,000,000,000,000,000,000 kg).

Dealing with a large range of scales is mentally challenging, and science has invented a tool, scientific notation, as a helpful crutch. We will often use scientific notation and the associated Greek prefixes. The primary ones we will use are:

- Giga- \(10^9\) a billion
- Mega- \(10^6\) a million
- Kilo- \(10^3\) a thousand
- Centi- \(10^{-2}\) 1/100
- Milli- \(10^{-3}\) 1/1000
- Micro- \(10^{-6}\) 1/1 million
- Nano- \(10^{-9}\) 1/1 billion
So when we speak of a centimeter, we mean 1/100 of a meter, and when we talk about a kilogram, we mean 1000 grams. You will have to always exercise care with these prefixes. They are the source of many student errors. You don’t, for example, want to confuse micro and mega…

As our study of the universe has advanced, science has pushed these limits, and so there are more of the prefixes. Going up, the next few are Peta- and Exa- \((10^{15} \text{ and } 10^{18})\), going down they are Femto- and Atto- \((10^{-12} \text{ and } 10^{-15})\).

You should notice that just about all of these are multiples of a thousand. The exception, Centi-, is useful because the basic distance, a meter, is a quite a bit bigger than our hands, and many things we work with are the sizes of our hands or smaller. The next scale down (the millimeter) is a little too tiny for many of the things we usually work with. So this is an accident of convenience.

Notice that with time, too, we define additional out-of-spec units which are imposed on our system by our circumstances. It’s impossible for most living things on Earth to ignore the day as a meaningful unit of time, or the year. People, who might live a hundred years, find the century \((10^2 \text{ years})\) to be more relevant than the millennium.

Of course the second was originally defined in reference to the day; as a subdivision of hours and minutes. Of course these units, though important to us, are surely important only on the Earth. They have no physically fundamental importance. What is a day or a year in seconds?

\[
\begin{align*}
1 \text{ day} &= 86,400 \text{ seconds} \\
2 \text{ year} &= 3.15 \times 10^7 \text{ seconds} (\text{or very close to } \pi \times 10^7 \text{ seconds})
\end{align*}
\]

Once we start counting in years, interesting coincidences emerge, like the fact that a typical 50 minute University of Michigan lecture is just about 1 microcentury. Some lectures seem much larger, some much shorter. You can make of that what you will.

Scientific notation is a tool which allows us to talk about an enormous range of sizes, times, and masses. There are about as many nanoseconds in a single second as there are seconds in your life; each is an almost unimaginably tiny period. And yet physicists on North Campus now routinely produce pulses of laser light which are "femto-seconds" long. A femtosecond is a millionth of a nanosecond. You will need to learn to manipulate these things, so if you are rusty with exponents, I suggest you review them.

Here are the basics:
\[
\begin{align*}
10^a \times 10^b &= 10^{a+b} \quad \text{so } (5 \times 10^8) \times (4 \times 10^6) &= 20 \times 10^{17} = 2 \times 10^{18} \\
10^a / 10^b &= 10^{a-b} \quad \text{so } (5 \times 10^8) / (4 \times 10^5) &= 5/4 \times 10^1 = 1.25 \times 10^1 = 0.125
\end{align*}
\]

The nature of things we might measure: scalars and vectors

When we set about quantifying the world, measuring things about it, we discover that not all things can be described in quite the same way. Many things we might like to measure
seem rather simple; they can be represented by a single number. A baseball has a mass. That mass is just a number (5 1/4 “ounces avoirdupois”, or about 149 gm, according to the rulebook). Everything there is to know about the baseball’s mass is represented in that number. It also has a circumference (officially “not less than nine nor more than 9 1/4 inches”, or about 23 cm). That number also tells you everything there is to know about its circumference.

Physical properties which can be represented by just a single number are known as **scalars**. They are quite common. In addition to mass and diameter, they might include temperature, density, pressure, metabolic rate, pH, age, or even cost. Scalars are properties which can be fully described by just one number. We sometimes say that scalars are properties of things which have only a **magnitude**.

Some things we want to measure, especially in physics, are more complex. If we want to describe the wind for a sailor, it’s not enough to simply list its speed. To usefully describe the wind, we need also to give its direction. Another example of a quantity like this is a force. To fully describe a force, to tell you everything you need to know about it, we have to give both its magnitude and its direction. There are other examples in physics, including displacement, velocity, acceleration, electric, magnetic, and gravitational fields. These properties are called **vector quantities**. These vectors are things which require us to specify both a **magnitude** and a **direction** to give a complete description.

The point of this discussion is to draw your attention to this important difference. If you’re analyzing something and the answer you seek is a scalar, you only have to figure out and report how big it is. But if the answer to your question is a vector, you will have to determine and report both its magnitude and its direction.

**Displacement as an example vector:**

Scalars are pretty familiar things, so they don’t need much further introduction. Vectors are considerably less so, as they were invented for and are largely used in physics. So we will take some time to talk about what vectors are and how we do things like add, subtract, and multiply them.

Let’s take as our example a displacement; a kind of instruction for a trip. To describe this trip, we have to say how far to travel, and also what direction you should go. One way to do this is to specify the magnitude of the vector, and describe its direction by measuring an angle relative to some reference direction. Here’s an example; you receive an instruction telling you to travel 30m in the direction North-East.
We could represent that trip graphically as a little arrow that looks like this.

Think a little about what this graph represents. Each of the two axes represent positions. The horizontal axis measures how far East or West you are from the (arbitrarily selected) origin, while the vertical axis measures how far North or South you are from the origin. The solid arrow shows the displacement “30m in the NE direction”. The dashed arrow also shows a displacement “30m in the NE direction”.

Since these two displacement vectors have exactly the same magnitude and direction, they are precisely equal to one another. This is an important point. Vectors are not tied to particular points in a space. They don’t say go from this particular spot to that; they just tell you how far to go and in what direction.

The reason for this is actually rather deep. The way things move can’t be affected by how we choose to draw our coordinate system. If they were we would never know what was going to happen until we defined a coordinate system. We’re talking about these displacement vectors because they’ll be useful in describing physics, so they’ll have to be independent of particular starting and ending points.

Q: If I tell you a vector displacement like this is 10m North, what would it look like on a NSEW plot?
Q: What if I tell you the displacement is -10m North? What does this mean?
Q: Is the displacement "10m North" tied to particular points?

**Vectors:**

Displacement is the archetype of a "vector". We denote this in typed text by making the symbol boldface, so while s might be a scalar a distance, the boldface \( \mathbf{s} \) is a displacement vector. On a chalkboard or in your written homework we would usually note this by putting an arrow over the symbol, \( \vec{s} \), which is actually easier to remember.

There are several ways we might want to manipulate vectors. The first is addition. Why we might to add vectors? If we take two trips, we undergo two displacements, and the sum of the two is the same as taking a single equivalent trip. Likewise, the sum of two forces is the same as a single equivalent force.
Rather like scalars, the sum of two vectors $\mathbf{a}$ and $\mathbf{b}$ is equal to a single third vector, $\mathbf{c}$, which is equivalent of doing $\mathbf{a}$ then doing $\mathbf{b}$. So we can plausibly write:

$$\mathbf{a} + \mathbf{b} = \mathbf{c}$$

The order doesn't matter, which if you remember your fifth grade math is called commutativity:

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

We see drawn here the first way to discover the sum of two vectors. This is called graphical addition, or the tip to tail method. It relies on the fact that vectors are NOT tied to particular points; they are only have magnitudes and directions. Because of this, you’re free to move them around and line up the tip of the first with the tail of the second.

Note that we could go a step further, and consider what happens if we add three vectors together:

$$\mathbf{a} + \mathbf{b} + \mathbf{c} = \mathbf{d}$$

OR:

$$\mathbf{d}$$

Notice from this example that the same resultant vector $\mathbf{d}$ is produced whether I take:

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} \quad \text{or} \quad \mathbf{a} + (\mathbf{b} + \mathbf{c})$$

This different kind of independence from order is called associativity.
**Multiplication of a vector by a scalar:**

A vector is a quantity specified by both a magnitude and a direction. A "scalar" is something specified by just a magnitude. So a distance "3m" is a scalar, and a displacement like "3m Northwest" is a vector. It is possible to multiply a vector by a scalar. To do this we just multiply the magnitude of the vector times the scalar number, leaving the direction unchanged.

![Diagram of vector multiplication](image)

**Vector Subtraction:**

Imagine I have executed a displacement. If I want to execute a second displacement which will eliminate the effect of the first, what new displacement must I execute?

Just as

\[ 5 + (-5) = 0 \]
\[ \mathbf{a} + (-\mathbf{a}) = \mathbf{0} \]

where \((-\mathbf{a})\) refers to a vector with the same magnitude as \(\mathbf{a}\), but the opposite direction, and \(\mathbf{0}\) refers to a “null vector” of magnitude zero.

This suggests how we should do vector subtraction:

\[ \mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b}) \]

which can be drawn graphically as:

![Diagram of vector subtraction](image)

This emphasizes the fact that two vectors are equal if their magnitudes are equal, and their directions equal, or if: \( \mathbf{a} + (-\mathbf{b}) = \mathbf{0} \)

**Addition of Perpendicular vectors:**

A particularly illustrative example is the addition of two vectors which are perpendicular to one another:

Magnitude of \(\mathbf{c}\):

\[ c^2 = a^2 + b^2 \] (Pythagorean Theorem)

\[ 5^2 = 3^2 + 4^2 = c^2 \]

Direction of \(\mathbf{c}\):

\[ \tan \theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{4}{3} = 1.333 \]

\[ \theta = \tan^{-1} 1.333 = 53.1^\circ \]
Components of vectors, orthogonal motions:

This last example suggests what turns out to be an extremely useful way of thinking about vectors. A vector like \( \mathbf{c} \) can always be thought of as "made up of" the sum of \( \mathbf{a} + \mathbf{b} \). It is in every way equal to this sum. The two things, \( \mathbf{c} \), or \( \mathbf{a} + \mathbf{b} \) are exactly the same.

Breaking a vector into two perpendicular parts like this is called "resolving it into components". We do this because it makes many calculations with the vectors much simpler.

Very often, a notational simplification is also made. If I set up a simple x-y coordinate system, I can define a "unit vector" for each direction. Each of these unit vectors points directly along the axis it corresponds to and has a length of one in the units of choice; hence the name unit vector. Usually unit vectors are written using the name of the axis, either as a bold vector or with a little “hat” symbol:

\[
\text{x unit vector} = \hat{x} \quad \text{y unit vector} = \hat{y}
\]

To talk about the unit vector in the x direction we would say “x-hat”. I can use these unit vectors to rewrite the vectors above:

\[
\begin{align*}
\mathbf{a} &= 3\hat{x} \\
\mathbf{b} &= 4\hat{y} \\
\mathbf{c} &= \mathbf{a} + \mathbf{b} = 3\hat{x} + 4\hat{y}
\end{align*}
\]

Initially, breaking vectors into components doesn't seem like much of a help; why would we want to replace one vector with two? Wouldn't that just complicate things? There are several reasons why this can make life easier.

1. The first is mostly practical, it is often easier to work with vectors which are broken into components than it is with the original vectors.
2. The second is somewhat deeper. It is often the case that the motion of an object along one direction is completely independent of the motion along another; the physics of the problem can "decouple" these two "orthogonal" motions. In this case, it is often advantageous to break vectors into components because this emphasizes the important features of the motion, hence making it easier to understand.

Adding vectors by components:

The first advantage of component notation is that it can often simplify vector manipulations. Consider the following example of the sum of three vectors. We’ll use the unit vector notation \( \mathbf{E} \) for East and \( \mathbf{N} \) for North.
Add three vectors:
5m East  = 5E
8m North  = 8N
6m 30° East of North = (6*\sin30)E + (6*\cos30)N = 3E + 5.2N
so the sum is:
(5 + 3)E + (8 + 5.2)N = 8E + 13.2N
We can work out the magnitude and direction of this final vector in the way we
did for adding perpendicular vectors above:
Magnitude \( m^2 = 8^2 + 13.2^2 \) or \( m = 15.4m \)
Direction \( \tan\theta = \frac{\text{opp}}{\text{adj}} = \frac{13.2}{8} = 1.65 \)
Or \( \theta = \arctan(1.65) = 1.02 \text{ radians} = 58.8° \)

Notice that this would have been just as easy if there were 30 vectors instead of three. So
whenever you have to add vectors, it is usually easiest to do it by components.

**Picking the right coordinate system:**

Now often you can *greatly* simplify a problem by using some feature of the arrangement
of elements in a problem to simplify its solution. When we talk about "picking the right
coordinate system" for a problem, this is usually what we mean. A couple of examples
will give the general idea:

First a simple one: Add two displacement vectors 4m NE and 3m SW. We could break
this into N and E components, buts it easier to add them along the direction NE/SW.
Then we immediately find:
Sum = 1m NE

Now a second, slightly more complicated example:

We could resolve this into components along horizontal and vertical x and y axes, but
that would be hard. Easier to think about a coordinate system rotated 30°
clockwise:

Writing these vectors in components along these x'-y' axes is simple:
Physics 135 Fall 2009: Lecture Notes

\[a = 4x'\]
\[b = -1x'\]
\[c = 4y'\]

So the sum of the three in this coordinate system is just:
\[r = 3x' + 4y'\]

Which is a vector with magnitude 5, and direction \( \theta = \tan^{-1}(\text{opp/adj}) = \tan^{-1} 1.33 = 53.1^\circ \).

Does it matter physically that we’ve defined this vector in an unusual coordinate system?
Not at all. It’s always the same vector no matter how we choose to measure it.

**Components and vector equality**

We say that two vectors are equal when both their magnitude *and* their directions are the same. It’s also true that if any two vectors are equal, each of their individual components (along the x, y, and z, axes for example) must be equal. So if we have two vectors \( \mathbf{A} \) and \( \mathbf{B} \) and they’re equal, we can write:

\[\mathbf{A} = \mathbf{B} \quad \text{or} \quad A_x = B_x \quad \text{and} \quad A_y = B_y \quad \text{and} \quad A_z = B_z\]

This alternate way of writing things will often be simpler to keep track of than the more general definition of vector equality. So a lot of times when we know two vectors are equal we’ll go ahead and write out three independent equations, one for each component. Since each equation is just a scalar equation, it’s simpler to work with.

**Velocity vectors:**

So now we have displacement vectors and we have some ideas about how to manipulate them. Apparently velocity also must be described with vectors, because we usually need to know both how fast we're going and in what direction. We can define an average velocity vector in a straightforward way from the displacement vector:

\[\mathbf{v}_{av} = (\Delta s/\Delta t)\]

if we determine this velocity over an infinitely short period of time, we speak of the instantaneous velocity:

\[\mathbf{v}_{inst} = \lim_{\Delta t \to 0} (\Delta s/\Delta t) = ds/dt\]

Notice carefully what this is. The velocity vector is really just a scaled version of the displacement vector. In other words it is just the displacement vector multiplied by a scalar number; the inverse of the time it took to make this displacement (1/\(\Delta t\)). What this means is that the velocity vector always points in the same direction as the displacement vector.

Because motion takes place in three spatial dimensions, many things we will use to discuss motion this semester will be vectors; including forces, accelerations, stresses, flow rates, etc. It is important that you understand vectors very clearly, and that's why we're expending so much effort on them now.
Vector Multiplication:

We have seen above how to multiply vectors with scalars. How do we multiply vectors with vectors? There’s no “obvious” way we should define vector multiplication, but it turns out there are two physically *useful* ways to do it. The first produces as the product of two vectors a scalar, and the second produces as the product of two vectors a new vector.

1. The scalar product: $\mathbf{a} \cdot \mathbf{b}$

   The scalar product of two vectors produces a scalar, just a number. It is defined so that the number produced expresses the degree to which the two input vectors are aligned with one another. The formal definition is:
   
   $$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \phi$$

   where $\phi$ is the angle between the two vectors.

   There are two ways of looking at this. It is either:

   The component of $\mathbf{a}$ along $\mathbf{b}$, times the magnitude of $\mathbf{b}$
   
   $$|\mathbf{a}| \cos \phi |\mathbf{b}| = (\mathbf{a} \cdot |\mathbf{b}|)$$

   or

   The component of $\mathbf{b}$ along $\mathbf{a}$, times the magnitude of $\mathbf{a}$
   
   $$|\mathbf{b}| \cos \phi |\mathbf{a}| = (\mathbf{b} \cdot |\mathbf{a}|)$$

   In either case, the scalar product is a kind of "colinear product" or a product of the colinear parts of a pair of vectors. Because the little ‘dot’ symbol is used to denote this operation, it is sometimes called ‘the dot product’.

   What kinds of questions will we use the scalar product for? Eventually we will want to keep track of how much objects move up and down. Imagine a bird flies through a displacement that we can write as a vector $\mathbf{d}_{\text{bird}}$. If we want to know how much higher the bird is at the end compared to the beginning we could use the scalar product. If we ‘take the dot product’ of this displacement with a unit vector which points straight up. Let’s call this unit vector $\mathbf{y}$. If we do this, we can write:

   $$\text{Distance the bird rises} = \Delta y = \mathbf{d}_{\text{bird}} \cdot \mathbf{y}$$

   Likewise, this gives a shorthand for finding the components of a vector $\mathbf{v}$:

   $$v_x = \mathbf{v} \cdot \mathbf{x}$$

   $$v_y = \mathbf{v} \cdot \mathbf{y}$$

   $$v_z = \mathbf{v} \cdot \mathbf{z}$$

   where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are unit vectors in the $x$, $y$, and $z$ direction.
2. The vector product: \( \mathbf{a} \times \mathbf{b} \)

The vector product takes two vectors and makes a third, new vector out of them.

\[
\mathbf{a} \times \mathbf{b} = \mathbf{c}
\]

where the magnitude of \( \mathbf{c} \) is given by:

\[
|\mathbf{c}| = |\mathbf{a}| * |\mathbf{b}| * \sin \phi
\]

and the direction of \( \mathbf{c} \) is perpendicular to the plane defined by \( \mathbf{a} \) and \( \mathbf{b} \) in a direction given by the right hand rule. The right hand rule says you should:

- Take your right hand
- Point your fingers in the direction of the first vector (\( \mathbf{a} \) in this case)
- Turn your hand until you can "curl" your fingers in the direction of the second vector (\( \mathbf{b} \))
- Now your thumb defines the direction of the vector \( \mathbf{c} \).

From this definition you can see that the vector \( \mathbf{c} \) is always perpendicular to both \( \mathbf{a} \) and \( \mathbf{b} \). The vector product is a kind of measure of the amount of perpendicularity of two vectors. Because the symbol ‘x’ is used to denote this operation, it is often called ‘the cross product’.

Note that this vector product has the special property that it does not commute. That is:

\[
\mathbf{a} \times \mathbf{b} \neq \mathbf{b} \times \mathbf{a}
\]

in fact it "anticommutes"

\[
(\mathbf{a} \times \mathbf{b}) = - (\mathbf{b} \times \mathbf{a})
\]

Where will we use the vector product in physics? One good example has to do with rotation. If you want to get something to start rotating, you have to apply a force to it. The ability of the force you apply to make the object rotate depends on both where you apply the force and in what direction you push. First we define a radius vector \( \mathbf{r} \) which goes from the center of rotation (the hinge of a door for example) to the point where the force is applied. Given this vector \( \mathbf{r} \) and the force vector \( \mathbf{F} \), we will quantify this ‘ability to create rotation’ by defining the torque \( \mathbf{\tau} \) with the vector product:

\[
\mathbf{\tau} = \mathbf{r} \times \mathbf{F}
\]

Don’t worry if this is confusing now. It’s just an example which you ought to recognize when we return to it later.

**Vector multiplication by components**

Remember that there are two kinds of vector multiplication: the **scalar product** and the **vector product**. In both cases, there’s a basic definition in terms of the magnitudes of the vectors and the angle \( \theta \) between them:
Scalar product: \[ \mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos \theta \]

Vector product: \[ \mathbf{A} \times \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \sin \theta \]
(with direction from the right hand rule)

Multiplying by components

It’s often the case that you’ll have two vectors written in terms of components. For example, you might have:

\[
\begin{align*}
\mathbf{A} &= A_x \mathbf{x} + A_y \mathbf{y} + A_z \mathbf{z} \\
\mathbf{B} &= B_x \mathbf{x} + B_y \mathbf{y} + B_z \mathbf{z}
\end{align*}
\]

Where \( \mathbf{x} \), \( \mathbf{y} \), and \( \mathbf{z} \) are unit vectors in the x, y, and z directions. If you have this, you can multiply the vectors in either the scalar or the vector product in a convenient way. This approach takes advantage of the fact that these operations are distributive, and that the scalar and vector products of the unit vectors are simple.

In particular:

\[
\begin{align*}
\mathbf{x} \cdot \mathbf{x} &= 1 \\
\mathbf{y} \cdot \mathbf{y} &= 1 \\
\mathbf{z} \cdot \mathbf{z} &= 1 \\
\mathbf{x} \cdot \mathbf{y} &= 0 \\
\mathbf{y} \cdot \mathbf{z} &= 0 \\
\mathbf{z} \cdot \mathbf{x} &= 0
\end{align*}
\]

Think about why this is the case. Remember the scalar product measures colinearity. Two identical unit vectors are perfectly collinear. Two perpendicular unit vectors are not collinear at all.

What about the vector products of unit vectors? Here the opposite is true. The vector product measures something about how perpendicular vectors are. The vector product of two identical unit vectors is zero; they aren’t perpendicular at all. The vector product of two perpendicular unit vectors has magnitude of one, but now it’s a new vector. In fact it’s a unit vector in the third direction! In particular:

\[
\begin{align*}
\mathbf{x} \times \mathbf{x} &= 0 \\
\mathbf{x} \times \mathbf{y} &= \mathbf{z} \\
\mathbf{x} \times \mathbf{z} &= -\mathbf{y} \\
\mathbf{y} \times \mathbf{x} &= -\mathbf{z} \\
\mathbf{y} \times \mathbf{y} &= 0 \\
\mathbf{y} \times \mathbf{z} &= \mathbf{x} \\
\mathbf{z} \times \mathbf{x} &= \mathbf{y} \\
\mathbf{z} \times \mathbf{y} &= -\mathbf{x} \\
\mathbf{z} \times \mathbf{z} &= 0
\end{align*}
\]

You should draw a little coordinate system like this and use the right hand rule to work check all the elements of this little table.
Using this for the scalar product

Let’s look at the scalar product first:

\[ \mathbf{A} \cdot \mathbf{B} = (A_x x + A_y y + A_z z) \cdot (B_x x + B_y y + B_z z) \]

Multiplying this all out you get:

\[ = A_x B_x + A_y B_y + A_z B_z \]

Wow! That looks like a real mess. But remember what you learned about scalar products of unit vectors above and you’ll see this reduces to:

\[ \mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y + A_z B_z \]

which is pretty simple. So here’s an example:

\[ \mathbf{A} = (3x - 8y + 5z) \quad \mathbf{B} = (2x + 1y + 3z) \]

What’s the dot product? This would be hard graphically, as they’re both in 3D. In component form it’s simple:

\[ \mathbf{A} \cdot \mathbf{B} = 3*2 + (-8)*1 + 5*3 = 13 \]

That’s it.

Using this for the vector product

This approach works the same way for the vector product. Let’s do an example for this:

\[ \mathbf{A} = (3x - 8y) \quad \mathbf{B} = (1y + 3z) \]

What’s the vector product of these two? Just expand it out and use the table above.

\[ \mathbf{A} \times \mathbf{B} = (3x - 8y) \times (1y + 3z) = 3(x \times y) + 9(x \times z) - 8(y \times y) - 24(y \times z) \]

\[ = 3z - 9y - 24x \]

Decoupled motions and vector components: an application

OK, so looking at vectors by their components is a useful convenience, a nice simplification of some problems. It is also useful because the motion of objects along different directions can often be independent, so that if we break it into components we can consider each motion independent of the others. One particularly nice example of this is the idea of relative velocity.
Relative velocity:

The "relative" we're talking about here is the constant velocity motion of some object observed by two different observers, who are themselves moving relative to one another. Start with a simple example:

Joe is on a train moving past the platform at 4m/s. He walks forward in the train with a speed of 1m/s relative to the train. What is Joe's speed relative to the platform? In vector form this problem is:

\[ V_{tp} = \text{speed of train relative to platform} = 4 \text{ m/s} \]
\[ V_{jt} = \text{speed of Joe relative to the train} = 1 \text{ m/s} \]
\[ V_{jp} = V_{jt} + V_{tp} = 4 \text{ m/s} + 1 \text{ m/s} = 5 \text{ m/s} \]

So it's fairly obvious how this works for colinear motion. What about in two dimensions?

A boat can travel at 3 m/s through the water. It steers straight across a river which flows past the shore at 5 m/s. What is the velocity of the boat relative to the shore?

\[ v^2 = 3^2 + 5^2 \quad \text{or} \quad v = 5.8 \text{ m/s} \]
And its direction is \( \theta = \tan^{-1}(5/3) = 59^\circ \) West of North
Physics for the Life Sciences: Fall 2008 Lecture #3

Newton’s Laws and Statics

Life on Earth faces many challenges. One of the most basic is dealing with the constant pull of the Earth’s gravity. Every living thing near the surface of the Earth (and every nonliving thing too) is constantly pulled downward. This downward pull is so steady and omnipresent we usually forget it’s there. But just one misstep on the staircase, one moment’s loss of balance on your bike, one slip of the cup off the edge of the table, and you’re reminded of the power of gravity with shocking suddenness. It’s not a stretch to claim that gravity is America’s number one killer, and it is certainly life’s number one mechanical challenge.

In addition to standing up to gravity, many living things have to move around, pick things up, and do things with them (carry them, chew them, throw them, etc.). To accomplish any of this, living things have to not only resist forces but apply them. So our first big task will be to understand forces and how they affect motion. This topic in physics is called “mechanics”, one of those otherwise everyday words which means something quite special in physics.

The core of our understanding of mechanics is contained in three terse laws first gathered together by Newton in the 17th century. Newton's laws provide the rules we need to understand how objects react to forces, and to describe motions as various as the orbits of the planets and the swimming of a bacterium. Their ability to analyze almost every mechanical situation observed at the human scale makes them a remarkable part of the collective human intellectual legacy. They’re also incredibly useful for understanding what’s going on around you, and hopefully by the time you finish studying them you will see the world in a new and richer way.

We will begin trying to understand in detail how to analyze cases in which objects aren’t moving, a subset of mechanics sometimes called “statics”. On Earth objects sitting still always experience a number of forces, but these are balanced, so that the total force on them is zero. Once we have that in hand, we’ll look at cases where the forces are not balanced, and learn to understand how these unbalanced forces make motions change.

Newton’s first law

Newton’s first law is this somewhat surprising assertion:

Any body continues in a state of rest, or of uniform motion in a straight line, unless it is compelled to change its motion by unbalanced forces imposed on it.
This is usually called the law of "inertia", or Newton's first law. It's not really Newton's, as even he would have freely acknowledged; everyone working seriously on motion at the time knew about it. Still, it is hardly obvious. The first bit is no surprise. Objects at rest stay at rest unless you do something to move them about. But the second part isn’t familiar at all.

Daily experience does not suggest that an object in motion tends to stay in motion. What do you have to do to keep an object in motion? You have to push all the time. Aristotle, noting this ubiquitous experience, assumed that the “natural state” of an object was to be at rest, and that to have an object in motion required a motive force. He believed that "motion implies a mover". But even he allowed a tantalizing exception; objects in free-fall seemed to fall only because it was in their nature to do so. No “mover” was required to create free-fall motion.

Aristotle gets a bad rap in physics, and we tend to dismiss Aristotelian beliefs too quickly. They in fact well describe a lot of ordinary experience. Seeing behind these "obvious" facts requires great care. It took a world of smart people thousands of years to see what you’re learning now.

It was Galileo Galilei, one of the delightful Italians of the 1600s, who first clearly exploded the Aristotelian idea of motion. His argument, which elegantly encapsulates the idealization which has proven so powerful in physics, went like this:

1. Imagine a wedge shaped track. Roll a ball down one side and it rises up the other to almost the same height
2. Carefully clean and polish it all and the ball rolls still more closely to the starting height, so we might ascribe any "loss of height" to friction between the ball and the track.
3. Now decrease the angle of the second side. The ball still rises to the same height from which it was launched, but now travels much farther along the ramp.
4. Carry this to its logical conclusion: if we lower the second side to be horizontal, the ball will travel forever, always attempting to rise again to its original height.

Notice the details here: the ball will roll forever with no help from anything, its “natural state” is to be in motion, and friction is the only thing which prevents that motion. Without friction everything that is moving would continue to move forever.
The reason it was so difficult for people in the 1800 years between Aristotle and Galileo to understand this is that friction is nearly ubiquitous. We will talk about friction quite a bit over the next few weeks, and it is interesting and extremely important practically. But today we are interested in seeing the world without it. Galileo couldn’t quite do this. He couldn’t actually make a world without friction. But he could imagine it, and it was this imagined idealization that allowed him to recognize the law of inertia.

Where might you have seen motion without friction? The wonderful feeling you get gliding along while ice skating or coasting on your bicycle comes because, unlike most of your motion, this is ‘effortless’. Nothing is required to keep you going. When you are almost free from the shackles of friction, you get to experience the law of inertia. A thrown ball or a fired arrow also moves along without obvious influence from friction. Aristotle had some trouble trying to explain this. What ‘mover’ was causing the motion of the arrow?

There aren’t many cases of clear frictionless motion in nature, but there are a few. Perhaps the most impressive is the motion of the planets, which have continued to circle the sun for billions of years without slowing or stopping to refuel.

This law of inertia is crucial for understanding motion. Take for example what happens if you smash your car into a brick wall. The front of your car strikes the wall. The force impressed on the car by the wall causes it to stop moving forward. But that force is not directly impressed on you. Unless something happens to you, you’ll just continue to move forward, in a continuous straight line, until something causes you to stop. Hopefully that will be your seatbelt or at least your airbag, but if not, it will be the dashboard, the windshield, or the wall. You will not stop (your motion will remain unchanged) until a force is applied to you.

The basic point is that “rest” is not really the basic (natural) state of things; uniform motion in a straight line is the natural state of things. Objects at rest are just a special case of this. The most important thing the law of inertia provides is a way to tell when unbalanced forces act. If you seen an object moving in a constant rate in a straight line, you know there is no unbalanced force acting. If, on the other hand, you see the motion of an object change, then you know for sure that an unbalanced force is acting on it. The first law tells you how to know that an unbalanced force has been applied.

Q: Is motion in a circle at a constant speed uniform? What did this tell Newton about the planets?

**Newton’s second law**

Newton’s second law is essentially of a quantification of the first. The first said that the motion of an object will not change unless an unbalanced force acts on it. The second tells us just how much force is needed to create a certain change in the motion. In a basic
sense, the second law is a quantitative definition of a force. If you want to know what force acted, you need only how it altered the motion of the object on which it acted.

To understand the second law, we first have to quantify how much motion an object has. How to quantify motion? Is motion a scalar, something with only a magnitude? Or is it a vector, something with both a magnitude and a direction? When an object is moving, it often seems natural to care about which way it’s moving. So for starters we’re looking for a vector measure of motion. A bit later we will see that it’s sometimes also interesting to just ask whether things are moving at all, without regard for the direction they’re going. This would be a scalar measure of motion.

Experience suggests that there’s something more to quantifying motion than just measuring how fast things are going. There is a difference between a ping-pong ball and a minivan when each is approaching you at 20 miles per hour. So in addition to knowing how fast things are going, we’ll also need to include how much stuff is moving.

With these general ideas in hand, we will define two different measures of motion; a vector measure and a scalar measure.

- Momentum: a vector measure of motion. Momentum is often denoted with the symbol \( p \), and it is calculated by multiplying the mass of an object by its velocity vector. Written as an equation we have \( p = mv \). Notice that since the mass is a scalar and the velocity is a vector, the momentum is always in the same direction as the velocity.

- Kinetic Energy: a scalar measure of motion. Kinetic Energy is usually denoted by the symbol KE (which you will note is not boldface; it’s just a scalar). The convenient definition for this scalar measure of motion is \( KE = \frac{1}{2}mv^2 \). The “\( v \)” in this equation is just the magnitude of the velocity vector.

OK, now that we have a way of quantifying motion, we can write Newton’s second law and quantify force. The usual way to write Newton’s second law is:

\[
F_{\text{total}} = \frac{dp}{dt}
\]

Put into words, the total force acting on an object is equal to the time rate of change of the momentum. When you see the momentum of an object changing, you can find the total force which is acting by examining how the momentum is changing. Because you’re probably not used to thinking about the derivative of a vector, it might be useful to think about this a bit.

One little note about terminology. Many physics texts talk about the “net” force on an object, by which they mean the total, vector sum of all the forces which act. To me, it’s much clearer to simply say “total” force, and we’ll usually do that.

Imagine the momentum of an object changes and we want to estimate how much force was required to change the momentum. We could write:
We know the force at each instant is given by the instantaneous rate of change of the momentum, so we can estimate it during the period while the momentum is changing by dividing the total change in momentum by the total amount of time this change took.

Let’s look at this with vectors.

\[ \mathbf{F}_{\text{est}} \cong \frac{\Delta \mathbf{p}}{\Delta t} \]

It’s useful to think about it in several ways just to get it clear.

- If you see the momentum changing very suddenly, \( \frac{dp}{dt} \) will be large, and there must be a large force acting.
- If you want to create a certain change in momentum \( \Delta p \), you can rearrange the second equation to see \( \Delta p = F \Delta t \). This emphasizes that to achieve a particular change in momentum you can either apply a large force for a short time, or use a small force for a long time. In fact, this quantity \( F \Delta t \) (which is equal to \( \Delta p \)) has a special name; it is called “impulse”. So it is sometimes said that you “apply an impulse” to achieve a certain change in momentum.

For the moment, we’ll set aside the second law, because we’re going to talk about static cases where the momentum doesn’t change. We’ll return to this in a week or two.

**Units for force**

What are the units of a force? From the second law, we see that a force is a change in momentum (which has the same units as momentum) divided by a time. So using the typical units of meters, seconds, and kilograms, we have:

\[ \text{kg} \ast (\text{meters }/ \text{second})/\text{seconds} = \text{kgm/seconds}^2 \]

Because forces are so important, this particular combination of the basic units is especially significant, and is given a name of its own. Because Newton was so important in understanding the importance of forces, this unit of force is named for him.

\[ 1 \text{ kilogram meter }/ \text{second}^2 = 1 \text{ Newton} \]

Though many other units of force exist, in this course we will measure forces almost always in these units.
Newton’s third law

The most important and perhaps least obvious fact about forces is that they are never isolated actions; they happen only through interactions. Pick apart that word a little. “Inter-actions” are actions that take place between things. There is never a force that comes from nowhere and pushes on something. Every force comes from one thing and pushes (or pulls, or whatever) on another.

Newton, realizing this, was the first to recognize a fundamental fact about interactions: they are always perfectly balanced. In absolutely every case, when one object applies a force to a second, the second applies a perfectly equal and opposite force on the first:

\[ F_{12} = -F_{21} \]

That is, every interaction has exactly two forces, equal in magnitude and opposite in direction, one acting on each of the two interacting bodies.

This 'third law' of Newton is often stated in words as:

For every action there is an equal and opposite reaction.

But it is perhaps more useful to rewrite this in more modern terms as:

If object A exerts a force on object B, then object B exerts and equal and opposite force on object A

Let’s be careful about the notation here. I have written \( F_{12} \). By this notation I mean the force applied by object 1 on object 2. Likewise the notation \( F_{21} \) stands for the force applied by object 2 on object 1. The minus sign in the way we have written the 3rd law just reflects the fact that while these two vectors are equal in magnitude, they have exactly opposite directions.

How can the third law be true?

The 3rd law is simple to state, but is quite surprising. No one before Newton ever recognized it. It almost seems it can’t be true. If every time I push on you, you push back equally on me, how can anything get anywhere? Don’t those two forces always cancel out?

Consider the way you throw a ball.

The 3rd law says that \( F_{pb} = -F_{bp} \). That is, the force the ball exerts on the person is equal and opposite to the force the person exerts on the ball. But the ball goes flying off, and the person does not. What's going on?

The reason for this asymmetry of outcome is that the force of the person on the ball is the only force acting on the ball, but the force of the ball on the person is not the only force acting on the person. The two sets of forces are drawn below:
The forces on the person are the force of the ball on the person ($F_{bp}$) and the force of friction between the person’s feet and the floor. So the total force on the person is zero, and her motion doesn’t change, but the total force on the ball is non-zero, and its momentum is suddenly increased. Note that I have not drawn this picture at a time after the ball was thrown, I have just drawn separate pictures of the ball and the person while it is being thrown. After it is released there is no interaction at all between the ball and the person.

Also note that to simplify this picture I’ve left out the force of gravity which pulls down on both the person and the ball, along with the counteracting upward forces that balance the pull of gravity. We’ll have much more to say about this a bit later.

**Third Law Thought Experiments:**

Since the third law was somewhat surprising, it’s useful to consider a few thought experiments. These are offered up as examples which might help to ease you into acceptance of this very important idea.

Consider what happens when a horse pulls a stone with a rope. In the process the rope is stretched. It will do whatever it can to "ease" the stretch, so it both pulls the rock forward, and pulls backward on the horse. The effect is that the force with which the horse pulls the rock forward is just equal to the force with which the rock pulls the horse backwards.
In this case the rope just “transmits” the force between the horse and the rock. We’ll see more of how ropes transmit forces in a bit.

Newton himself put forth the following example: Imagine two objects A and B which attract each other through empty space (magnets if you like). Now put a piece of paper C in between:

![Diagram](A C B)

The force of A on C is just $F_{BA}$ (the force of B on A), and the force of B on C is just $F_{AB}$ (the force of A on B). If these two forces were NOT equal and opposite, there would be a non-zero net force on C, and the whole arrangement would accelerate off in some direction. Newton did not just take it for granted that this would not happen; he did the experiment. He set up a pair of little boats with magnets in them (boats so that the friction would be low), and placed between them a slip of paper. When they didn’t zip off in some direction he had new experimental evidence for his third law.

Later in the course we will see that this third law, so surprising at first, has a very deep origin in physics. It is ultimately related to the simple fact that the laws of physics are the same everywhere.

**Types of forces:**

These three laws will, it turns out, be adequate to allow us to understand a truly extraordinary range of phenomena, including everything to do with the structures of living things and how they move around. In a sense, they tell us that to understand motion, we need to pay attention to the forces. To do this, we’re going to spend considerable time pondering different kinds of forces. We begin with some generalities.

It's often helpful to talk about forces in two different dialectics. Every force can be put into one or the other category in each of these two dichotomies. The first way of splitting forces is to talk about active and passive forces.

Active forces are those whose magnitudes are determined by some external factor. Good examples of these are pushes and pulls (where we decide how large the forces will be), the gravitational force (which is always the same for a body near the Earth's surface), and electric and magnetic forces.

Passive forces are different from these. Passive forces are those which arise, and adjust themselves, in response to active ones. A good example emerges when you push on a wall. The wall pushes on you just enough to counteract your active push. If you lean gently against the wall, it pushes only a little. If you race at the
A wall and smash into it, it exerts a large force back on you. The key point with passive forces is they are *whatever they have to be*, and what they have to be is determined by the active forces in the problem. There are limits to passive forces. If you run into the wall hard enough, it will not be able to create a large enough force to stop you. Instead it will apply the biggest force it can and then break.

There is a second, independent, way to distinguish between forces; we can classify them as **contact** and **non-contact** (or long range) forces.

Contact forces are those that arise from physical contact between two bodies. They act only when the bodies touch in the usual sense of having their surfaces approach one another at atomic scales. Examples of these very common contact forces are those we associate with pushing against a wall or placing a book on the table, the forces which occur in a collision, and the frictional forces which allow you to walk across the room.

Non-contact forces are forces which can act even when the two bodies are not touching one another. The only non-contact force we’ll talk about extensively in this class is gravity. You don't have to be touching the ground for the force of gravity to act on you. Just step off a chair and you will see what I mean. In fact gravity can act on you even when you are very far from the Earth, through completely empty space. There are other non-contact forces. One which you have probably seen a little is the magnetic force, which like gravity can act even through empty space.

We’ll recall these divisions a little as we move through our discussion of mechanics. Keeping these ideas in mind will help you to know how different forces act.

**A first force: weight, the force exerted on you by the gravitational attraction of the Earth**

The first specific force we will talk about is weight. By “weight”, we mean the downward force exerted by gravity on any object near the surface of the Earth. The magnitude of this force depends only on the mass of the object of interest. The direction is always down, toward the center of the Earth. These facts can be written in our first force law:

\[ F_w = mg \]

Notice here that I have used the symbol \( g \), which is the vector acceleration due to gravity. This is a vector with magnitude \( 9.8 \text{m/s}^2 \) directed towards the center of the Earth. In words this equation just means that the magnitude of the force is always \( mg \), and its direction is always towards the center of the Earth. There are several aspects of this we need to explore.
The first of these has to do with the third law. It says that if the Earth is pulling you down with a force $F_{\text{Earth-You}}$, you must be pulling the Earth upward with a force $F_{\text{You-Earth}}$ which is equal and opposite to the first. That is:

$$F_{\text{Earth-You}} = -F_{\text{You-Earth}}$$

While the Earth pulls you down, you pull it up. That action-reaction pair is always there.

Why doesn’t the Earth come rushing up to meet you when you jump off a cliff? After all, you’re pulling it upward just as hard as it’s pulling you down. The reason for this disparity, which emerges always in unequal matches like this, is the relatively enormous mass of the Earth. Your weight ($F_{\text{Earth-You}}$) is a big enough force to quite easily change your motion. But a force of the same size ($F_{\text{You-Earth}}$) is much too small to make any appreciable change in the motion of the Earth. So although these two forces act to pull you and the Earth together, it’s you who does all of the moving.

The second thing worth discussing in some detail is the sensation of weight. What is it we feel when we feel our own weight? Can you feel the force of gravity upon you?

Imagine what happens when you jump off a chair. For a moment you are floating freely in the air. Do you "feel" a force tugging on you? While you’re in the air, there is no sensation of force at all. Try this if you dare, but please be careful! So what is the sensation of weight? What is this feeling you get?

When you are standing on the floor, the sensation that you feel is of pressure on your feet. This sensation is exactly what you would feel if, while you were lying down, someone pressed with a board on your feet. Now think about what you feel when you sit in your chair. Is there any pressure on your feet? Now the pressure seems to be on your backside. And if you stand on your head you "feel your weight" on your hands and head.

So could what we feel really be the “weight”? Is what you feel really the force exerted on you by the gravity of the Earth? In fact it is not. The sensation you feel when you talk about weight is actually the force which something else applies to you to resist the downward pull of gravity, to prevent you from falling downward.

When you are standing still your weight (the force of gravity on you) pulls you down. In order to remain stationary, some other force must balance this. That force is provided by the object which you are in contact with. The sensation you feel as weight is just the force of the floor (or your chair, or whatever) pushing up on you to resist your weight. Take away the floor, or the chair, and your sensation of weight would vanish, but the weight itself (the downward pull of gravity) would not.

**Contact forces and a second force: the normal force**

Most of the forces we encounter in our lives fit are contact forces. They arise from the direct "touching" interaction between two bodies. When two more or less solid bodies are in such direct, atom-to-atom contact, it will usually be useful to talk about the force between them as being composed of two parts. We will break the total contact force
between objects into a component perpendicular to the plane of contact between the surfaces (the normal force) and a component along the plane of contact between the surfaces (the friction force). The first of these is the part which prevents one object from moving through the other. Let’s look at an example.

When I put a book on a table, the book's weight tries to pull it downward. To move downward it would have to pass through the table. The table prevents the book from doing this by pressing back up on it with a force which keeps it in place. This force (which acts to prevent the objects from passing through one another) always acts perpendicular to the plane of contact of the objects. Because of this it is called the “normal” force. It is NOT normal in the sense of "usual", but normal in the mathematical sense of “perpendicular” to the surface between the two objects.

How does this normal force occur? Ultimately all of the forces (save gravity) which we will talk about in this course are consequences of electromagnetism. Electromagnetic interactions determine whether two atoms placed close together will resist being pushed closer together or attract one another (perhaps bonding together). In the second semester of this course, Physics 235, you will learn quite a lot about the nature of these electromagnetic forces. For now, we will encode a lot of complicated atomic interactions in a few simple phenomenological rules.

There’s a real mystery here. How can an inanimate object like a table exert a force? More important, how can it “decide” how much force to apply; knowing to push upward very lightly on a pencil while providing a much larger force for lamp? The answer lies in the passive nature of the normal force. It exists only in response to some other force.

Begin by thinking about a cushioned chair. When you sit on it, the chair compresses to the point where to compress it further would require a force larger than your weight. In other words, as it is "squashed", it pushes back up on you, harder and harder, until it is pushing up on you with a force equal to your weight. This squashing, this distortion, is what allows the chair to push back up on you. When you push atoms closer together, they push back.

Now imagine something "harder" than this cushioned chair. If you stand on a plastic chair, it too is distorted until it pushes back on you just enough to balance your weight. Take this to its conclusion; when you stand on the floor, the floor actually distorts until it pushes back on you with a force just large enough to prevent you from falling through it. This distortion, which always accompanies forces applied to solid objects, is perfectly real, even when it is not apparent.
So this "normal force" prevents objects from passing through one another. How big is this force? What is its magnitude? The basic answer is "whatever it has to be". For this reason, the normal force is our first good example of a passive force. Passive forces have magnitudes which are not determined in advance. They arise, and adjust themselves, in response to active forces. Their magnitudes are determined based on the restrictions which give rise to them. They can usually be any size up to some limit at which the object which is creating them breaks.

So, if I push down on a table, it pushes back up on my hand with a force just equal and opposite to my own. If I push harder the resisting normal force increases. If I stop pushing it goes away. The force adjusts itself to be just as large as it needs to be to prevent my hand from moving through the table.

Free body diagrams

Notice what we did there. In order to understand what happens to two bodies while they interact, we have drawn each of the bodies separate from the others, so that we can understand fully the forces on each one. Let's look at a couple of other examples of this. What happens when a book sits on the table? What are the forces on it: First we might draw the circumstance:

Now, in order to understand it, I draw a free body diagram for each part of the problem. First consider the book. What are the forces acting on it? It experiences a weight, the gravitational force of the Earth pulling it downward. Since it is sitting still, it must also experience some other force which balances this. This is the force with which the table pushes back up on the book. This is called the normal force.

Now in order for it not to move we know that: \( F_N = F_W = m_b g \). These forces must be equal in magnitude (and opposite in direction). So now we know that \( F_N = m_b g \).

What does Newton's third law say about this? It says that for every force there must be an equal and opposite reaction force. What are they here? The table pushes on the book with a force: \( F_N = m_b g \), so the book must push on the table with a force \( F_{BT} = -F_N = -m_b g \).
What about the weight? The Earth pulls on the book with a force $F_W = -m_b g y$, so the book must equally pull on the Earth with a force $F_{BE} = m_b g y$. So to understand the (non)motion of the book, we have to consider two different third law pairs:

$$F_{TB} = -F_{BT} \text{ (normal force)} \quad \text{and} \quad F_{EB} = -F_{BE} \text{ (weight, or gravitational force)}$$

And to draw all these forces and keep track of both interactions I would have to draw three objects:

Is this all there is? NO! I have not drawn all the forces acting on either the table or the Earth. The only body I have completely analyzed here is the book. The table must have other forces acting on it, or it would accelerate away. Likewise, the Earth must have other forces acting on it or it would accelerate (however slowly) up towards the book.

It is very important that you should understand this third law, and be able to identify third law pairs in a problem like this correctly. If you can't identify third law pairs with confidence it will be very difficult for you to correctly analyze even slightly complex systems.

**Examples of determining the normal force:**

First the most familiar one, the book on the table:

When I set the book down on the table, the table distorts slightly, continuing to do so until the force which it exerts back up just balances the force which gravity exerts on the book. So in this case

$$F_N = F_W$$
Now imagine that instead of just this, I push down on the book with some force $F_p$. What is the normal force now?

Since the book remains in place, I know that $F_N - F_W - F_p = 0$, or $F_N = F_W + F_p$. How does the table do this? How does it adjust its force to just the right value? It does this by having the structure of the table push back with a bigger and bigger force the more it is distorted. So it just continues to distort until the force it is applying is just big enough to balance the book.

What if, instead, I pull up somewhat on the book?

Now I know that:

$$F_p + F_N = F_W$$

or

$$F_N = F_W - F_p$$

Notice what happens though, as I gradually increase $F_p$, $F_N$ continually decreases until it becomes zero.

What happens after this? Does $F_N$ become negative?

So we have seen simple cases in which $F_N$ for an object on a table can be either larger or smaller than the weight of the object. In fact the special case in which $F_N = F_W$ is just a special case, and not any kind of general rule. This is because the weight and the normal force are not third law partners. The third law partner for the upward force $F_N$ is the downward force of the book on the table, and the third law partner for the downward force $F_W$ is the upward gravitational force of the book on the Earth.

Since $F_N$ and $F_W$ are not third law partners, there is no reason that they must be the same. Sometimes they are, but the certainly don’t have to be. To determine $F_N$ in a particular problem, you just have to figure out how large a force is required to keep the objects from moving through one another.

In fact there’s another, really simple and obvious way to tell that $F_N$ and $F_W$ cannot be third law partners in this problem: both forces act on the same object! Remember, the third law is about forces which are exchanged between two objects. So the two parts of a third law pair can never act on the same object.

Copyright: Regents of the University of Michigan, 2009
The other part of the contact force: friction

The normal force is what prevents objects from passing through one another. It is the part of the total force between two surfaces which is perpendicular to their plane of contact. The rest of the force between two bodies is that part of the force which is parallel to the plane of contact. This force acts to prevent one object from slipping over the other. It resists their relative motion. We call this part of the interaction between two objects the force of friction.

We will have a lot more to say about friction and how it works next time. For now let’s just concentrate on the manner in which friction acts. Friction acts in an attempt to prevent relative motion *along* the plane of contact between two objects.

Let's look at a simple example; an object sitting at rest on a slope:

![Diagram of object on a slope](image)

What are the forces on this? There is a weight acting straight down. Then there is some interaction between the book and the surface of the wedge it sits on. This total contact interaction has two parts; a normal force perpendicular to the surface, and a frictional force along the surface. I know (because its motion doesn’t change), that the total force $F_{total} = F_N + F_f + F_W = 0$. So now I can write the forces:

![Forces diagram](image)

Very often in a problem like this it is useful to work in a coordinate system which defines directions along and perpendicular to the surface between two bodies. Such a coordinate system is shown above. Now we know that since it doesn't move the sum of the forces must be zero, and that in turn means that the sum of the forces in each direction must be zero. Now let's add up these forces in each direction:

\[
\sum F_{\parallel} = F_w \sin \theta - F_f = 0 \\
\sum F_{\perp} = F_N - F_w \cos \theta = 0
\]

So in this case we know that:

$F_f = F_w \sin \theta = mg \sin \theta$ and $F_N = F_w \cos \theta = mg \cos \theta$
Once again we see that the normal force is not equal to the weight (it rarely is), and in addition we see how we can use the motion (or lack of it) to figure out how large this "frictional force" must be.

Now this picture, a block on an “inclined plane”, is the very icon of traditional introductory physics courses. Poor students have been learning to analyze these for literally hundreds of years, and they’ve always seemed completely unconnected from everyday experience. After all, most of you stopped playing with blocks some time ago. But in fact this example, while drawn in an abstract (and easy to draw) way, is very much an everyday experience. Here are two examples.

The first is standing on a slope: Every time you stand on a slope a situation very like what we just described happens. If there is not enough frictional force preventing you from sliding down the slope you will slip downward. No doubt this is something you’re quite aware of, and once snow and ice arrive you’ll be careful about standing on slopes.

The second example is the slope itself! Each layer of a hill would slip downward if not held in place by a frictional force. Sand dunes, like Sleeping Bear dune in Northwest Michigan, provide a vivid example of this. When they become too steep, the force required to hold the sand in place becomes larger than the frictional force available, and top layers of sand begin to slide downward. The maximum angle for a pile of sand is often given the poetic name “the angle of repose”. More impressive and dangerous versions of this are seen all the time in avalanches and landslides.
**A way to transmit force; ropes and tension**

We have seen how simple contact forces can occur due to the compression of bodies, as when a book sits on a table. It is also possible for bodies to exert forces when they are “stretched”. This process of attempting to stretch a body is called putting it “in tension”.

A simple example of how an object in tension behaves is given by a mass hanging on a rope.

Imagine a mass hanging from a rope attached to the ceiling. We can draw three free body diagrams:

What do we know about these? We know that for the mass at the bottom:

\[ \sum F_y = F_{rm} - m_r g = 0 \]

so \( F_{rm} = m_r g \)

and for the rope:

\[ \sum F_y = F_{cr} - F_{mr} - m_r g = 0 \]

so \( F_{cr} = F_{mr} + m_r g \)

We also know that some of these are third law partners, so their magnitudes must be equal. In particular:

\( F_{cr} = F_{rc} \quad F_{rm} = F_{mr} \)

So now we can say:

\[ F_{cr} = F_{mr} + m_r g = F_{rm} + m_r g = m_r g + m_r g = (m_m + m_r) g \]

So the force which the ceiling must exert to support the rope and the weight is just equal to the weight of the mass plus the weight of the rope. Not too surprising.

**What about the forces within the rope?**

Now let’s think about what’s really happening in the rope….Picture the little piece of the rope at the bottom. The mass is pulling down on it with a force \( m_m g \). This little piece has some mass \( \Delta m_r \), so we can write the same kind of free body diagram for it…. 
So for this little piece we find:

\[ \sum F_y = F_{r\Delta m} - \Delta m g - m_r g = 0 \quad \text{so} \quad F_{r\Delta m} = (m_m + \Delta m_r)g \]

As we gradually move up the rope, this force inside the rope gradually grows, until just at the top, where it must support the full weight of the rope below it, the internal stretching force in the rope is what we calculated above.

Think about this a little. What's happening is that each little part is being pulled equally both up and down; these forces are trying to tear the rope apart. It is the ability of the rope to hold itself together against this “tension” that allows it to "transmit" the weight of the hanging mass to the ceiling above.

Now it is also perfectly possible to transmit forces with compression. That's what I do when I push something along with a stick. And there's no need for the thing in tension to be a rope. I would just as well hang the mass from typical solid like a wooden meter stick and the analysis would be exactly the same. There would be a tension in the stick.

Most solids can support loads either in tension or in compression. Some others, especially those with more interesting internal structure like ropes, tendons, and flesh, are better at supporting loads in tension than compression. As we will see, your body is basically a framework of bones capable of supporting your weight in compression, with flesh and organs hanging from the skeleton supported by tension.

**Carrying a suitcase**

Let’s apply this to a real world example. You’re headed to the airport with a suitcase to check. It has the maximum allowable weight (50 lbs, or about 23 kilograms), and you carry it with your arm straight up and down. The picture is to the right. The analysis is the same as above:

\[
F_{\text{shoulder-arm}} = F_{\text{arm-shoulder}} = W_{\text{arm}} + W_{\text{suitcase}}
\]

\[
F_{\text{arm-shoulder}} = F_{\text{Earth-arm}} + F_{\text{suitcase-hand}} = W_{\text{arm}} + W_{\text{suitcase}}
\]

\[
F_{\text{hand-suitcase}} = F_{\text{suitcase-hand}} = W_{\text{suitcase}}
\]

So the force your shoulder applies to your arm is larger than just the weight of the suitcase. Your shoulder has to hold up both the suitcase and your arm.
How much does this matter? To know you have to figure out how much your arm weighs. How might you estimate this? Let’s say your arm is a cylinder about 10 cm in diameter and 0.8 m long. This would have a volume:

\[ V_{\text{arm}} \approx (\pi \times 0.05^2) \times 0.8 = 6.3 \times 10^{-3} \text{ m}^3 \]

To find the mass, we multiply this volume by a density. What’s the density of your arm? You probably know that you’re mostly made of water, so a reasonable estimate is the density of water, which is about 1000 kg/m³. So the estimate for the mass of your arm is:

\[ m_{\text{arm}} \approx \text{density} \times \text{volume} = 1000 \text{ kg/m}^3 \times 6.3 \times 10^{-3} \text{ m}^3 = 6.3 \text{ kg} \]

The force your shoulder applies to your arm is \( W_{\text{arm}} + W_{\text{suitcase}} = m_{\text{arm}}g + m_{\text{suitcase}}g = 287 \) N. To just hold up the suitcase the force would be less: \( W_{\text{suitcase}} = 225 \) N.

**Ropes and sending forces around corners**

One important reason that ropes and tendons are important is their ability to transmit a force around a corner.

Consider the following variation:

What happens here? Here the weight of the mass pulls down on the rope, stretching it, putting it in tension. The nifty thing is that the rope allows this force to be transmitted around the corner, so that ultimately it is supported by the wall on the left.

This happens very often in organisms, and for our purposes this will be the most important kind of application. A good example is the human shoulder, where a complex series of tendons transmit forces generated by muscles in your back and arms around the flexible corner of your shoulder. Muscles pull straight along their length, while the tendons (which do not themselves generate any forces) transmit the forces generated by the muscles to where they need to be applied.
Why would physicists talk about massless ropes:

Let's go back a step to the simple hanging mass problem. We found that the force exerted by the ceiling on the rope was:

\[ F_{cr} = (m_m + m_r)g \]

In other words the rope isn't a perfect force transmitter; it isn’t just transmitting a force from the ceiling to the hanging mass. Some extra force is needed to support the weight of the rope.

How important is this? The answer depends on the details of the problem. If the mass of the rope is very much less than the mass of the hanging object \( m_r << m_m \), then we can say with some accuracy:

\[ F_{rc} \approx m_m g \]

In this case the rope is a very good force transmitter. It allows the ceiling to exert a force on the object below without adding anything to it.

What if \( m_r >= m_m \)? Now the force the rope exerts on the ceiling is at least twice would it would be if we directly attached the mass. So in this case the rope is a really poor force transmitter.

For some of the problems we will do we will assume that the mass of a rope is "small enough" to "not matter". If the rope is assumed to be "massless" in this way, then it becomes a perfect force transmitter. Any force applied to one end is directly transferred to the other end no matter what the circumstances. In technological cases, people always try to use ropes where this is the case, choosing a rope strong enough to support the load, but light enough to be able to transmit almost all of the force to the load. Very often it is a reasonable approximation. The one case where this becomes very difficult is with very long ropes, like those used to sample material at the sea floor, or the cables used in the constructions cranes which have become so common in Ann Arbor.

When we assume such a massless rope, the force exerted on one end is directly transmitted to the other. This force which is transmitted is what we call the "tension" in the rope. So, if the tension in a rope stretched between two objects A and B is 50N, this means that object A is pulling on object B with a 50N force, and object B is pulling on object A with a 50N force. This force is perfectly "transmitted" by the rope, with no loss.

**Tension and force transmission:**

Consider the following situation. A block \( m_b \) hangs on an essentially massless string.
Considering the block as we did above we find:

\[ F_{sb} = T \]
\[ F_{Eb} = W = mg \]

The fact that the block is at rest implies that \( T = mg \). With what force does the string pull on the wall?

Now consider a slight variation. Instead of attaching the string to the wall, I hang a second, identical weight off a second pulley. What is the tension in the string now?

It is still \( T = mg \). Remember, all the string does is transmit a force from one end to another. It doesn't matter if that force comes from the wall or another block, it is still just transmitted.

**Reducing forces required with simple machines**

There is one last general topic to raise before moving on. We have been discussing the forces required to cause objects to move in various ways. We found, for example, that to lift an object at a constant rate straight up we have to apply a force just equal to its weight.

I want you to think a little about what limits the actions people can take. When a person tries to do something, tries to make something happen, they are generally limited by the size of the maximum force they can apply. This is true whether you're lifting a large stone or trying to open a jar; people are generally force limited. So how can we do things when the force required is too large for us to create?

We have to get some help, and there is a general class of ‘tools’ which have long been known to help evade these limits. These tools are generally called ‘simple machines’, and they are probably best thought of as force magnifiers.
A first machine: the inclined plane

Let’s start with the simplest simple machine: the inclined plane. If I want to lift an object straight upward at a constant rate without help, I have to apply a force equal to its weight:

$$\Sigma F_y = F_{lift} - mg = 0$$
$$F_{lift} = mg$$

If I’m not strong enough to supply this large a force, I can take advantage of the simplest simple machine: the inclined plane. To slide a box up such a plane at a constant speed, I will have to push with a force $F_{push}$ such that:

$$\Sigma F_{along} = F_{push} - mgsin\theta - F_{friction} = 0$$
$$F_{push} = mgsin\theta + F_{friction}$$

$$\Sigma F_{perp} = F_N - mgcos\theta = 0$$
$$F_N = mgcos\theta$$

So long as I can keep the friction low, I can ‘lift’ an object of arbitrary mass with my small maximum force by making the angle $\theta$ small and sliding the block up the plane. This kind of simple machine was obviously known to the early Egyptian pyramid builders who combined large numbers of laborers (increasing $F_{max}$) and shallow inclined planes to lift quite enormous blocks of stone.

If you think about this a little more generally, this trick has allowed you to ‘magnify’ your force by a factor of $1/sin\theta$. You apply a force $F_{push}$ in a direction along the slope and you are able to generate a force approximately perpendicular to this which has a magnitude $F_{push}/sin\theta$.

The force magnification of the inclined plane is at work in many other systems as well. An excellent example is a knife. What are you doing when you cut? You’re trying to tear a material apart, to pull the two sides of the material apart with so much force that the bonds holding it together come apart. Doing this generally requires a large force. By forcing in a very thin knife blade you can generate a large sideways force.

In this case, if you push in with a force $F_{push}$, you generate a sideways force with a magnitude on the order of $F_{push}/sin\theta$. This is what allows you to smoothly cut apart an object which would otherwise be almost impossible to rip apart.
How does a screw work? It’s basically a twisted inclined plane. You apply a downward force, and it generates a magnified force $F_{\text{down}}/\sin \theta_{\text{pitch}}$ sideways, forcing apart the wood into which you are inserting the screw. A nail is similar, except that the pitch angle is usually larger (just at the tip of the nail), so you have to apply a large force to push it in.

Isn’t this somehow cheating; getting something for nothings? Not really, because there is a significant price paid. Think of the original inclined plane problem. Yes, I only need to apply a force $mg \sin \theta$, but if I want to raise the block a distance $\Delta h$, I have to apply this reduced force over a much larger distance ($\Delta h/\sin \theta$) than I would if I lifted it straight up. So although I can use a smaller force, I must apply it over a greater distance. For most purposes this is fine, because people are far more strength limited than they are endurance limited.

It’s interesting to note that the product of force applied times distance:

$$F_{\text{push}} \times \text{distance} = mg \sin \theta \times \Delta h/\sin \theta = mg \Delta h$$

Is the same as if we lift it straight up:

$$F_{\text{lift}} \times \text{distance} = mg \Delta h$$

As we will see in a while, this product of force times distance represents something very important. It measures a transfer of “energy”. Lifting the block always requires the same total amount of energy, but we may supply it all at once or a little at a time.

**A second simple machine: the block-and-tackle**

Let’s consider a second kind of popular force magnifier, the block and tackle. Compare these cases:

\[
\begin{align*}
\sum F_y &= T - Mg = 0 \\
T &= Mg \quad \sum F_y &= 2T - Mg = 0 \\
T &= Mg/2 \quad \sum F_y &= 4T - Mg = 0 \\
T &= Mg/4
\end{align*}
\]

What’s going on in each case? In the first, the rope which I’m holding onto supports the full weight of the block; no force magnification here. In the second case, there are two ropes attached to the block, each pulling up with tension $T$. We know the tensions are both $T$ because if they weren’t, the rope would move around the pulley until they were. I pull on the rope with force $T$, and the ceiling pulls the other way with force $T$. So the ceiling actually supports half of the weight. In the next case we’ve just extended this.
Now the block is held up by four ropes pulling with tension T; one is supported by me, and three are supported by the roof.

You can see how you would multiply the number of attachments to increase the magnification. Here again, there is a penalty. First, you have to keep friction in the pulleys low. Also, you have to pull farther. Consider the middle case. Here if you pull a distance $\Delta x$, the block will rise a distance $\Delta x/2$. This is because you must shorten both ropes which hold up the block to make it rise. In the third case you have four ropes to shorten, so $\Delta h = \Delta x/4$.

Again we find that the product of force times distance is constant, no matter what we do. You’re not cheating with these devices, just changing the way you pay the price. There are other important simple machines, including hydraulic devices and various levers. We will revisit these when we have better addressed the physical principles which underlie them.
Physics for the Life Sciences: Fall 2006 Lecture #4

Torque and rotational statics

We're going to extend our discussion of objects at equilibrium; things which aren't accelerating. We know from Newton's second law for translational motion that:

\[ \sum F = \frac{dp}{dt} \]

so if \( \frac{dp}{dt} = 0 \), then \( \sum F = 0 \). This is the first condition for equilibrium: if the momentum of an object is not changing, the vector sum of the forces on the object must be zero. For point objects this is all we need to know.

Now it is time to go beyond this, and begin thinking about how extended bodies will behave under the influence of forces. What happens if I have the following arrangement; a bar with a rope holding it up on one end?

This is a body for which \( \sum F_x = \sum F_y = 0 \). Do you think it will remain at rest? No, it will begin to rotate. Whenever a body is extended, larger than a "point" object, it is necessary to know both what forces act on it, and also where the forces are applied.

You all know from experience what we would have to do to prevent rotation in the system described above; we'd just hang it from two ropes:

Why is this case stable when the other wasn't? Because here we have one force which "tends" to make the object rotate clockwise, and one which "tends" equally to make the object rotate counterclockwise. This is the basic idea of our second condition of equilibrium; the forces applied to an object at equilibrium must be applied in such a way that their tendency to make the object rotate cancels out.
Why haven't we talked about this before? A lot of the time we have considered point objects, and any object which does have extension really can't rotate. They're just points, and a point can't rotate. So any time I apply a force to a point, it can only translate. When we have considered problems involving extended objects, we have just ignored rotation. Not because it wasn't possible, just because we hadn't gotten to talking about it yet. We will look at some examples of how rotation comes into problems like those we have done later today.

**Quantifying the ability of a force to make something rotate:**

It should be clear from this little discussion that the ability of a force to make an object rotate depends not only on how large the force is, but also on where it is applied. If you want to open a door, to make it rotate around its hinges, you can either apply a large force close to the hinge, or a small force very far from the hinge. The direction you apply the force also affects the result. If you push on a door along a line which passes through the hinge, the door will never begin to rotate.

The number which quantifies this "ability to cause an object to begin to rotate" is the called the "torque". A first definition of the torque which will generate rotation around a center \( c \) is:

\[
\tau_c = \mathbf{r} \perp \mathbf{F}
\]

where \( \mathbf{r} \perp \) is called the "moment-arm" of the applied force and \( \mathbf{F} \) is the magnitude of the force. This "moment-arm" is illustrated in the following figures:

There are three lines drawn in each picture. In each the solid line is a vector representing the force \( \mathbf{F} \). The dot-dashed line is a vector drawn from the center of rotation to the point at which the force is applied. We call this position vector \( \mathbf{r} \). The dotted line perpendicular to the force \( \mathbf{F} \) in each case represents the "moment-arm" associated with this force, \( \mathbf{r} \perp \).

One thing to note is that I can move the force forward or back along its direction and produce exactly the same \( \mathbf{r} \perp \). A second thing to note is that the rotation which is produced by a force has a particular direction, it can cause rotation one way or the other. We record direction of rotation by using the "right hand rule". One way of stating this rule says that if you curl the fingers of your right hand in the direction of motion of the rotation, your thumb points in what we define as the direction of rotation.
Experience, and a second way of looking at torque

Does this agree with experience? Let's look at a few limiting cases.

1. $\mathbf{F}$ is perpendicular to $\mathbf{r}$. Now $|\mathbf{r}| = r_\perp$ and the torque is just $\tau = rF$. This is the familiar case of opening a door by pushing perpendicular to its surface. I can create the same torque by either pushing with a large force close to the door hinge, or pushing with a small force far from the door hinge. The torque is the product of these two.

2. Force $\mathbf{F}$ is parallel to (or anti-parallel to) $\mathbf{r}$. In this case the force points towards or away from the center of rotation. In this case $r_\perp$ is zero, and the torque associated with this force is zero. Applying a force like this can never cause rotation about this center of rotation.

3. Force $\mathbf{F}$ is applied at the center of rotation. This is really a subset of the previous example, and it also generates no torque and causes no rotation.

These facts suggest a second, equivalent way of looking at torque. We can take each force which acts on a body and break it up into two components, one directed along the line to the center of rotation and one perpendicular to it. The component along the line through the center of rotation will generate no rotation. Only that component of the force which is perpendicular to the line through the center will cause rotation. This suggests another way to determine the torque generated by a force:

$$\tau = rF_\perp$$

So now we have two alternate ways of looking at it:

$$\tau = r_\perp F = rF_\perp$$

Compare these on the drawing:

Torque and the cross-product

There is a general way to see what the magnitude of the torque will be:

$$\tau = r_\perp F = rF_\perp = rF\sin\theta$$

where $\theta$ is the angle between the vectors $\mathbf{r}$ and $\mathbf{F}$, as shown in the drawing:

You can see this by noticing that $F\sin\theta = F_\perp$, or by seeing that $r\sin\theta = r_\perp$ as shown below:
This generic way of denoting the torque is often written in a mathematical short-hand:
\[ \tau = r \times F \]

This is the first of two kinds of "vector multiplications" which we defined earlier in the class. It's called a "cross product", and we describe it by saying "torque is equal to r cross F". What it means is that the torque is a vector perpendicular to both \( r \) and \( F \), which has a magnitude \( rF \sin \theta \). The direction of this torque vector comes from the same right hand rule we described above. Because the cross-product of two vectors produces a new vector it is also called the vector product.

Point the fingers of your right hand along \( r \), then curl them towards \( F \), and your thumb points in the direction of the torque.

Notice one essential feature of this: You cannot calculate how large a torque some force produces until you specify exactly what center of rotation you’re talking. Sometimes the center of interest will be obvious, other times, like in examples below, it will not.

**Conditions for equilibrium**

So now there are two conditions required for equilibrium. In addition to requiring the sum of the forces being zero, we must also require the sum of the torques to be zero

\[ \sum F = 0 \quad \sum \tau = 0 \]

Now there is a key feature to this second condition. It says that the sum of the torques around every possible center of rotation must be equal to zero. If this were not so, if there were some center of rotation around which the torque was not zero, the object would begin to rotate around that center. This is a powerful fact. When you are analyzing some situation and finding the forces and torques on a body in equilibrium, you can sum the torques around any convenient center of rotation and you know they must always be equal to zero. Judicious choice of which center to use can very much simplify the calculations required in many statics problems.

**Weight and the center of gravity**

Most of the forces we’ve been talking about so far are contact forces. Each is applied at a particular place on the surface of a body. It is clear where they act, and hence easy to calculate what kinds of torques they create.
The only non-contact force we are concerned with right now is gravity. Where does gravity act? How do we determine what kinds of torques the weight of an object produces?

For extended bodies, we can always treat the force of gravity as if the entire force were being applied at a particular spot in the body called the “center of gravity”. For objects near the surface of the Earth this is in the same "mass weighted average position" of the object which we might also call the center of mass:

$$R_{cg} = \frac{\sum m_i r_i}{\sum m_i} = \frac{\int r \, dm}{\int dm}$$

In this definition the sums are taken over all the little pieces which make up the object. If the object is not near the surface of the Earth the CG will be different. Usually this only happens for really large objects, like the Earth and the Moon. In our considerations of living things we won’t need to worry about this.

For homogenous (all made of one kind of material) and symmetrical objects, the center of gravity is always at the most obvious center of the object. For a sphere it is at the center, for a hoop at the center, for a square at the center, a cube at the center, etc. If the object is either not homogeneous or not symmetrical the CG will be “pulled” toward the parts which are more massive.

For example, consider the CG of a 30 cm long barbell with a 10 kg mass on one end and a 20 kg mass on the other. Where is the CG of this object? Taking the origin to be the location of the 10 kg mass, we have:

$$R_{cg} = \frac{\sum m_i r_i}{\sum m_i} = \frac{(10 \text{ kg} \times 0 \text{ m} + 20 \text{ kg} \times 0.3 \text{ m})}{(10 \text{ kg} + 20 \text{ kg})} = 0.2 \text{ m}$$

Rather than being in the center of the barbell, the CG is pulled toward the larger mass, and is instead 2/3 of the way down from the end.

There is an important lesson here: Because gravity acts equally on all parts of an object, it cannot actually cause rotation around the center of mass. Contact forces, on the other hand, always act on particular parts of the body, and hence can generate rotation.

**An application: stability and balance**

A fine application of the idea of equilibrium is to stability and balance. Thinking about this is especially useful as it helps to illustrate the different kinds of equilibrium. There are three kinds of equilibrium:

1. Stable equilibrium: if the object is slightly disturbed it will return to equilibrium
2. Unstable equilibrium: if the object is even slightly disturbed it will move far away from equilibrium
3. Neutral equilibrium: a small displacement leaves the object in a new equilibrium position

These three are illustrated by thinking about a cone. When the cone is standing on its bottom, it is in stable equilibrium. Tip it a bit and it falls back into place. If it’s standing on its tip, it is in unstable equilibrium. Tip it a bit and it falls over completely. If you lay the cone down on its side it is in neutral equilibrium. Roll it over a bit and it just lies there.

For simple cases of balance, like those which apply to many organisms the following rules apply:

1. If a tilt away from equilibrium raises the center of gravity, the object is in stable equilibrium
2. If a tilt away from equilibrium lowers the center of gravity, the object is in unstable equilibrium
3. If a tilt away from equilibrium leaves the height of the center of gravity unchanged, the object is in neutral equilibrium

This is because of the torque exerted on the object by the force of gravity. If you’re raising the center of gravity when you tip it, the force of gravity will tend to pull it back down. If you’re lowering the center of gravity, the force of gravity will tend to pull it away from where it started.

When will an object "tip" over? Any time the center of gravity of an object is not above the supporting surface of the object it will tip. This is because once the CG moves past the support point, gravity exerts a torque which tends to tip the object over.

Consider these pictures to get an idea of how this works. As we tip this stiff little person over he is, at first, stable. We’re still raising the CG. Eventually, the CG moves outside the support point, and is now moving downward. This is unstable.
A quantitative example of rotational equilibrium: a truck on a bridge

A first useful example comes from analyzing a truck driving over a bridge which is supported on its two ends. What might we like to know about this? Imagine we know the weight of the truck, and we want to know how much force must be applied by the right and left hand supports as the truck moves across the bridge. This is just the sort of thing an engineer or an architect might need to know to make sure the bridge is safely constructed.

Consider the forces on the bridge slab. There is an upward force on each end, a downward force of the weight of the bridge, and a downward force due to the mass of the truck.

We can find the magnitudes of $F_L$ and $F_R$ by using the equations of equilibrium.

$$\sum F = 0 = F_L + F_R - m_t g - m_b g$$

and

$$\sum \tau = 0 = -m_t g x - m_b g L/2 + F_R L$$

where for this second equation I have calculated the torques with a center of rotation at the left hand end of the bridge.

Why do I choose this center? After all, I could calculate the torques around any center and they must always be equal. This choice is made purely for convenience. By picking a center through which one of the two unknown forces passes ($F_L$ goes through this spot) I know that only one of the unknown forces will appear in the torque equation I obtain. This just makes the algebra a little simpler than it would otherwise be.

Solving the torque equation yields:

$$F_R = m_t g x / L + m_b g / 2$$

And plugging this back into the first equation gives us the other unknown force

$$F_L = m_t g (1 - x / L) + m_b g$$

What does this mean? It means that the upward force exerted by the right hand bridge support is half the weight of the bridge plus a fraction of the weight of the truck that varies as it drives across the bridge. When it is over the left hand support, all of the truck’s weight is supported on the left, when it reaches the middle, half of it is supported by each, and when $x=L$, all the weight is supported on the right. The upward force exerted by the left hand support makes up for the rest of the weight of the bridge and truck. Added together, $F_L + F_R = m_t g + m_b g$, all the time. This is a nice answer, which we might have anticipated, worked out using our equilibrium conditions.
Example 2: Weight lifting

Your bodies, and the structures of other living things, obey these same principles of statics. These simple ideas, that in a static situation the sum of forces and sum of torques must be zero, is all it takes to understand a lot about how your body works. Let's start with a simple example: lifting a one dumbbell:

Imagine that the mass of your forearm is \( m_{fa} \), and the mass of the dumbbell is \( m_d \). How do you hold this up? You have a bicep which attaches to your forearm just a few centimeters from the joint in your elbow. If you poke around the inside of your elbow now, you can feel the tendons which connect this muscle to your forearm. So the picture looks something like the situation on the right.

Remember, the only way to analyze forces is to consider only a single object in a free body diagram. So let's look just at the bone in the forearm:

If you just had the three forces we mentioned (weight of forearm, weight of barbell, and upward bicep force \( F_b \)), what would happen? Your forearm would rotate. So, there must be another force. What is it? It's the force of the end of your upper arm bone pushing down on the end of your forearm. This sort of thing is always happening when your body supports weight; you do this by pulling with muscles and pushing with bones. The combination of the two is what allows your full range of movement, and you need them both: muscles must have something to pull against.

To determine the size of these various forces, we have two facts to work with:

\[
\sum F = 0 \quad \text{and} \quad \sum \tau = 0
\]

Here there are only \( y \) forces, so:

\[
\sum F_y = F_b - F_{ua} - m_{fa}g - m_dg = 0 \quad \text{and} \quad F_{ua} = F_b - m_{fa}g - m_dg
\]

If we sum the torques around the end, we find:

\[
\sum \tau = xF_b - (L/2)m_{fa}g - Lm_dg = 0 \quad \text{or} \quad F_b = (L/x)[1/2m_{fa}g + m_dg]
\]
What happens in the limits?

\[ x = L/2 \text{ (bicep supports at the center of the forearm):} \]
\[ F_b = m_\text{b}g + 2m_\text{d}g \]
\[ x = 0 \text{ (bicep supports exactly at the end of the arm):} \]
\[ F_b = \infty ! \]

You can't do this, so there must be some extension away from the joint.

Note that \( F_b \) scales with \( L/x \). What does this mean for people of different sizes? If I make a person larger, what happens to \( L \)? What happens to \( x \)? So how does the force they must apply change? Not at all. This would suggest that lifting a dumbbell of mass \( m_\text{d} \) would require the same force for a small person or a large person.

Now buried in this statement is the assumption of isometry. Isometry assumes that all dimensions of the person change by exactly the same factor. This is another way of saying the scale change is isomorphic, and preserves the shape. The shape of the person is the same, only the size of the person is different. If this is violated, let’s say because one person is short and stocky and the other is tall and lanky, the simplest scaling will not hold.

For the short and stocky person, \( L/x \) will be relatively small. For the tall and lanky person, \( L/x \) will be relatively large. Hence it will be easier (require less bicep force) for the short and stocky person to hold up the barbell than the tall and spindly person. This is in perfect accord with our sense that short stocky people seem stronger than tall and lanky people. Short, stocky people actually can lift more with muscles of the same intrinsic strength.

Understanding the physics of the distribution of these forces is what makes it possible for you to appreciate these simple scalings, which explain a lot of what we know about how people move.

Returning to the first equation:
\[ F_\text{ua} = (L/x)[1/2m_\text{b}g + m_\text{b}g] \]
\[ - m_\text{b}g - m_\text{d}g = (L/2x - 1)m_\text{b}g + (L/x - 1)m_\text{d}g \]
Note that this too goes to \( \infty \) when \( x \) goes to zero. If your body is going to work, the tendons from your bicep must be attached some small distance from your elbow.
Standing on tiptoe

Here’s another very common situation we can understand with statics principles: standing on tiptoe. How do the forces work for this:

We’ll start with a few simplifying assumptions. First, ignore the weight of the foot itself. This is probably OK because your body weighs so much more than your foot. Second, assume all three forces shown act straight up and down. In reality the forces applied by your shin ($F_{shin}$) and your Achilles tendon ($F_{Ach}$) don’t act quite straight up and down, but they nearly do.

Let’s sum the forces:

\[ \Sigma F_x = 0 \]
\[ \Sigma F_y = F_{Ach} + F_{toe} - F_{shin} = 0 \]

And sum the torques around the point where the shin bone contacts the foot:

\[ \Sigma \tau_{around \ shin \ contact} = F_{toe} L_{toe} \sin \theta - F_{Ach} L_{ach} \sin \theta = 0 \]

These two (useful) equations contain three unknowns. So we can’t solve them with more information. What information is there? We know that the floor must support your total weight. If we assume you’re standing on tiptoe with both feet, then each must support half your weight, and we know that:

\[ F_{toe} = \frac{m_{you} g}{2} \]

Using this additional fact we can find:

\[ F_{Ach} = \left( \frac{L_{toe}}{L_{ach}} \right) \times F_{toe} = \left( \frac{L_{toe}}{L_{ach}} \right) \times \left( \frac{m_{you} g}{2} \right) \]

And from this we find

\[ F_{shin} = \left( \frac{m_{you} g}{2} \right) \times \left( 1 + \frac{L_{toe}}{L_{ach}} \right) \]

There are some interesting things to note. First, the force applied by the Achilles tendon is independent of the angle of your foot. This is a little surprising, but if you try it out you’ll see it’s approximately true. It is nearly as easy to stand on tiptoe no matter what angle your foot is at, and you can move up and down with very little additional effort.
Second, the force applied by the Achilles tendon is larger than half your weight, usually by quite a large amount. Just holding a ruler up to my foot I find $L_{\text{ach}} \approx 5\,\text{cm}$ and $L_{\text{toe}} \approx 20\,\text{cm}$. So the upward force applied by each of your Achilles tendons is about four times as large as half your weight, or twice your total weight!

This is why your calf muscles are probably substantially larger than your biceps. You rarely lift twice your weight with your biceps, but you do it all the time with your calves, with every step you take.
Physics for the Life Sciences: Fall 2008 Lecture #5

Last time we learned about the two conditions for equilibrium. At least in some cases, these rules allow us to understand how organisms stand up to the pull of gravity. Today we will see that these two rules are, almost always, not enough to really figure out what’s happening. To really understand organisms and other structures, we need to understand how the materials of which they are made apply forces. How can a bone, or a chair, or a string, apply a force? Answer that and you have the crucial piece for understanding equilibrium.

The problem with our two equilibrium conditions:

What’s the problem with our equilibrium conditions? The condition for equilibrium requires: 
\[ \Sigma F = 0 \quad \text{and} \quad \Sigma \tau = 0 \text{ (around every center)} \]

We have noted these conditions before, when we first started to consider rotational motion. Note that since each of these is a vector equation, there are really six different equations here. These include the forces in each of the three directions, and torques around each of three perpendicular axes. Remember the equilibrium problems we did:

In each of these cases, we looked at the sum of the forces in each direction, and at the sum of the torques around any center, and used the constraints of equilibrium to determine the required sizes of unknown forces.

Let’s try a new example to remind you how this works. Consider a ladder resting on a floor which exerts a frictional force on it but which leans on a frictionless wall:
How large is the frictional force applied by the floor? Use the constraints:

\[ \Sigma F_x = F_N - F_f = 0 \quad \Sigma F_y = F_{NF} - m_Lg = 0 \]

or

\[ F_{NF} = m_Lg \]

What about the torques? Let’s calculate the sum of the torques around the point at the top. Defining counterclockwise rotation as positive, I have:

\[ \Sigma \tau = F_{NF}L\cos\theta - m_Lg(L/2)\cos\theta - F_fL\sin\theta = 0 \]

or

\[ m_Lg(1/2)\cos\theta - F_f\sin\theta = 0 \quad \text{or} \quad F_f = m_Lg/(2\tan\theta) \]

As the angle gets smaller, you’re going to need a larger and larger amount of friction to prevent the ladder from slipping. Your intuition will tell you that there’s a limit here. If you try to stand the ladder up at too steep an angle it will slip. No big surprise in this.

Now let’s consider a just slightly more complicated case. The same situation, but now imagine that there is some friction with the vertical wall. Of course there would be in any real case.

How are the equations changed?

\[ \Sigma F_x = F_N - F_f = 0 \quad \Sigma F_y = F_{NF} + F_{FW} - m_Lg = 0 \]

and still

\[ \Sigma \tau_{top} = F_{NF}L\cos\theta - m_Lg(L/2)\cos\theta - F_fL\sin\theta = 0 \]

But now I can no longer identify \( F_{NF} \) with \( m_Lg \), because part of the weight may be supported by friction with the wall! What’s going on here? I have three equations to work with, and four unknowns:

\( F_{NF}, F_f, F_{NW} \) and \( F_{FW} \)
This problem cannot be solved without more information. It has many solutions, and nothing we have done will tell us which of the many will be correct.

We can illustrate this ambiguity by considering two possibilities. If I take the ladder and cram it down into the corner, there will be a large frictional force at the top of the ladder pushing it up. This will reduce the frictional force pushing it in on the bottom. If, on the other hand, I pull it up some on the ladder, trying to lift it away from the corner, there will be a net frictional force at the top preventing it from sliding up. In this case I will have to increase the inward frictional force at the bottom.

This is quite generally the case in statics problems. You have only the six equations which the equilibrium conditions provide, and in most cases this will not be enough to determine the forces in an object. Another simple example is a cow. There are four legs, each of which supplies an upward force. In this case you can use one force equation (up and down) and two torque equations (tilting side to side and front to back) to constrain things, but there are still four forces, so you’re out of luck.

A more extreme example is the suspension bridge. Here there are three equations, and potentially thousands of forces you need to know. If you’re a bridge designer and you incorrectly determine even one of these forces, the entire bridge may fail.

How can such a problem be solved? Nature solves such problems without difficulty. When you assemble a bridge like this it has particular forces in each element. When you sit on your chair, there are in fact four particular forces on the legs. What additional information is needed to predict what these will be?

The answer lies hidden in the ways objects actually supply forces. Remember, an object like a chair or a table, or the bone in your leg, can only apply a force when it bends a little, when it changes shape. An object like this will push on you only when you push on it. To understand a structure like your chair, you have to know exactly how the chair changes shape when you apply a load to it. It’s no longer enough to assume that its shape is fixed, since it actually never is.
Structures and Materials

Understanding how objects support loads is an ancient endeavor. People need to be able to build things which will stand up, and not land on their heads. Over the centuries, a tremendous amount of empirical knowledge about how structures behave was built up. This empirical knowledge is what enabled the construction of such magnificent buildings as the Cathedral at Chartres in France, the Hagia Sophia in Instanbul, and the Taj Mahal in India. Theoretical understanding of how structures support loads was developed only very much later. Some important aspects of how even common materials support loads were not understood until well into the 20th century.

A Little History: Hooke and how objects support loads

Newton (1642-1727) understood that when you hand a mass from a string its weight must be supported by a tension in the string. Newton knew what happened in a case like this, but didn’t discuss how the inanimate string could apply a force to the block. Perhaps Newton wasn’t especially interested in such practical questions.

Robert Hooke, Newton’s contemporary (1635-1703) and sometime competitor, was intensely interested in practical things. He built the first air pump (for Robert Boyle), discovered the diffraction of light and used it to promote a wave theory of light, and was the first person to explicitly note the expansion of materials when they are heated. Hooke also published a book called “Micrographia”, an enormously influential (and beautiful) series of images obtained through his microscope. In these images he revealed, among many other things, that a drop of water is alive with microorganisms, and that living
tissues are made of many tiny “cells”. In fact Hooke invented the name. In many ways, this book was a starting point for the life sciences.

You can look at some of this exceptionally beautiful book online at:

For our purposes today though, Hooke’s most important discoveries had to do with how objects support loads. His main realizations are part of what is now called Hooke’s law.

1. Objects can support loads only by yielding to them. If you push on an object, it will push back by distorting.
2. Most solids are elastic as opposed to plastic. This means that if you squash them and let them go, they spring completely back to their original shapes. Elastic objects (like rubber) do this. Plastic objects (like clay) change shape permanently when you distort them.
3. Hooke quantified all of this by noting that the amount of deformation is, in many solids, proportional to the load.

These observations are encoded in a general form as a very simple law:

\[ F = -k \Delta x \]

If you deform a solid (change its length by an amount \( \Delta x \)), it will push back with a force which resists your deformation. The solid is trying to return to its original shape. The constant “\( k \)” tells you how hard it will push for a given deformation \( \Delta x \). Each object has a different constant, which you can determine by apply a force \( F \) and seeing how far it distorts (\( \Delta x \)). If the constant \( k \) is small, the object is easy to deform; you would call it flexible. If the constant \( k \) is large, the object is hard to deform, you would call it stiff. The minus sign in this equation just means that if you stretch an object out, so that \( \Delta x \) is positive, the force the object exerts on you will be negative, in the opposite direction.

**Limitations to Hooke’s picture: it only works for individual objects**

The problem with Hooke’s law for practical purposes is that it makes no predictions about what \( k \) will be for a particular object. If I’m interested in one particular object I can measure \( k \) for that object, and then predict exactly how it will deform under a load. But this is not too practical in making a new building for example. You sort of have to build it first before you can see whether it will work. This was, in fact, how cathedrals were built in Europe. A fair number were constructed, fell down, and then were rebuilt with more extensive supports.

The basic problem is that Hooke focused on individual objects, rather than the materials of which they were made. He could tell you about an individual spring or beam, but not usually about a new one.

To give a specific example, Hooke could not answer the question: “I know the constant \( k \) for this particular beam. What will the new constant \( k' \) be if I make a beam of the same material which is twice as long?”
Focusing on materials instead of objects

What’s the key to answering this? To begin, let’s think about a toy model of what a solid object is like. Solids are built of atoms more or less locked in place by bonding with their neighbors. You can think of this as an array of atoms held together by rather stiff springs.

If I apply a force which is spread out over the top of this solid, it will have to squash all of these springs. If I apply the force to just a little spot, it will have to squash only a few springs. This is easier to do, so that same force applied to a small spot will squash things more; it will create a larger distortion.

To quantify this, consider a 1 and 2 spring model, where each individual spring has the same spring constant $k$. This would be the case, for example, for each of the little springs that connect the atoms in a solid. If I apply a force $F$ to one spring, it is compressed a distance $\Delta x = F/k$. If I put two such springs in parallel, the combined spring constant will be $2k$, and the same force $F$ will compress them only half as much $\Delta x_{2\text{ springs}} = F/2k$.

You can see from this example that what matters is the force per spring. If I measure the force per spring, all objects which are made of this material (whether big or small) will behave the same way. Now we can’t actually measure the “force per spring” unless we know exactly how far apart the atoms are. So instead, we just account for this by asking whether the force is spread out over more springs or fewer. We can do this by just measuring the area over which the force is applied. If we double the area, there will be twice as many springs, etc.

So instead of just measuring the force applied to an object, what we will care about is the force per unit area. In this application the quantity force per unit area is called the stress, and it’s defined as:

$$\text{stress} = \frac{\text{Force}}{\text{Area}} = \frac{F}{A} = \sigma$$

Here $F$ is the total force, and $A$ encodes the area (proportional to the number of atoms) this force is applied over. The symbol $\sigma$ (“sigma”) is usually used for stress. Notice that stress is really the same as the more familiar “pressure”. There’s no clear distinction between the two, though the term pressure is used more often in cases of fluids, as for air pressure or hydrostatic pressure.
Now think about the displacements in a real material. One spring again displaces by a distance $\Delta x = F/k$. Now imagine I have two springs in series, stacked on top of one another. In this case each spring will compress by a distance $\Delta x$, because the same force will be transmitted through both. The combination of two springs will compress twice as much as one. This is somewhat tricky, so think about it carefully.

This suggests that what matters for displacement is the compression per spring. Again, we can’t really do this per spring without knowing exactly how far apart the atoms are, so instead we use the thickness of the material as a way of tracking whether there are more or fewer atoms. If the material is twice as thick, there will be twice as many atoms and the material will compress twice as far.

To keep track of this, we measure not just the change in length of the material ($\Delta x$) but the fractional change in length, which is called the strain. It is defined as

$$\varepsilon = \frac{\Delta L}{L} = \text{strain}$$

Here $\Delta L$ is the total displacement, and $L$ keeps track of the number of atoms which are compressed in series.

**Stress and Strain: avoiding the obvious confusion…**

The words “stress” and “strain” are great examples of the problem with using ordinary words to describe physics concepts. In everyday language, both stress and strain mean something similar. But in physics they mean very particular, quite different things.

The “stress” is a measure of how much force per unit area is applied to an object. It has units of N/m$^2$. The “strain” is our measure of how much an object is distorted by the stress which is applied to it. Strain, measured as the ratio of the distortion $\Delta L$ to the total size $L$, is dimensionless.

You’ll have to find a way to remember, unfailingly, the difference between these two. One way is to practice the mantra: “stress causes strain…stress causes strain…stress causes strain…” until you can’t think of it any other way. If you invent some other good mnemonic for this please share it with your peers and your instructor. It’s been a problem for students throughout the ages. Maybe you can solve it.

**Stress causes strain: a material dependent version of Hooke’s Law**

For every different kind of material, there will be some relation between stress and strain. If you apply a stress $\sigma$, you will measure some amount of strain $\varepsilon$. Very often it is the case, at least for small stresses, that the stress and strain will follow a simple, linear relation. Now, although you would probably do the experiment by applying a stress and measuring a strain, people usually plot it by showing the strain as a function of the stress.
This harks back to Hooke, who wrote his law $F = -k\Delta x$. He wrote it as if the force was a function of the distortion $\Delta x$. Here we say

\[ \text{Stress} = \text{constant} \times \text{strain} \quad \text{or} \quad \sigma = Y\varepsilon \]

Note that this equation does not describe an object; it describes a material. If you make an object of a material you have studied in this way, you can predict how it will behave.

\[ \sigma \]
\[ \varepsilon \]

Remember too that this linear relation definitely does not describe the relation between stress and strain in all materials. It is true that many materials have such a linear relation, but quite a few, and especially materials which are used by living organisms, do not. We’ll talk more about this later.

The ideas of stress and strain, very important ones, were first discussed by Thomas Young, a remarkable 19th century scientist who we will encounter later coming up with the concepts of work and energy. The notion of stress and strain as important for understanding structures was later codified and developed by a series of French theoreticians like Cauchy.

What is the constant in this equation?

\[ \text{constant} = \frac{\text{stress}}{\text{strain}} = \frac{F/A}{(\Delta L/L)} \]

This constant is called the ‘modulus’ of a material. Modulus just means “little measure” and it’s called this because it is a property of the material. Since the strain is unitless, the modulus has the same units as the stress. In this case, the units are $\text{N/m}^2$.

**Young’s Modulus**

For the simple case where the stress is either tension or compression, stretching or squashing, the modulus is called the “Young’s modulus” of a material, and is usually denoted with a capital ‘E’. Its units are $\text{N/m}^2$ (the same as the stress).

Some Young’s moduli:

<table>
<thead>
<tr>
<th>Material</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tendon</td>
<td>$0.6 \times 10^9 \text{ N/m}^2$</td>
</tr>
<tr>
<td>Oak</td>
<td>$14 \times 10^9 \text{ N/m}^2$</td>
</tr>
<tr>
<td>Marble</td>
<td>$50 \times 10^9 \text{ N/m}^2$</td>
</tr>
</tbody>
</table>
Physics 135 Fall 2009: Lecture Notes

Steel 200x10^9 N/m²  
Diamond 1200x10^9 N/m²

What does this mean? Consider the simplest case of a rod in tension:

If you double L, you double ΔL
If you double A, you halve ΔL

Consider wood, with E ≈ 10^{10} N/m². If I hang a 1000N weight on a 1 cm x 1 cm x 10 cm piece, I will get a stretching:

ΔL = (F/AE) * L = 1000N / (0.01*0.01*10^{10}) * 0.1 = 10^{-4} m = 0.1 mm

A 1000 N weight is something with a mass of about 100 kg, a bit heavier than most of you. So even if you stand on a piece of wood only 1cm x 1cm square and 10 cm long, you will compress it only 0.1mm.

In general the very large Young’s moduli of solids imply that it is quite difficult to either stretch or compress a solid. It’s this property that human engineers like, and why they choose such “stiff” materials for constructing things. An engineer would like to know that a certain part will always stay the same size (at least very nearly) regardless of what you do to it. As we will see in a bit, most biological materials are quite different from this, and have many attractive features that our engineering materials often lack.

**Stress, strain, and the scaling of living things**

Now that we have recognized the importance of stress for understanding structures, we can resolve an enormous range of interesting questions about life. One of the most obvious differences between large organisms (think dogs to elephants) and small ones (think aphids to birds) is sturdiness. A rhino is a solid thing, with thick legs held straight beneath its body. If I showed you a picture of such a beast, even without any scale information, you’d know this was a massive creature. A spider, by contrast, has absurdly long spindly legs, stretching way out to the sides. You never see such ungainly structures in creatures our size.

The reason for this difference lies in scaling laws. Imagine a simple rhino model with four straight legs. As we argued in the first lecture, the volume (and therefore mass and weight) of the rhino will vary like the size of the rhino cubed. Meanwhile the area of its
legs will vary like size$^2$. This implies that the stress in the rhino’s legs, which is force per unit area, will vary like $F/A \propto \text{size}^3/\text{size}^2 = \text{size}$. If you increase the size of a rhino by ten times, the stress in its leg bones will increase by a factor of 10. Remembering that the material making up the bones stays the same, you can see how just making an organism larger becomes risky very fast.

For this reason, large animals have different shapes from small ones. They have evolved proportionately thicker legs. They have also adopted postures which tend to keep the legs straight and under the body. On the other end of the size spectrum, tiny organisms have no trouble at all keeping the stress in their limbs low. As a result, they can adopt a much wider array of shapes, with long spindly legs that allow them to walk uninterrupted over the extremely varied terrain they see at their size.

So you can see that an understanding of how materials support loads, of the importance of stress, helps us to understand a lot about the diversity of form we see in living things. Organisms don’t simply choose the shapes they take. These forms are, in a very real way, imposed on them by physical constraints. There is much more we could say on this topic, and you’ll have an outside reading$^2$ to give you a broader perspective on it.

**Other kinds of stress and strain**

There are several other kinds of stress and strain. Since they affect the “springs between the atoms” differently, they have to be accounted for differently. The first new kind of stress is called “shear”. Shear is what happens when you try to shove the top of something sideways relative to the bottom. As an example, consider laying a textbook on the table, then sliding the front cover of a textbook to the left while pushing the bottom to the right.

---

Stress = \frac{F}{\text{Area of top}} \quad \text{Strain} = \frac{\Delta x}{L}

This is called ‘shear stress’ and ‘shear strain’. Usually this would be written:

\sigma_{\text{shear}} = S \varepsilon_{\text{shear}} \quad \text{or} \quad \frac{F}{A} = S\left(\frac{\Delta x}{L}\right)

For every material there is a constant associated with the response to this stress. For shear stresses this is called the “shear modulus” and usually denoted S. Shear stress tries to make one layer of a material slide over another. Note that the shear modulus and the Young’s modulus can be very different. So to know how a material will respond, you need to know what kind of stress is applied to it.

A third kind of stress and strain is called bulk stress and strain, or “hydrostatic” stress and strain.

![Diagram of bulk stress and strain]

\text{Stress} = \frac{F}{A} \quad \text{Strain} = \frac{\Delta V}{V}

This is the kind of stress and strain encountered when the object is under pressure which squeezes in from every direction, like when it is deep under water. For this we write:

\sigma_{\text{bulk}} = B \varepsilon_{\text{bulk}} \quad \text{or} \quad \frac{F}{A} = B\left(\frac{\Delta V}{V}\right)

For every material there is a constant associated with the response to this stress. For shear stresses this is called the “bulk modulus” and usually denoted B. You can see how all three of these stress/strain relations are really the same. They’re all just expressions of the basic model of a solid as a collection of springs connecting atoms.

**Limitations to the “Hooke’s law” model of linear stress/strain relations**

Hooke’s law, the \textbf{linear} model of how stress creates strain, is an empirical, phenomenological, law. Many materials behave like this under modest stresses. We certainly expect it to break down eventually. It can’t be right when the stress becomes large enough to break the object. Likewise, we know that there are materials with more exotic behaviors. Let’s talk about these limitations in turn.

1. **Strength**: what happens if the stress becomes too large, and you stretch the object too much?
a. First, it stops behaving elastically, and no longer returns to its original shape when you remove the stress. For each material, this happens at some stress which is called the ‘elastic limit’ of the material. If the stress rises above this, it will become ‘plastic’ and permanently deforms. This behavior might seem familiar to you if you think about bending something made of metal, like a paper clip. Bend it a little, like when you use it normally, and it springs back to its original shape. Bend it a lot, and it stays permanently deformed.

b. Eventually the material breaks. We sometimes talk about this happening at the ‘breaking stress’ of the material. Actually, what happens when things break is much more complicated. We will look at some features of tearing and shattering after we discuss energy.

c. Note that since stress is related to strain in materials, we could talk about either the breaking stress or the breaking strain of a material. Which one we use may depend on the application. We can divide materials up in several interesting ways based on how they break:

- “Strong” material: high breaking stress (supports big loads)
- “Weak” material: low breaking stress (can’t support big loads)
- “Stiff” material: low breaking strain (can’t be stretched much before breaking, no matter how large that stress is…)
- “Flexible” material: high breaking strain (can be stretched a lot before breaking)

Here are some illustrative examples of various kinds of materials:

- Steel: stiff, and strong
- Biscuit: stiff, but weak
- Nylon: flexible and strong
- Jello: flexible and weak

2. Nonlinear, non-Hookian flexible materials: Most man-made objects are made of relatively stiff materials that have linear stress/strain relations extending to a large fraction of their breaking stress. They are mostly dry, hard, solids which behave more or less according to Hooke’s law.

By contrast, most biological structures are often made of wet, squishy stuff which doesn’t obey a linear Hooke’s law stress/strain relation. These materials are flexible, able to stretch a lot before breaking, but still very much elastic. They easily deform and then spring back to their original shape. Nonetheless, they are often quite strong, with high breaking strains.
For many of these squishy materials, like your flesh or tendons, the stress vs. strain curve takes a shape called the J curve:

What does this ‘J curve’ stress-strain relation imply?

- Small stress ($\sigma$) creates a large strain ($\varepsilon$) in the beginning
- Gradually, the additional stress required to increase the strain becomes larger and larger. The “stiffness” increases as the strain becomes larger. This might be expressed as being proportional to the slope of this curve of the stress vs. strain curve at each value of strain. Remembering that the Young’s modulus is also the slope of the stress vs. strain curve, this makes a lot of sense.
- At large strain, the material becomes very stiff indeed. To stretch it any further you have to increase the strain very substantially.

You can confirm this general kind of behavior for one biological material rather easily; your earlobe. Try tugging on it. At first, it stretches a lot even when you pull just a little. Eventually though, you reach the steep part of the J curve, and you have to pull a lot harder to get it to stretch even a little bit further.

Why does nature do this? What are the advantages?

- The easy ability to reach large strains eliminates the requirement for very fine tolerances in construction. To pick something up, I don’t need a hand of exactly the right size. A small force allows my hand to distort to fit snugly whatever I try to pick up. Try picking up a can of soda with a pair of metal tongs and you’ll see what I mean.
- These materials are robust against shattering; imagine a window which could ‘give’ and freely change shape when hit by a baseball. Such a window would be able to spread the force of impact over a large area and hence wouldn’t shatter.

It is possible that this kind of ‘organic’ design will expand its influence in human design. For example, perhaps one day cars will become squishy, protective things. Certainly flexible plastic bumpers and air bags are steps in this direction.

Steven Vogel, a Professor of Engineering at Duke University, has written a very good book about the relation between human design and evolved form. Called
“Cat’s Paws and Catapults”3, it compares fundamental elements of human and natural technology, noting that:

“Natural and human technologies differ extensively and pervasively. We build dry and stiff structures; nature mostly makes hers wet and flexible. We build of metals; nature never does. Our hinges mainly slide; hers mostly bend. We do wonders with wheels and rotary motion; nature makes fully competent boats, aircraft, and terrestrial vehicles that lack them entirely. Our engines expand or spin; hers contract or slide. We fabricate large devices directly; nature's large things are cunning proliferations of tiny components.”

If you’re interested in learning more about the differences in these two technologies, Vogel’s books on the topic are a great place to start.

---

Physics for the Life Sciences: Fall 2008 Lecture #6

Now that we have learned a bit about statics, about the balances of forces and torques needed to keep objects in place, it’s time to delve a little deeper into how some of these individual forces work. This time we will concentrate on friction, a very important player in the lives of organisms, and in all of our technology. Without it there is no way we could get around at all.

Place of friction in motion

Friction is the tendency of objects in contact to resist relative motion, to stick together. It is the dominant factor in most motion which we observe in our world. Its tendency to bring any moving body to rest is what so reinforced the Aristotalean view that "motion implies a mover". Because of the resistance which friction provides, experience suggests that a continuous force is required to keep something moving. To uncover the real principles of dynamics, Galileo had to imagine a world without friction. In such a world motion could be perpetual, and no force at all would be required to maintain it.

Constant motion is as natural a state as rest. Because of this, Newton focused our attention on forces as the cause of changes in motion. He also showed that to predict the motion of an object, all you need to do is to understand the forces which act on it. So if we wish to understand the influence of friction on the motion of bodies, we need to understand how to predict the magnitude and direction of the frictional force.

We will start by examining one particular kind of friction, the "sliding" friction which occurs when two dry, solid surfaces slip across one another. This is the force which causes a book to slow to a halt when you slide it across the table.

To understand forces like friction, we often will often seek a "force law". Such a force law will tell us how large a force on a body would be if it had a particular set of properties (such as mass, composition, surface condition etc). Establishing force laws is a basic task in physics, and the force laws which we derive tend to fall into two categories:

- **Fundamental forces**: (like gravity). These laws seem to reflect an underlying reality at a level which suggests that they are "True" with a capital T. Their behavior tends to become simpler and simpler as we look at them more closely. These laws approach the underlying basics. It turns out there are only four of these basic kinds of forces in nature, and even some of these are closely coupled.

  The four fundamental forces are:

  - Gravity: Every object with mass attracts every other. This attraction holds planets, stars and galaxies together, and keeps you on the surface of the Earth.
• Electricity & Magnetism: This combination is responsible for chemistry, and all the bonding between atoms that makes matter interesting. Every force you’ll see in this class, except gravity, ultimately arises just from E&M.

• Strong Nuclear Force: This very short range force holds atomic nuclei together, allowing for all the various elements of the periodic table.

• Weak Nuclear Force: This short range force is responsible for the radioactive decay of some atomic nuclei.

While both the strong and weak force are crucial for the existence and nature of chemical elements, they are also remote, in the sense that about all they do is help create the periodic table. You won’t see them acting more directly in your lives.

Phenomenological forces: (like the force of friction). While all forces ultimately arise due to the four fundamental forces, this is often far from clear. When forces are more complex, we describe them with “laws” that would more appropriately be called models. In them we attempt to quantify an often very complicated set of phenomena by a series of approximations. The distinguishing feature of phenomenological models is that the more closely you study the phenomena, the more complicated the law you are using to describe it becomes. This is considered evidence that the understanding you have is ad-hoc, approximate, and not fundamental.

That doesn’t mean these models are not accurate reflections of reality or that they aren’t “true”. It’s just that by acknowledging that they gloss over details, we confine them to being “true” with a lower-case t. We know for sure that there are other details hiding beneath these general principles.

To find a force law, we first try to describe the basic behavior. We try to predict correctly the approximate size and direction of forces. We try to understand approximately how these forces will change when we change the properties of the objects in question. This is called understanding the problem in the “first approximation". Then, once we have a handle on the basic behavior, we look at things at the next level of detail (in the second approximation), and so on. Phenomenological laws are never perfect, but they can be enormously useful, and in complicated cases like friction they are absolutely necessary.

It’s worth noting here that the structure and behavior of living systems is often extraordinarily complex. As a result, such systems almost always require description with these kinds of phenomenological approaches. Quantitative models of biological systems almost always begin simple and gradually add complexity, and accuracy, in this way. This is called “mathematical modeling”, and these days it’s a very important part of life science research. Research in these areas is very hot, and is reflected at Michigan in programs like Mathematical Biology in the Math Department, the Biostatistics program in the School of Public Health, and the Complex Systems Program in Physics.
How does friction act?

Friction always acts so as to resist relative motion between two surfaces which are in contact. Let's consider two examples to see what this means. First the simplest: imagine I slide a block over the table. I push it for a bit and then let it slide to a halt. While it is sliding to a halt the friction between the surfaces will generate the force which acts to prevent this motion, and which decelerates the object to rest.

In this case the frictional force acts in a direction opposite to the direction of motion. This is kind of frictional effect that brings objects to rest, and which led Aristotle to conclude that the natural state of objects was to be at rest.

Now let's consider another case. A heavy block sits at rest on a surface. I touch it on the side with my finger and apply a force; but the block doesn’t move. It remains at rest. Why?

Since we see the block remain at rest, some force must balance the push I applied, keeping the net force on it zero. This balancing force, which appears when it’s needed to oppose my force, is also a frictional force.

For reasons which we will describe in a minute, we think of these two examples as involving two different kinds of friction. The first case, in which the two objects are already in motion relative to one another, is called "kinetic" friction, because it refers to objects which are in motion. Once they are in motion, this kind of friction acts to oppose their relative motion. The second kind of friction, acting before the objects begin to slip, is called "static" friction. Static friction acts to prevent objects from beginning to slide over one another.
Two tendencies of sliding friction:

The size of sliding frictional forces can be described reasonably well by two laws first uncovered by Leonardo da Vinci. Leonardo was an eclectic genius, interested in both the practical and esthetic arts. He studied friction for a typically complex reason.

During his time, it was thought that the planets were held in their orbits by a set of solid concentric spheres, each of which rotated at a different rate. From classical times it was thought that the true scholar, understanding the beauty of this arrangement, would hear the "music of the spheres", a kind of divine symphony. Leonardo thought perhaps this music was generated by friction between these spheres as they moved relative to one another. His attempts to understand worldly friction were driven, in part, by his desire to experience this divine music.

These laws are basic "rules" about the forces which resist the sliding motion of two solid bodies over one another. The rules are simple:

1. The force of friction resisting the relative motion of two bodies is directly proportional to the normal force between the two bodies.
2. The force of friction resisting the relative motion of two bodies is independent of their area of contact and the rate of motion.

The first assertion sounds reasonable; the harder you squeeze two objects together, the more difficult it is to slide them over one another. The second is somewhat surprising; we will see a bit later why this is the case. Remember that these two rules are nothing like absolute laws; they're just general principles that capture the main features of friction in some cases, especially for two dry flat surfaces.

The "coefficient of friction"

These two rules for friction allow us to construct a mathematical model for how large the friction force is. We’re hoping to write an equation which says:

$$F_{\text{friction}} = ?$$

The first rule says that $F_{\text{friction}}$ is directly proportional to the size of the normal force between the two objects. The second rule says that $F_{\text{friction}}$ does not depend on the area of contact or on the rate of motion. So we know the area of contact and rate of motion won’t appear in the equation. This allows us to write:

$$F_{\text{friction}} \propto F_N$$

We know $F_{\text{friction}}$ is proportional to $F_N$, but what determines the proportionality constant? The magnitude of the frictional force also depends in on the properties of each of the surfaces. For surfaces made of any pair of materials, like steel and wood, a single proportionality constant is usually enough:

$$F_{\text{friction}} = \mu F_N$$

Where $F_{\text{friction}}$ is the force of friction between two bodies, $F_N$ is the normal force between the two bodies, and $\mu$ is the "coefficient of friction".
This coefficient of friction depends mostly just on the composition of the two surfaces which are in contact. Here are a few examples:

- **Brick on wood:** $\mu \approx 0.5$
- **Ice on Ice:** $\mu \approx 0.02-0.03$
- **Copper on copper:** $\mu \approx 0.8-1.0$

This coefficient of friction is a measure of how freely surfaces of two materials stick to one another. When the coefficient is high, the surfaces stick together freely. When it is small, they stick together very little. Using the example of brick on wood, we would find:

$$F_{\text{friction}} = \mu_{\text{brick-wood}} \cdot F_N = 0.5 \cdot F_N$$

So in this case, the size of the force of friction would be about half the size of the normal force pushing the two materials together.

**Frictional forces and practical examples of motion:**

To understand how friction acts, it’s useful to start with a simple case where the motion is on a horizontal surface. Imagine you take a book on a table and push on it with a horizontal force. If the force is too small, static friction will hold the book in place. When you push very lightly, there is a small frictional force holding it in place (because it’s all that’s needed). When you push harder, the frictional force becomes larger. Eventually, your force becomes large enough, and the book will break loose and start to slide. Once this happens, you will usually find that the force required to keep it moving is less than that required to get it started. Try this yourself and you’ll see what I mean.

What is there to notice here? First, you can see the difference between static vs. kinetic friction. Static friction is a passive force, it adjusts its value to be just what it has to be to keep the object from moving. As I increase the force with which I push, the static frictional force increases just in step with me to prevent the object from moving. It will keep increasing up to some maximum point, at which the object breaks free and begins to move.

We find that, generally, this maximum static frictional force depends on the nature of the surfaces in contact, together with the normal force, in just the manner described above. That is:

$$F_{\text{f max}}^s = \mu_s F_N$$

In this equation, $\mu_s$ is the "coefficient of static friction", and $F_N$ is the normal force preventing the book from moving through the table. Remember that this equation describes the *maximum* static friction force. Because static friction is passive, the actual $F_f^s$ is *whatever it has to be* to prevent the object from sliding. Be very careful about this distinction when you attempt to understand problems involving friction.

Once we apply a force large enough to make the book begin to slide, once we overcome static friction, kinetic friction becomes the relevant force. Kinetic friction is generally...
better considered an active force than a passive one. Its magnitude is determined all the
time by the equation:
\[ F^k_f = \mu_k F_N \]
In this equation \( \mu_k \) is the "coefficient of kinetic friction" and \( F_N \) is again the normal force
between the two objects. So once an object is moving, the frictional force becomes
independent of the size of the force I push with and independent of the rate of motion.

Let's think about this for a moment. I apply a force to get an object moving. If I start out
with a small force and gradually increase it, the object will first remain at rest, as the
static frictional force gradually increases to match my push. Then, when my push exceeds
\( \mu_s F_N \), the object will begin to move. Once it is moving, regardless of the force with which
I push, the friction force which resists motion will always be
\[ F^k_f = \mu_k F_N \]
Since this frictional force is now constant, different things can happen depending on how
large the other applied forces are.

If I apply a force less than \( F^k_f \), perhaps by no longer pushing, the unbalanced frictional
force will decelerate the object and slow it down. If I apply a force larger than \( F^k_f \), my
unbalanced force will accelerate the object forward. Only if I apply a force exactly to \( F^k_f \),
no more and no less, will the object move along the surface at constant speed, because
only then will the net force along the surface be zero.

This is just what happens when you try to push something heavy, let’s say a cabinet,
across the floor. When you first push, it goes nowhere. Then you shove harder, and
eventually it breaks loose and starts to slip. Once this happens you adjust your force (not
too hard, not too weak) so that it slips along at a constant rate. The rate of motion now is
not set by how hard you push (you’re always just balancing the friction, which is
independent of how fast the couch moves) but rather by how quickly you’re prepared to
move along with it.

**Coefficients of friction, static and kinetic**

The constants in these equations, the \( \mu \) ’s, determine the size of frictional forces. Their
values depend primarily on the composition of the two surfaces which are placed in
contact. In detail, they depend on many other things, such as how the surfaces are
prepared (are they rough or smooth), or the temperature of the surfaces. But for starters
we will stick with the most important factor, what the surfaces are made of. Some
examples:

<table>
<thead>
<tr>
<th>Materials</th>
<th>( \mu_s )</th>
<th>( \mu_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steel on Steel</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Rope on wood</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Tires on dry concrete</td>
<td>1.0</td>
<td>0.75</td>
</tr>
<tr>
<td>Tires on icy concrete</td>
<td>0.3</td>
<td>0.02</td>
</tr>
<tr>
<td>Teflon on teflon</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Copyright: Regents of the University of Michigan, 2009
A larger table can be found here:
http://www.roymech.co.uk/Useful_Tables/Tribology/co_of_fric.htm#coef
and in many other places online.

There are several things to note. First, the static coefficient $\mu_s$ it larger than the kinetic coefficient $\mu_k$ for almost every set of materials. It is harder to start something sliding than to keep it moving. If you have ever tried to move furniture you will have experienced this. You push hard trying to get something sliding, then once it "breaks free", it slips along more easily.

The one exception to this general rule shown here is teflon, for which $\mu_s \approx \mu_k$. Because of this, it is as easy to start sliding motion with teflon as it is to maintain it, there is no sudden jerk as the two surfaces break free. This property makes teflon on teflon contacts very useful for artificial joints, like knee repair. Try moving some of your joints, bending your elbow for example. There is no apparent need to apply an extra large force to get the motion started. This is because your joints, unlike the dry interfaces discussed here, are nicely lubricated.

Second, there is a relatively large range of coefficients of friction, another hint that what's really going on here is very complicated. And remember, all we're talking about here is the contact between clean, dry surfaces. Imagine how many more different coefficients we would have to know if we wanted to predict the frictional forces for surfaces with varying degrees of contamination, or surfaces which are wet or otherwise lubricated.

This is what I mean when I say the laws of friction are "first approximation" and "phenomenological". They are not a wild guess, real objects do behave in roughly this way. But such phenomenological laws are only capable of giving us a ballpark idea of what's going on, and we must use them with caution. With these basics in hand, let’s consider some examples of how friction acts.

**Examples of the role of friction:**

Let's consider two typical examples of friction. The first involves the slippage of a block down a slope. If I have a block sitting on a board like this and I gradually increase the angle which the board makes with the vertical, what happens? The force with which gravity pulls the object down the slope gradually increases. For a while, the static frictional force increases in step with this, causing the body to remain in place. But at some point, the pull of gravity down the slope overcomes the maximum possible static frictional force, and the block slips.
Now let’s analyze this situation. At a particular angle $\theta$, we sum the forces along and perpendicular to the slope and find:

Before the block begins to slide we have:

$$\sum F_{\text{along}} = mgsin\theta - F_f = ma_{\text{along}} = 0 \quad \text{or} \quad F_f = mgsin\theta$$

$$\sum F_{\text{perpendicular}} = F_N - mgsin\theta = ma_{\text{perpendicular}} = 0 \quad \text{or} \quad F_N = mgsin\theta$$

But remember, there is a maximum static frictional force, it cannot be larger than $\mu_sN$. From this information we can calculate the angle at which it will start to slip. This will happen when:

$$F_f = \mu_s F_N = mgsin\theta$$

Or

$$\mu_s(mgsin\theta) = mgsin\theta \quad \text{or} \quad \mu_s = tan\theta$$

Notice what’s going on here. There are two effects. First the frictional force required to hold it on the slope is increasing as we increase the angle $\theta$. Second, the maximum available static friction is decreasing as we increase $\theta$, because the normal force between the surfaces is decreasing. Both these facts tend to make objects slip down slopes more easily.

This application of the laws of friction tells us that for a particular coefficient of friction, there exists some maximum slope beyond which the object will slip. Surely this is familiar to you from standing on slopes. If the slope is too steep, you slip downward. You probably have also experienced the important dependence of this critical slope on the nature of the two materials. Stand on a slope in sneakers and you can avoid slipping to quite steep angles. Do it in dress shoes (or on ice!) and you can avoid slipping to quite steep angles. Do it in dress shoes (or on ice!) and the slope you can manage is much less steep.

This result also suggests that the critical angle at which things will slip is independent of the nature of the object on the slope. If you stand on a slippery slope with a two year old, you’ll both start to slip at the same angle. This might be somewhat surprising, but it’s a clear prediction of this result.

Going back to the original example, once static friction is overcome and the block starts slipping, then the friction becomes purely kinetic, and you know exactly how large it is:

$$\sum F_{\text{along}} = mgsin\theta - F_f = ma_{\text{along}} = \frac{dp_{\text{along}}}{dt} \neq 0$$

$$\sum F_{\text{perpendicular}} = F_N - mgsin\theta = ma_{\text{perpendicular}} = 0 \quad \text{or} \quad F_N = mgsin\theta$$
Now we solve the motion along the plane using:
\[ F_f = \mu_k F_N = \mu_k mg \cos \theta \]
So that:
\[ mgsin \theta - \mu_k mg \cos \theta = \frac{dp_{along}}{dt} \]
or:
\[ \frac{dp_{along}}{dt} = mg(sin \theta - \mu_k \cos \theta) \]
This relation tells us the rate of change of momentum of the block with time.

Here is another example for you to think about:

When you push a book against the wall with a force like this, will it slip up the wall, down the wall, or remain in place? The only way to be sure is to work this out. Since we don’t (until we work it out) know what’s going to happen, let’s imagine that the friction acts upward. If we find it is different, the answer we get for it will contain a minus sign.

First we sum the forces along and perpendicular to the wall:
\[ \sum F_y = F \cos \theta + F_f - mg = ma_y = 0 \text{ or } F_f = mg - F \cos \theta \]
\[ \sum F_x = F \sin \theta - F_N = ma_x = 0 \text{ or } F_N = F \sin \theta \]

Now we can apply what we know about static friction to write:
\[ F_{\text{max}} = \mu_s F_N = \mu_s F \sin \theta \]
Let’s think about the limits. If F is small, then friction must help to support the book, to keep it from sliding down. This is the case we guessed and drew in the diagram above.

There is some limit, some minimum force F for which the book will not slip down the wall. This occurs when:
\[ F_f = \mu_s F \sin \theta = mg - F \cos \theta \]
Or when
\[ F = \frac{mg}{\mu_s \sin \theta + \cos \theta} \]

There is also a point at which the frictional force required is zero. This happens when:
\[ F_f = mg - F \cos \theta = 0 \text{ or } F = mg \cos \theta \]
Finally, as we push up harder and harder, we are attempting to push the book up the wall. In this case the friction acts downward, to oppose the slippage of the book. In this case:
\[ F \cos \theta > mg \quad \text{so} \quad F_r = mg - F \cos \theta \quad \text{is negative, pointing down} \]

There is another limit here. If we push too hard, friction cannot resist our pushing, and the book will begin to slip upward. This occurs when
\[ \sum F_y = F \cos \theta - F_r - mg = ma_y = 0 \quad \text{or} \quad F_r = F \cos \theta - mg \]
\[ F_r = \mu_s F \sin \theta = F \cos \theta - mg \quad \text{or} \quad F = mg / [\cos \theta - \mu_s \sin \theta] \]

Notice what I did with the signs! We have postulated the situation where the friction is holding the book down, rather than up. So I have redrawn it with the force down, and added forces appropriately.

Now this is a subtle, interesting relation. It says that the force required to make the book start slipping upward is given by:
\[ F = mg / [\cos \theta - \mu_s \sin \theta] \]

This tells us some interesting things. What force is required if the denominator of this expression is zero? In that case an infinite force would be required. This happens when:
\[ \cos \theta - \mu_s \sin \theta = 0 \quad \text{or when} \quad \mu_s = \cot \theta \]

So for an angle \( \theta \) of 60°, it would be impossible to make the book slide up the wall if the coefficient of friction is more than \( \mu_s = 0.58 \), and if the angle \( \theta \) is 85°, the required \( \mu_s = 0.08 \).

What's interesting about this is that it isn't completely obvious. Without this careful analysis we might never have guessed that it is impossible to make an object slide up the wall under certain circumstances, and we almost certainly would not have guessed that the limiting condition would be independent of the mass of the book we're trying to slide.

**Origins of friction: adhesion and surface roughness**

The fundamental origin of this rich, important, and complicated force is of great intellectual and practical interest, and I will give you the briefest introduction to it here. To understand friction, you have to begin with an idea of what the surfaces of ordinary objects look like. Are typical objects smooth or rough? It depends on how you look at them.

Is the Earth smooth? You might not say so. After all, the largest features on the Earth are about 9 km high. But the radius of the Earth is about 6.4x10⁶ m, so the Earth is smooth to about 0.1% of its radius. This is far more smooth, in this fractional sense, than a typical marble.
But when we look at objects on an atomic scale, they are very rough, with peaks and valleys that are typically many 100s of atoms tall. We can do this now (look at surfaces on the atomic scale), especially using instruments like Scanning-Tunneling Microscopes and Atomic Force Microscopes. What we find when we image typical surfaces is something that looks like the Alps.

So when we put two surfaces in contact, it's like taking the Alps in Switzerland and turning them over on top of the Alps of Bavaria. In such a case, the actual microscopic area of contact is a tiny fraction of the total area of the two objects. At these points where the objects do meet, the atoms actually bond, and the materials stick to each other (or "adhere"). This happens through the same kind of chemical bonds which hold the object themselves together. This "sticking" of these points, is what we perceive as friction.

With this picture in mind, we can start to understand da Vinci’s two rules and some other features of friction.

**Stick-and-slip friction, \( \mu_s \), and \( \mu_k \)**

The first thing we want explore is why \( \mu_s \) and \( \mu_k \) are different, and why \( \mu_s \) is typically larger. When I try to move a stationary object I gradually stretch these “merged mountaintops”, causing them to distort and resist the force which I apply. Once the object breaks free and starts to actually move, what happens goes something like this:

- The distorted object breaks free at the surface, releasing the stretched points of contact which spring forward until the "catch up" with the bulk of the object
- Then the two surfaces are again essentially at rest. New points of contact bond, generating a new source of frictional force, grabbing ahold of the surface below.
- As we continue to pull, these new contacts are stretched, until finally they break loose, allowing the contact points to "jump" again.

This cycle, which is known as "stick-and-slip" friction, is the reason why static and kinetic friction are related in the way they are. Kinetic friction is really a bunch of repeated applications of static friction. In each step the static frictional force builds up from zero to it’s maximum value, then breaks free and starts again. Each time you break free the bonds between these two surfaces, the top material jumps forward. After this, you have to build up the force to overcome static friction in this new spot. This makes the average value of kinetic friction somewhat less than the maximum for static friction.

If you want to feel this stick-and-slip friction in action, try putting the eraser of your pencil down on the desk, while you push down fairly hard. As you drag it across the table, it will perform just this kind of stick-slip motion. The horrible shriek of chalk on a blackboard is also just this stick-slip happening, now at a high frequency, so it generates a high pitched noise.
How large can $\mu_s$ be?

There is, in principle, no limit to the size of $\mu_s$. It often seems that since $F_f = \mu F_N$, that the frictional force can never be larger than $F_N$. But this is not the case. Friction depends on adhesion, so it is possible for the friction between objects to be much larger than $F_N$. This is just what we use glue to do, to make the adhesion large enough to dramatically increase the maximum static friction. Then objects will not slip over one another.

The laws of friction and independence of area

Ultimately all friction is caused by this bonding between atoms, and all such bonds are ultimately due to electromagnetic interactions between the atoms. The same interactions that hold matter together create the stickiness which underlies friction. How large this effect is depends on both the materials you use and the nature of the surfaces (polished so that many atoms come into contact, or Alps-on-Alps so that very few come into contact). But since most surfaces are actually quite rough, the simple laws da Vinci first discovered give a pretty accurate estimate of what will happen.

Probably the most surprising thing about sliding friction is that it is independent of contact area. How can this be, particularly in light of the fact that friction is really due to adhesion? The trick to this, which was not understood until at least the 1950s, is that friction does depend on the area of contact, but not on the apparent area of contact. It depends instead on the tiny bit of contact at the tops of those mountains, what we might write $A_{\text{contact microscopic}}$. With this knowledge we can write

$$F_f \propto A_{\text{contact microscopic}}$$

And for most solids, which are very rough this microscopic area of contact depends on how hard you squeeze the two surfaces together:

$$A_{\text{contact microscopic}} \propto F_N$$

Hence

$$F_f \propto F_N$$

This is a great example of how going deeper, looking beneath the first level of phenomenological laws, can be revealing. Understanding the reason why friction is independent of contact area makes it possible to better appreciate the limitations of this general rule, as we will see in a moment.

Breaking the rules of friction

As I have several times said, these rules are basic, and generally work quite well for dry, solid surfaces. They don't work for a lot of practically important cases, especially for ones involving biological materials. A nice example is your fingertips (Need to do a little experiment on friction with fingertips!)

Another area in which substantial violations of these rules occurs is with materials that are neither dry, nor rigid. Many biological cases are like this, in your joints for example. In these lubricated cases, frictional forces can be very different from what is described.
here, depending on factors like temperature and velocity, instead of just on the nature of the surfaces and the normal force.

A good technological example, drawn from biology, is rubber shoes and tires. Rubber is a substance which distorts pretty easily, following the typical biological J-curve stress vs. strain relation. The fact that it can distort so easily means that I can make the microscopic area of contact between the rubber and a floor very large without pushing down on it too hard. This large area of microscopic contact means a lot of adhesion, which in turn means large friction.

You can see the efficacy of this as you jump across the floor. Your rubber shoes can provide the relatively enormous frictional force required to stop you. If instead you place a sheet of paper on the floor, and jump onto it, your feet will slide and you’ll land on your can. You may have experienced this with dress shoes.

We can see how this works with a simple model. If we approximate this effect by saying that:

\[ A_{\text{contact micro}} \propto F_N^2 \]

We would expect to find:

\[ F_f \propto F_N^2 \]

Which is approximately what we see.

Other violations of these general rules come from surfaces which are unlike the "typical" surfaces considered here in other ways. It is possible to make extremely smooth and clean surfaces. This is often done in machining to make something called "gauge blocks" out of metal. If I clean to of these gauge blocks carefully and put them together, they will form such a large area of contact that they will essentially merge into one block in a process known as "cold welding". All the atoms from one surface bond with all the atoms from the other, and the surfaces essentially disappear.

### Rolling friction

Another practically important example is rolling friction. Why is it that a wheel can roll along for so long, apparently unaffected by friction, while a box that you slide so rapidly skids to a halt? The very small friction associated with a wheel comes about because, perhaps surprisingly, the point of contact between the wheel and the ground does not move. Since there is no relative motion between the wheel and the ground, there is no sliding friction. How can this be? Consider the picture below:
In this picture, we see a rolling wheel rolling to the right at three different moments. As it rolls, the spot on the wheel to which the arrow points is set down on the ground, then lifted off again. While it is in contact with the ground, that spot on the wheel is not moving relative to the ground at all. This is why the friction between the wheel and the ground is so small, and consequently why the invention of the wheel is such a big deal. Wheels, balls, and other rolling objects move across other surfaces with extremely little friction. Taking advantage of this makes it much easier and cheaper to move things around. Imagine how difficult it would be to get a car from here to Cleveland if you had to overcome sliding friction the whole way!

**Motion through a fluid (like air or water)**

When an object moves through a fluid (like air or water) it also experiences a force of friction. This is familiar enough if you imagine sticking your hand out the window of a moving car. Why is there friction? It turns out there are two sources of this friction.

The first is the fact that fluids have mass. As your hand moves through the air, it has to exert a force on the air to "move it out of the way". When it exerts this force on the air, the air exerts an equal and opposite force on the hand. This kind of fluid friction, dependent mostly on the density of the fluid, is important for large things moving fast, and is totally dominant for something like a skydiver falling through the air.

The second side of fluid friction is internal to the fluid, and relates to how difficult it is for the fluid to flow. This effect is small in air, larger in water, and very important in something like syrup. This second kind of fluid friction is most important for small things moving slowly, and is totally dominant for something like a cell swimming through water.

What does this friction depend on? Well just as with sliding friction, fluid friction is very complicated, but for common circumstances, where the objects moving are "relatively large" and move "relatively fast", the force of this friction depends on the size of the object, its speed, and the density of the fluid, in the following way:

\[ F_{\text{fluid}} \approx \frac{1}{2} C \rho A v^2 \]

Where \( \rho \) is the density of the fluid (in \( \text{kg/m}^3 \)), \( A \) is the "cross-sectional" area of the object, and \( v \) is the speed of the object relative to the fluid. The remaining term \( C \) is what's called the "drag coefficient" and its value depends on the object shape and the properties of the fluid.

So fluid friction increases linearly with area of the object (more air must be moved out of the way), linearly with the density of the fluid (more mass must be moved out of the way) and as the square of the speed of the object. The faster an object goes, the larger the force which resists it. Let’s think about how fluid friction affects motion.
Consider what happens in free fall:

\[ F_f = \frac{1}{2} C \rho A v^2 \]

\[ F_w = mg \]

What happens to an object like this?

Now if we sum the forces in the y direction:

\[ \sum F_y = F_f - F_w = \frac{1}{2} C \rho A v^2 - mg = \frac{dp}{dt} \]

Initially \( F_f = 0 \), and the object begins to gain momentum downward. Then, as it picks up speed, the frictional force increases, until eventually it reaches a speed where the frictional force just balances the weight. Then the net force on the object is zero, and it’s momentum stops changing. When this happens, \( \frac{dp}{dt}=0 \), so:

\[ F_f - F_w = 0 = \frac{1}{2} C \rho A v^2 - mg \quad \text{or} \quad v = \sqrt{\frac{2mg}{C \rho A}} \]

The speed where this happens is called "terminal velocity". We’ll have quite a bit more to say about cases like this in a few lectures.
Today’s class will be a bit of a pause. For a while now we have been examining forces and how they play against one another in cases of equilibrium (either rest or uniform motion). Today we’ll begin to discuss how motions change, and what happens when forces acting on objects are not balanced. To do this, we must first focus a bit on refining our description of motion.

We’ll spend a good bit of today on what is called “kinematics”; the pure description of motion. We will see that if we know the full path of an object, its position at each instant of time, we know everything about the motion. To understand that motion, we will need to speak of the velocity of the object (how rapidly its position is changing), and its acceleration (how rapidly the velocity is changing). For starters, we’ll talk just about motion along a straight line. Later we will see that motion in two and three dimensions is a rather straightforward extension of one dimensional motion.

Let’s discuss a simple example first, just to get a sense of where we’re headed. Picture in your mind a sprinter prepared to run the 100m dash. Before the start she is still on the starting blocks. During this time her position remains the same from instant to instant, her speed is zero, and since her speed is not changing her acceleration is zero. Then the gun fires, she bursts forward from the blocks, heading quickly for her top speed. During this period, her position changes from instant to instant. In addition, her speed changes from instant to instant, becoming larger and larger. This changing speed means that she is accelerating as well.

After just a few seconds, our sprinter is going full out, running at absolutely her top speed. During this period, her position continues to change from instant to instant, but her speed does not. Since her speed is not changing her acceleration is now zero.

After bursting through the finish tape, our sprinter cruises to a stop. During this period, her position continues to increase, always moving farther from the starting blocks. Now her speed is gradually decreasing, and this changing speed implies an acceleration.

Finally, she stops, hugging her coach in victory. Now her position remains the same from instant to instant, her speed is zero, and since her speed does not change, her acceleration is zero as well. In today’s lecture we will develop in more detail the tools we need to describe this motion fully. We will speak of position, velocity, and acceleration, each determined instant-by-instant along the path.

**Position, and intervals of distance:**

The motion of real, extended objects is complex. You can see this by considering what we mean by a cloud traveling at 10 mph, or a horse racing at 35 mph. When we talk informally about this our meaning is clear (the cloud is ‘on average’ moving along at 10

Copyright: Regents of the University of Michigan, 2009
mph), and to begin we will simplify complicated motions in this way. We will begin by ignoring any internal structure in an object and treat it as a "point object". If we do this we can talk about a complicated motion (like a horse racing down the track with its legs churning along and its rider bobbing up and down) in a simple way.

To describe a motion we begin by setting up a reference frame, a standard scale against which to measure.

\[
\begin{array}{cccccccc}
-3 & -2 & -1 & 0 & 1 & 2 & 3 \\
\end{array}
\]

To describe motion we simply record the position of the object at every instant. We'll use this kind of labeling scheme.

\[
s = \text{position at a particular instant} \\
s_1 = \text{position at the instant } t_1 \\
\Delta s_{21} = s_2 - s_1 = \text{interval between the two instants } t_1 \text{ and } t_2 = \text{displacement}
\]

Notice the notation, the symbol \( \Delta \) (the Greek letter "delta") is used to denote a change in a quantity.

Q: Is \( s_1 \) a distance? No, it's just a location. We merely label it with its distance from the origin.

Q: What do the signs of \( \Delta s \) mean?

- What if \( \Delta s \) is positive? The object is moving to the right.
- What if \( \Delta s \) is negative? The object is moving to the left.
- Can you have \( s_2 < 0, s_1 < 0, \text{ and } \Delta s > 0? \) Yes. You should think of an example.
- Does \( \Delta s = 0 \) imply that no distance has been traveled at all during this period?

**Instants and Intervals of Time:**

Time too must be measured. How do we do this? We measure time by comparison to something which happens “regularly”. What do we compare to? Over the years, many steady timekeepers have been used. The oldest are astronomical, including the rotation of the Earth and its orbit around the Sun. These allow us to mark off days and keep track of the years. To measure shorter periods of time requires something which repeats more often. For this purpose, many different tools have been used, including the pulse, water clocks, pendula, and more recently the very regular, rapid oscillations of quartz crystals and atoms.

In a manner very similar to the way we described positions and intervals of distance, we also talk about instants and intervals of time:

\[
t_1 = \text{time of a particular instant when something happens (an “event”)}
\]

Copyright: Regents of the University of Michigan, 2009
\( t_2 = \) time of a second instant when something happens
\[
\Delta t = t_2 - t_1 = \text{interval of time between two events}
\]
Like a position, an instant is a "location" in time. It's really just a marker, without units. Only an interval, the time between two instants, has units of seconds.

**Motion: s vs. t "histories"**

Since each position \( s_i \) corresponds to a particular instant \( t_i \), we can represent the series of events which makes up the motion of an object graphically:

If the motion is "smooth" we can reliably fill in the history between these points with a curve "describing" the motion. With no extra information, other histories are possible. Particularly when motions change very fast the sampling of positions must be very dense in time to have an adequately accurate history.

What does the above picture describe? This object starts out at a position which is negative (to the left of the zero point), moves farther to the left, stops briefly, moves towards the right, passes the zero point, travels a little farther, and then stops and stays at the same place for a while.

Interpreting position time graphs properly can be very helpful. In class we will do quite a bit of practicing. Now we want to extend our discussion of motion to include the idea of speed.

**Rate of Motion:**

What do we mean by speed? Note that we’ll distinguish, somewhat loosely, between speed and velocity. By speed, we mean the magnitude of the velocity vector, and by velocity, we mean the full vector, including both magnitude and direction. Speed is a good example of something we talk about all the time without care. It clearly has something to do with how fast you go, how much distance you travel in some short period of time, but what's a "short period of time"?

For a cross country trip, it might make sense to consider the entire length of the trip. If we use this in determining speed we will learn an **average speed**. If, instead, we wish to
know the speed at an instant, we must examine the distance traveled in an infinitesimally short period. This will tell us the **instantaneous speed**.

We use the notation for spatial intervals, or displacements, and time intervals to define these:

\[ v = \frac{\Delta s}{\Delta t} \]

**Average speed**: Uniform speed required to travel distance \( \Delta s \) in time \( \Delta t \)

\[ v_{av} = \frac{\Delta s}{\Delta t} \]

Average speed tells you nothing about variations in speed along the way

**Instantaneous speed**: Uniform speed at which the object would continue if all change ceased

\[ v = \lim_{\Delta t \to 0} \left( \frac{\Delta s}{\Delta t} \right) = \frac{ds}{dt} \]

Instantaneous speed tells you nothing about speed at other times

The average speed is easy to understand, it’s just distance traveled divided by time, but it fails to take into account variations in speed. In cases where there are no variations, its fine. The instantaneous speed is just the derivative of the position with respect to time, \( ds/dt \). This idea of “instantaneous rate of change” is the central idea of calculus. And of course it was invented by Newton and Leibnitz for just this purpose, to properly describe the motion of objects.

For simple cases, where the speed of the object is uniform and regular, we can use these relations to tell us things about how the object moved. For instance:

1. I have 1 hour to travel 30 km. How fast, in m/s, must I travel?

\[ 30 \text{ km} \times (1000 \text{ m/km}) = 30,000 \text{ m} \]
\[ 1 \text{ hour} \times (60 \text{ min/hr}) \times (60 \text{ s/min}) = 3600 \text{ s} \]
\[ v = \frac{\Delta s}{\Delta t} = 30,000 \text{ m} / 3600 \text{ s} = 8.3 \text{ m/s} \]

2. I can jog at about 2.2 m/s, and I continue for 3 hours, how far do I get?

\[ 3 \text{ hours} \times (3600 \text{ s/hour}) = 10,800 \text{ s} \]
\[ v = \frac{\Delta s}{\Delta t} = 10,800 \text{ s} \]

so

\[ \Delta s = v \Delta t = 2.2 \text{ m/s} \times 10,800 \text{ s} = 23,760 \text{ m} = 23.8 \text{ km} \]

How fast is fast? We will usually describe speeds in this class in units of meters per second.

A useful fact: Most of us are familiar with the mph scale because we use cars all the time. By watching the speedometer and looking at motion, you get some visceral notion of velocity. As a result, it is often helpful to note that:
Here are some example speeds, just to give you a sense of the range we might deal with:

- Sea floor spreading: $1 \times 10^{-9}$ m/s (~3 cm/yr)
- Grass growing: $2 \times 10^{-8}$ m/s
- Glacier: $3 \times 10^{-6}$ m/s
- Walking: 1.3 m/s
- Car: 25 m/s
- Sound in air: 330 m/s
- Earth's motion around the sun: $2.9 \times 10^4$ m/s
- Sun's motion around the Milky Way center: $2.2 \times 10^5$ m/s
- Approximate speed of an electron orbiting in Hydrogen: $2 \times 10^6$ m/s
- Speed of light in empty space: $2.998 \times 10^8$ m/s (~1 ft/ns)

**Speed-Time Graphs and finding them from position time graphs:**

If there are variations in speed, then to accurately describe the motion we have to consider the instantaneous speed.

Note that using the instantaneous speed means we have a measure of the speed at each instant. So now our description of motion includes a position $s_i$ at each instant $t_i$, and a speed $v_i$ at each instant $t_i$. This means we should be able to make a speed time graph, just like we have a position time graph.

What is the relation between the two graphs? The instantaneous speed $\frac{ds}{dt}$ also defines the slope of the position-time curve, "rise over run", so:

So, to create a speed time graph from a position time graph, you just have to examine the slope of the position time graph at each point in time, and put that slope on the velocity time graph. The next page shows some simple examples.
Constant slope implies constant speed, positive slope means positive speed, then zero speed.

Positive speed which increases and decreases.

Approximately constant positive speed.
So the slope of the position time graph determines the speed time graph. In particular, any "linear" position time graph corresponds to a constant speed.

**Finding distance from speed-time graphs:**

Is there a way to reverse this? Given the speed time graph, can I determine the position time graph? If I know how fast you are going at each instant, I can determine how far you went in each little period of time. Take speed in each second, multiply by 1 second to get how far you went, and add this to the total.

What does this mean on the speed-time graph?

\[ \int v(t) \, dt = \int \frac{ds(t)}{dt} \, dt \]

So, we have a pair of relations:

\[ s(t) = \int v(t) \, dt \]
\[ v(t) = \frac{ds(t)}{dt} \]
What if the speed is negative?

![Graph showing velocity over time]

In this case, the part with negative velocity means motion in the negative direction, so it adds in "negative area" to the total:
- 0-2 seconds: $-10 \text{m/s} \times 2 \text{s} = -20 \text{m}$
- 2-4 seconds: $0.5 \times -10 \text{m/s} \times 2 \text{s} = -10 \text{m}$
- 4-6 seconds: $0.5 \times 10 \text{m/s} \times 2 \text{s} = 10 \text{m}$

Total distance $= -20 \text{m} + -10 \text{m} + 10 \text{m} = -20 \text{m}$

Negative area; what does that mean? Well it isn’t really negative area, it’s just distance traveled to the left instead of to the right.

Let's consider a slightly more complicated case:

If $s(t) = 15 + 0.5t + 0.1t^2$ what is $v(t)$?

$v(t) = \frac{ds(t)}{dt} = 0.5 + 0.2t$

If $v(t) = 0.5 + 0.2t$ how far does the particle travel from $t=0$ to $t=4$ s?

$s(t) = \left[ v(t) dt = (0.5t + 0.1t^2) \right]_0^4 = (0.5 \times 4 + 0.1 \times 4^2) - (0.5 \times 0 + 0.1 \times 0^2) = 3.6 \text{m}$

Changes in speed: acceleration

We have talked about ways to relate the position, speed, and time, for motion of objects. Earlier in the course, Newton’s first law taught us that there is no real need to explain motion, but that it is necessary to explain changes in motion. To describe these changes in motion we will need to include a way to describe changes in speed.

We described the rate of change in position with time using speed:

$v = \lim_{\Delta t \to 0} (\Delta s/\Delta t) = ds/dt$

And we describe the rate of change in speed with time as acceleration:
\[ a = \lim_{\Delta t \to 0} (\frac{\Delta v}{\Delta t}) = \frac{dv}{dt} \]

What are the dimensions of this? It's a change in speed (which has units of distance/time) divided by a time, so acceleration has dimensions of (distance/time) / time = distance / time\(^2\). In the usual units acceleration is expressed in meters per second\(^2\).

We begin our consideration of acceleration by thinking about our sprinter moving in a straight line. For this discussion \(\Delta s > 0\) means displacement to the right \(\Delta s < 0\) means displacement to the left.

The sprinter begins at rest. She speeds up until she is moving at a constant speed, which she continues for a while. Then after crossing the finish, she gradually slows down and stops. How do we describe her motion? Think about this motion in terms of your car. You start on a straight road, speed up to some speed, travel at a constant speed for a while, then slow to a stop. What does your speedometer read during this time? What in your car measures the distance? Is there anything in your car which measures acceleration?

We now have a three-fold history of the motion of the object. Note that although the speed is always positive, the acceleration is both positive and negative at different times. This is an essential point: that although velocity is always along the direction of motion, acceleration can be in any direction.

**The important general rules to take away: what's really going on...**

What can we say about the relations between distance, speed, and acceleration when the acceleration is not a constant? What are the general relations that always apply? What we *always* know is the relation between the position-time, speed-time, and acceleration-time graphs. There are four parts to remember:

1. The speed is the slope of the position-time graph (\(ds/dt\))
2. The acceleration is the slope of the speed-time graph (\(dv/dt\))
3. The distance traveled is the area under the speed-time graph
4. The change in speed is the area under the acceleration-time graph

Copyright: Regents of the University of Michigan, 2009
**Constant acceleration in a straight line:**

Most motion is complex, but as usual we will begin by analyzing in detail a very simple case, motion in which the acceleration is constant. This is often described in introductory physics courses as a common situation. Is it? What happens to something which experiences a constant acceleration? Pretty quickly, it’s going really fast, and then of course it doesn’t stay around long. So in fact, motion with constant acceleration is exceptionally rare.

Why talk about it then? There are two reasons really:

1. Because it sometimes happens for a little while, at least approximately
2. Because it is easy to solve analytically (this is the main reason the example is so popular in introductory physics courses…)

Remember that we have already considered one special case of motion with constant acceleration: motion in which the acceleration is zero, so the speed never changes. Now we want to consider cases where the acceleration is not zero, but still constant.

\[ a = \frac{dv}{dt} = \text{constant} \]

In this case, with constant acceleration, there is no difference between average acceleration and instantaneous acceleration. We can write:

\[ a = \frac{dv}{dt} = \frac{\Delta v}{\Delta t} = \frac{(v_f - v_i)}{\Delta t} \]

or:

\[ v_f - v_i = a\Delta t \quad \text{or} \quad v_f = v_i + a\Delta t \]

In this simple case we can determine how the speed of an object changes completely. Taking a sample case in which the initial speed is 5 m/s, and the acceleration is 3m/s². What is the speed after 10 seconds?

\[ v_f = v_i + a\Delta t = 5 \text{ m/s} + 3 \text{ m/s}^2 \times 10 \text{s} = 5 \text{ m/s} + 30 \text{ m/s} = 35 \text{ m/s} \]

We might instead ask: if a car can accelerate at 5 m/s², how long does it take it to accelerate from a stop to 35 m/s?

\[ v_f = v_i + a\Delta t \quad \text{or} \quad (v_f - v_i) / a = (35 \text{ m/s} - 0 \text{ m/s}) / 5 \text{ m/s}^2 = 35/5 \text{ s} = 7 \text{ s} \]

So now we know how the speed changes. What about distance traveled? If I start a some speed \(v_i\), and accelerate at a constant rate to some speed \(v_f\), I can draw the speed-time curve for this as:
Now remember that because for each little period $\Delta t$, the distance traveled is $\Delta s = v(t)\Delta t$, the total distance traveled is equal to the area under the speed-time curve. This area can be determined by breaking the above into two parts: a lower rectangle which has area:

$$\text{area} = v_i \Delta t$$

and an upper triangle which has an area

$$\text{area} = \frac{1}{2}(v_f - v_i)\Delta t$$

So the total distance traveled is:

$$\Delta s = v_i \Delta t + \frac{1}{2}(v_f - v_i)\Delta t = v_i \Delta t + v_{av} \Delta t$$

This should not be too surprising. The first term in this equation is just the distance the object would have traveled if the speed were not changing. The second term represents the additional distance traveled because the speed is increasing.

It is often useful to rearrange these relations, expressing them in different ways. The relation we just wrote is very useful if you know the initial and final speeds, and you want to know how far you have traveled. If instead you know your initial speed and your acceleration, and you want to know how far you go, it is more useful to restate the above relation in terms of acceleration:

$$a = \frac{(v_f - v_i)}{\Delta t} \quad \text{or} \quad (v_f - v_i) = a \Delta t$$

this lets us replace this part of the above relation to find:

$$\Delta s = v_i \Delta t + \frac{1}{2}a(\Delta t)^2$$

Remembering that $\Delta s = s_f - s_i$, we can write this in a commonly used, and explicit form:

$$s_f = s_i + v_i \Delta t + \frac{1}{2}a(\Delta t)^2$$

**The same thing, but with calculus**

Let's determine this more directly, by using some simple calculus

$$a = \frac{dv}{dt}$$

so that
\[ \frac{dv}{dt} = adt \quad \text{or} \quad \int dv = \int adt \]

but \( a \) is constant so:

\[ \int dv = a \int dt \quad \text{or} \quad v = a(t_f - t_i) + C \]

where \( C \) is a constant of integration, something undetermined by the mathematics.

In physics, such constants must always be determined by the physical circumstances, by the "boundary conditions" of the problem. In this case, we have to know what the velocity was at time \( t = t_i \). If this is written \( v_i \) then we know that:

\[ v = a \cdot 0 + C = v_i \]

so:

\[ C = v_i \]

And we write:

\[ v_f = a(t_f - t_i) + v_i \]

In a similar way we write:

\[ \frac{ds}{dt} = v = \frac{ds}{dt} \quad \text{or} \quad ds = vdt \]

So

\[ \int ds = s = \int vdt = \int (at + v_0)dt = \frac{1}{2}a(t_f - t_i)^2 + v_i(t_f - t_i) + C \]

applying the boundary condition that \( s = s_i \) when \( t_f = t_i \) in a manner similar to the above, we have:

\[ s_f = s_i + v_i \Delta t + \frac{1}{2}a \Delta t^2 \]

Now we have two kinematic equations to work with in situations where the acceleration is constant:

\[ v_f = v_i + a \Delta t \]

And

\[ s_f = s_i + v_i \Delta t + \frac{1}{2}a \Delta t^2 \]

It is possible to combine these two to eliminate \( \Delta t \) and set up a third kinematic equation:

\[ v_f^2 - v_i^2 = 2a(s_f - s_i) \]

The derivation of this third equation is straightforward:

\[ \Delta t = \frac{(v_f - v_i)}{a} \]

\[ s_f - s_i = v_i \left( \frac{(v_f - v_i)}{a} \right) + \frac{1}{2}a(v_f - v_i)^2/a^2 \]

or

\[ 2a(s_f - s_i) = (v_f - v_i)(2v_i + (v_f - v_i)) = (v_f - v_i)(v_f + v_i) = (v_f^2 - v_i^2) \]

to finally give:

\[ v_f^2 - v_i^2 = 2a(s_f - s_i) \]

**A caution**

Very often in physics we follow a line of reasoning which begins with a simple assumption and progresses so far beyond this that we forget the assumption. What have we assumed about the acceleration to derive this set of equations? We have assumed that the acceleration is constant throughout!
Do these equations apply if the acceleration is not a constant? No. Why do we talk about the constant acceleration case? Mostly because it is easy to analytically solve, but also because it does sometimes happen, at least for a while. As a result, it can be a useful element for a model meant to describe something more complex. It’s another of our “spherical cow approximations”.
Physics for the Life Sciences: Fall 2008 Lecture #8

Force and the alteration of motion:

In the last lecture we learned how to describe motion with position time graphs, rate of motion with speed-time graphs, and changes in rate of motion with acceleration-time graphs. Now that we have the tools in place to describe motion we’re going to learn about what causes motion to change, and how rapidly that change occurs.

Recall what Newton said, an object in motion remains in motion "unless compelled to change its motion by forces impressed upon it". The law of inertia tells us that every unmolested object remains in uniform motion. But right here Newton is telling us how motion changes; motion changes from uniform because of unbalanced forces.

Mass and momentum:

OK, forces alter motion. How do they do this quantitatively? We know from experience that the way in which a force acts on an object depends on the nature of an object; it is easier to stop a running 2 year old than a 300 lb football player. The proper way to express this inertia of motion, the ease or difficulty with which something is stopped, is through momentum, which is defined as:

\[ p = mv \]

This is the mass of an object times its velocity.

What to notice about this "momentum":

- It is a vector, you need to be aware of both its magnitude and its direction. Changes in it will be vector changes. That is, they can be changes in magnitude, or direction, or both.
- Momentum depends "linearly" on the mass of the object. Double the mass and you double the momentum. Reduce the mass by 1/10th and you reduce the momentum by 1/10th.
- It depends linearly on the magnitude of the velocity, double the velocity and you double the momentum.
- It is always in the direction of the velocity (and hence always along the path of the object’s motion).
**Force and changes of momentum:**

This notion of how difficult it is to change the motion of an object is quantified in Newton's second law:

> The force exerted on a body equals the resulting change in its momentum divided by the time elapsed in the process.

Expressed as an equation, this amounts to a quantitative definition of the force vector:

\[ \mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m \mathbf{v})}{dt} \]

The force is equal to the change in momentum of an object divided by the time it takes to make that momentum change. The dimensions of force are \([M(L/T)]/T = ML/T^2\) or in the usual units \(\text{kgm/s}^2\). This unit, \(1 \text{ kgm/s}^2\), is called a Newton for the obvious reason. How large is \(1 \text{ N}\) of force? It's about equal to the weight of a modest sized apple, like the one which fell on Newton's head in the (probably apocryphal) story.

What is a force really? We don't know, but we can say a lot about how what they do. In fact the study of forces and how they behave is a very large fraction of physics, and in the end, they *are* what they do. What do I mean? We can tell there is a force present only by seeing it act on an object:

I don't really have to know what happens in the oval to describe the force.

These forces are external things; they come from *outside* the object whose momentum is changing. To alter my motion something outside me must act. I cannot literally pull myself up by my own bootstraps.

**Relativity and momentum:**

There is an important point to note here. The value of the momentum of an object is relative. I can measure the momentum of a table in two reference frames, one at rest with respect to the room and one moving across it at \(2\text{ m/s}\). I will get two very different values. This is because velocity is a relative thing. The absolute value of the momentum is not important. Note that this is related to the law of inertia; objects in motion continue in
motion; the only relevant thing is changes in motion, not the absolute amount of motion something has.

So the value of momentum is relative, but changes in momentum are not. If a train passes by the platform carrying a girl with a ball, a person on the platform will measure a different momentum for the ball than a person on the train. But if the girl throws the ball, causing its velocity to change, both the person on the platform and the girl on the train will see the velocity change by the same amount, both will see the same change in momentum, and hence both will know exactly what force acted on it. It is change that is important in physics.

**Duration of force and impulse:**

There is an alternate way of looking at Newton’s second law which is quite instructive. If I rearrange this equation as:

\[ F \Delta t = \Delta p \]

I can explicitly see that to obtain a certain change in momentum I can either use a large force for a short time, or a small force for a long time. The product of force*time is what creates momentum change. This product is called “impulse”, and in some ways is easier to understand than force.

We can turn this around productively. If I see a certain momentum change, I know how large the impulse was. Then, if I know about how long the interaction took place, I can estimate the average size of the force.

Here’s a simple example. I run to the East at 2.5 m/s. Then, rather quickly, in a half second or so, I stop, turn around, and run back to the West at 2.5 m/s. About how large is the force which must act on me to alter my motion in this way? Newton’s second law gives us what we need to find out:

\[ F = \frac{\Delta p}{\Delta t} \approx \frac{\Delta p}{\Delta t} \]

How does my momentum change?

\[ \Delta p = p_f - p_i \]

My initial velocity is 2.5 m/s East. Let’s call East the positive direction, and West negative. My mass is about 80 kg. So my initial momentum is:

\[ p_i = 80 \text{ kg} \times 2.5 \text{ m/s} = 200 \text{ kgm/s} \]

The final momentum has the same magnitude, but is in the opposite direction:

\[ p_f = 80 \text{ kg} \times -2.5 \text{ m/s} = -200 \text{ kgm/s} \]

Putting these together, we get

\[ \Delta p = p_f - p_i = -200 \text{ kgm/s} - 200 \text{ kgm/s} = -400 \text{ kgm/s} \]

In this answer, the minus sign simply means in the negative, or West, direction. Now we get the force:

\[ F = \frac{\Delta p}{\Delta t} = \frac{-400 \text{ kgm/s}}{0.5 \text{ s}} = -800 \text{ kgm/s}^2 = -800 \text{ N} \]
This just means 800 N in the West direction. It’s not surprising that to change my motion from going East to going West requires a force in the West direction. This is a pretty large force too. Remembering that 1 N is about the weight of an apple, this is the weight of 800 apples. It’s also about equal to my weight, which is 80 kg * 9.8 m/s² = 784 N.

If I’m going to run forward at 2.5 m/s (6.25 mph, a modest jog) and turn around in half a second, there has to be a force equal to my weight acting in the opposite direction for about a half a second. Where does this force come from? I push my feet against the ground, and the ground pushes back on me. It’s the force of the ground pushing on me that turns me around and sends me back the other way. Of course that frictional force only appears because I push on the ground.

**Force and acceleration:**

This view that force is coupled to changes in momentum is the fundamental way of stating Newtonian mechanics, and it is the most correct. But there is another, often useful, way of looking at force.

\[ \mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt} = m\frac{d\mathbf{v}}{dt} = ma \]

Now to derive this we have had to assert that the mass doesn't change with time, so that \( \Delta(m\mathbf{v}) = m\Delta\mathbf{v} \). Is this true? For a very long time, everyone thought so, until early in this century when the theory of relativity, and more important its experimental confirmation, demonstrated that it was not. When objects approach the speed of light, the only correct formulation is

\[ \mathbf{F} = \frac{d\mathbf{p}}{dt} \]

And it is impossible to say that the mass does not change.

I want to make it completely clear that \( \mathbf{F} = ma \) is perfectly acceptable for everything we will do in this class, but I want you to understand why it is that we emphasize \( \mathbf{F} = \frac{d\mathbf{p}}{dt} \). This conception of the second law, which was Newton's original conception, is now known to be the only correct one for all speeds.

We will work a lot from now on with this \( \mathbf{F} = ma \) formulation. It’s important, because it allows us to determine forces from observed accelerations, and accelerations from known forces. This will be the subject of the next two classes.

**Weight: the force exerted by gravity**

This second definition (\( \mathbf{F} = ma \)) lets us see clearly what weight is all about. Way back in the 17th century Galileo discovered that if you can remove air friction, all objects are accelerated at the same rate by gravity. This equation, \( \mathbf{F} = ma \), tells us that if we know the acceleration of an object, and its mass, we know the force which is acting on it. So, every object feels a force exerted on it by gravity:
\[ \mathbf{F}_g = m \mathbf{a}_g = -mg = \mathbf{W} \]

The weight of an object is just its mass multiplied by 9.8 m/s\(^2\) and it is always directed towards the center of the Earth.

A 1kg mass has a weight of \(1\text{kg} \times -9.8\text{m/s}^2\) = -9.8kgm/s\(^2\) = -9.8N. A typical person has a mass of about 70kg, and hence a weight of about \(70\text{kg} \times -9.8\text{m/s}^2\) = -690N. It may be useful to remember these approximate figures when you think about forces. 10N is the gravitational force on a 1kg weight, 700N is the gravitational force on a person.

**Free Fall:**

Galileo’s example is the one common case of approximately constant acceleration: free fall, or the unimpeded motion of an object which is dropped. The question of why things fall when dropped is as old as our ability to form it. It touches on our most basic understanding of the universe.

Early Greek theory:

Free fall is a "natural" motion. Everything moves to its natural place:

- Earth + Water => Gravity
- Air + Fire => Levity

Notice the parallel to words like sobriety and frivolity.

The idea here is that objects fall because it is in their nature to do so. The motion is intrinsic to the object.

What do we now believe this motion to be due to? Newtonian theory suggests that an object falls because of an interaction between the Earth and the object. All of Newtonian theory relies on this idea of interaction. Instead of the object falling on its own, it falls because the Earth makes it fall.

Aristotelian theory held that object fall with constant speeds proportional to their weights. This was not a crazy idea. In the presence of air friction it is true that dropped objects reach a "terminal" velocity which depends on their weight and geometry. A piece of paper falls in very different ways if it is flat or crumpled up.

Galileo was responsible for successful public refutation of this idea. To do it he used several techniques:

1. **Idealization:** As I have emphasized before, it is often very important to consider ideal, difficult to realize, cases in trying to understand motion. Galileo could see that air resistance was confusing the problem, and studied ways to remove it.
2. Experiment: How were measurements made in the 17th century? How could Galileo measure time? Would it be possible for him to measure the time it takes to drop a ball? He had to find ways to "dilute" the acceleration due to gravity.

3. Thought experiments: Consider three identical balls of clay. They will all fall at the same rate. Now stick two of them together. Will the double ball now fall twice as fast as the single ball?

What he found was that all objects, independent of their mass, were accelerated at the same rate independent of their mass. Gravity creates in any unsupported object an acceleration which is independent of the properties of the object. This is an extraordinary fact, almost bizarre. How can it be that all objects, independent of their properties, behave the same way? Egalitarianism indeed. We will return to this surprising fact several times later in the class.

The value of this acceleration which gravity produces near the surface of the Earth is 9.8m/s², and it is always directed towards the center of the Earth. We usually use the symbol g to denote this. Always remember, g has magnitude 9.8m/s², and its direction is towards the center of the Earth.

Consider the basics of throwing a ball up in the air. This is 1D motion under constant acceleration. If we use a reference frame in which up is the positive direction, then the object starts out going up (Δs > 0) with positive velocity (v_i > 0). Gravity is accelerating the object down all the time, so its velocity decreases as it goes up. Eventually the velocity goes to zero, then becomes negative as the object begins to descend. This is illustrated in the following motion graphs:
Free fall and motion under constant acceleration:

Last time we discussed general motion under constant acceleration and developed a set of equations to describe this motion:

\[ v_f = v_i + a\Delta t \]
\[ \Delta s = v_i \Delta t + \frac{1}{2}a\Delta t^2 \]
\[ v_f^2 - v_i^2 = 2a\Delta s \]

Galileo's basic observation was that in the absence of air resistance all objects, independent of their properties, are accelerated towards the center of the Earth at the same rate. The magnitude of this acceleration is 9.8m/s², and it is directed towards the center of the Earth. We can use this fact, and the equations shown above, to discuss the motion of a body moving straight up or down quantitatively.

A simple example of this motion is when I toss a ball up into the air. This motion is begins with some positive speed which is continually, regularly, reduced by the constant negative acceleration until at the top of the path, the speed becomes zero. Then as it descends it has a continually, regularly, increasing negative speed. All the while the acceleration has a constant value of 9.8m/s² downwards.

It is sometimes difficult to believe that there is still acceleration at the moment when the ball comes to a stop at the top of its path. It's perhaps worthwhile to think of the following thought experiment. Imagine you are driving your car up a hill and you slip it into neutral, coasting up the hill. You gradually slow to a stop. What happens if, at just the instant when your car stops, you abruptly put on the brake? What if, instead, you are coasting on a flat road and put on the brake at just the moment at which you stop? The fact that you experience a "jerk" when coasting up a hill is evidence that by putting on the brake you actually change your acceleration (from a constant negative value to zero) at that instant.

If I drop a ball from rest from a height \( y=10 \text{m} \), how long does it take to strike the ground at height \( 0 \text{m} \)?

\[ \Delta s = y_f - y_i = v_i \Delta t + \frac{1}{2}a\Delta t^2 \quad \text{or} \quad -10 = \frac{1}{2}(-9.8)\Delta t^2 \]

or

\[ \Delta t = \sqrt{\frac{2\times 10}{9.8}} = 1.4 \text{ s} \]

it strikes the ground with speed:

\[ v_f = v_i + a\Delta t = (-9.8)(1.4) = -14 \text{m/s} \]

If, instead, I throw the ball downward at 10m/s, how long does it take to reach the ground?

\[ \Delta s = -10 = v_i \Delta t + \frac{1}{2}a\Delta t^2 = -10 = \frac{1}{2}(-9.8)\Delta t^2 \]

or

\[ 4.9\Delta t^2 + 10 = 0 \]

Copyright: Regents of the University of Michigan, 2009
Notice that this has the familiar form:

\[ ax^2 + bx + c = 0 \]

This is a quadratic equation, which has roots given by

\[ \Delta t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \]

which in this case are:

\[ \Delta t = \frac{-10 \pm \sqrt{(10^2 - 4(4.9)(-10))}}{2(4.9)} \]
\[ = \frac{-10 \pm 17.2}{9.8} = 0.73 \text{ or } -2.8 \]

The solution we seek is the positive one, the one with \( \Delta t = 0.73 \text{s} \)

At this point the object has speed:

\[ v_f = v_i + a\Delta t = -10 \text{m/s} + (-9.8 \text{m/s}^2)\Delta t = -17.2 \text{ m/s} \]

Two points are important to make here. First, if you are uncomfortable with solving quadratic equations in this way, I again encourage you to review this math. You'll have to be able to do this over the coming months. Second, this equation, like every quadratic equation, has two different roots. In this case one is negative and one positive. The physical meaning of this is as follows:

- We asked the question: "If the motion of this object is governed only by the uniform acceleration of gravity, and if we know it passed a point \( y=10 \text{m} \), traveling downward at \( 10 \text{m/s} \) at time \( t=0 \), at what time did it pass the point \( y=0 \text{m}?""
- The answer we seek is the one which occurs \text{after} the ball is thrown

But it is perfectly possible for the object to have passed this point at an earlier time, while all the time executing motion described completely accurately by these equations. How does this happen?

These same conditions could be created by launching the ball from the ground (\( y=0 \text{m} \)) at a time equal to \( t = -2.8 \text{s} \), with a speed equal to an opposite the speed we have calculated it will hit with. So if we launch a ball upward, with speed 17.2 m/s at time \( t = -2.8 \text{s} \), it will pass the point \( y = 10 \text{m} \) at time \( t = 0 \), traveling down at -10 m/s, and strike the ground at time \( t = 0.73 \text{s} \), traveling down at -17.2m/s.

Notice that even in the first problem, in which I just dropped a ball from 10m up, there were really two solutions: we found that \( \Delta t = \sqrt{2.0} = +/- 1.4 \). We ignored the negative solution at that point, but it is mathematically allowed. In this case it just indicates that we could have thrown the ball up with a speed of 14 m/s at time \( t = -1.4 \). In either case we will have the motion from \( t = 0 \) to \( t = 1.4 \) be the same.

This emphasizes a basic symmetry in the motion of objects moving under constant acceleration. When I throw a ball up and down, there are two equal and opposite parts of the motion; a continual slowing as the ball rises, and a continual speeding up as it falls. So when I launch a ball in the air with speed +10 m/s, I know that it will return to the ground with speed -10 m/s.
Once we have an understanding of this, we can use it to look at all problems involving constant acceleration: If I launch a ball upward with speed 25 m/s, how long will it take it to come back down? What's the “easy” way to approach a problem like this? In the absence of air resistance, I know that it will reach the ground with a “final” speed of exactly -25 m/s, so I could use:

\[ v_f = v_i + a\Delta t \quad \text{or} \quad -25 \text{m/s} = 25 \text{m/s} + (-9.8 \text{m/s}^2)\Delta t \]

Or

\[ \Delta t = \frac{-50 \text{m/s}}{-9.8 \text{m/s}^2} = 5.1 \text{s} \]

How else could we figure this out?
- Use the fact that at the top of its flight \( v=0 \), and that this must be half-way through its entire flight
- Use the equation \( \Delta s = v_i\Delta t + \frac{1}{2}a\Delta t^2 \), with \( \Delta s=0 \)

Any of these methods would be correct. Often the key in physics is selecting the correct approach which admits easiest analysis of the problem. Very often the simplifying point involves taking advantage of symmetry in the problem.

**Another force calculation example**

A 200g ball is dropped from a 2m height. When it reaches the ground, it reverses its direction, bouncing off the ground and traveling back up to about where it started. Estimate the force exerted on it by the floor.

First, what is the momentum change? When it reaches the floor it has been accelerated by gravity through 2m of distance. The final velocity it has can be estimated from one of our equations for object moving under constant acceleration:

\[ v_f^2 - v_i^2 = 2a\Delta s \quad \text{or} \quad v_f^2 = 2(-9.8 \text{m/s}^2)(0 \text{m} - 2 \text{m}) = 39.2 \]

or

\[ v_f = -6.3 \text{m/s} \]

After it hits the floor it will bounce back with just the opposite velocity.

So:

\[ \Delta p = p_f - p_i = (200 \text{g})(6.3 \text{m/s})\hat{y} - (200 \text{g})(-6.3 \text{m/s})\hat{y} = (1.3 \text{kgm/s})\hat{y} + (1.3 \text{kgm/s})\hat{y} = (2.6 \text{kgm/s})\hat{y} \]

Notice that throughout we have to keep careful track of the fact that this is a vector!

OK, we have the momentum change. That's equal to the impulse, so:

\[ F\Delta t = (2.6 \text{kgm/s})\hat{y} \]

What direction is the force in? It must be up, because that's the direction of the impulse. Is this surprising? Not really, because in order to make the ball change direction from moving down to moving up, you have to push up.

How large is the force? To estimate this you have to estimate the amount of time during which the force acts. Is it 1s? Is it 1/1000 of a second? It's probably somewhere in between, say 1/20 of a second. So, we would estimate that the force is about:
\[ F = \frac{\Delta p}{\Delta t} \approx \frac{(2.6 \text{kgm/s}) \cdot y}{0.05 \text{s}} = (52 \text{kgm/s}^2) y \]

So this force is about equal to the weight of 50 medium sized apples, a pretty sizeable impact.

What if I made the impact shorter, by making the ball harder for example? How would the force have to change if the impact where shorter? What if I made it longer, by bouncing it off a trampoline for instance? How would the force required change if the impact were longer?

This is how airbags, tennis shoes, and shoulder pads work. By allowing the momentum change to take place over a longer time, these devices let smaller, less harmful forces produce the same change in momentum.

**Not so free fall: falling with air friction**

In fact, when objects fall they are not only acted on by the force of gravity. As they move through the air, there is a frictional force applied to them. When an object moves through a fluid (like air or water) it experiences a force of friction. This is familiar enough if you imagine sticking your hand out the window of a moving car. Why is there friction? As you hand moves through the air, it has to exert a force on the air to "move it out of the way". When it exerts this force on the air, the air exerts an equal and opposite force on the hand. This is why airplanes need to run their engines constantly in flight; the force exerted by the engines just balances the force exerted by the air friction.

What does this friction depend on? Well just as with sliding friction, fluid friction is very complicated. There are two useful limits to consider. Which applies depends on the details of the circumstances.

One case is where the objects moving are "relatively large", moving "relatively fast", and the fluid flows “relatively freely”. In this case the frictional force depends on the size of the object, its speed, and the density of the fluid, in something like the following way:

\[ F_{\text{fluid}} \approx \frac{1}{2} C \rho A v^2 \]

Where \( \rho \) is the density of the fluid (in kg/m\(^3\)), \( A \) is the "cross-sectional" area of the object, and \( v \) is the speed of the object relative to the fluid. The remaining term \( C \) is what's called the "drag coefficient" and its value depends on the shape of the object and the properties of the fluid. This drag coefficient is, in some sense, where we hide the details. Why is there still a factor of a half in front of this equation? We could simply absorb this in the unknown factor \( C \), but the convention is to leave it separate like this.

So this kind of fluid friction increases linearly with area of the object (more air must be moved out of the way), linearly with the density of the fluid (more mass must be moved out of the way) and as the square of the speed of the object. If an object goes twice as fast, the force which resists it becomes four times as large.
The other case of fluid friction occurs when the objects moving are "relatively small", moving "relatively slowly", through a fluid which flows "relatively poorly". This last part is expressed by something called the “viscosity” of the fluid. Viscosity is a measure of how much internal friction there is in the flow of the fluid. Viscous fluids are things like honey, less viscous materials are things which flow freely, like air.

In this case of small, slow things in sticky fluids, the force of friction depends on the size of the object, its speed, and the viscosity of the fluid, in something like the following way:

\[ F_{f,\text{fluid}} \approx -12\pi \eta Dv \]

Where \( \eta \) is the viscosity of the fluid, \( D \) is a measure of the diameter of the object, and \( v \) is the speed of the object relative to the fluid. Notice that this kind of friction contains no “drag coefficient” fudge factor. The reason is simple. In this small-slow case, the drag is almost entirely due to friction in the flow of the fluid itself, and not to the inertia of the fluid. The amount of flowing which has to happen as the object moves through the fluid is not very dependent on the object.

How does fluid friction affect motion? Consider what happens to something like a ball in free fall:

\[ F_{f} = 1/2C\rho A v^2 \]
\[ F_{w} = mg \]

Now if we sum the forces in the y direction:

\[ \sum F_y = F_f - F_w = 1/2C\rho A v^2 - mg = ma \]

Initially it’s not moving through the air, \( F_f = 0 \), and the object accelerates downward with an initial acceleration \( g \). Then, as it picks up speed, the frictional force increases, until eventually it reaches a speed where the frictional force just balances the weight. Then the net force on the object is zero, and it ceases to accelerate. When this happens, \( a = 0 \), so:

\[ F_f - F_w = 0 = 1/2C\rho A v^2 - mg \]
\[ v_{\text{terminal}} = \sqrt{(2mg/C\rho A)} \]

The speed where this happens is called "terminal velocity". Notice what this tells us. It says that if we increase the area of the object, \( v_t \) is reduced. If we increase the density \( \rho \) the \( v_t \) is reduced. But if we increase the weight, \( v_t \) is increased. More important, it tells us how rapidly these things change. Double the weight and you get \( \sqrt{2} \) times as large a \( v_t \) (at least when the friction takes this form).

There is a connection here to the scaling laws we have been emphasizing, and to the LaBarbera article in particular. If we just take an organism and scale it up in size, it’s
volume will increase like size\(^3\) (increasing the mass), while its cross-sectional area will increase like size\(^2\). As a result the terminal velocity will increase like size\(^{1/2}\). This is a factor in LaBarbera’s reminder that ‘the bigger they are, the harder they fall’, and in his assertion that King Kong would ‘splash’ onto the streets of New York.

If, instead, the object was small, traveling slowly, or moving through a more viscous fluid, the friction force law would differ in detail. But an important point is that the motion would be qualitatively identical. When you drop the object, it starts out with a downward acceleration \(g\). Then as it speeds up through the fluid, the frictional force resists its motion, gradually reducing the acceleration until eventually it reaches a stable “terminal velocity” where the frictional force balances the downward pull of gravity. Using the force law we wrote above, we’d find:

\[
12\pi \eta D v_{\text{terminal}} = mg \quad \text{or} \quad v_{\text{terminal}} = mg / 12\pi \eta D
\]

Notice how this is different from what we found above. For these small, slow things in sticky fluids, \(v_t\) depends linearly on the mass. Something with twice the mass will reach a terminal velocity twice as large.

**Scales, free fall, and life**

Life lives largely in fluids, either air or water. The interactions between living things and these fluids play a huge role in life, allowing motion, providing oxygen, and carrying sound and scent to our senses. This example of fluid friction and free fall is just the first encounter we’ll have with this very rich topic. One thing it illustrates at the outset is that the way our fluid environment affects life will be very dependent on scale. The behavior of very small things (individual cells and the world of tiny multicelled creatures) in fluids is very different from that of the world of macroscopic animals.

Just to give one example, there is a class of microscopic things living in the ocean which are called “plankton”. These are organisms so small that the frictional forces applied to them by the water they live in can easily overwhelm the downward forces gravity applies to them. Any little flow of water simply carries them with it. In practice, many organisms have evolved means to enhance this, developing shapes which increase fluid drag. This allows them to become somewhat larger while retaining their ability to take advantage of the free motion riding along with the fluid provides.

Plankton are not entirely limited to water. There are things which, for at least part of their life cycle, take advantage of fluid friction in the air to further their ends. Perhaps the most obvious are seeding strategies. Fungi and plants, for the most part, cannot move. If their
seeds are to spread they will have to be carried by something else. Quite often, plants enlist mobile animals for this purpose. But many also take advantage of the air, releasing very small seeds or spores, and often equipping them with mechanisms for enhancing the friction between the fluid and the seed (think of the dandelion, cottonwood, or the little maple seed propeller). Another quite beautiful example is the ballooning of baby spiders, which perhaps you will remember from the end of E.B. White’s “Charlotte’s Web”.

Cottonwood seeds  Maple seed  Ballooning spiders
Summing up motion in one dimension: speeding up and slowing down

Last time we considered cases in which unbalanced forces act only along or opposite the direction of motion. This includes all kinds of cases of speeding up and slowing down while traveling straight. As always, predicting what will happen in this kind of motion requires knowing what the forces are. If you know the net force on an object, you can use Newton’s second law to predict its acceleration ($F = ma$). Once you know the acceleration, you can use the kinematic relations between acceleration, velocity, and position to predict exactly how it will move. So, if you know the forces, you know what will happen.

We talked about two special cases as well. One was motion under the influence of constant acceleration. The position, velocity, and acceleration as a function of time have very simple forms for this situation. This makes it a nice model for those parts of motion in which the total force acting on an object is approximately constant. As an example of this we looked at the motion of a ball dropped or tossed up in the air. While this model is usually not precise (because the net forces are rarely exactly constant) it is a useful first approximation in many cases. This sort of ‘start with a basic model’ should bring to mind the spherical cow we used to figure out scaling relations at the start of the class.

As an example of something more realistic, we looked at the case of falling through a fluid in which friction is present. In this case, an object released from rest starts out feeling a downward force equal to its weight. This causes it to accelerate downward. As its speed increases, the frictional force resisting its downward motion gradually increases. As a result, the net downward force is gradually reduced, and the downward acceleration decreases with it. This process continues until finally the upward frictional force just balances the downward weight. At this point, the net force is zero, and the object continues to fall at a constant speed. This final, constant speed is called terminal velocity.

Fortunately, you rarely experience this precise kind of terminal velocity motion. It usually ends badly…But most of the motions you do experience are very like this. You start from rest, and a force applied to you creates an acceleration. This acceleration continues for a while, gradually becoming smaller as the forces which resist your motion build up, until at last the forces pushing you forward are just balanced by those pushing you backwards. From then on your velocity remains constant. What are some examples of this?

- Walking: you push on the floor with your foot, it pushes on you, and you start to accelerate. That forward force is pretty quickly balanced by backward forces
which come (mostly) from the landing of the foot you put forward on the ground. The average force becomes zero, and you walk along at constant velocity.

- **Running**: this is similar, except that you reach high enough speeds for friction with the air to make a significant contribution to the force resisting your forward motion.

- **Bicycling**: By riding a bicycle, you reduce the ground friction resisting your forward motion quite substantially. As a result, you can travel much faster, and your final speed is limited by air friction more than friction with the ground. Think about what you do when you cycle on level ground. You pedal just hard enough to keep going at constant speed. To go faster, you adjust how hard you pedal to match the resistive force at that particular speed. Going up or down a hill you have an additional force from gravity either aiding or resisting your motion.

- **Driving a car**: Again, in this case you can adjust the force propelling the car to overcome the resisting force and increase the speed, to balance the resisting force and travel at constant speed, or to be smaller than the resisting force and slow down. Here the changes are really obvious, as they are controlled by just pushing harder or letting up on the gas pedal. You can’t travel along at 70 mph without keeping your foot on the gas. You must be pushed forward with a force equal to the friction which is resisting your motion.

Notice that in all these cases the final constant velocity part involves feedback. In cycling, you pedal just hard enough to counteract the resistive force. If you want to travel faster, you pedal harder and speed up until the resistive force matches this new, stronger, forward force. If you’re going faster than you want to, you ease up on pedaling and let the bike slow down.

What if you want to actively start and stop, rather than coast to a stop? In starting normal forward motion your feet (on the bike) or the motor (in the car) makes the wheels push backward against the ground, trying to slip over it. Static friction between the wheel and the ground then pushes back, driving the wheel (with the bicycle and rider) forward. When you want to stop, you do the opposite, using your brakes to slow the wheel’s rotation. This would tend to make the wheels skid over the ground. So the wheel pushes forward on the ground, and the frictional force with the ground pushes backward on the wheel, slowing the wheel (and the bicycle and rider).

In all these cases the ability to *actively control* the forward and backward force which acts on you is what allows you to control your motion, speeding up and slowing down at will. The source of this active control is friction, and in particular the passive, static friction form of friction. Because it is static, it has an adjustable value, and you can control it.

Notice what happens when it’s no longer static friction, but kinetic friction which acts. You loose the ability to control the forces on you. If you start to slip on ice, then the
frictional force between you and the ice becomes a fixed kinetic friction value. Nothing you can do changes this, so you lose control of your motion and can only continue to skid until you come to a stop.

2D motion: some general comments

While motion in one dimension captures many interesting cases, it leaves out the rather important fact that we live in a three-dimensional space. We need to understand not only how an organism might speed up and slow down, but also how it might change direction. Indeed the most general case involves a combination of speeding up or slowing down (changing the magnitude of the velocity) and turning (changing the direction of the velocity). Just because it’s simpler to draw, we’re going to talk mostly about two-dimensional motion, rather than the full 3D. The extension to motion in three dimensions instead of two is relatively easy to see.

There are some very powerful things purely descriptive kinematics can tell us about absolutely any 2D motion. Consider motion along a simple curve at constant speed (constant magnitude of the velocity):

The velocity is changing in this motion. Its magnitude stays the same, but the direction is changing all the time. The velocity, as we have several times stressed, is always along this path of motion. In what direction is the acceleration? To find the acceleration we can draw the velocity vector at two successive moments, and find the change in velocity \( \Delta v = v_f - v_i \)

The acceleration vector is \( a = \Delta v/\Delta t \), a vector in the direction of \( \Delta v \). So for the curved motion which is shown, the acceleration is directed into the curve. In fact it has to be directed into the curve. If you think about it, the velocity is always turning into the curve, so that's always the direction of its change.

How large is the acceleration? For motion at constant speed along the curve, the size of the acceleration depends on how suddenly it curves. This can be expressed in terms of the
"radius of curvature"; the sharper the curve, the smaller this radius, and the larger the associated acceleration.

What if there is acceleration along the curve (speeding up or slowing down) as well as perpendicular to it? The total acceleration will then have a component into the curve (like the one shown above) as well as a component along it, so that the net acceleration is the vector sum of these two components. But no matter what, if there is curvature in the path of an object, there is some non-zero component of the acceleration towards the center of the curve.

How to handle the general case of a path you observe?

How do we find the acceleration associated with a particular path? Let’s imagine you know (from a video for example) exactly where a wolf is at every moment while it chases a deer. This means you know its path \( s(t) \). Notice that I wrote this position as a vector \( s \). This is a way of describing a general three-dimensional position. At each moment the animal is at a particular point in three dimensions (x, y, and z for example).

The “position vector” \( s(t) \) is a vector which goes from the origin of the coordinate system to the location of the animal at that moment. As it moves from one position to another, the change in position is a displacement vector given by the equation:

\[
\Delta s = s(t_f) - s(t_i) \quad \text{or} \quad s(t_i) + \Delta s = s(t_f)
\]

And can be shown in a figure like this:

![Diagram showing displacement \( \Delta s \) and position vectors \( s(t_i) \) and \( s(t_f) \).](image)

Given the displacement \( \Delta s \) for this little time interval, you can find the velocity for it, just by dividing by the time the displacement takes \( v = \Delta s / \Delta t \). One thing to remember about this is that the direction of the velocity is always the same as the direction of the displacement \( \Delta s \). The velocity is always along the motion.

Tracking the changes in velocity from moment to moment allows you to find the accelerations: \( a = \Delta v / \Delta t \). Notice since the change in velocity is not always in the
direction of the velocity, the acceleration is *not* always in the direction of motion. This is true for the curved motion described above. The velocity is always along the motion, but the acceleration (the change in velocity) may be partly perpendicular to the motion.

Any time you see the whole motion of an animal (or any other object) you can work out from the path \(s(t)\) the velocity and acceleration as a function of time. From these, you can deduce exactly the size of the total force acting on it at each instant. If the motion is complicated, the forces which act will have to be complicated. There are, however, a couple of cases where curved 2D motion is relatively simple: circular motion at constant speed, and motion under the influence of a constant force. As usual, we’ll use these simple cases as starting models for understanding more complex motion.

**One important 2D motion example: Circular motion at a constant speed**

What is the acceleration of an object traveling in a circle at a constant speed? This is another very restricted physics situation which has a tidy, analytic solution. Consider the following diagram:

![Diagram](image)

In the diagram on the right we have used the fact that if the angle \(\Delta \theta\) is the one shown on the left, it is also the angle between the two vectors \(v_i\) and \(v_f\). This is because the radius of the circle is always perpendicular to the velocity vector \(v\). The acceleration \(a = \Delta v/\Delta t\). In the limit where \(\Delta \theta\) is small we can look at the little triangle on the right and write:

\[
\Delta \theta = \frac{\Delta s}{r} \approx \frac{\Delta v}{v}
\]

so

\[
\Delta v = v \frac{\Delta s}{r}
\]

but

\[
a = \frac{\Delta v}{\Delta t} = \frac{v \Delta s/r}{\Delta s/v} = \frac{v^2}{r}
\]

What direction is the acceleration in? As we noted before, since there is no acceleration along the motion (it doesn’t speed up or slow down along the circle), the acceleration must be perpendicular to the motion and directed toward the center of the circle. What does this mean? Any time you see an object moving in a circle at a constant speed, you know its acceleration must have a size given by \(v^2/r\), and it must be directed towards the center of the circle.
There are two important comments to make about this. First, you might have guessed that the acceleration would be towards the center of the circle. If it was not always perpendicular to the motion, the object would either speed up or slow down along its direction of motion. Second, you can use this to approximate the acceleration perpendicular to the motion for any object traveling in a curve, especially when it’s traveling at approximately constant speed. Just estimate the "effective radius of curvature" of the curve at that point and calculate:

\[ a_{\perp} = \frac{v^2}{r_{\text{effective}}} \]

Note though, this won't tell you about acceleration along the curve. How does this work? Consider the picture drawn below:

Imagine this is the path of a horse which moves along a road at constant speed. At each of the three points (A, B, and C) shown, the direction of the horse is changing, so there must be an acceleration. How large is each? Since the speed along the path is always the same, we’ll just call that v. At point A, the horse is turning with a “radius of curvature” r_A, so the acceleration it experiences is approximately \( v^2/r_A \). You can similarly estimate the acceleration at B and C.

Notice that if the radius of curvature is small, the horse is turning quickly. This quick turn requires a larger acceleration. You can see this from the equation because when it turns suddenly (as at B), the radius of curvature is small, and the resulting acceleration must large. To have this larger acceleration, there must also be a larger force.

This ought to be a familiar fact. When you want to turn a corner sharply, you need a lot of force. You can do it if you wear good, sticky sneakers and push off really hard against the floor. But you can’t do it if large frictional force you need is not available, perhaps because you’re wearing socks on a hardwood floor, or because you’re trying to turn your car on an icy road.
Two dimensional motion under the influence of a constant force

Another important special case of two dimensional motion occurs under the influence of a force which is constant in both magnitude and direction. The famous example of this is a projectile near the Earth’s surface. At least to the extent that we can ignore air friction, such an object feels a constant downward force, but no horizontal force at all. What would the motion of such an object be like?

We know how that motion would change already. The constant downward force would create a constant downward acceleration. This acceleration would continually change the vertical component of the velocity. Whatever vertical velocity it begins with will continuously become more and more negative.

What about the horizontal velocity? Since there is no force in the horizontal direction, there will be no horizontal acceleration. As a result, whatever horizontal motion the object has to start will stay exactly the same throughout the motion. This problem is not purely theoretical. For reasons of military history, this kind of motion is usually called projectile motion. Before we work out the details of such motion, let’s look at a little history

Projectile Motion and Superposition in Physics

One of the reasons for resolving a vector into its components is that the motion of objects along different directions can be independent. This allows us to consider each motion, each component of the total vector, independent of the others.

The oldest application of this idea in physics, which was first performed by Galileo, is to projectile motion: the motion of an object we throw through the air. The motion of such a projectile includes several parts. The object rises, falls, and travels across the room while doing this. This much (and more) was well known to lots of people before Galileo.

The crucial contribution Galileo made to our understanding of how to describe this motion is contained in the following quotation from his book 'Two New Sciences'. In it he speaks of 'naturally accelerated' motion. By this he means motion accelerated by gravity.

I propose to set forth those properties which belong to a body whose motion is compounded of two other motions, namely one uniform and one naturally accelerated… This is the kind of motion seen in a projectile… Imagine any particle projected along a horizontal plane without friction… This particle will move along this plane with a motion that is uniform and perpetual, providing the plane has no limits. But if the plane is limited and elevated, then the moving particle… will, on passing over the edge of the plane, acquire, in addition to its previous uniform and perpetual motion, a downward propensity due to its own weight; so that the resulting motion… is compounded of one which is uniform and horizontal, and one which is vertical and naturally accelerated.
Galileo is making a quite surprising assertion about the physics of motion here. He is saying that an object’s horizontal motion has NO effect on how it falls, and further, that it’s falling has NO effect on its horizontal motion. He says that the motion is a \textbf{superposition} of two independent motions. This idea of superposition is a very important one in physics, which we will return to again and again.

Now this is NOT a statement about vectors or math. It’s a statement about physics, about how the world REALLY works. It cannot be proven by deduction, but only by experiment. Also note carefully that there are two statements here:

1. A horizontal velocity does not affect the way an object falls
2. The fact that an object is falling does not affect its horizontal velocity

They are not the same, and Galileo experimentally tested each. His confirmation of this independence supported his idea of the principle of superposition, and helped to introduce this important idea into physics.

\textbf{Projectile Motion: the details}

Now we're going to look at this motion in detail. Imagine an object which starts out with some initial velocity $v_0$, which has a magnitude $v_0$ and a direction an angle $\theta$ above the horizontal:

We can write this initial velocity vector in terms of components, in the fashion we have learned to do, as:

\[
v_0 = v_{0x} \hat{x} + v_{0y} \hat{y} = v_0 \cos \theta \hat{x} + v_0 \sin \theta \hat{y}
\]

How does the motion progress after the launch? As we have seen, the two motions, horizontal and vertical should be completely independent, so we really have two separate motions to describe:

Horizontal: The horizontal motion experiences no acceleration, so it continues at constant horizontal speed, and the distance traveled is described by:

\[
\begin{align*}
v_x &= v_{0x} = v_0 \cos \theta \\
\Delta x &= v_{0x} \Delta t = v_0 \cos \theta \Delta t
\end{align*}
\]
Vertical: The vertical motion experiences just the acceleration due to gravity that we have been analyzing so far in class, so its motion is described by:

\[ v_y = v_{oy} + a\Delta t = v_0\sin\theta - g\Delta t \]
\[ \Delta y = v_{oy}\Delta t + 1/2a\Delta t^2 = v_0\sin\theta\Delta t + 1/2(-g)\Delta t^2 \]

This is exactly what we analyzed above when we talked about an object thrown upward with a certain speed.

We can use these two equations \( x(t) \) and \( y(t) \), to determine the path \( y(x) \).

\[ \Delta t = \frac{\Delta x}{v_0\cos\theta} \]
\[ \Delta y = \left[\frac{(v_0\sin\theta)}{(v_0\cos\theta)}\right]\Delta x + 1/2*(-g)[\frac{\Delta x}{(v_0\cos\theta)}]^2 \]
\[ \Delta y = \tan\theta\Delta x - \frac{g}{2v_0^2\sin^2\theta}\Delta x^2 \]

Notice what this says. The distance traveled in the y direction (\( \Delta y \)) is a quadratic function of the distance traveled in the x direction (\( \Delta x \)). This tells us that the motion is parabolic: the shape of the path is a portion of a parabola.

That's it really, we now know how "projectiles" move, and we can write their motion as follows:

Position: \( \mathbf{r} = \Delta xx + \Delta yy = (v_0\cos\theta\Delta t)x + (v_0\sin\theta\Delta t + 1/2(-g)\Delta t^2)y \)

Velocity: \( \mathbf{v} = v_{x}x + v_{y}y = v_0\cos\theta x + (v_0\sin\theta + (-g)\Delta t)y \)

Acceleration: \( \mathbf{a} = a_{x}x + a_{y}y = -gy \)

What sort of questions might we like to ask about this motion? Let's do a simple example:

If I fire a projectile with initial velocity \( v_0 = 50\text{m/s} \) at an angle \( 30^\circ \) above the horizontal, how far does it travel before returning to the ground?

Start by considering the vertical motion. It has an initial y velocity of \( v_{0y} = v_0\sin\theta = (50\text{m/s})\sin30^\circ = 25\text{m/s} \).

How long does it stay in the air? This is exactly like the problem we did earlier, in which we asked how long does a ball stay in the air when we throw it up at 25m/s. We used the fact that it will return to the ground traveling at the same speed it left to find:

\[ v_{fy} = -v_{0y} = v_0\sin\theta - g\Delta t \]
\[ \Delta t = 2v_{0y}/g = 2v_0\sin\theta/g \]

And

\[ \Delta t = 50\text{m/s} / 9.8\text{m/s}^2 = 5.1\text{s} \]

OK, so the projectile is in the air for 5.1s, how far does it travel in this time? The horizontal motion is governed by:

\[ \Delta x = v_{ox}\Delta t = v_0\cos\theta\Delta t = (2v_0^2\sin\theta\cos\theta)/g \]

where we’ve taken the result for \( \Delta t \) derived above and inserted it here. Putting in the numbers we find:
\[
\frac{(2(50\text{m/s})^2 \sin 30^\circ \cos 30^\circ)}{9.8\text{m/s}^2} = 220\text{m}
\]

Note that there's a nice way to rewrite this using the fact that:
\[2\sin \theta \cos \theta = \sin 2\theta \quad \text{(a trigonometric identity)}\]
and we find the famous "range equation":
\[
\Delta x = \left(\frac{v_0^2}{g}\right) \sin 2\theta
\]
This expression tells us that, if we launch an object on level ground with some initial speed \(v_0\), at some angle \(\theta\) above the horizontal, how far it will go. What is it good for? It tells us how the result (the range) depends upon the input variables. We see that if we double the initial speed, we quadruple the range. We also see that the maximum range occurs when the launch angle is 45°, because the maximum of \(\sin 2\theta\) occurs at 45°.

A caution: this "range equation" and others like it throughout this class, applies ONLY when the conditions for which it was derived are satisfied. First of all, this requires that the only force acting is the downward force due to gravity. If air friction is important for your projectile, the range predictions provided by this model will be quite inaccurate. If, however, the forces due to air friction are always small compared to the downward force of gravity, this range equation might be OK.

There are other detailed assumptions here. This derivation assumes you are firing from level ground at an angle \(\theta\) above the horizontal. It does not apply if, for example, you are firing onto or off of or at a wall or cliff. I strongly urge you to always begin from the basic equations and use what you know about the problem to derive the solution.

Note what we have done here, we've realized that the vertical motion (in this case) determines the duration of the flight, and we've used that to determine the horizontal distance traveled.

Here we see some of the basics of these problems: because the motions are coupled, we can usually use one of the motions to constrain the other.

**Baseball Pitches:**

The second obvious application of projectile motion is to sports. An analysis using these tools suggests why baseball is such a difficult sport. A pitcher may throw at 100mph, and the ball must travel 60ft to the batter:

\[
100\text{mph} \times \left(\frac{1\text{m/s}}{2.24\text{mph}}\right) = 45\text{ m/s}
\]

\[
60\text{ft} \times \left(\frac{1\text{m}}{3.28\text{ft}}\right) = 18.3\text{m}
\]

\[
\begin{align*}
\text{45 m/s} & \\
\text{18.3 m} & \\
\end{align*}
\]

How long does it take to travel this distance?
Δx = 18.3m = v₀ₓΔt = 45m/s Δt  or  Δt = 18.3m / 45m/s = 0.41s

How far does it drop in this time?

Δy = v₀ᵧΔt + 1/2aΔt² = 1/2(-9.8m/s²)(0.41s)² = 0.82m (≈32")

When the ball is half-way, it has dropped only one fourth as much:

Δy₁/₂ = 1/2(-9.8m/s²)(0.2s)² = 0.20m (≈8")

What happens if the initial speed is reduced to 85mph? (37.9m/s)

Δt = 18.3m / 37.9m/s = 0.48s (only 15% different)

Δy₁/₂ = 0.28m (≈11")

Δy = 1.1m (≈44")

So, the batter has to look out, see how much the ball drops at the half way point. Then in the remaining 0.2 seconds, the batter must adjust for a pitch which may differ in location by 12", all the while swinging a bat which is "not more than 2 3/4" in diameter at the thickest part".

What's the slowest initially horizontal pitch thrown from a height of 2m which will cross the plate before it hits the ground?

Δy = -2m = 1/2(-9.8m/s²)Δt² or  Δt = 0.64s

Δx = 18.3m = v₀ₓΔt  or  v₀ₓ = 18.3m / 0.64s = 28.6m/s (≈ 64 mph)

Projectiles subject to air friction:

Real projectiles on Earth are never immune from air friction. Instead, they feel a friction force which always acts opposite their direction of motion, and which increases as the velocity of the projectile increases. Exactly how the friction depends on velocity depends on the details. As we saw last time, it will typically be proportional to the speed v if the object is small and moving slowly, or proportional to v² if the object is moving rapidly or is large.

In either case, the friction will resist the motion. This will create a force which slows the motion in both the x and y directions. The effect of this is different in the horizontal and vertical directions. In the x direction, this resistance to motion may gradually reduce the x velocity to zero, because there is no force acting to keep the motion going.

In the y direction, friction will always act to resist the motion. While an object is moving upward, this resistance acts in concert with the weight to slow the object down more rapidly. This means the object will not rise as high as it would without air friction. Once the projectile turns around and starts to fall downward, the weight acts to increase the downward velocity while the air friction acts opposite this. Eventually, the object speeds up until the friction force equals the weight, then it travels at “terminal velocity” as we discussed last time.

What does the path look like? Instead of traveling horizontally at a constant rate, the horizontal motion slows, covering less and less distance in each unit of time. Meanwhile,
the vertical velocity is everywhere slowed, the object doesn’t rise as high and drops more slowly. The net result is a shorter, stubbier path, like this:

friction, the ball only travels about 150m, or close to 500 feet. Smashing it harder helps, but very much less than the range equation would suggest, because increasing the launch speed also increases the importance of the friction.
Physics for the Life Sciences: Fall 2008 Lecture #10

Interactions, Systems, and State

We’re going to embark now on a substantial expansion of what we’ve been doing so far, putting together tools which allow us to discuss a much broader range of phenomena. In the most general sense, what we want to talk about how one ‘thing’ affects another. We have so far concentrated on a few simple ‘effects’ one object can have on another: it can make it either accelerate or deform. We invented a name for the means by which one object creates acceleration or deformation in another, we called this a ‘force’.

But there are many other ways in which objects affect one another. If I place a hot object next to a cool one, the hot one cools and the cool one heats up. If I place a metal in acid, it dissolves, and hydrogen gas is released. In the most general way, we can speak of all these ‘effects’ as interactions.

The interactions we have been concerned with so far, those associated with forces, we call mechanical interactions. Interactions involving heat we call ‘thermal’. Interactions which give rise to changes in chemical composition we call ‘chemical’ interactions. When you rub a balloon on the rug and then stick it to the wall we talk of ‘electrical’ interactions. Magnets stick to the refrigerator through a ‘magnetic’ interaction. Often what we see, especially living things, is a mix of many of these. If I plant a seed in the ground and water it, a tree will grow. This complicated effect is built of many more fundamental interactions. Yet it remains just a very elaborate example of the same essential elements.

In each case we talk about an interaction because we can see a change take place in the properties of the systems involved. There’s always clear, repeatable evidence that something has happened. This is an important point. The interactions which concern us in physics are brought to our attention through the clear and repeatable effects which they have on the world around them.

This is an important point, because there are people who talk about interactions which don’t bring about such clear and repeatable effects; things like ESP, rain dances, and astrology. The reason these subjects are outside the bounds of science is not that scientists are prejudiced against them. It’s that they don’t create the kind of repeatable, reliable effects on their surroundings which make it useful for science to discuss them. If they did, believe me, scientists would be the first to study them, because discovering something new is the coin of the realm in science.

To talk about interactions we need to say a bit about the ‘things’ which are interacting. We’ll call these, in a general sense, ‘systems’. In many situations we have analyzed, the ‘system’ is one of the objects we would have drawn a free body diagram for; a block or a ball, a monkey or an elephant. Note that the definition of a system is really arbitrary. We
get to draw the lines around these systems where we want. It is only important that we are
careful to be clear about what we really mean in any particular discussion.

The way in which we perceive interactions between systems is through changes in their
properties. Typical mechanical properties include the momentum, position, or mass.
Systems also have other properties we have so far not discussed, like temperature,
pressure, chemical composition, electric charge, magnetization, and physical state (solid,
liquid, gas, or plasma). For a given situation, we can describe the system by listing the
values for all the relevant properties which describe it. In a certain sense, the system is
the set of properties which describe it. If there were something more to it, we would
simply add that to the description of properties.

So, to do physics, we need to talk about interactions between systems, which alter their
state. We’ve already done a lot of this for mechanical interactions between objects which
alter their positions, velocities, and accelerations.

As we move into this more general description of what’s happening in the world, it’s
clear that we won’t be able to describe everything in terms of just forces and
accelerations. We must begin to include more. Today we take a first step down that path,
as we begin to introduce the idea of energy. We will find that ‘energy’ is another property
of a system, something which we must know to describe its state. It is a very important,
highly general thing. You can’t understand the world, and especially not life, without
understanding energy.

**Energy, an essential missing piece**

So far our study of physics, our study of change in the physical universe, has
concentrated almost entirely on the Newtonian ideas of Force and Momentum. With
these ideas we can understand a tremendous amount of motion. We have explained the
motion of objects thrown through the air, understood the influence of friction and weight,
and noted the way in which motion in curved paths requires the presence of forces. From
the time of Newton until well into the 19th century, this is really all there was to physics.

But from the beginning there were thoughtful people who realized there must be more to
change than just force and acceleration. Christian Huygens, a brilliant Dutch
contemporary of Newton, was very aware of this problem. He recognized that if, for
example, a bomb goes off and pieces fly in every direction, there is no net force acting.
Since all the forces are internal, when one part of the bomb pushes on another, the second
pushes back with an equal and opposite force. So somehow, though there was no net
force applied to this system, there was clearly a very big change in motion.

Huygens' recognized that there was something to motion other than just the forces. He
saw that this new thing didn't depend on direction; it was a "scalar" quantity. You could
have a net force of zero in a system, yet still have changes in motion. But the changes you
see would not be in the overall vector momentum, but only in this new scalar measure of
motion. He wanted to invent a 'quantity' of motion which describes in some general way
whether things are moving or not, without regard to what direction they're traveling. He called this quantity "vis viva", or the motion of life. Today we refer to this scalar measure of motion as "kinetic energy".

Energy is a much more complicated idea than momentum. Unlike momentum, objects can ‘store’ energy in a wide variety of forms, while momentum just appears as organized motion. As a result when we consider energy, the accounting of how much there is and how it changes will be more complex than it is for momentum.

**What is kinetic energy and how does it change?**

Just what is this scalar quantity, this ‘vis viva’ or energy, and how does it change?

As usual in physics, this idea that there was another important concept lurking around was not only due to Huygens. Others had glimpsed it before, including the Greeks, and (of course) Galileo, particularly in connection with the simple machine (like block-and-tackles and inclined planes). Simple machines allow you to do something which would usually require a large force acting over a short distance, by instead applying a small force for a long distance.

The notion that there was something important about the application of a force over a certain distance was finally codified by Thomas Young, a medical doctor of remarkable mental agility in the early 19th century. He called this quantity (force times distance) "work", and suggested that work is what changes “vis viva” or energy of an object.

It is interesting that someone outside the usual path of physicists made the discovery. Perhaps this is because it is such an original idea. The problem with understanding work and energy has nothing to do with mathematically complexity. The mathematics is simple, but the ideas are subtle and abstract. Perhaps this is why so much of the progress in this area came from people outside the physics community; from doctors, engineers, and even winemakers. Physicists now view the world very much from this energy perspective. Everything that happens is governed by the flow of energy.

The idea of energy is especially essential in understanding life. So we will spend the central part of this class focusing on the idea of energy, and seeing life as a manifestation of the natural flow of energy across the Earth.

**Work and changes in kinetic energy contrasted to impulse and changes in momentum**

We have previously recognized that when a force acts for some period of time, it generates an ‘Impulse’ (see Lecture #3). This impulse changes the momentum of an object:

\[ \int F \, dt = p_f - p_i = \Delta p \]
Here we are summing the product of a vector force with a scalar time. This gives a vector result, which is the vector change in the momentum. This is the part that Newton understood.

Young recognized that there was also something important about the product of a force and the distance over which it acts. We want to consider something like:

\[ \int F \, ds = 'Work' \]

There is a complication though. The displacement ds is a vector, which may or may not be in the direction of the force. To calculate this we have to decide how we want to multiply these two vectors F and ds. It turns out that if we consider the ‘scalar’ product, or dot product (Lecture #2), of F and ds, we will produce a scalar result. That is, this integral will tell me about a scalar property of the system.

Remember what this is. If the angle between F and ds is \( \theta \), then:

\[ \int F \cdot ds = \int F \cos \theta \, ds = \int F \text{along} ds \]

It’s the sum of the component of the force along the direction of motion times the displacement. **Only forces along the direction of motion do work!**

Let’s work out what this ‘work’ is equal to:

\[ \int F \cdot ds = \int F \cos \theta \, ds = \int \text{macos} \theta \, ds \]

Now we know that \( a = \frac{dv}{dt} = \frac{dv}{ds} \frac{ds}{dt} = v \frac{dv}{ds} \) by the chain rule. So:

\[ \int \text{macos} \theta \, ds = \int \text{mcos} \theta * (v \frac{dv}{ds}) \, ds = \int \text{mv} \cos \theta \, dv \]

Consider the simple case where the force is along the direction of the motion, so that \( \cos \theta = 1 \) everywhere. In this case:

\[ \int F \cdot ds = \int \text{mv} \, dv = 1/2mv_f^2 - 1/2mv_i^2 = \Delta(1/2mv^2) = \Delta(KE) \]

What this is saying is that the integral of the scalar product of F and ds gives a result which is the change in a scalar quantity. This scalar quantity, a quantity describing the *amount* of motion, but ignoring its direction, is called kinetic energy:

\[ KE = 1/2mv^2 \]

What we have found is that the work, defined as \( W = \int F \cdot ds \), is equal to a change in the kinetic energy of an object. Just as the product of the force times the time over which it acts is the change in the momentum, the product of the force times the distance over which it acts is the change in the kinetic energy.

| Impulse | \( \int F dt \) | \( \Delta p \) | changes the vector quantity of motion |
| Work | \( \int F \cdot ds \) | \( \Delta KE \) | changes the scalar quantity of motion |

What is this kinetic energy? Remember that we want to talk about interactions between systems which have states. Kinetic energy is one of the parameters which describes the state of a system. We have found here that interactions between systems cause changes in KE which are determined by calculating the work done on the system. This is completely parallel to our discovery that impulse and changes in momentum are related. It you want to know how the momentum of a system changes, calculate the impulse applied to it. If you want to know how the kinetic energy of a system changes, calculate the work done on it.
Determining work: some examples

Consider first the case of a force applied to an object along its direction of motion. Nothing fancy here, just imagine a constant force accelerating an object from rest.

\[
\text{Work} = F \cdot \Delta s
\]

Clearly such a force affects a change in the motion. This change is both a change in momentum and in energy.

What if the force is opposite the direction of motion? Then \( F_{\text{along}} \) is negative, and the work done on the object is negative. To repeat, work is positive if the force is in the direction of motion, and negative when the force is opposite the direction of motion.

The units of work are Newton\( \times \)meters, or \( \text{km} \times \text{m}^2/\text{s}^2 \), which we will soon begin calling "Joules". Since work creates changes in energy, these Joules are also the units of energy.

What happens if there is a force which does not act purely along the direction of motion? Imagine an object traveling around in a circle. Are there any unbalanced forces acting on it? Yes there are. Do these forces change the motion of the object? Yes, they alter the direction of its motion. Do they do any work? No, because the direction of motion is always perpendicular to the direction of the force. The object never moves along the direction of the force. This is a beautiful example, because the force is not zero, and the motion does change, but the work \( (F_{\text{along}} \Delta s) \) is zero, and so in this case the change in energy is also zero.

What about an intermediate case, in which the force is partly along the direction of motion?

In this case \( F_{\text{along}} = F \cos \theta \) where \( \theta \) is the angle between the force and the direction of motion.

This relation \( \text{Work} = \int F \cdot ds = \Delta KE \) is called the work-energy theorem, and it plays the central role in tracking the changes in kinetic energy of a system.
"Work" in physics and in the vernacular

Now this idea of work is often confusing. Unfortunately we use the word “work” all the time in life, as you have long before you ever took a physics class. In everyday life it has a familiar, if vague meaning. But now we want to take its very broad and general meaning, the meaning we use in ordinary life all the time, and replace it with a very specific and exact meaning. You must remember this when you think about work in physics.

As an example consider carrying a suitcase through the airport. The force you apply is always perpendicular to the direction of motion (up vs. across), so you do no ‘work’. Nevertheless, you get very tired carrying this suitcase from one terminal to another. This little problem illustrates how work in physics and the "work" which you usually talk about are not the same thing. When you usually think about work, you think about how tired you get trying to accomplish something. This differs greatly from the "work" we're discussing here, primarily because of the way you muscles function.

Remember from the Vogel reading on muscle function the way muscles are triggered (the key parts are on page 23). The active muscle you use cannot just contract and stay contracted. It “twitches”, and each fiber provides a brief jerk. To apply a constant force, the fibers must be triggered to twitch in a continuous and staggered way. Each twitch has a muscle applying a force along its length, in the direction of motion, so real work is done in each twitch. To keep your arm bent, with the load sitting still, your muscle is continuously doing work in tiny little steps.

Energy is relative, but changes in energy are universal

There is a small but essential point lurking here. What we care about in physics is how energy changes, not what its absolute value is. You can see that this is true because KE = 1/2mv^2. If we measure an object in different reference frames moving relative to one another at constant speed we would find different energies.

For an example, imagine a passenger in a car (speeding along at 70 mph) who reaches out the window and drops a beer bottle. To the passenger, the energy and momentum of the bottle is zero. To the pedestrian watching from the sidewalk, the energy and momentum of the bottle is dangerously large. Who is right?

In fact they both are! Nothing in physics depends on the absolute value of the energy or momentum. To see why, consider what would happen if the bottle, just after being released, struck a rock along the road. The moving observer would see the bottle shatter suddenly rocket off behind her, undergoing a sudden, large change in momentum. The stationary observer would see the bottle zooming toward him shatter and suddenly stop moving forward, undergoing a sudden, large change in momentum. They would both see the same change in momentum (and energy), and hence infer the same force between the
Using the Work-Energy Theorem: simple examples

We can use the work energy theorem in a variety of ways. For example: If we know the force which acts on a system and we know the distance through which it travels, we can calculate changes in energy:

If I slide a 3kg book across the table, the kinetic frictional force between it and the table is 20N, and it slides 2m in coming to rest, what was its initial speed?

\[ W = F \cdot \Delta s = -F_f \Delta s = -20N \times 2m = -40J = \Delta KE = 1/2mv_f^2 - 1/2mv_i^2 = -1/2mv_i^2 \]

So

\[ 40J = 1/2mv_i^2 \]

Or

\[ v_i = \sqrt{(2 \times 40J) / 3kg} = 5.2 \text{ m/s} \]

Here is another example. If we know how much the energy of a system changes and we know how far it travels, we can find the net force which acts on it. If a sled starts out traveling at 5 m/s and comes to a halt while sliding through a distance of 20m of level ground, what is the coefficient of kinetic friction between the sled and the snow?

\[ W = F \cdot \Delta s = F \Delta s = 1/2mv_f^2 - 1/2mv_i^2 = -1/2mv_i^2 \]

\[ F_f \Delta s = -\mu_k mg \Delta s = -1/2mv_i^2 \]

Or

\[ \mu_k g \Delta s = 1/2v_i^2 \]

\[ \mu_k = v_i^2 / 2g \Delta s = (5m/s)^2 / (2 \times 9.8m/s^2 \times 20m) = 0.063 \]

Work with a force varying in magnitude

Calculating the work done by a constant force on an object moving in a straight line is simple. Now we want to consider two variations. First, what is the work done by a varying force on an object moving in a straight line?

\[ W(s) = \int F \cdot ds = \int F_{\text{along}(s)} ds \]

The work done as the object travels through each little distance is just the force at the point times the distance traveled. So if we have a picture of the force applied at each position during the motion, we can calculate the work done.
Consider a simple motion. I start out pushing a block hard, and gradually decrease my force as it accelerates away from me. This is how you usually get something (like a bicycle) going when you push it. So the force vs. distance curve might look like:

The total work done by you during this period is just the sum of $F(s)\Delta s$ from the point where it starts until the end. This sum is just the area under the curve. So in general when we’re dealing with a force that varies while the object travels, the work done is just the area under the force vs. distance curve.

**Work and objects traveling in curved paths.**

In the same way, we can calculate the work done by a force acting on an object traveling in a curved path. Consider walking up a hill:

As you go up this hill, gravity always acts straight down. For each little part of the path $ds$, the work done by gravity is:

\[ dW = F_g dscos \theta \]

But $dscos \theta$ is just $-dh$, how much higher the object goes. So:

\[ W_{gravity} = \int Fdscos \theta = -mg\int dh = -mg\Delta H \]

Notice what we’ve done here. Our definition of work is $\int F \cdot ds$. We interpreted this as $\int F_{\text{along}} ds$, the force along the direction of motion times the displacement. There is another way to look at it. You could also say that the work is $\int F_{\text{along}} ds$, that is, the force times the displacement along the direction of the force. What we’ve done here is to see that the work done by this gravitational force depends just on the magnitude of the weight and the amount by which the objects height changes. We will revisit this idea of the work done by gravity next time when we will relate it to potential energy.
Power

Power is the rate at which work is done. For some purposes, this is a useful quantity to keep track of. More generally, power is the rate at which energy changes. Since work represents a change in energy, the rate at which work is done is a rate at which energy changes.

\[
\text{Power} = P = \frac{dW}{dt}
\]

The units for this are Nm/s = Joules/sec. This unit, 1 Nm/s, is called a Watt (after James Watt, steam engine inventor).

For a \textit{constant force} we can write a special form by recalling that the work done is given by \(dW = F \cdot ds\), so

\[
P = \frac{dW}{dt} = \frac{d(F \cdot ds)}{dt} = F \cdot \frac{ds}{dt} = F \cdot v
\]

So in this special constant force case, power is the dot product of force and velocity.

Quantifying energy: some comparisons

Energy makes everything happen, including life. As a result, energy is something we pay for. Oddly, the electrical energy you pay for is usually not accounted in Joules, but in a funny mixed unit called “kilowatt-hours”. For one kW-hour of electrical energy, you might pay 8-9 cents.

What is this unit really? One kilo-watt hour is 1000 Watts * 3600 seconds = 3.6x10^6 Joules; 3.6 million Joules. That’s actually a lot of energy. Perhaps a comparison will help put it into some kind of context. If you weigh 80 kg, you’d have to be moving at 300 m/s to have this much kinetic energy. That’s about 675 mph, a healthy (or maybe unhealthy) speed indeed! If someone said they’d speed you up to 675 mph for 9 cents, you’d probably think it pretty cheap.

Right now, people in the US use, on average, about 8000 kilowatt-hours per year. That’s about 22 kilowatt-hours per day. The cost at current rates is a few dollars a day, and hence not prohibitive.

Another comparison might be helpful. To stay alive, you have to eat regularly. A typical adult diet might contain 2000 Calories a day. The “Calorie” of nutrition is equal to about 4184 Joules, so the 2000 Calories a day you need provides about 8.4x10^6 Joules of energy. What would this daily consumption of energy be in the usual units we pay for, kilowatt-hours? The 2000 Calories a day you must consume is about equal to two kilowatt-hours. If you paid for your food energy as you do your electrical energy, your daily meals would add up to about 20 cents a day.
You can see that the price we pay for electrical energy is very low compared to what we pay for food energy. There are a lot of reasons for this, but it’s worth noting that electrical energy is largely extracted from the long ago stored resources of fossil fuels, while your food comes ultimately from the annual collection of sunlight by plants and the animals that eat them. In this very real sense, food is a renewable energy resource. The higher cost of your daily 2000 Calories of food has many origins, but a good part of it is due to the fact that it is fundamentally renewable.

It is useful to remember some simple conversions that help you to place power and energy in a context:

- The power obtained burning one gallon of gas in an hour = 39,000 watts
- The power to run a light bulb = < 100 watts

So if you want to save energy, don’t bother to turn off the lights, just stop driving your car. When you drive, you use energy at a rate hundreds of times larger than you do when you light a lamp.

The table below gives an idea of the relative cost of various energy sources. This is changing fast, but the basic pattern, that “renewable” food is 30-40x as expensive as our more familiar “energy sources” is pretty robust. Now food is not the cheapest way to harvest renewable energy. Among other things, we pay a lot to keep it tasty. Electricity generated from renewable resources is currently more like several times more expensive than electricity from the coal and nuclear sources we use now.

It may shock you to see how much those batteries really cost, Joule for Joule…

<table>
<thead>
<tr>
<th>Type</th>
<th>Unit</th>
<th>Cost/Unit</th>
<th>Cost per Mega-Joule</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electricity</td>
<td>1 Kwh</td>
<td>$0.10</td>
<td>$0.027</td>
<td>appliances, motors</td>
</tr>
<tr>
<td>Gasoline</td>
<td>1 gallon</td>
<td>4.20</td>
<td>$0.032</td>
<td>transportation</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>1 Therm</td>
<td>0.60</td>
<td>$0.011</td>
<td>heating</td>
</tr>
<tr>
<td>Milky Way candy bar</td>
<td>1 bar, 1.9 oz</td>
<td>0.90</td>
<td>$0.94</td>
<td>food</td>
</tr>
<tr>
<td>AA battery</td>
<td>1 battery</td>
<td>0.40</td>
<td>$80.00</td>
<td>portable electronics</td>
</tr>
</tbody>
</table>
Let’s recap what we’ve been saying. We have begun considering an alternative way of tracking the interactions which occur between objects; one which is based on Impulse and Work as the arbiters of change. We have simple definitions for these things:

\[
\text{Impulse} = \int F \text{dt} = \Delta p = m v \\
\text{Work} = \int F \cdot ds = \Delta \text{KE} = \frac{1}{2}mv^2
\]

Notice again that there is nothing here dependent on the absolute value of the momentum or KE. Those are both arbitrary and dependent on the choice of reference frame. Only changes in these quantities are described by physics.

Let’s consider two simple examples as ways of illustrating how Impulse and Work describe different things. In the first case a ball is dropped from some height into a pile of sand. It comes to a halt after imbedding itself some distance \( \Delta s \) in the sand:

How does this interaction between the ball and the sand look from the perspective of Impulse and Work? First let’s look at the impulse. Since we know the change in momentum, we can find the total impulse:

\[
\text{Impulse} = \int F \text{dt} = \Delta p = p_f - p_i = -mv_0 = mv_0
\]

From this we can find an average force. We get this by estimating the time of interation \( \Delta t_{\text{stop}} \):

\[
F_{\text{av}} = \frac{\Delta p}{\Delta t_{\text{stop}}} = (mv_0/\Delta t)y
\]

What is \( \Delta t_{\text{stop}} \)? We might estimate it by saying \( \Delta t_{\text{stop}} = \Delta s/ v_{\text{av}} \), where \( v_{\text{av}} \) is the average velocity while it is slowing down, and \( v_{\text{av}} = v_0/2 \), so

\[
\Delta t_{\text{stop}} = \Delta s/ v_{\text{av}} = 2\Delta s/ v_0
\]

Now consider the work done. We can find this because we know the change in kinetic energy:

\[
\text{Work} = \int F \cdot ds = \Delta \text{KE} = 1/2mv_f^2 - 1/2mv_0^2 = -1/2mv_0^2
\]

From this too we can find an average force:
Physics 135 Fall 2009: Lecture Notes

\[ F_{av} = \Delta KE/\Delta s = 1/2mv_0^2/\Delta s \]

How do these two estimates compare?

The impulse tells me: \( F_{av} = mv_0/\Delta t_{stop} = mv_0^2/2\Delta s \)

The work tells me: \( F_{av} = mv_0^2/2\Delta s \)

So they are completely consistent.

Now consider a bit more complicated example: an elastic collision between a ball and the floor. A ball falls downward and bounces back off the floor, so these are the initial and final states:

Before                                    After

In this case the Impulse is:

\[ \text{Impulse} = \int F dt = \Delta p = p_f - p_i = mv_0y - mv_0(-y) = 2mv_0y \]

\[ F_{av} = (2mv_0/\Delta t_{bounce})y \]

And the work is

\[ \text{Work} = \Delta KE \text{ or } \int F \cdot ds = (1/2mv_f^2 - 1/2mv_o^2) = 0 \]

What’s going on here? To understand we have to see how the ball actually moves during the collision (which lasts for a total time \( \Delta t_{bounce} \)):

The ball reaches the floor (shown on the left), squashes down until it comes to rest (in the center), then springs back in the opposite direction (on the right). The force which alters the direction of the ball is always acting upward. The motion \( \Delta s \) is first downward (as the ball squashes) and then back upward (as it springs back). As a result this force first does negative work on the object, reducing its KE to zero, and then does positive work on it, restoring its KE.

So again, both pictures provide a consistent picture of the interaction. This will of course always be true. Learning to see how it is true in various situations will refine your understanding of work and impulse.

**Work done by gravity and potential energy**

Last time when we discussed work done on an object moving in a curved path we looked at the specific case of work done by the force of gravity as an object moves from one place to another. In this case we found:

\[ W_{grav} = \int F_{grav} ds = -mg(y_f - y_i) \]

There is an odd feature about this which we pointed out. The work done by gravity as you go from some initial location to some final location depends only on where you start and
stop, and not at all on how you get there. Another, fancier way to note this is to say that the work done by gravity when moving from one place to another is “path independent”.

We need to think about what this means. Imagine an object traveling only under the influence of gravity. As it moves, gravity does work on it, changing its kinetic energy:

\[ W_{\text{grav}} = -mg(y_f - y_i) = \Delta KE = KE_f - KE_i \]

Since the work done by gravity depends only on position and not on path, the final kinetic energy of this object is determined completely by where it is:

\[ KE_f = KE_i + W_{\text{grav}} = KE_i - mg(y_f - y_i) \]

Let’s repeat that: for an object acted on only by gravity, energy is simply a function of position. You tell me where the object is, and I’ll tell you what its energy is.

Here’s an example: when I toss a ball into the air it begins with large kinetic energy. As it rises, gravity does negative work on it, taking away some of its kinetic energy. Then as it falls, gravity does positive work on it, giving that energy back.

Since we know that gravity will do this, since it can restore the energy every time, it is useful to think of the work done by gravity as "storing up" some of the energy of the ball, keeping it fully available for return to the system. This kind of stored energy is something we call potential energy. We throw the ball up. Its kinetic energy is converted to potential energy, at the top when the kinetic energy is instantaneously zero, and then as it falls the potential energy is turned back into kinetic energy.

How shall we measure this potential energy? It turns out the appropriate definition is:

\[ \Delta PE_{\text{grav}} = -W_{\text{gravity}} \]

Pay close attention to the minus sign in this definition. It’s really crucial. It says that when gravity does positive work on the system, when gravity increases the kinetic energy of the object, the potential energy decreases. This is as it should be. To increase the kinetic energy, some of the energy stored as potential energy must be used up.

By contrast, when gravity does negative work on the system, it reduces the kinetic energy of the object, then the potential energy of the system increases. Again, this makes sense. Now I’m taking away kinetic energy and storing it as potential energy.

How does this changing potential energy relate to position?

\[ \Delta PE_{\text{grav}} = -W_{\text{grav}} = mg(h_f - h_i) \]
\[ \Delta PE_{\text{grav}} = mgh_f - mgh_i = mg\Delta H \]

As the height of an object changes, it’s potential energy changes. If it goes higher (\( \Delta H \) positive), the potential energy is becoming larger. If it goes lower (\( \Delta H \) negative), the potential energy is becoming smaller.

Notice too that we’re not defining any absolute potential energy. This definition only allows us to track changes in the potential energy. That’s consistent with our overall understanding that physics cares only about changes in energy.
**Energy when only gravity acts**

Now if no force but gravity does work, we can use the work-kinetic energy theorem to notice something interesting:
\[ \Delta KE = \sum W_{all} = W_{\text{gravity}} \]
or, moving this to the other side:
\[ \Delta KE - W_{\text{gravity}} = 0 \quad \text{or} \quad \Delta KE + \Delta PE = 0 \]
As long as no other forces do work, any change in kinetic energy is just balanced by a change in potential energy. I can freely transfer the energy back and forth between potential and kinetic.

There is another instructive way of writing this:
\[ \Delta KE + \Delta PE = KE_f - KE_i + PE_f - PE_i = 0 \quad \Rightarrow \quad KE_f + PE_f = KE_i + PE_i \]
In this form you can clearly see that this is a conservation law. There is some quantity, the sum of kinetic and potential energy. Whatever this total is at the initial time, the total of the two remains exactly the same at any later final time. This quantity, the total of kinetic and gravitational potential energy, is **conserved**; it never changes with time.

Remember though, to apply the work-kinetic energy theorem as we did here, we need to be sure that gravity is the *only* force doing work on the object. If air friction, for example, is present, it too would do work on the object. Then we would write:
\[ \Delta KE = \sum W_{all} = W_{\text{gravity}} + W_{\text{friction}} \]
And
\[ \Delta KE - W_{\text{gravity}} = \Delta KE + \Delta PE = W_{\text{friction}} \]
What would this mean? In this case, instead of just converting potential energy into kinetic energy, some of it would be disappearing in the energy lost to friction. More on this below…

**Constraining forces and energy**

It is often the case that an object will be constrained to move on a particular path by a force. A good example of this is a pendulum. The tension which keeps the pendulum moving in a circle *ALWAYS* acts in a direction perpendicular to the direction of motion, so it never does work on the object. Since this force never does work on it, the only force which does do work is gravity. So a pendulum goes back and forth, trading energy between potential energy at the top and kinetic energy at the bottom. The other force which acts, the tension, does no work at all, and hence never alters the energy.

Likewise, if I take a ball and roll it inside a bowl, the force which keeps it from moving through the bowl is the normal force. Again this normal force always acts perpendicular to the motion, so it can never do work, it can never change the energy of the ball. So again, it just rolls back and forth, trading energy between kinetic and potential, and back again.

Some situations which do this:

1. Roller-coasters (the very name "coaster" evokes this)
2. Ball rolling in bowl
3. Pendulum
4. Car driving on a hilly road

We have left something obvious out of course. In each of these cases, there is a force other than gravity and the force of constraint acting: friction. In these cases, friction always acts opposite the direction of motion. Because of this, it always does negative work. This negative work gradually drains away the energy which would otherwise be trading back and forth forever between kinetic and potential. Eventually, the negative work done by friction drains away all the energy, and the object comes to rest, sitting at the lowest point.

**Gravitational Potential Energy**

So what have we done? We’ve said that since we can ‘store’ energy in the work done by gravity; it’s useful to talk about this as a form of energy. How much potential energy does an object have?

It is impossible to say what the ‘absolute’ potential energy of anything is. Just as the absolute value of kinetic energy is observer dependent, so too the absolute value of potential energy is meaningless. Remember how we defined potential energy:

$$\Delta PE_{grav} = -W_{grav}$$

We never defined the absolute potential energy, but only changes in it. So it is meaningless to talk about the absolute potential energy. **It is only useful to speak of the change in potential energy as an object moves from one location to another!**

**Example: Velocity of a pendulum**

Consider the motion of a pendulum. We will release it from rest at some angle \(\theta_0\) from the vertical:

$$\Delta h = L - L\cos\theta_0 = L(1 - \cos\theta_0)$$

What will its velocity be at the bottom? To answer this question using Newton’s laws is difficult, because the tension T is constantly changing as it progresses. This question is easy to from the energy approach though.

$$\Delta KE = W_{total} = W_{tension} + W_{gravity}$$

The tension does no work because it is a force of constraint (it always acts perpendicular to the motion). So:

$$\Delta KE = W_{gravity} = -\Delta PE \quad \Delta KE + \Delta PE = 0$$

$$KE_f - KE_i + PE_f - PE_i = 0$$

Let’s call the point at the bottom the final position, and measure our y distances from there.
Physics 135 Fall 2009: Lecture Notes

\[ \frac{1}{2}mv_f^2 - 0 + mg*0 - mgL(1-\cos\theta_0) = 0 \quad \text{or} \quad v_f = \sqrt{2gL(1-\cos\theta_0)} \]

What if I ask for the velocity of the pendulum at some other angle, call it \( \theta_1 \)?

\[ \frac{1}{2}mv_1^2 - 0 + mgL(1-\cos\theta_1) - mgL(1-\cos\theta_0) \]

or

\[ v_1 = \sqrt{2gL(\cos\theta_1 - \cos\theta_0)} \]

It is often the case that a very ‘hard’ problem in the force picture can become very easy in the energy/momentum picture. This is a fine example. In a sense this is because the energy approach concentrates on what’s really important in the problem, ignoring things (like forces of constraint) which don’t play a crucial role.

**Potential Energy and Conservative forces**

This idea of potential energy is a very powerful one. It allows us to understand the work done (or which will be done) by some force in a very simple way: to know the work done all we have to know is where things begin and end. Changes in energy are then just a function of position, and not of history.

Gravity is not the only force which we can define potential energy for. In fact we can define a meaningful potential energy for any force which is a "conservative" force. You can tell that a force is conservative if the work done by it when traveling between two points is independent of the path taken between these points. If this is true, the work done is a function only of where you are, and a potential energy can be meaningfully defined.

There is an alternate way to state this requirement. If the total work done by a force as you travel in a closed path (starting and ending at the same point) is zero, the force is conservative.

The conservative forces we will encounter in this class include gravity and the elastic force associated with deforming a solid. These are both cases where the obvious macroscopic nature of the force obeys the requirements outlines above: work done by these forces is path independent.

You will see later in your study of physics other forces which are conservative, like the electromagnetic force. They too will be associated with potential energies, as a way of keeping track of how energy is flowing. In each case, these are forces which allow you to use the work done by the force to "store" energy which you can then completely recover.

In the end you’ll find that all fundamental forces are conservative like this, and that energy is actually always conserved. But this is not always obvious when we’re talking about complicated, extended objects made up of many parts. For this reason we will sometimes speak of “non-conservative” forces.

These non-conservative forces are those doing work which is path dependent. As a result you cannot easily recover the energy these forces extract. Imagine I slide a book from
point A to B on a table. The force of friction as I slide is always \( \mu k F_N \), and it always directly opposes the motion. So the work done by it is just:

\[ W_{\text{friction}} = -\mu k F_N \text{dist}_{AB} \]

The amount of work done definitely depends on which path I take from point A to B. If I move the book from A to B, and then back from B to A, friction does negative work on it during both parts of the motion, always extracting energy from the object. Thinking about the second definition of a conservative force, you can see that the total work friction does around a closed path is not zero. Though I can take away energy through the action of friction, I really can’t get it back, at least not through the action of the friction force. So the work done by friction is not ‘storing’ energy in a way which we can easily recover. It’s not a ‘conservative’ force.

Does this mean that energy is really not conserved? Is energy used up, or destroyed when friction acts? In fact energy, like momentum, is always conserved. When we speak of a “non-conservative force” in this sense, we’re just saying that when this force acts, it converts the energy taken from the object into a form from which is difficult to recover. Every bit of the energy removed from the object by a non-conservative force still exists; it has just been converted into a form which makes it difficult to recover.

Notice that all these non-conservative forces (friction, the force you push something with, etc.) are described by phenomenological force laws, rather than fundamental ones. We will find, later in the class, that all of the fundamental forces are conservative. These ‘non-conservative’ forces we encounter are really nothing more than complex manifestations of underlying conservative forces.

**Elastic potential energy of a deformed material**

Let's consider briefly another way of storing energy, another form of potential energy. If I fire a ball at a spring it will arrive at the spring with some kinetic energy. The spring will gradually slow it down until it has no kinetic energy (it comes to a stop), and then it will push back against the object, speeding it up until eventually it leaves the spring with just the energy it arrived with. Note the similarity to gravity. The spring absorbs the energy of the moving ball, and then restores it to the ball as it flies off. Let's see how this fits in our work/KE language.

The force applied by a spring when you deform it depends on how much you squash or stretch it. Deform it a little and it applies a little force, deform it a lot and it applies a lot of force. For many materials, we can express this numerically as Hooke’s law:

\[ F = -kx \]

Where \( x \) is how much you stretch the spring and \( k \) is the 'spring constant', a number which describes how stiff the spring is. For a stiff spring, \( k \) is large, for a smushy spring \( k \) is small. Remember that the minus sign means the force is opposite the direction of the deformation “\( x \)”. So if you squish an object inward, it pushes back out. If you stretch it outward, it pulls back in.
How much work does such a force do? \( W = \int F \cdot ds \). Let's draw a picture of the magnitude of the force exerted by a spring at each point. We start with \( x=0 \) at the point where the spring is un-stretched.

![Diagram of force exerted by a spring](image)

The work done by the spring through some little displacement from \( x_i \) to \( x_f \) is the force there (-kx) times the distance over which it acts \( (x_f - x_i) \). This work is a little bit of the area beneath the curve shown. So if I add up all the work done by the spring between \( x=0 \) and \( x=x_f \), I will just get the area underneath the curve, or:

\[
W_{\text{spring}} = -\frac{1}{2}kx^2
\]

Note the minus sign. This means that as the object moves from \( x_i=0 \) to \( x_f = x \), the force is opposite the motion, always doing negative work on it.

How does this relate to the storage of energy in the spring? What is the 'potential energy' for the spring? We know by definition that:

\[
\Delta PE = -W
\]

so for a spring:

\[
\Delta PE \text{ (measured from } x=0 \text{ at the rest point!)} = \frac{1}{2}kx^2
\]

Now this is an interesting equation. Note that because it involves \( x^2 \), the potential is always positive. Any time I move away from \( x=0 \), I remove energy from the object and store it in the spring. Any time I either stretch or squash a spring, I store energy which the spring can then return to me by returning to its unstretched length.
Non-linear materials and stored energy

Earlier in the course, way back in Lecture #5, we noted that many biological materials deform under stress in ways which differ from the linear (Hooke’s Law) response described in the last section. How does this affect the energy stored in the deformation? The basic principle is exactly the same, with the energy stored in the deformation being equal to the work done to stretch the material. Let’s start by recalling the nature of the typical strain vs. stress curve:

![Graph showing strain vs. stress curve](image_url)

We called this, the “j curve” because of its shape. Remember, it means you can stretch the material a lot by applying just a small force, but that if you want to continue to stretch it further, you’ll have to apply a larger and larger force. What’s the energy stored in this kind of deformation?

Just like in the spring case, we find this energy by determining the work done by the material as we stretch it. The work going \( x_i \) to \( x_f \) around some average position \( x_{av} = (x_i + x_f)/2 \) is the force \( F(x_{av}) \) times the displacement \( \Delta x = x_f - x_i \). Adding this up from \( x=0 \) to \( x_f \) again just gives the area under the curve.

But notice how this differs from the Hooke’s law case. With this j curve shape, you can deform a material quite a lot (to large \( x \)) while the force \( F \) remains small. The area under the curve is tiny, and the energy stored in the material remains small. Why is this important?

Stored energy, fracture, and toughness

All of us have broken things: smashed a water glass, snapped a piece of string, knocked an icicle off the gutter. You may have noticed that when things break, it seems to happen “by itself”. When you throw a ball into a window, you don’t just punch a ball-sized hole in it; the whole thing shatters. How do things break, and how does this relate to the energy stored in deformation we’re just talking about?

To break a solid object you have to pull apart atoms which begin bonded together. Doing this requires energy, and that energy has to come from somewhere. As an example, here’s what happens when you drop a waterglass. It accelerates toward the floor. When it hits
the floor, the floor pushes up on it with a large force. This force deforms the glass, squashing it just as a ball deforms when it strikes the ground. Deforming the glass stores energy! That stored energy can be used to break the bonds between atoms in the glass. If enough energy is stored, there will be enough to send cracks propagating through the whole thing and it will shatter.

How about other things, like snapping the string or breaking the icicle? In these cases too, breakage occurs after you store energy in the material by deforming it. You stretch the string, or push the bottom of the icicle to the side. The energy you store by doing this is then released when the object breaks.

If you think a bit, you’ll realize that biological materials are a lot tougher than some of these manmade or crystalline materials. When you trip and fall you may bruise, but fortunately you don’t shatter. The j-curve is one of the important reasons for this. Many biological materials can stretch a lot while storing only a small amount of energy. This allows your tissue to deform smoothly while spreading the applied force around, avoiding the concentrations of stress which occur in stiffer materials.

There are other reasons why biological materials are especially tough. These ideas will be explored in an additional reading.

**Oscillations of a mass on a spring**

Consider the motion of an object on a spring. We start it out by stretching the spring to a maximum value of $x_i$ and releasing it from rest at that point. Imagine we arrange this so that no other forces do work on the mass. When I release the mass it oscillates back and forth through the equilibrium position. How is energy flowing in this system?

- When it starts all of the energy is in potential energy of the spring
- This is converted gradually to kinetic energy until, just as it passes through the equilibrium position, it is all kinetic
- Then it is traded back to potential as it comes to an instantaneous stop at the far end
- After this the process reverses and repeats

So this is a system in which energy is repeatedly traded back and forth between kinetic and potential forms. We’ll have a lot more to say about oscillations about equilibrium a little later in the class.
All matter is made of atoms. Despite appearances, these atoms are always in motion; rattling around, banging into one another. In a liquid or a solid, they repeatedly stretch and release the chemical bonds which hold them together. In a gas, they zoom around freely, crashing into each other once in a while, but not sticking together. This energy associated with this continual motion is what we measure and perceive as temperature.

Our principle goal for the next two weeks is to understand how thermal energy, the random motion of atoms related to their temperature, is ultimately just kinetic and potential energy. Once we have these basic ideas in place, we will spend several weeks working to understand how this thermal energy powers much of life.

We begin today by talking about collisions. Studying collisions will help us to understand the interactions between atoms, especially in gases. A gas is made of atoms or molecules which aren’t bound to one another at all, but rather just fly freely through space, occasionally colliding with one another. To understand a gas, it’s useful to consider collisions in general.

**Introduction to collisions**

What is a collision?

An event (a period of time) during which several bodies, isolated from others, interact relatively strongly. The interaction among these bodies during this period should be much stronger than the interaction between any body and something outside the system.

Some examples:

- Ball striking floor
- Baseball on bat
- Cars crashing
- Atoms colliding in a gas

Consider the case of two projectiles colliding in midair. Before the collision their motions are independent. During the brief period of the collision only the interaction between them is important. During that time, the force due to gravity can be briefly ignored. This is so because, during the collision, the forces associated with the collision are much larger than the force of gravity. After the collision they again move as free projectiles.
During the collision, the only thing which alters the momentum of the objects is the forces between them. By definition all other forces are unimportant there. Let’s put these collisions in a familiar language:

\[ \text{Impulse} = I = \int F \, dt = \Delta p \]

When two objects collide we know that \( F_{12} = -F_{21} \), just from Newton’s third law.

Since the forces are equal and opposite, we know the two impulses, and hence the two momentum changes, are equal and opposite. So:

\[ I_1 + I_2 = \Delta p_1 + \Delta p_2 = 0 \]

Whatever momentum change particle one experiences, particle two experiences and exactly equal and opposite momentum change. When you add up these two momentum changes, they perfectly cancel, and the total momentum doesn’t change at all.

Momentum is always conserved in collisions. It doesn’t matter at all what kinds of forces act. This is one of the facts about collisions which students often get confused about.

Why study collisions? Mostly because they happen a lot, interactions are often ‘contact’ interactions which occur only for short times. While they are occurring, they are usually vastly more important than other forces in the problem, making the collision approximation a good one.
A simple worked example: collision of two identical balls

What do we know about this? Momentum is always conserved, so:

\[ mv_{1i} = mv_{1f} + mv_{2f} \]

So what is \( v_{1f} \)? It’s impossible to say without further information. Even in this extremely simple case we can’t determine the outcome of the collision using only momentum conservation. We need to know something more.

One kind of additional information we might have about a collision is a statement about what happens to the energy. What could we say if the KE of the system is conserved?

\[ \frac{1}{2}mv_{1i}^2 = \frac{1}{2}mv_{1f}^2 + \frac{1}{2}mv_{2f}^2 \]

With this constraint, we now have two statements, that the initial velocity equals the sum of the final velocities, and that the square of the initial velocity is equal to the sum of the squares of the final velocities. These can only both be true if either \( v_{1f} \) or \( v_{2f} \) is equal to zero. Unless ball one moves through ball two, we must have \( v_{2f} > v_{1i} \), so the only solution is:

\[ v_{2f} = v_{1i} \quad v_{1f} = 0 \]

This is the special case where both momentum and energy are conserved in a collision which happens along a single straight line.

This example illustrates the problem; to understand a collision we have to know something more about the motion than just momentum conservation. In simple two body collisions like this there are two extreme cases to consider:

1. “Elastic” collisions: in these the forces which act between the two objects are purely conservative (like springs) and energy is conserved in the collision. In other words \( \Delta KE = 0 \).
2. “Totally inelastic” collisions: in these the forces which act are purely nonconservative (like friction). For these collisions the change in KE is as large as possible.

Most collisions are in between, and hence might be called partially inelastic.

How can we minimize the final kinetic energy in a linear collision like this (with two identical objects)? From momentum conservation we know that:

\[ v_{1i} = v_{1f} + v_{2f} \]

\[ v_{1f} = v_{1i} - v_{2f} \]

and
\[ \Delta KE \propto v_{i1}^2 - v_{i1}^2 - v_{2f}^2 = v_{i1}^2 - (v_{i1} - v_{2f})^2 - v_{2f}^2 = v_{i1}^2 - 2v_{i1}v_{2f} - v_{2f}^2 - v_{2f}^2 = -2v_{i1}v_{2f} - 2v_{2f}^2 \]

If we want to maximize the change in KE by picking the right \( v_{2f} \), we take the derivative and set it equal to zero:

\[ \frac{d\Delta KE}{dv_{2f}} \propto -2v_{i1} - 4v_{2f} = 0 \quad \text{or} \quad v_{2f} = \frac{v_{i1}}{2} \]

OK, that’s tells us what ball 2 is doing. What’s the other ball doing? \( v_{1f} = v_{i1} - v_{2f} = \frac{v_{i1}}{2} \)

So both balls move off at the same velocity. They stick together!

This is generally true. The way to lose the maximum amount of kinetic energy is to have the final objects stick together. So, in totally inelastic collisions, collisions which lose as much kinetic energy as possible while still conserving momentum, the objects always stick together.

**Collisions between unequal bodies**

Now consider the same linear collision, but between objects with different masses:

\[ m_1v_{i1} = m_1v_{1f} + m_2v_{2f} \]

Now if this is an elastic collision then:

\[ \frac{1}{2}m_1v_{i1}^2 = \frac{1}{2}m_1v_{1f}^2 + \frac{1}{2}m_2v_{2f}^2 \quad \text{or} \quad m_1(v_{i1}^2 - v_{1f}^2) = m_2v_{2f}^2 \]

factoring the left hand side we have:

\[ m_1(v_{i1} + v_{1f})(v_{i1} - v_{1f}) = m_2v_{2f}^2 \]

or

\[ m_2v_{2f}(v_{i1} + v_{1f}) = m_2v_{2f}^2 \]

or

\[ v_{2f} = v_{i1} + v_{1f} \]

putting this in with the momentum constraint above

\[ m_2(v_{i1} + v_{1f}) = m_1(v_{i1} - v_{1f}) \]

or

\[ v_{1f} = \left( \frac{(m_1 - m_2)}{(m_1 + m_2)} \right) v_{i1} \]

and so

\[ v_{2f} = \left( \frac{2m_1}{(m_1 + m_2)} \right) v_{i1} \]

To get an appreciation of what these equations are telling us, consider some limiting cases:

\[ m_1 >> m_2 \quad v_{1f} = v_{i1} \quad v_{2f} = 2v_{i1} \]
\[ m_1 = m_2 \quad v_{1f} = 0 \quad v_{2f} = v_{1i} \]
\[ m_1 << m_2 \quad v_{1f} = -v_{1i} \quad v_{2f} = 0 \]

The middle case is what we just did originally, and of course you get the same answer.

The last case is like what happens when a ball bounces off a wall. In this case the \( m_1 \) is \textit{much} less than \( m_2 \), and ball one just bounces straight back, leaving ball two sitting there unchanged.

The first case, where \( m_1 \) is much larger than \( m_2 \), is perhaps more surprising. In this case object one continues on essentially as if nothing had happened, while the little object two, which started at rest, jumps forward with basically twice the velocity of object one.

There are cases when you’ve probably seen this happen. Imagine what happens when you hit a ping-pong ball with your paddle. The paddle is swinging along at some speed, and it continues essentially unchanged after the collision. The ball, on the other hand, “jumps off” the paddle, moving ahead of it with a speed which, you now know, is about twice the speed of the paddle.

Another important example of this, which we will return to soon, is what happens when a piston pushes down on a gas of atoms. In the gas, atoms are always rattling around, bouncing off the walls. Normally, the walls are at rest, and they bounce off with exactly the velocities they arrive with. But if the wall is actually a piston, and it is moving toward the atoms, they bounce back (like the ping-pong ball) with a larger velocity than they came in with. As we will discuss in more detail soon, this increase in their velocities corresponds to a change in temperature of the gas. This is why the temperature of a gas increases when you compress it.

**Collisions involving non-contact forces**

You don’t have to have contact to have a collision. Consider the collision of a satellite with a planet. Gravity is all that acts, and this is conservative, and this will be an elastic collision.

\[ v_{\text{Jupiter}} \]
\[ v_{\text{Satellite}} \]
What is $v_{\text{satellite}}$ after the collision? There’s a trick to figuring this one out. Imagine you see this happen while sitting on Jupiter:

What you see from this vantage point is the satellite come in with velocity $v_s + v_{\text{Jupiter}}$, and you see it go back out with velocity $v_s + v_{\text{Jupiter}}$. Now just transform this back into the ‘rest frame’, where you again see Jupiter moving:

You should be able to see the close similarity to what we worked out before. In a ‘piston and atom’ collision, the atom leaves the collision gaining 2x the velocity of the piston. That’s all that’s happening here.

This kind of collision with a heavy particle can be used to accelerate a light one. The same thing happens with electrons and nuclei in a process known as ‘inverse Compton scattering’.

Now think about the reverse process:
What happens here is a kind of braking. These two techniques are called gravity ‘assists’. NASA and the European Space Agency ESA use these gravity assists to accelerate satellites up to the speeds they need to travel through the solar system. Without gravity assists of this kind it would be nearly impossible for us to get satellites to the outer planets like Jupiter and Saturn. They also use the reverse process (the gravity braking) to slow satellites down, though more often this is done using friction with the atmosphere of the planet.

**The ballistic pendulum: a classic physics problem**

Consider the following contrived situation in which a bullet strikes a target pendulum:

Could you figure out how the angle $\theta$ and the masses and initial velocity $v_i$ related? You can find this by solving the problem in steps:

1: Conserve momentum in the collision:
   $$m_bv_0 = (m_b + m_t)v_1 \quad v_1 = m_b/(m_b+m_t)v_0$$

2: Conserve energy as the pendulum rises:
   $$\frac{1}{2}(m_b+m_t)v_1^2 = (m_b+m_t)g(h_f-h_i)$$
   or
   $$v_1^2 = 2g(h_f-h_i) = (m_b/(m_b+m_t)v_0)^2 \quad \text{or} \quad v_0 = [(m_b+m_t)/m_b]\sqrt{2g(h_f-h_i)}$$

Is energy conserved in the collision? No, it can’t be, because they stick together. The important point is that we don’t really care. While most people encounter this as an example in their physics class, this method is actually used to measure the muzzle velocity of bullets from guns sometimes.

**How to solve collisions in 3D**

So far we’ve only worried about collisions in one dimension. What happens if we try to extend this analysis to 3D?

What do we know? Momentum is conserved, so: $\Sigma p_i = \Sigma p_f$. This yields three equations:

$$\Sigma p_{ix} = \Sigma p_{fx} \quad \Sigma p_{iy} = \Sigma p_{fy} \quad \Sigma p_{iz} = \Sigma p_{fz}$$

What else do we know about the collision? Typically nothing, but we might know something like
Physics 135 Fall 2009: Lecture Notes

- It’s elastic: \( \Delta KE = 0 \)
- The objects stick together

For two objects there are many things we need to know:
- 2 masses
- 2 \( \mathbf{v}_1 \) (3 components each)
- 2 \( \mathbf{v}_f \) (3 components each)

So there are, effectively, 14 things we need to know. Momentum gives us three equations. Some statement about energy may give a fourth. So to fully solve the problem we will still need to be supplied with a lot of information.

An example:

\[ \theta_1 \]

\[ \theta_2 \]

Imagine we are given here: \( m_1, m_2, \mathbf{v}_{1x} \) and told that all of
\[ v_{1y} = v_{1z} = v_{2x} = v_{2y} = v_{2z} = v_{1x} = 0 \]

So we’re given the ten things. Then if it’s elastic, we have:
\[ m_1 v_{1x} = m_2 v_{2x} \cos \theta_2 + m_1 v_{1f} \cos \theta_1 \]
\[ 0 = m_2 v_{2f} \sin \theta_2 + m_1 v_{1f} \sin \theta_1 \]

and
\[ \frac{1}{2} m_1 v_{1x}^2 = \frac{1}{2} m_1 v_{1f}^2 + \frac{1}{2} m_2 v_{2f}^2 \]

But this is 13 things given, and a total of 14 unknowns.

What can we say about this? Well, if the masses are the same then I have three equations:
\[ v_{1x} = v_{1f} + v_{2f} \] or \( v_{1x}^2 = v_{1f}^2 + 2 v_{1f} v_{2f} + v_{2f}^2 \)
\[ 0 = v_{1y} + v_{2y} \] or \( 0 = v_{1y}^2 + 2 v_{1f} v_{2y} + v_{2y}^2 \)
\[ v_{1x}^2 = v_{1f}^2 + v_{2f}^2 \] or \( v_{1x}^2 + 2 v_{1f} v_{2f} + v_{1f}^2 + v_{1f}^2 + v_{2f}^2 + v_{2f}^2 \)

which leaves:
\[ 2 v_{1f} v_{2f} = v_{1y}^2 + v_{2y}^2 \]

rearranging the second equation gives
\[ v_{1y}^2 + v_{2y}^2 = -2 v_{1f} v_{2f} \]

so
\[ 2 v_{1f} v_{2f} = 2 v_{1f} v_{2f} \] or \( v_{1f}/v_{1y} = -v_{2f}/v_{2f} \) or \( \tan \theta_1 = -\cot \theta_2 \)

Recalling that \( -\cot(\theta) = \tan(90^\circ - \theta) \) we have
\[ \tan \theta_1 = \tan(90^\circ - \theta_2) \] or \( \theta_1 = 90^\circ - \theta_2 \) or \( \theta_1 + \theta_2 = 90^\circ \)

In collisions between two equal mass objects, they will move off at \( 90^\circ \) relative to one another.

More important though, this general problem still can’t be solved. All I just showed was that the final angles would be related. I wasn’t able to say what either angle actually was. If, however, I know one of the two angles \( \theta_1 \) or \( \theta_2 \), I can solve the problem completely.
Very often, then, collisions like this are solved by understanding the distribution of results which will occur as I vary the angle $\theta_1$ which the input particle comes out with.

So basically I can’t tell you what will happen in a collision without more information even than just momentum and energy conservation. But I can tell you that if you see particle 1 come out at an angle $\theta_1$, you will completely determine the motion of particle 2.

**Final example: 2 body decay and the neutrino**

As a final example, consider the decay of a neutron. In 1928 this was thought to involve a neutron turning into a proton and an electron.

\[
\begin{align*}
\text{KE}_0 &= 0 \\
\text{p}_0 &= 0
\end{align*}
\]

It was known at the time that the energy released in this “decay” came from “mass energy”, the kind of energy described by Einstein’s most famous equation $E = mc^2$. Since the mass of the neutron is a little bigger than the sum of the mass of the proton and the mass of the electron, energy is available in the neutron to release into the proton and the electron after the decay. The amount of available energy is:

\[
E_{\text{decay}} = m_n c^2 - (m_p + m_e) c^2
\]

We know that momentum is conserved, and believe that energy is conserved:

\[
0 = m_e v_e - m_p v_p \quad \text{or} \quad v_e = \frac{(m_p/m_e)v_p}{1 + m_p/m_e} \quad \text{or} \quad v_p = \frac{(m_e/m_p)v_e}{1 + m_e/m_p}
\]

\[
E_{\text{decay}} = m_n c^2 - (m_p + m_e) c^2 = \frac{1}{2}m_e v_e^2 + \frac{1}{2}m_p v_p^2
\]

Obviously they should come out back to back. What will be the energies of the proton and electron?

\[
E_f = \frac{1}{2}m_e v_e^2 \quad \text{or} \quad E_{\text{proton}} = \frac{E_f}{1 + m_p/m_e}
\]

\[
E_f = \frac{1}{2}m_p v_p^2 \quad \text{or} \quad E_{\text{electron}} = \frac{E_f}{1 + m_e/m_p}
\]

This is required by momentum and energy conservation. It says that if you measure the energy of electrons emerging from such a decay, you should find just one value for their energy. What do you really find?

![Expected and Observed](image-url)
This was a big problem for physics. How could they explain this? To Enrico Fermi, there seemed to be three choices:

1. Momentum is not conserved (a terrible possibility)
2. Energy is not conserved (even more terrible)
3. There is a third (undetectable) body in the collisions (surprising, but nicer...)

With three bodies in the decay, $E_{\text{electron}}$ can be anything from 0 to $E_{e}\text{max}$. This ‘missing’ particle, the neutrino, was predicted by Fermi to exist to fill this gap. He called it a neutrino because it was a “little neutral one”, and he was Italian. The neutrino was not observed independently for 27 years after Fermi’s prediction of it. Why so long? Because neutrinos hardly ever interact. Right now there are about $10^{11}/\text{cm}^2\text{s}$ passing through you. Essentially none interact.

How do we know they’re real? Well they sometimes do interact, and by building large enough experiments, and watching carefully enough, we can observe them. But Fermi was able to divine their existence by analyzing the decay of the neutron and insisting that momentum and energy should be conserved.
Physics for the Life Sciences: Fall 2008 Lecture #13

For the next two classes we’re going to spend time describing in some detail the motion of objects which oscillate around an equilibrium position. This kind of oscillation happens whenever an object held in place by some force (like an atom in a solid) is disturbed a little, moved away from its equilibrium position. When this happens, it gets pulled back toward that position of rest, often overshooting it and then oscillating back and forth around it.

Why is this kind of motion important for life?

The most basic reason is that it is intimately connected with temperature and thermal energy. When we say that a solid is hot, and has a temperature $T$, what we mean is that the atoms in that solid are moving. They aren’t going anywhere on average, they’re just oscillating around some average rest position. If the temperature is higher, they’re moving more, and in fact the average energy associated with this shaking around is just proportional to the temperature. If we’re going to understand the flow of energy through living systems, which is what makes all of life possible, we need to understand a lot more about this.

There is another important aspect of oscillating systems relevant for life. When oscillating systems are disturbed by outside forces, they exhibit a behavior called ‘resonance’. Their response to action from the outside is selective. If you disturb them in just the right way you may get an extremely strong response, while other disturbances have little impact. This phenomenon of resonance is used in many ways by living things, so it’s another aspect of oscillations we need to introduce.

Harmonic Motion

Most motions in the real world are neither constant velocity of constant acceleration. In fact most objects spend their time sitting at equilibrium. Motion of an object around equilibrium will be oscillatory. We’ll review why in a minute. To describe this I want to talk about bit about the ‘kinematics’ of harmonic motion. What kind of language do we have to use to describe motion which repeats over and over?

Any motion which repeats itself we will call harmonic (or periodic) motion. Examples include:

- Pendulum or swing
- Ball on spring
- Rotation of Earth
- Orbit of moon
- Motions of and sounds produced by musical instruments
- Motion of atoms
- Water waves
The basic feature of all of these things is that they have some measurable property (position, velocity, height, electric field strength, density, pressure, temperature….) which goes through a pattern of change with time that is repeated.

Does such oscillatory motion usually continue forever? No, usually it gradually fades away. We’ll get to this ‘damping’ next time.

**Describing periodic motion**

To describe any kind of periodic motion we have to specify a few minimal things:

1. How often does it happen? **Frequency** = $f =$ how many times a second does the pattern repeat. The units of frequency are 1/s, which are called Hertz for Heinrich Hertz, the great experimentalist who showed that light was a traveling electromagnetic oscillation.
2. How long does each cycle take? **Period** = $\Gamma =$ how long does each cycle take. The period has units of seconds and is just the inverse of the frequency: $\Gamma = 1/f$.
3. How large is the oscillation? **Amplitude** = $A$ describes this. It has units of whatever is oscillating.

This description is completely general, not referring to any particular kind of oscillatory motion, but just to oscillations in general.

**Oscillations around equilibrium**

Equilibrium points are those at which an object feels no force. There are three kinds of equilibrium points:

- **Stable**: if you move the object a little away from equilibrium, it is pushed back towards it (the bottom of a valley)
- **Unstable**: if you move the object a little away from equilibrium, it is pushed farther away from it (the top of a hill)
- **Neutral**: if you move the object a little away from equilibrium, it isn’t pushed either back toward or farther away from it (the middle of a flat plane)

Just how unstable, how stable, etc depends a lot on the details of the system, but the nature of the equilibrium points does not.

**Simple Harmonic Motion**

One particular kind of harmonic motion is what’s called “Simple Harmonic Motion”. In this kind of motion the position of the object varies sinusoidally with time. We can describe this motion as:

$$x(t) = A\cos(\omega t)$$
Once we know this, we can find the velocity and acceleration of the particle from their definitions:

\[
v(t) = \frac{dx(t)}{dt} = \frac{d}{dt}(A\cos(\omega t)) = -A\omega \sin(\omega t)
\]

\[
a(t) = \frac{dv(t)}{dt} = \frac{d}{dt}(-A\omega \sin(\omega t)) = -A\omega^2 \cos(\omega t)
\]

These three are all shown in the figure on the next page. There’s something to notice about this. The equation for the acceleration is almost the same as the equation for the position:

\[
a(t) = -A\omega^2 \cos(\omega t)
\]

\[
x(t) = A\cos(\omega t)
\]

so

\[
a(t) = -\omega^2 x(t)
\]

What are the properties of this motion? We said every harmonic motion should have a frequency, period and amplitude. For this motion, the frequency is encoded in this “angular frequency” \(\omega\). It’s related to the regular frequency according to:

\[
f = \frac{\omega}{2\pi}
\]

and the period

\[
\Gamma = \frac{2\pi}{\omega}
\]

How can you see this? When the argument of the cosine function goes from zero to \(2\pi\), the value of the cosine goes from 1 through a cycle to \(-1\) and returns to 1. Since the period is the time this takes to happen, \(\omega \Gamma = 2\pi\), or \(\Gamma = 2\pi/\omega\). Meanwhile, the amplitude of oscillation is the parameter \(A\) in front of the equation.

An example, if the equation of motion is \(x(t) = (12.5\text{m})\cos(23t)\), then the amplitude of motion is 12.5. That is, the object oscillates between +12.5m and -12.5m. The angular frequency of oscillation is 23 rad/s. From this we can find the frequency of oscillation \(f = \frac{\omega}{2\pi} = 11.5/\pi\) Hz and the period of oscillation is \(\Gamma = \pi/11.5\) seconds.
Linear, Hooke’s law like forces cause simple harmonic motion

Remember the Hooke’s law force, in which the force returning an object to equilibrium gets larger as you move farther away from equilibrium. The equation which describes this is:

\[ F = -kx \]

We also know from Newton’s second law that, if this is the only force that acts, we can write

\[ F = ma = -kx \]

Or

\[ a = -(k/m)x \]

Notice that this looks very similar to the equation we wrote down relating position and acceleration for simple harmonic motion:

\[ a = -\omega^2 x \]

So, an object moving under the influence of a linear Hooke’s law force will move in simple harmonic motion with an angular velocity \( \omega^2 = k/m \) or \( \omega = \sqrt{(k/m)} \).

Notice what this means. This object will oscillate rapidly, with high frequency, if the spring constant \( k \) is large. If this \( k \) is small, the oscillations will be slower. But it’s not just the strength of the restoring force which sets this. The inertia of the object experiencing the force affects this too. If the mass of the moving object is large, it will slow the frequency of oscillation. If the mass is small, it will oscillate more quickly.

A key point is that the frequency will always be determined by a balance of inertia and restoring force. This general relation is true even if the restoring force is not linear. Increasing the strength of the restoring force (\( k \)) will always speed the oscillations, while increasing the inertia (\( m \)) will always reduce them.

Any restoring force will cause HM, SHM if it is linear…

Referring to Hooke’s law tends to make people think that the only time you get simple harmonic motion is with a mass on a spring. This isn’t the case at all. In fact any time an object is at a point of stable equilibrium you can get this kind of motion. There are limits of course, and they’re basically limits on how big the amplitude of oscillation can be. If you make an atom in a solid oscillate too far from its position of rest, it will break out of the solid and not oscillate at all.

But for small enough oscillations, the motion around any position of rest will be simple harmonic motion.
Examples of harmonically oscillating systems

Mass on spring: \( F = -kx \) this is just what we’ve already described: balance between inertia of mass and spring strength. \( \omega = \sqrt{\frac{k}{m}} \)

A Simple Pendulum (where the mass is all at the bottom, a distance \( L \) from where it swings):

In this case we’re going to consider only small oscillations around equilibrium, where the angle \( \theta \) never becomes large. In this case we’ll be able to use the small angle approximation, which says that for small angles (remember, this only works for angles measured in radians!):

\[
\sin(\theta) \approx \theta
\]

How small do the angles have to be for this to be a good approximation? We can check by putting in different values. What we find is that it’s a good, better than 5%, approximation to surprisingly large angles, like \( 30^\circ \).

<table>
<thead>
<tr>
<th>Angle</th>
<th>Sin(\theta)</th>
<th>( \theta )</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi/16 ) (11.25\°)</td>
<td>0.1951</td>
<td>0.1935</td>
<td>0.8%</td>
</tr>
<tr>
<td>( \pi/8 ) (22.5\°)</td>
<td>0.3827</td>
<td>0.3927</td>
<td>2.6%</td>
</tr>
<tr>
<td>( \pi/4 ) (45\°)</td>
<td>0.7071</td>
<td>0.7854</td>
<td>10.0%</td>
</tr>
</tbody>
</table>

With that in hand, let’s work out how big the force is which returns the pendulum ball to its rest position. Referring to the drawing at the right, we can see that the force which pulls it back to the middle is approximately

\[
F_{\text{return}} = -mg\sin \theta = -mg \theta = -mg(\sin \theta)
\]

Along this line we used the small angle approximation \( (\sin \theta) = \theta \) and the definition of the angle \( \theta = s/L \) in radians. We can rewrite this as:

\[
F_{\text{return}} = -(mg/L)s
\]

What about the direction of this force? You can see from the picture that it if \( s \) is positive (you push it to the right), the force will be negative (it will be pushed back left) and vice versa. This is a restoring force. That’s the origin of the minus sign in front of the equation.

This force law looks exactly like Hooke’s law, with a ‘spring constant’ \( k_{\text{effective}} = (mg/L) \). This means the pendulum will show oscillations in the offset parameter \( s \) like:

\[
\omega = \sqrt{\frac{k_{\text{effective}}}{m}} = \sqrt{\frac{g}{L}}
\]

This one is funny, because the pendulum will oscillate the same way no matter what mass we put on it. Why? Because both the inertia and the restoring force depend on mass. Any gravitationally restored oscillator will share this property.
Notice some things about what this says; the period of oscillation of a pendulum is only dependent on its length. Not on its mass, the material it’s made of, or anything; just its length. This is why, throughout your childhood, you could swing on a swingset ‘in synch’ with your friends, or even your parents, regardless of how much each of you weighed.

Oscillating rod:

Another example is the oscillating rod. In this case you “shear” the rod sideways by some small amount $\Delta x$, then release it and let it oscillate. Here you might get

\[ \frac{F}{A} = -S \Delta L / L \] (where $S$ is the ‘shear modulus’)

Or \[ F = -(S A/L) \Delta L \]

Here you have a $k_{\text{effective}} = (S A/L)$. This means it will oscillate with:

\[ \omega = \sqrt{k_{\text{effective}} / m} = \sqrt{(S A/L) m} \]

In fact, we can write this another way, using the fact that $m = \rho V = \rho A L$ to find:

\[ \omega = \sqrt{(S A/L) m} = \sqrt{(S A/L \rho A L)} = \sqrt{(S/\rho L^2)} \]

Is this surprising? A stiffer rod, with high shear modulus $S$, will oscillate faster. A longer rod will oscillate more slowly. If the rod is denser, it will also oscillate more slowly.
Physics for the Life Sciences: Fall 2008 Lecture #14

Today we’ll extend the discussion of these harmonic oscillators in several additional ways. First, we’ll talk about energy in oscillators, and how when an object is oscillating back and forth it’s repeatedly trading kinetic and elastic potential energy. Then we’ll talk about what happens when there is friction in an oscillator, gradually draining away the energy which would be steadily trading between kinetic and potential.

Finally we’ll talk about the surprisingly important case of an oscillator driven by a force which is itself periodic. This seems like a bizarre and unlikely case, but is in fact incredibly common. It comes about because you so often have one oscillator next to another, as you do with atoms in a solid. If you disturb one atom, it starts to oscillate. In doing this, it bangs repeatedly, and periodically, into its neighbors. This kind of periodic driving is why knowing what happens to an oscillator when something pushes on it regularly is so important.

Energy in Harmonic Oscillators

When an object oscillates it takes energy stored in the elastic stretching of its ‘spring’ and converts this to kinetic energy, shooting through equilibrium, and then turns that kinetic energy back into elastic potential. Then the process repeats over and over.

What can we calculate about the energy?

\[ KE = \frac{1}{2}mv^2 = \frac{1}{2}m[A^2\omega^2\sin^2(\omega t)] = \frac{1}{2}kA^2\sin^2(\omega t) \]
\[ PE = \frac{1}{2}kx^2 = \frac{1}{2}kA^2\cos^2(\omega t) \]

All we did to calculate these was to plug in the equations described before for \( x(t) \) and \( v(t) \). Energy oscillates from KE to PE and back again. Notice that the PE is at a maximum when the position \( x \) is at a maximum; when the spring is fully stretched out. The kinetic energy, on the other hand, is at a maximum when the velocity is at a maximum. This happens just when the object passes through equilibrium.

Notice something else; neither the kinetic or potential energy involved in this oscillation depends on the mass of the object on the spring! At first this is surprising. Wouldn’t a brick oscillating on a spring carry more energy than a tennis ball on the same spring? The solution to this paradox is clear when you think about the details. First, the energy you put into the system shows up at one point entirely in the stretching of the spring. At this point, all the energy is there in the spring. This stored spring energy will then be converted to kinetic energy in the object, but the amount of energy depends only on how much you stretch the spring. If you have a brick on the spring, it will convert all this into potential energy, but can do this while traveling more slowly. If it’s a tennis ball, it can still take all the energy but will have to travel faster. So the energy in the oscillating system depends only on the amplitude and spring constant, and not on the mass of the object which is moving.
Here’s an example of the trading of energy in such an oscillation:

![Diagram of oscillation with KE and PE](image)

What is the total energy during this cycle?

\[
KE + PE = \frac{1}{2}kA^2\sin^2(\omega t) + \frac{1}{2}kA^2\cos^2(\omega t)
\]

\[
= \frac{1}{2}kA^2(\sin^2(\omega t) + \cos^2(\omega t)) = \frac{1}{2}kA^2
\]

In this last line we used the trigonometric identity \(\sin^2 A + \cos^2 A = 1\). This is really interesting! It says that the energy stored in an oscillator is \(\text{constant}\). That’s not too surprising, as we imagined that no other forces act. More important, it says that the total energy is determined by \(k\) (a property of the oscillator) and \(A\) (a property of the oscillation). Notice that it doesn’t depend \(\text{at all}\) on the mass of the object attached to the oscillator! That might be a little surprising, but if you remember that all the energy is really coming from this restoring force it becomes perhaps less so.

**First variation: damped oscillations**

The first important correction involves introducing a quick dose of reality. No known systems oscillate forever. They all eventually stop. This happens because friction always resists the motion, gradually draining away the energy which was so happily trading between kinetic and potential forms.

If there is friction around in an oscillator, it may have either a very small effect on the system (simply making the oscillations very gradually fade away), or it may have a very large effect, preventing oscillations completely and just barely allowing the restoring force to gradually drag the object back to equilibrium. Before doing any mathematical analysis, let’s consider what these different limits might look like.

**Weak damping**

Imagine first the case of friction which is relatively weak. Weak relative to what? If, during a typical oscillation, the size of the frictional force is always small compared to the size of the restoring force, the friction is weak. Another way to say this is to insist that
the amount of energy drained from the system during one cycle should be small compared to the overall energy of the system.

Systems like this are called “underdamped”. They will oscillate many times before coming to rest, each time losing some small fraction of their energy to friction. So in these systems, we expect to see the amplitude of oscillations gradually decrease.

Does this damping affect the frequency of oscillation? In fact it does. Think about the first cycle. Without damping, the restoring force would accelerate the mass back toward equilibrium and it would take some time to get there. With damping, there is a resistive force always working against the restoring force. This resistive force slows the motion, making the frequency of oscillation lower.

**Overdamped motion**

If the frictional force is really large, the system will never oscillate at all. This is what happens. You pull the object away from equilibrium and let it go. The restoring force starts to pull it back to equilibrium, the object speeds up a little, and suddenly the frictional force is large enough to completely balance the restoring force. When this happens, the object moves along at constant speed; in a motion completely analogous to terminal velocity. It’s a little different from that case though, because the restoring force isn’t constant. As the object gets closer to equilibrium, the restoring force gets smaller. The frictional force needed to balance that weaker restoring force is smaller, and the ‘terminal velocity’ is smaller.

This balance gives rise to a gradual, smooth return to equilibrium, without any overshooting or oscillations. You pull it away from equilibrium, and it just gets smoothly dragged back.
Notice how in this case the slope of this position time graph starts out steep (with large negative velocity) and then becomes more and more shallow (with velocity decreasing). This shows how a linear restoring force differs from the terminal velocity case we considered before.

**Working out a mathematical model for damped oscillators**

How can we include this in our mathematical model for this motion? What we have to do is add some new term to the force equation and see what we get. One possibility is that the friction would be proportional to the speed $v$, and always in a direction opposite to it.

Such a friction force would have the form:

$$F_f = -bv$$

Where $b$ is just some parameter describing how large this friction is. For a small sphere moving slowly in a fluid this would be: $F_f = -12\pi\eta Dv$ where $D$ is the diameter of the sphere and $\eta$ is the “viscosity”, kind of the thickness, of the fluid. For this kind of friction, we’d have $b_{\text{eff}} = 12\pi\eta D$.

If we happen to have friction of this form:

$$F = -kx - bv$$

Here the friction is proportional to $v$. This is approximately true for many systems, but is only one possibility. The details of other possibilities will be a little different, but this will give us a good example.

To solve this we can write it as a ‘differential equation’. Don’t worry if you don’t know how to do this, I just want to show you how it works in general.

$$F = ma = m(d^2x/dt^2) = -kx - b(dx/dt)$$
Rearranging we find:
\[ m(d^2x/dt^2) + b(dx/dt) + kx = 0 \]
or
\[ d^2x/dt^2 + (b/m)dx/dt + (k/m)x=0 \]
which we often write:
\[ d^2x/dt^2 + (b/m)dx/dt + \omega_0^2x=0 \]

A “solution” to this linear differential equation is just some function \( x(t) \) which obeys this equation. The general form for this solution can be written:
\[ x(t) = Ae^{-bt/2m}\cos(\omega't+\phi) \]

Where
\[ \omega' = \sqrt{(\omega_0^2 - (b/2m)^2)} \]

How do you find this solution? It turns out there are a variety of techniques for finding the solutions to differential equations. You’ll learn about them when you take more advanced math courses. Let’s check to see that this solution works. First, find the derivatives:
\[
\frac{dx}{dt} = -(b/2m)Ae^{-bt/2m}\cos(\omega't+\phi) - A\omega'e^{-bt/2m}\sin(\omega't+\phi)
\]
\[
\frac{d^2x}{dt^2} = (b/2m)^2Ae^{-bt/2m}\cos(\omega't+\phi) + (b/2m)\omega'Ae^{-bt/2m}\sin(\omega't+\phi)
\]
\[ + (b/2m)\omega'e^{-bt/2m}\sin(\omega't+\phi) - A\omega'^2e^{-bt/2m}\cos(\omega't+\phi)
\]
\[ = (b/2m)^2Ae^{-bt/2m}\cos(\omega't+\phi) + (b/m)\omega'Ae^{-bt/2m}\sin(\omega't+\phi)
\]
\[ - A\omega'^2e^{-bt/2m}\cos(\omega't+\phi) \]

Now put these into the differential equation above:
\[ \frac{d^2x}{dt^2} + (b/m)\frac{dx}{dt} + \omega_0^2x=0 \]
\[ (b/2m)^2Ae^{-bt/2m}\cos(\omega't+\phi) + (b/m)\omega'Ae^{-bt/2m}\sin(\omega't+\phi) - A\omega'^2e^{-bt/2m}\cos(\omega't+\phi)
\]
\[ - (b/m)(b/2m)Ae^{-bt/2m}\cos(\omega't+\phi) - (b/m)A\omega'e^{-bt/2m}\sin(\omega't+\phi) + \omega_0^2Ae^{-bt/2m}\cos(\omega't+\phi) = 0 \]

simplifying
\[ -(b/2m)^2 - \omega'^2 + \omega_0^2]Ae^{-bt/2m}\cos(\omega't+\phi) = 0 \]

You can see that equation will hold if \( \omega'^2 = \omega_0^2 - (b/2m)^2 \). So this solution works. It will describe the motion of an object under the influence of the force \( F = -bv - kx \).

There are two things to note about this solution. First, the amplitude of the oscillation decreases exponentially with time. This is shown by the term \( Ae^{-bt/2m} \) in front of the oscillatory cosine function.

Second, the frequency of the oscillations is altered from the undamped (\( b = 0 \)) case, decreased by an amount that depends on damping. This coefficient \( b/m \) is sometimes
called \(1/\tau\), where \(\tau\) is the ‘decay time’ of the system. This is because the amplitude of oscillation \((Ae^{-t/(2\tau)})\) drops by a factor of \(e^{-1/2}\) in time \(\tau\).

These solutions have three different “regimes”, three different general behaviors which depend on the specific choices for the parameters \(b\), \(m\), and \(k\):

- **Underdamped:** \(k/m > (b/2m)^2\) or \(\omega_0^2 > (b/2m)^2\)
- **Overdamped:** \(k/m < (b/2m)^2\) or \(\omega_0^2 < (b/2m)^2\)
- **Critically damped:** \(k/m = (b/2m)^2\) or \(\omega_0^2 = (b/2m)^2\)

Think about the overdamped case. Since in this case \(\omega_0 > b/2m\):
\[\omega' = \sqrt{(\omega_0^2 - (b/2m)^2)} = \sqrt{\text{negative number}}\]

Here we have \(\omega'\) being the square root of a negative number! So we will have
\[\omega' = \pm i\sqrt{(b/2m)^2 - \omega_0^2} = \pm iS\]
and since
\[\cos \theta = 1/2(e^{i\theta} + e^{-i\theta})\]
we have
\[\cos(\omega't + \phi) = 1/2(e^{i(St+\phi)} + e^{-i(St+\phi)}) = 1/2(e^{-St}e^{i\phi} + e^{St}e^{-i\phi})\]

This makes the whole solution:
\[x = A'e^{-bt/2m} + B'e^{-bt/2m} = A'e^{-\mu t} + B'e^{-\kappa t}\]

(the addition factors \(e^{i\phi}\) and \(e^{-i\phi}\) get absorbed into the constants \(A'\) and \(B'\) with:
\[\mu = (b/2m - \sqrt{(b/2m)^2 - \omega_0^2})\]
\[\kappa = (b/2m + \sqrt{(b/2m)^2 - \omega_0^2})\]

In other words, a decaying double exponential function. In this solution the displacement just decays away to zero.

Think about the critically damped case. Here we have:
\[\omega' = \sqrt{(\omega_0^2 - (b/2m)^2)} = 0\]
so the full solution is just:
\[x = A'e^{-bt/2m}\]

This one is just a pure single exponential decay.

In the third, underdamped case, \(\omega'\) is a real, positive number, and the solution is the full:
\[x = Ae^{-bt/2m}\cos(\omega't + \phi)\]
In this case there are two changes from ‘normal’ undamped oscillations. First, the amplitude decreases exponentially with time. In a sense we can say:
\[A' = Ae^{-bt/2m}\]
The second change is that the frequency of oscillation is lowered (and hence the period lengthened). \(\omega' = \sqrt{(\omega_0^2 - (b/2m)^2)} < \omega_0\)

As \((b/2m)^2\) gets closer to \(\omega_0^2\), the frequency of oscillation falls, until at the limit \(\omega_0^2 = (b/2m)^2\), we have zero frequency and an infinite period. So remember that two things happen with damped oscillations;

- The amplitude drops exponentially with time

Copyright: Regents of the University of Michigan, 2009
The frequency is shifted lower

Now this set of mathematics give specific predictions for the particular case where friction is proportional to velocity. But these basic facts, that in a damped oscillator the amplitude will gradually drop to zero and the frequency will be shifted lower, are true for any kind of damping. That’s why we introduced the basic ideas before doing this one mathematical example.

Natural frequencies of oscillation

We have seen that oscillators, if disturbed, will oscillate with a frequency that is determined by a balance of the strength of the restoring force and the inertia of the system. In the derivation above, we have added to that an adjustment based on the strength of the damping experienced by the system. But it is still the case that, for any oscillator, there is a ‘natural frequency’ at which it will oscillate if you disturb it. How do you find this natural frequency? This is simple, just disturb it and watch to see what frequency it oscillates with. For example, if you want to know the natural frequency of a playground swing is, you just get on and push off. You’ll oscillate back and forth and can measure the natural frequency.

An additional variation: driven oscillations:

As we mentioned at the start, it is often important to consider what will happen to an oscillator if you hit it with a periodic force. Most often this happens because two oscillators are sitting next to one another (like two atoms in a solid) and once one starts oscillating it repeatedly strikes the next.

How do we handle this mathematically? We could start with the equation we wrote for a damped oscillator and now include a new ‘driving’ force in the equation.

\[ F = ma = md^2x/dt^2 = -kx - bdx/dt + F_{driving} \]

So

\[ md^2x/dt^2 + bdx/dt + kx = F_{driving} \]

or

\[ d^2x/dt^2 + (b/m)dx/dt + \omega_0^2x = F_{driving}/m \]

What should we take for this driving force? The answer of course depends on what we're doing. An interesting case is:

\[ F_{driving} = F_0 \cos(\omega_0 t) \]

Solving this is a complicated business, best left to a mathematical methods class. The general solution is variable in time. But there is a simple, physically motivated prediction we can make. If we wait long enough, the system will settle down and oscillate with the frequency of the driving force.
Why is this? If the system starts out oscillating with some frequency that’s not equal to the driving frequency, this part of the oscillation will eventually die out due to damping. These oscillations must die out because there is nothing in the system putting new energy into these oscillations. What won’t die out are those oscillations which happen at the frequency of the driving force, because these oscillations are continually restored.

So the ‘long time’ solution will be a final oscillation which has the frequency of the driving force. How long it takes to settle down depends on the size of the damping. If the damping is large, this will happen quickly. If the damping is small, this may take a long time. But we will always end up with oscillations at angular frequency $\omega_d$, a fairly simple and perhaps not too interesting solution. What is interesting is the surprising relation between the amplitude of the final oscillation, the frequency of the driving force $\omega_d$, the natural frequency of the undriven oscillator $\omega'$, and the size of the damping $b$.

You probably know all about this from your playground experience. If you go to push your friend Mary on a swing, you know that to build up a nice, large amplitude motion, you have to push at the same frequency with which Mary would naturally swing. If, for example, you push her back and forth really quickly, like five times a second, she will shake back and forth, but she’ll never build up any large amplitude. Likewise, if you push her back and forth very slowly, say once every 30 seconds, again, she will build up no large amplitude. But if you push her at just the right rate, at just the natural frequency of the swing, large amplitudes will quickly appear.

How do you know what frequency to push with, and moreover, how do children know this in kindergarten? It’s easy. You want to push with the natural frequency of the swing. To find this, you just give a shove and watch Mary swing. Once you see this natural “free” oscillation, you know how often to push.

How large an oscillation will we get? If we drive the system at its "natural frequency" $\omega'$ (from above) then it oscillates with large amplitude; how large depends on the damping. If we drive it at a frequency different from this, it oscillates at the new frequency, but with smaller amplitude; essentially the oscillation always fights the driving force. The relation between the frequency of the driving force and the amplitude of the final oscillation is called the resonance curve, which shows what final amplitude you’d get if you drove the oscillator at each driving frequency.
Not surprisingly, if you have damping, the maximum amplitude you can reach while driving the oscillations is less. In fact the shape of the ‘resonance curve’ changes in several ways as the damping increases. First, the peak frequency (the natural frequency) shifts lower. This is just the effect we saw already ($\omega^2 = \omega_0^2 - b^2/2m^2$). Second, the maximum amplitude of oscillation decreases. Since energy is being drained out more quickly, you just don’t build up as large an amplitude. Finally, the overall width of the observed resonance curve increases with damping.

So for driven systems, you get oscillations at the driving frequency, and their amplitude depends on the relation of the driving frequency to the "natural" frequency of the system. If there is a lot of damping, oscillations will never be large. If there is little damping, and you drive the system at the right frequency, the amplitude can be very large.

**Examples of resonance**

The playground swing is the most obvious and familiar. There are many similar examples though. If your Frisbee is stuck in a small tree you might shove the tree repeatedly, trying to shake it loose. In doing this, you will without thinking about it note the natural period of oscillation of the tree and shove in synch with this. Working in this way allows you to drive the oscillations of the tree to large amplitude, and hopefully knocks loose your Frisbee.
Another familiar example involves walking with a cup of coffee. You can measure the natural frequency of the system $\omega'$ can be measured by disturbing it. If you walk at that frequency, you will drive these oscillations in resonance, and the coffee will start sloshing and may spill all over you.

NaCl ionic crystal:

One kind of oscillation this might have is when you move all the + charges to the right and all the - charges to the left. How do you make this happen? If you have an oscillating electric field it will push the positive charges one way and the negative charges the other. What is an oscillating electric field? Light!

So if I shine light of various frequencies on this salt, it will cause oscillations of varying sizes. The appearance of these oscillations means that the light is absorbed. If I plot the transmission of light through a salt crystal as a function of frequency, I will find something like:

It is easy to have large amplitude oscillations when we “drive” this crystal with light at just the natural frequency it would like to oscillate with. Since large amplitudes build up there, a lot of light will be absorbed there. This creates the kind of resonant absorption lines which are used in spectrometers to determine the composition of materials.
Physics for the Life Sciences: Fall 2008 Lecture #15

For the next several weeks, we’ll be talking about energy and how the flow of energy from one form to another enables life. We begin by revisiting the basics of energy as a concept, and then turn to understanding heat energy.

Energy and its forms

The idea of energy was not part of Newton’s physics. There are many reasons for this, not least that the energy concept is subtle and complicated. The first form of energy to be accounted for in the framework of Newtonian mechanics was kinetic energy. Thomas Young introduced KE as an important quantity to track in physics, and showed that KE is changed when work is done.

\[ KE = \frac{1}{2}mv^2 \]

\[ W = \int \mathbf{F} \cdot d\mathbf{s} = \Delta KE \]

We have also discussed the idea of potential energy; the notion that it is useful to account for the potential effects of certain forces by defining a potential energy as a function of position, as we did for gravity and the energy associated with elastic deformation. These are our first two forms of energy: kinetic and potential.

It turns out there are many others as well. The fact that many forces can do work (changing kinetic energy) without being “conservative” forces associated with potential energies is strong evidence for this.

Perhaps it is best to consider an example. Imagine that you push a book across a table. It starts out with kinetic energy, then gradually slides to a halt as the force of friction does negative work on it. What happens to the energy which the book had? It is not lost, or “used up” by the friction force. Instead, it is converted to another form. The energy which the book started out with is instead converted into heat in the book and the table.

To understand this process, we need to figure out how to include heat, which is clearly a form of energy, in the tidy framework we established for studying kinetic and potential energy.

Conservation of Energy

The conservation of energy, the notion that it can neither be created nor destroyed, was originally simply an observation. Preserving this observation initially required a kind of tricky bookkeeping. In the little example above, we preserved energy conservation by asserting that the kinetic energy originally present was somehow turned into heat energy. That kinetic energy is gone, sure, but look, here it is in a new form we weren’t worrying about before. That might seem like slight of hand; if it disappears we just give it a new name. Absent any more fundamental reason for believing in the conservation of energy, this would certainly be a valid criticism.
Richard Feynman, an influential 20th century American physicist, described the situation this way:

“There is a fact, or if you wish, a law, governing natural phenomena that are known to date. There is no known exception to this law—it is exact so far we know. The law is called conservation of energy; it states that there is a certain quantity, which we call energy that does not change in manifold changes which nature undergoes. That is a most abstract idea, because it is a mathematical principle; it says that there is a numerical quantity, which does not change when something happens. It is not a description of a mechanism, or anything concrete; it is just a strange fact that we can calculate some number, and when we finish watching nature go through her tricks and calculate the number again, it is the same..."

If energy conservation were based only on observation, if we only thought it was conserved because we’d always seen it conserved, you might hope to violate energy conservation. In this case, you might be able to create energy, to make it without cost. Obviously a world with this kind of free energy would be marvelous, and some people continue to seek it. In fact, there are people who will sell you stock in their free energy company online.

Fortunately, (unfortunately?) there is a deeper understanding of the origin of energy conservation, one which provides a convincing theoretical foundation for the conservation we observe. It was discovered in the first part of the 20th century by a remarkable mathematician named Emmy Noether. She was able to connect conservation laws (like the conservation of energy and momentum) in a rigorous way to symmetry principles. What does this mean?

First, what’s a symmetry in a physical law? Saying that a physical law has a symmetry just means that the law is unchanged when you alter it in some way. For example, the laws of physics have a spatial symmetry. They are the same whether you study them in Ann Arbor or Uttar Pradesh. In fact, they are known to be the same both here and in the most distant parts of the universe. As far as we know, the laws of physics also possess a time symmetry. They are the same now as they were in Newton’s time. They seem not to have changed since the big bang, and we expect them to be the same tomorrow.

What Noether proved was that if the laws of physics possess such symmetry, there must be some conserved quantity associated with each symmetry. Any violation of the symmetry would require a violation of the conservation law and any violation of the conservation law would require a violation of the symmetry.

As it happens, the conservation of energy derives from the time symmetry of the laws of physics. Saying that energy is conserved is equivalent to saying that the laws of physics we learn today were correct yesterday, and will be the same tomorrow. This is one of the key reasons why you should be extremely skeptical if someone tries to sell you free energy.
Temperature and heat energy

We know from experience that when friction (a “non-conservative” force) acts, things heat up. For example, you probably rub your hands together to warm them on a chilly morning. We’re saying now that this “heating up” is due to the deposition of energy taken from motion by friction. It seems clear that heat energy is associated with temperature, but what exactly is the connection? To see this we need first to take a step back and talk about how temperature scales are defined.

There are two important temperature scales in use in science. Each is defined by fixing the temperature at which water freezes and boils, then dividing the temperature range between these into 100 equal intervals called degrees. The Celsius scale sets the freezing point to 0°C and the boiling point to 100°C.

Measurement of temperature relies on the fact that many properties of matter change in a fully repeatable way with temperature. To take just one of many examples, liquids and solids expand and contract as temperature changes. This is the basis of mercury thermometers. Once calibrated, changes like this can be used to build devices which reliably measure temperature.

While expansion of any solid or liquid can be used to define a temperature scale in this way, there are drawbacks to this approach. The scale you get doing this is very specific, and will be a somewhat different if you use a different material. It turns out that a much more universal scale of temperature can be defined using the properties of dilute gases. What do we mean by a “dilute” gas? Basically, we just need to make sure that the typical distance between particles in the gas is very large compared to the size of the particles. This condition is easily met by all the gases you usually encounter. Such dilute gases have remarkably simple, and similar, relations between physical properties and temperature. We will see why in a bit, but first let’s explore the basic phenomena.

A simple model for a gas

In a gas individual atoms or molecules move about freely. Usually they’re just flying through space, though occasionally they run into one another or into the walls of the vessel in which they’re contained. When they do, they bounce off one another elastically. In an “ideal” gas, there’s no interaction among the atoms at all. They never hit one another, though they do still hit the walls of the vessel. Each collision of an atom with the wall exerts a tiny, very brief force. Taken together, this enormous number of continuous, tiny impacts creates a steady average force pushing out against the walls. The force is spread nearly uniformly across the walls, and so is described as a pressure.

Pressure and the equation of state

Unlike solids and liquids, gases will expand freely to fill whatever contains them. In fact, they always push outward on the vessel which contains them, exerting a uniform pressure.
on the walls of the container. This pressure, a force per unit area, is measured in N/m², and is uniform throughout the gas.

Early in the 19th century careful experiments in England showed that, if you keep a gas at a fixed temperature and alter its volume, the resulting pressure obeys “Boyle’s Law”:

\[ PV = \text{constant} \]

If you increase the volume, the pressure goes down. If you decrease the volume, all the time holding the temperature constant, the pressure goes up. Note that this means if you completely remove the pressure, taking away the walls of the vessel, the gas will expand to infinite volume. Only the presence of the walls prevents this from happening.

Further experiments later showed another connection. If you keep the pressure constant and vary the temperature (measured on the Celsius scale), the volume of the gas changes according to a rule called Charles’s law:

\[ \frac{V}{T+273.15} = \text{constant} \]

This discovery prompted the invention of a new temperature scale, called the “Kelvin” scale, which is defined by:

\[ T_{\text{Kelvin}} = T_{\text{Celsius}} + 273.15 \]

Using this temperature scale, Charles law can be written:

\[ \frac{V}{T} = \text{constant} \quad \text{or} \quad V = \text{constant} \times T \quad (\text{at constant} \ P) \]

From experiments, it’s possible to find these constants in Boyle’s and Charles’s Laws. Remarkably they are the same for all dilute gases, with one exception. To know what the constant is, you have to know how many atoms are in the gas you’re looking at. It makes no difference what kind of atom or molecule they are; you only need to know how many.

Moles and numbers of atoms

How do you know how many atoms are in the gas you’re looking at? It turns out the number of atoms in any reasonable amount of gas is very large, so you can’t just count. Instead, we take the mass of each atom (or molecule if it’s molecular gas), and divide the total mass by the mass of each atom. Masses of atoms and molecules are measured in “atomic mass units”, or amu. The scale for this is 1 amu = 1.66x10⁻²⁷ kg. If your molecule weighs 28 amu and you have 100 gm of the stuff the number of atoms you have is:

\[ 0.1 \text{ kg} / (28 \times 1.66 \times 10^{-27} \text{ kg}) = 2.15 \times 10^{24} \text{ atoms} \]

which is a lot of atoms.
It is useful to define a particular number of atoms such that the total weight in grams is just equal to the particle weight in amu. For example, the Hydrogen molecule H₂ has a molecular mass of 2 amu, or about $3.32 \times 10^{-27}$ kg. If we had 2 grams of H₂ we would have a total number of atoms:

$$0.002 \text{ kg} / 3.32 \times 10^{-27} \text{ kg/atom} = 6.024 \times 10^{23} \text{ atoms}$$

This particular number of atoms is called a “mole” of atoms. It’s just a number, arbitrary really, that is useful for comparing the same number of atoms of different substances. It roughly reflects the number of atoms you might find in a typical modest sized sample of some material. In this sense, it tells us the rough conversion between some ordinary thing and the number of atoms which make it up.

Using this definition, the mass of one mole of a substance is:

$$\text{Mass of one mole} = 6.024 \times 10^{24} \text{ atoms} \times \text{number of amu} \times 1.66 \times 10^{-27} \text{ kg/atom} = \text{number of amu} \times 10^{-3} \text{ kg}$$

Using this definition, for a molecule with a molecular weight of 28 amu (like N₂), the mass of one mole is 28 grams.

The ideal gas law

As it turns out, you can combine Charles’ and Boyle’s Laws into a single “ideal gas law”:

$$PV = nRT$$

Where P is pressure, V is volume, T is temperature in the Kelvin scale, n is the number of moles of gas in your sample, and R is a constant called the “universal gas constant” for rather obvious reasons. Experiments show that its value is:

$$R = 8.3 \text{ J / mole*K}$$

The units here are Joules per mole per Kelvin. Do these make sense? What are the units of PV? Pressure is force / area (N/m²) and volume is m³, so PV has units of Nm, which is the unit of energy, the Joule.

This law is quite accurate for all gases under ordinary conditions. It allows us to define a temperature scale in a more fundamental way. Take a fixed number of gas atoms (a number of moles n), put it in a vessel of volume V and measure the pressure P. Then the temperature T is given by:

$$T = PV/nR$$

This scale provides a material independent temperature scale. Use any gas you like (so long as it is dilute) and you’ll get this result. When talking about this, we’ll sometimes say “atoms” or “molecules” or “particles”. It doesn’t really matter, as the result is independent of that detail, at least to the extent that the gas behaves ideally.

This gas temperature scale allows us to build a thermometer which is more general than a mercury or alcohol thermometer. We don’t even have to know precisely what it’s made of for it to work. This generality is very attractive to physicists, mostly because it hints at something very general going on, and provides a hint about where to seek the fundamental nature of temperature.
Origins of the ideal gas law: atoms and kinetic theory

In our atomic picture, we imagine that a gas is made up of an enormous number of tiny particles, each traveling freely except when they collide with one another or the wall. In this model, the space between particles is very large compared to their size, so they spend most of their time moving about freely, and only a small fraction of their time colliding with one another or the wall.

What happens when they strike the wall? Each time a particle hits the wall it bounces off elastically. The wall applies an impulse (a force for some period of time: $F \Delta t$) to the particle, and the particle applies an equal and opposite impulse to the wall. When you average the effect of a huge number of impacts, the wall experiences some average force per unit area (a pressure). If you change the conditions, you can change this pressure. Increase the number of atoms in the vessel, and they will strike the walls more often, increasing the pressure. Increase the volume of the vessel and they will strike the wall less often, decreasing the pressure.

To see quantitatively how this works, consider a simple model of an atom rattling around in a cubic box with edge length $L$. Each time it hits the wall on the right, there is a change in momentum $\Delta p = 2mv_x$ (because it bounces back). This happens once in every time $\Delta t = 2L/v_x$. Putting these together, we get an estimate of the average force exerted by one atom on the right hand wall: $F_{\text{av}} = \Delta p/\Delta t = 2mv_x/(2L/v_x) = mv_x^2/L$. Multiply this by the number of atoms ($n_{\text{atoms}} = n_{\text{moles}}N_A$) in the box to get the total average force on the right hand wall, then divide by the area of the right hand wall to get the average pressure $P = F/A$. Rearranging this gives:

$$PV = nN_A(mv_x^2)$$

The details are here:

$$F_{\text{one atom}} = \frac{\Delta p}{\Delta t} = \frac{2mv_x}{(2L/v_x)} = \frac{mv_x^2}{L}$$

$$F_{\text{Total}} = nN_A F_{\text{one atom}} = \frac{nN_A}{L}(mv_x^2)_{\text{average}}$$

$$P = \frac{F}{A} = \frac{nN_A}{AL}(mv_x^2)_{\text{average}} = \frac{nN_A}{V}(mv_x^2)_{\text{average}}$$

$$PV = nN_A(mv_x^2)_{\text{average}}$$

This is an interesting result, because this $mv_x^2$ looks like part of the kinetic energy of the atoms. To see how it relates to the total kinetic energy, consider how the magnitude of $v_x$ relates to the magnitude of $v_y$ and $v_z$ on average for these atoms.
If they’re bouncing around over and over in this box, there’s no reason for any one of the components (x, y, or z) to be any different from the others. So we expect on average:

\[(v_x^2)_{av} = (v_y^2)_{av} = (v_z^2)_{av}\]

This allows us to relate \((v_x^2)_{av}\) to the average total velocity squared:

\[(v_{tot}^2)_{av} = (v_x^2)_{av} + (v_y^2)_{av} + (v_z^2)_{av} = 3(v_x^2)_{av}\]  
or  
\[(v_x^2)_{av} = \frac{(v_{tot}^2)_{av}}{3}\]

Plugging this into the above relation, we find:

\[PV = n_NA(mv_{tot}^2)_{av} / 3 = (2/3)n_NA(1/2mv_{tot}^2)_{av} = (2/3)n_NA(KE_{per \ atom})_{av}\]

This is a remarkable relation, because we can compare it to the ideal gas law (PV = nRT) and write:

\[RT = (2/3)N_A(KE_{per \ atom})_{av}\]  
or  
\[(KE_{per \ atom})_{av} = (3/2)(R/N_A)T\]

If you tell me the temperature, I can tell you the average kinetic energy per atom in a gas.

Turning this around, we can use it to tell us what temperature really is. Temperature is really a measure of the average energy per atom. It doesn’t matter what kind of atoms you have, heavy or light, the temperature is a measure of the average kinetic energy of the particles in the gas.

This equation also suggests a new constant, called the Boltzmann constant:

\[k_B = (R/N_A) = 1.38\times10^{-23} \text{ J/K}\]  
so  
\[(KE_{per \ atom})_{av} = (3/2)k_BT\]

Since typical temperatures are around 300 K, this tells you that the typical kinetic energy of an atom in a gas is:

\[(KE_{per \ atom})_{av} = (3/2)k_BT = (3/2)(1.38\times10^{-23} \text{ J/K})(300 \text{ K}) = 6.2\times10^{-21} \text{ J}\]

This seems like a tiny energy. But remember, atoms are very light. Let’s imagine this is a Helium atom. What would its velocity be? The mass of a Helium atom is about 4 amu = 6.6\times10^{-27} \text{ kg}, so:

\[1/2m_{He}v^2 = KE\]  
\[v_{He} = (2KE_{He} / m_{He})^{1/2} = 1366 \text{ m/s}\]

So at room temperature, atoms inside a Helium balloon are zipping along at around 1366 m/s. That’s a bit more than 3000 mph, so they’re pretty peppy. If the gas is made of heavier atoms (Argon say, or N₂ molecules), the velocities are lower. But they’re always quite high compared to the macroscopic velocities we’re used to.

**So what is temperature really?**

This “kinetic theory” discussion of ideal gases provides a way to show that, for this case, temperature is a measure of the average kinetic energy of particles in a gas. This is a very simple case. In particular, for an ideal gas like this kinetic energy is the only form of energy the atoms can have. In more complex cases, like a solid, atoms can have both
kinetic and potential energy. Inside a solid, atoms oscillate around their positions of equilibrium. During these oscillations they have both kinetic energy and potential energy. In these more complex cases, temperature is still a measure of the average energy of the atoms in a material, but now that energy appears in more forms than just kinetic energy.

Two objects which “have the same temperature” have the same average energy in each of their tiny parts.

This doesn’t imply that all the atoms in the material have exactly the same energy. In fact the energy of each atom changes all the time as they bounce around against one another. What stays the same is the average energy of the atoms. At any given moment, the atoms in a gas will have a distribution of energies, some higher, some lower. Since there are typically so many atoms, this distribution is actually very stable.

The picture at the right shows, schematically, what the distribution of particle velocities might look like. As you increase the temperature, the average speed increases, and the distribution of speeds becomes broader. As time goes on, atoms change speed, some going faster, some slower. But the distribution of speeds, and so its average, remain the same.
Physics for the Life Sciences: Fall 200 Lecture #16

In our effort to make our notions of energy more inclusive we began to discuss heat energy. We have seen that temperature is a measure of the average amount of energy in each of the particles which makes up a material. Every atom in a material at temperature $T$ will have an average kinetic energy of $(3/2)k_BT$. At room temperature this is about $6 \times 10^{-21}$ J.

These thermal motions are a fundamental fact about matter. Atoms are always in motion. In gases they fly about freely, most of the time completely unconnected to one another. In liquids they slip around over one another, exchanging positions all the time but not completely escaping. In solids they remain mostly locked in place, but oscillate violently around their equilibrium positions.

Thermal energy, and this ubiquitous thermal motion, is also incredibly important for life. Life is based on a continuous flow of energy. Living things, yourself included, exist by continuously taking in energy, changing its form, and sending it back out again. Much of what you take in is converted to thermal energy, and many of the mechanisms in your body rely heavily on this simple thermal motion to happen. Before we explore the details of this, we need to address a deep and fundamental question about thermal motion, conservation laws, and what happens in the world.

**Why do things happen the way they do: particle velocities**

If we examine the velocities of these particles in a gas, we find they fill out a very particular distribution which, if there are even a modest number of atoms, is very stable. This fact raises a really big question. What is it that “makes” these atoms take on this particular distribution of velocities? Why does this distribution always emerge?

The conservation of energy and momentum don’t require this distribution. They only insist that the total momentum and energy remains the same. So they’re just as happy with all the energy and momentum in two atoms, with all the others at rest. Conservation laws don’t determine this distribution of velocities.

This question is actually extremely broad. There are many things which could happen in physics, but which never do. These are processes allowed by the conservation of energy and momentum, but which never occur. Sliding a book to a stop across a table is an example.

When you do this, the kinetic energy of the book is converted to thermal energy in the table and the book. Conservation laws allow this thermal energy to come back together and reenter the book. If this happened, you might set your book on the table, have it pick up thermal energy from the table and start sliding away. That process is allowed, but it never happens. Sliding the book to a halt is something we’d call an “irreversible” process.
Irreversibility and chance

Energy flow constraints allow many transformations which never occur. How do we determine which ones will and which won’t? Doing this is the subject of statistical physics, an enormously important branch of physics which answers just this question. How do you figure out not only what can happen, but what will happen?

Statistical physics relies on one fundamental assumption:

All possible outcomes are equally likely

To see how this very reasonable sounding assumption might be consistent with the irreversibility we just discussed, in which some outcomes never happen while others always do, we’ll consider a simple model system.

Imagine a cube filled with some ideal gas atoms rattling around in it. This is a closed box. Neither atoms nor energy ever enters or leaves it.

We begin by defining microstates and macrostates for the system. The “microstate” is simple. It’s just the precise condition of all the atoms in the box: a complete detailed description at some instant. The idea of a macrostate is a little more subtle. A “macrostate” is some collective feature of the distribution of positions or velocities of the particles. Examples include:

- Average particle kinetic energy (temperature)
- Average pressure on the wall
- Average position of the atoms
- What fraction of the atoms are on the top half of the box

In general, macrostates like this are the sort of thing we might actually observe about such a box. There’s no way to measure exactly where each atom in a gas is, along with exactly how each is moving. But it is perfectly possible to measure the temperature, determine approximately how much gas is on the top of the box, etc.

Next we need to think a bit about timescales. We want to see what will happen if we let this system evolve, so we have to wait long enough to let the system rearrange itself. Each time we check to see what the system has done (where the atoms are) we should give it time to rearrange itself before we check again.

How long do we wait? The typical time for rearrangement is comparable to the time it takes an atom to cross the box: \(L/v_{av}\). We’ve seen that atoms travel quite fast in a gas, so we won’t have to wait long. If our box is 10 cm across, we might need to wait 0.1 m / 1000 m/s = 10^{-4} \text{ s} or so. The only reason to raise this point is that some systems, big ones like galaxies, take much longer to rearrange themselves. What happens is very similar, the same really, but the timescales are a lot longer.

Now we start looking at the box over and over again, each time waiting long enough for the atoms to rearrange themselves. If we do this, every possible arrangement of particles,
every microstate, will eventually occur. In fact, they’re all exactly equally likely. What about the macrostates? How often, for example, will we find all the particles in the top half of the box?

To figure this out, we need to find out what fraction of all the equally likely microstates corresponds to this macrostate. The probability that this macrostate will be observed is then equal to

$$P_{\text{top}} = \frac{\text{(# of microstates for this macrostate)}}{\text{(total number of microstates)}}$$

Let’s work this out for our simple “all particles on the top of the box” macrostate. The answer depends on the number of particles. Consider one atom first. This atom can be on top or on the bottom. One of these two possible microstates corresponds to all atoms on top, so:

$$P_{\text{top}}(1 \text{ atom}) = 0.5$$

If there are two atoms, it becomes more complicated. Now you could have

- Both on top (1 arrangement)
- One on top, one on the bottom (2 arrangements, one with each on top)
- Both on the bottom (1 arrangement)

So now the probability of having them all on top is:

$$P_{\text{top}}(2 \text{ atoms}) = 0.25$$

You should do the same for 3 and four atoms in the box. Label them A,B,C, etc. and figure out how many total arrangements there are for three and four in the box. You should find that:

$$P_{\text{top}}(3 \text{ atoms}) = 0.125 \quad \text{and} \quad P_{\text{top}}(4 \text{ atoms}) = 0.0625$$

In fact, there is a general rule for this:

$$P_{\text{top}}(N \text{ atoms}) = \frac{1}{2^N}$$

So this particular macrostate, with all the atoms on the top, becomes increasingly unlikely as the number of atoms increases. The microstate which makes this macrostate is not intrinsically less likely than any other microstate. In fact it’s exactly as likely as any other microstate. But as the number of atoms increases there are more and more different microstates, only one of which corresponds to this macrostate. This particular macrostate gets very unlikely as the number of atoms increases.

There is another macrostate, however, which becomes increasingly likely. Think back to the 4 atom case. How often do you have half the atoms on top and half on the bottom? There are 16 total microstates for these four distinguishable atoms. Six of these correspond to half on top and half below. So the probability of finding the half-and-half macrostate is $6/16 = 0.375$. It’s much more likely that the all-on-top macrostate.

This is the key idea in statistical mechanics. All microstates are equally probable. Every allowed outcome is equally likely. But some macrostates correspond to many microstates, while others correspond to very few. Random wandering among the possible
microstates then guarantees that some macrostates will happen more often than others. In this sense, having all the kinetic energy in a sliding book is a macrostate which involves few microstates, while having that energy spread out as heat in the table involves many microstates. Once you move from the rare, all concentrated in the book state, into one of the incredibly many ways you can have the energy spread out, you never wander back into the rare state of having all the energy in the book.

How can this be irreversibility?

Nothing we have said would seem to explain irreversibility. According to this, all we have to do to see the book take off across the table is wait long enough. In our four atom example, we’d only have to wait about 16 rearrangement times to see all the atoms on top. How long might this be? Imagine our atoms rearrange themselves a million times a second. For our 4 atom system, we would, on average, see them all on top every 16 microseconds. This would happen about 62,000 times a second. It’s pretty common.

What if we increase the number of atoms to a reasonable number? If you have 80 atoms, the probability of finding them all on one side is 1/1200000000000000000000000, or about $8.7 \times 10^{-25}$. These atoms, zooming along, rearranging themselves a million times a second, will all be on one side once every $10^{18}$ seconds or so. That’s about 30 billion years…Unfortunately the universe is only about 14.5 billion years old so far, so it probably hasn’t happened yet.

To be more realistic, 1 cm$^3$ of air contains about $2 \times 10^{19}$ atoms. How often will you find them all on one side of the box? Never.

So the following simple process, push all the atoms in a 1 cm$^3$ box to one side, then let them go, is, despite the equal probability of every microstate, completely irreversible. The sliding of a book to a halt, while more difficult to calculate, is just the same. There’s only one way to have all that energy in the book, with all its atoms sliding along together. But there are quadzillions (that’s the technical term) of ways to spread that energy out in random motion of atoms in the book and the table. So sliding to a halt is irreversible. Not because it can’t reverse, but simply because it won’t.

Using this in statistical physics

This toy example gives us a good idea of how statistical physics works. To predict what will happen you define macrostates you want to compare, then figure out how to count the allowed microstates which correspond to each macrostate. Comparing then tells you how likely each macrostate is. Those macrostates which are made from many microstates are the most likely. Usually the tricky part is counting the microstates.

In general what you find is that any state which spreads the energy out equally among all the parts is vastly more likely than a state with all the energy concentrated in a subset of the matter. This idea, that energy will spread out as much as possible, is called the principle of equipartition. It isn’t meant to suggest that energy is spread perfectly equally,
but that it is spread out in some calculable distribution (some atoms have more energy, some less, and the distribution is stable).

To summarize: Random chance is what decides which among the allowed possible outcomes will actually occur. How long will it take the final outcome to emerge? This depends on how rapidly the system rearranges itself. The progress from order (like all on one side) to disorder (like evenly spread) happens by chance, not by force or design. In any system you can measure, it is so likely as to be practically inevitable. Once you get into this kind of disordered state, you won’t leave it. It is a stable equilibrium state.

There are cases, lots of them, which involve change from disordered to order. But in every case, this happens because the system is open, and able to exchange matter and energy with the rest of the universe. If you bring energy into a system, you can cause it to move into a more ordered state. For example, we could push that gas into the upper half of our 1 cm$^3$ block. But if you leave it alone, if the system is truly isolated, it will inevitably progress toward disorder.

**Entropy and disorder**

You have probably heard about “entropy”, and heard that entropy always increases. There is a quantitative connection between the ideas here and entropy. In fact, the clearest definition of the entropy of a system in a particular macrostate is:

\[ S = k_B \ln(\text{# of microstates equivalent to this macrostate}) \]

So when people say entropy increases, they just mean that the system will, by chance, migrate toward macrostates represented by many microstates.
Physics for the Life Sciences: Fall 2008 Lecture #17

We have seen in the last class that completely random processes, like atoms rattling around in a cube, can lead to inevitable conclusions. These things happen automatically, not by force or design. As it happens, life uses natural progression through random processes very extensively. We’ll begin by looking at two related processes, diffusion and osmosis, where this is obvious. Later we’ll discuss more generally how this harnessing of random motion, of what’s going to happen no matter what, is behind all of life.

**Diffusion: moving things around with random motion**

We have seen how random motion causes atoms released in just part of a box to inevitably spread uniformly throughout it. Life takes advantage of the spreading caused by random motion very extensively to move things around. The process is called “diffusion”, and it relies on the random thermal motions of particles.

When particles are at a temperature $T$ they have a typical velocity (taken here as the root-mean-square velocity) which depends on the temperature. We find this from:

$$\frac{1}{2}mv_{av}^2 = \frac{3k_BT}{m_{atom}}$$

This is shown in another form, written for the molar mass, at the right.

$$v_{\text{rms}} = \sqrt{\frac{3k_BT}{M_{\text{mole}}}}$$

For any material, this $v_{\text{rms}}$ varies like the square root of the temperature and molar mass. To double $v_{\text{rms}}$ you have to quadruple the temperature. To cut it in half you can increase the molar mass by a factor of four. Most of life exists in a very narrow temperature range, something like $273 \, \text{K} < T < 373 \, \text{K}$, so that doesn’t vary too much. But the molar mass of things which might be diffusing does change a lot. Here is a table:

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$m$ (g)</th>
<th>$M$ (g/mol)</th>
<th>$v_{\text{rms}}$ (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrogen</td>
<td>$3.35 \times 10^{-24}$</td>
<td>2.016</td>
<td>1880</td>
</tr>
<tr>
<td>Helium</td>
<td>$6.65 \times 10^{-24}$</td>
<td>4.002</td>
<td>1340</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>$4.65 \times 10^{-23}$</td>
<td>28.02</td>
<td>506</td>
</tr>
<tr>
<td>Oxygen</td>
<td>$5.32 \times 10^{-23}$</td>
<td>32.0</td>
<td>474</td>
</tr>
<tr>
<td>Mercury</td>
<td>$3.33 \times 10^{-22}$</td>
<td>200.6</td>
<td>186</td>
</tr>
</tbody>
</table>

| Macromolecules | $1.67 \times 10^{-20}$ | $10^4$ | 26 m/s |
|                | $1.67 \times 10^{-18}$ | $10^6$ | 2.6 m/s |
| Viruses        | $1.67 \times 10^{-16}$ | $10^8$ | 26 cm/s |
|                | $1.67 \times 10^{-14}$ | $10^{10}$ | 2.6 cm/s |

Copyright: Regents of the University of Michigan, 2009
From this table you can see that $v_{\text{rms}}$ varies a lot, from 500 m/s for N$_2$ to 2.6 cm/s for a virus.

**Particle size and mean free path (MFP)**

To figure out how quickly random motion will get things from one place to another, we need to know not only how fast things move, but also how far they go between collisions. If they move fast, but are continually running into other things, they won’t get far. On the other hand, if they *never* hit anything else, they will zoom along at constant speed. Now a truly ideal gas is like that. The atoms in such a gas never interact. But real atoms do. We can work out how far they go between collisions from the ideal gas law:

$$V = \frac{nRT}{P} \quad \text{or} \quad V_{\text{mole}} = \frac{RT}{P}$$

$$\frac{V}{\text{Particle}} = \frac{V_{\text{mole}}}{N_A} = \frac{RT}{N_A P}$$

$$V_{\text{swept out}} = \pi r^2 d_{\text{MFP}} = \frac{V}{\text{Particle}}$$

$$d_{\text{MFP}} = \frac{RT}{N_A P \pi r^2}$$

$$d_{\text{MFP}} = \frac{8.3 \frac{J}{\text{mole} \cdot ^\circ \text{C}} \times 273 \text{K}}{6 \times 10^{23} \times 10^5 \frac{\text{N}}{\text{m}^2} \times \pi \times (10^{-10} \text{m}^2)^2}$$

$$d_{\text{MFP}} \approx 10^{-7} \text{m}$$

Basically, this derivation finds the volume per particle, then says this volume is equal to the volume “swept out” by the particle as it travels from one collision to the next. This provides an estimate of the distance between collisions, what’s called the “mean free path” of the particles. You can see that it depends on temperature, pressure, and the radius of the atom $r$. For typical values in a cold gas, this mean free path is around $10^{-7}$ m; not very far at all. You get the typical time between collisions from $t_{\text{mean}} = d_{\text{MFP}} / v_{\text{rms}}$, and putting in typical numbers you find times around $10^{-10}$ s.

So when you picture atoms rattling around in a gas you should imagine they have around 10 billion collisions a second, each time traveling around 0.1 μm between collisions. Atoms moving like this conduct what’s called a “random walk”, something like what’s illustrated in this picture. During this walk, they gradually move away from where they started. Since it is a random process, you can’t predict *exactly* how far or in which direction any particle will move, but you can statistically predict how far particles will move on average.

It turns out this distance is characterized by the “diffusion coefficient”, which for an ideal gas is sometimes defined as:

$$D = \frac{1}{2} v_{\text{rms}} d_{\text{MFP}}$$

Notice that it depends both on how fast things are moving and how far they go between collisions. In fact the diffusion coefficient for different situations is often measured,
rather than calculated from first principles. The units of the diffusion coefficient can be seen from the definition to be m²/s.

Using this definition for D, the rate at which particles move away from where they start due to this random walk is given by:

\[ r_{\text{rms}} = (6Dt)^{1/2} \]
\[ x_{\text{rms}} = (2Dt)^{1/2} \]

In these relations, \( r_{\text{rms}} \) is the root-mean-square distance away from the starting point in three dimensions, while \( x_{\text{rms}} \) is the root-mean-square distance in a single direction (like \( x \) for example).

What’s important to notice about this? First, the distance traveled varies like the square root of the time. So if you want a particle to diffuse twice as far it will take four times as long. Second, the distance depends on this diffusion coefficient, and hence on \( v_{\text{rms}} \) and \( d_{\text{MFP}} \). As a result, small, light atoms will diffuse faster than large atoms. In addition, diffusion will happen faster when the temperature is larger.

Here are some tables of diffusion coefficients:

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Temp (C)</th>
<th>D (m²/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂</td>
<td>0</td>
<td>6.4x10⁻⁵</td>
</tr>
<tr>
<td>H₂O (Vapor)</td>
<td>8</td>
<td>2.4x10⁻⁵</td>
</tr>
<tr>
<td>O₂</td>
<td>0</td>
<td>1.8x10⁻⁵</td>
</tr>
<tr>
<td>CO₂</td>
<td>0</td>
<td>1.4x10⁻⁵</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Temp (C)</th>
<th>D (m²/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O₂</td>
<td>20</td>
<td>1.0x10⁻⁹</td>
</tr>
<tr>
<td>Glucose</td>
<td>20</td>
<td>6.7x10⁻¹⁰</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>20</td>
<td>6.9x10⁻¹¹</td>
</tr>
<tr>
<td>DNA</td>
<td>20</td>
<td>1.3x10⁻¹²</td>
</tr>
</tbody>
</table>

Given these numbers, you might expect a molecule of O₂ to diffuse about 1 cm in a second in air, while it would go only 70 μm diffusing in water for the same one second.

So diffusion is a process which can move atoms and molecules from one place to another just by allowing them to rattle around. No mechanism is needed to make it happen. This mechanism is used very extensively by life, but it is limited to working over relatively short distances.

One particularly important example is O₂, which is used in your body for metabolism. O₂ can be delivered by diffusion, but if it is diffusing in water, it takes:

- 2x10⁻⁴ s to go 1 μm
- 170 s to go 1 mm
- 4.6 hours to go 1 cm
- 5.4 years to go 1 meter

Since your cells can’t wait five years for oxygen to travel from your lungs to your legs, some method of transport other than diffusion is required, and that’s why you have this complex circulatory system which moves bulk amounts of oxygen loaded blood around.
To have diffusion be effective, you need to keep the distance scales short. Diffusion is the primary method of transport within a cell. Inside the cell, you get key molecules from one place to another just by letting them go. They diffuse around, ramming into things randomly, until they get where they’re supposed to go.

The fact that diffusion is so important in cellular transport is the reason cells are so small. If they were larger, say 1 cm in size, it would take forever for important materials to diffuse from one place to another. If an organism wants to grow bigger, it has to develop methods of transport which carry things long distances much more effectively than diffusion. Most cells are on the 1-10 μm size scale so that they can efficiently use diffusion as a mechanism of transport.

**As easy as breathing: getting oxygen to your cells**

The transport of oxygen from the atmosphere into an organism, then through it to the cells where it must be used, is a great example of how different approaches to a fundamental physical constraint have been explored by evolution. Let’s consider the steps.

1: First, to get oxygen from the atmosphere into your bloodstream it must diffuse across a membrane. That’s easy, just make a thin membrane and let it go. There is a problem though, at least for us large organisms. Remember that surface area is proportional to size², while volume is proportional to size³. The amount of O₂ you can get per unit time depends on surface area, while the amount you need depends on volume. So as organisms get larger, they need to somehow manufacture more and more oxygen diffusing surface area.

There are at least three general solutions. One is to stay small. This works for insects and other smaller organisms. They mostly don’t have to breathe. They can just wait around for O₂ to diffuse in. A second approach is gills. These big feathery structures have enormous amounts of surface area given their volume. Hanging them out in the water is enough to allow lots of O₂ to diffuse in. Our approach, along with most land animals, is to make internal lungs which are packed full of tiny sacs called alveoli. The reason for the sacks is to maximize the surface area which is available for O₂ diffusion. These little sacks create enormous amounts of area and allow red blood cells to get really close to the surface so that oxygen can get to them.

2: Once you get the oxygen into a carrier (like a red blood cell) you need a very efficient transport system for moving it to where it’s needed. This is what circulatory systems are for. If you didn’t need one, you certainly wouldn’t have one, because they fail so easily. Your arteries bring together a huge number of capillaries to allow efficient pumping off the remote parts of the body.
3: Once there, you again have to put the oxygen carrying red blood cells close enough to the target cells to allow diffusion to deliver the oxygen. That’s what the incredibly fine capillary system through the rest of your body is for.

Another way to look at diffusion

In the discussion above, we considered how diffusion, the random motion associated with thermal energy, can cause an atom to wander away from its starting point. This wandering can be used for transport, but to be useful it must happen where there is a density gradient.

To see why, imagine a box filled with N₂ molecules. As time goes by, each molecule will wander away from its original position according to the equations described above. But since every molecule is wandering, the average density of molecules at any point will remain exactly the same. No net transport will take place. This is just another way of saying that this uniform distribution is an equilibrium condition; the average properties, the macrostate of this, will not change.

To have some real transport, some flow of molecules, which is driven by diffusion, you need to have density gradients. The simplest case is the old cube of gas. If you start all the atoms on one side so the density on one side is high and on the other it is zero, you will get diffusion driven net flow until the density is uniform.

More generally you might have a case with a smooth density variation like this. In this case you would measure the density gradient by seeing how the density changes with position at each point. In the general case you would write the density as a function of position \( \Phi(x,y,z) \) and then measure how rapidly it changes in each direction. In this case you would describe the density gradient measuring the derivative of the density along each direction using what are called “partial derivatives”. You will learn how to do this when you learn vector calculus.

To see basically how this works, let’s consider a one dimensional version. Imagine a case of a tube along which the density of a substance varies as \( \Phi(x) \) and remains fixed in time. This is what happens in steady state diffusive flow. In this case we expect the net flow of material to be described by Fick’s Law:

\[
J = -D \frac{d\Phi}{dx}
\]

In this equation \( D \) is the diffusion coefficient we encountered above, \( d\Phi/dx \) is the density gradient (the rate of change of density with position), and \( J \) is the flow rate or current of particles. \( J \) is measured in units of number / area*second. Checking the units we have
# / m²s = (m² / s) (# / m³) = # / m²s
So this checks out.

What does Fick’s Law tell us? First, the rate at which particles flow due to diffusion depends on the diffusion coefficient D. Recall that D depends on temperature and the properties of what’s diffusing through what material. Second, Fick’s Law tells us that there will be flow only where there is a density gradient, and the net flow will be largest where the gradient is largest.

**Diffusion across membranes and osmosis**

A key task for life is to separate itself from its environment. By doing this, life can control its conditions, ensuring that sudden changes in the outside world do not instantly affect it. Life does this isolation with membranes. There are many kinds of biological membranes. Some surround cells, isolating them to maintain a controlled environment inside. Others define parts of the cell, like peroxisomes, the nucleus, and the Golgi apparatus. Most are constructed of lipid bilayers; back-to-back sheets of fatty molecules. Typical thicknesses are range from 0.5-4.0 nm, so they are very thin.

The main purpose of membranes is to control the environment within by selectively allowing material to cross. Biological membranes are selectively permeable; they allow some substances to pass through them while blocking others. They are often called “semipermeable” membranes, though this name is perhaps misleading. This property, when combined with diffusion, allows osmosis, and it has a number of interesting properties.

To get a sense of this, consider a toy case in which such a semipermeable membrane separates two compartments in a cylinder. On the right, the pressure is created by both water and sugar molecules striking the walls. On the left the pressure is created only by water molecules hitting the walls. The pressure starts out equal, so there is a higher density of water on the left hand side. There must be, to make up for the pressure made by the sugar molecules on the right. Now imagine what happens if the membrane in the middle is semipermeable, and lets through water but not sugar.

At the start, there will be a density gradient of water across the membrane. This density gradient will allow diffusion to transport water from the left to the right, until the density of water will be equal on both sides. At this point the pressure created by water molecules will be the same on both sides; they will strike the membrane equally from left and right. But the total pressure on the two sides will now be different. On the left you have both water and sugar molecules hitting the walls, while on the right you have only an equal number of water molecules hitting the walls. The pressure on the right will be higher by the amount contributed by the sugar molecules.
This final state is described by writing:
\[ P_{\text{inside}} = P_{\text{outside}} + P_{\text{osmotic}} \]
This \( P_{\text{osmotic}} \) is just the extra pressure created by the material which can’t diffuse across the membrane (the sugar in this case), and the symbol \( \pi \) is commonly used for \( P_{\text{osmotic}} \).

Can we determine how large this extra pressure is? Doing this precisely is tricky, but as long as the solution is dilute (there is little sugar in the water), we can actually use the ideal gas approximation and come pretty close:
\[ P_{\text{sugar}} = \pi = \frac{nRT}{V} = \frac{n}{V}RT = cRT \]
Where \( c \) is the molar concentration, the number of moles per unit volume \( n/V \). From this we see that the osmotic pressure increase directly with the solute concentration \( c \), and also with temperature.

**Consequences of osmosis**

Many membranes are permeable to water but not to larger molecules. When this happens, water diffuses across the membranes selectively as we have seen above. Sometimes this happens in “hypotonic” (too much tension) environments, where the concentration of the solute is larger inside the membrane. In these cases, osmotic pressure builds inside the cell. This can, in extreme cases, lead to membrane rupture and cell death. In “hypertonic” (too little tension) environments, the opposite is true, and water diffuses out of the cell. When the solute concentration is balanced, conditions are “isotonic” and the flow of water into and out of the cell is balanced.

**Osmosis in culture**

Interestingly, the word osmosis has been picked up as a popular term. Here’s the definition of the popular version from the OED: “A process resembling osmosis, esp. the gradual and often unconscious assimilation or transfer of ideas, knowledge, influences,
etc.” Things that are adopted “by osmosis” somehow happen by themselves, without being forced or planned. This is a remarkably apt way of remember what’s happening in diffusion and osmosis. These aren’t processes driven by pumps or will, they occur as the inevitable outcome of a lot of random chance.

The association of osmosis with some kind of mysterious influence goes back to its discovery in the 19th century. At this time, the atomic nature of matter was not completely accepted, and certainly not understood, so osmosis seemed almost like magic. You can get the sense in this sentence, from British scientist Thomas Graham in 1854: “This flow of water through the membrane I shall speak of as osmose, and the unknown power producing it as the osmotic force”. This unknown power is the idea that caught the popular mind and led to adoption of this phrase.
Today we’ll expand our discussion of thermal energy and consider what happens when heat energy is added to liquids and solids. In these cases, where atoms are not freely flying but are attached together by interatomic forces, the introduction of heat is a little more complicated than it is with gases, though all the same main ideas apply.

Temperature and thermal expansion

In ideal gases, temperature appears only as motion, and as we’ve seen the energy put into a gas as heat spread smoothly throughout, giving rise to a distribution of velocities with the average kinetic energy of each atom being \((3/2)k_BT\). In a solid too, heat energy added will spread out among the atoms, but now they are not free to fly around. Instead they oscillate around their positions of equilibrium. In these oscillations, each atom trades energy between kinetic and potential energy, on average splitting that energy equally between the two. So in addition to kinetic energy, each atom has also potential energy associated with the stretching of the bonds which hold the atoms together.

This oscillation around equilibrium, which increases with temperature, has many consequences. The first is thermal expansion.

You can think of atoms in a solid as being attached by springs. If two such atoms are connected by a perfect Hooke’s law spring, then the average distance between them will be the same whether they oscillate or not, because the oscillations will move them apart and together in exactly equal amounts. In fact though, the forces which hold atoms together are not perfect, linear, Hooke’s law forces. Instead they are asymmetric. Usually, it is harder to push two atoms closer together by a distance \(\Delta x\) than it is to stretch them apart by the same distance. It’s a little easier to pull the atoms apart than it is to squash them together.

When atoms start oscillating in such a material, the average distance between them becomes larger. The solid expands. Exactly how much depends on just how asymmetric the “springs” connecting the are, which depends on the nature of the chemical bonds.

This thermal expansion of a solid rod with length \(L\) can usually be well described by a phenomenological relation:

\[
\Delta L = \alpha L \Delta T
\]

where \(\alpha\) is the “coefficient of thermal expansion” for this material, and \(\Delta T\) is the change in temperature of the material. Notice that the change in length \(\Delta L\) depends on the length \(L\). This is because each bond in the material becomes a little longer. So if there are more bonds, there is a larger increase in length. As a result, it’s often more useful to consider the fractional change in length \(\Delta L/L = \alpha \Delta T\).
The table to the right lists these thermal expansion coefficients for a few materials. Notice that they are small. If you change the temperature by 1°C, then a stainless steel rod will undergo a fractional change in length $\Delta L/L = \alpha \Delta T = 1.7 \times 10^{-5}$, or about 0.0017%.

Another important thing to notice is that they vary a lot. Why is this important? If you build something made of tightly fit parts of glass and aluminum, then change the temperature, the aluminum will expand about three times as much as the glass. If you hadn’t counted on this in your design, parts may come apart or break. The stresses associated with thermal expansion can be quite spectacular. You can see this by comparing the thermal expansion described here with another way of changing the length: stress.

Recall the stress-strain relation: stress = Young’s modulus * strain. Putting this into symbols:

$$F/A = Y(\Delta L/L)$$

Now imagine you have a bar of aluminum fit tightly into a bracket. If you heat the bar it will expand according to the thermal expansion law above. If the bracket does not expand, then it must squeeze the bar inward. How hard must it push to prevent the bar from expanding? The stress-strain relation tells us:

$$F/A = Y(\Delta L/L) = Y(\alpha \Delta T)$$

For aluminum, $\alpha = 23 \times 10^{-6}/K$, and $Y = 70 \times 10^9$ N/m², so this relation becomes:

$$F/A = (7 \times 10^{10} \text{N/m}^2)(23 \times 10^{-6}/K) \Delta T = (1.6 \times 10^6 \text{ N/m}^2) \Delta T$$

The bracket has to apply a pressure of $1.6 \times 10^6$ N/m² to keep the aluminum from expanding if its temperature rises by just one degree! Because of this, differential thermal expansion is a major consideration for engineering in any case where an object has to survive significant changes in temperature.

One last thing to know; thermal expansion coefficients vary with temperature, often rather strongly, so as an engineer tries to deal with this issue they have to be careful to account for this as well.

**Volume expansion**

Thermal expansion doesn’t just happen in one direction, instead materials expand in every direction. This makes their volume change. To track this, we’d like to know how the volume changes with temperature. This can be captured in a volume expansion coefficient so that:

$$\Delta V = \beta V \Delta T$$

How does this volume coefficient $\beta$ relate to the linear expansion coefficient $\alpha$? To see, just calculate how the volume of a cube changes with temperature. Since we expect changes in length $\Delta L$ to be small, we can drop terms in $\Delta L^2$ and $\Delta L^3$ along the way.
So the volume expansion coefficient for most solids is just about three times the linear expansion coefficient.

**Heat capacity**

We know the temperature of a material is related to the kinetic energy shared out amongst all its atoms. How much energy must we add to a material in order to change its temperature? The answer lies in our already established understanding of statistical physics.

When we dump energy into a material what will happen to it? Any energy we add will eventually get shared out equally among all the different locations and forms it can take. Each atom in the material will take some of the energy. Meanwhile, temperature is a measure of the average energy in each atom of the material. Since each atom takes energy, we to raise the temperature of the material we must raise the average energy of all the atoms, raising the temperature of a material with more atoms will require more energy. A material with more atoms will require more heat to raise the temperature.

We express this cost (how much energy to raise the temperature) by defining the heat capacity of a material. This is defined in two ways, the “molar heat capacity” and the “specific heat capacity”. They’re simply related, but you have to keep the two straight.

The first comes from measuring how much heat you must add ($\Delta Q$) to a given number of moles $n$ if you want to raise the temperature by an amount $\Delta T$:

$$\Delta Q = nC_{\text{molar}}\Delta T$$

or

$$C_{\text{molar}} = \frac{1}{n}(\Delta Q/\Delta T)$$

The second comes from measuring how much heat you must add ($\Delta Q$) to a given mass of material $m$ if you want to raise the temperature by an amount $\Delta T$:

$$\Delta Q = mc_{\text{specific}}\Delta T$$

or

$$c_{\text{specific}} = \frac{1}{m}(\Delta Q/\Delta T)$$

The two are related in a simple way, because a mass $m$ includes a number of moles which is $n = m/M_{\text{molar}}$. This implies:

$$c_{\text{specific}} = C_{\text{molar}} / M_{\text{molar}}$$

In general, the molar heat capacity is more fundamental. It essentially measures how much heat per atom you have to add to raise the temperature. The specific heat involves both that and how many atoms per unit mass the material has. As a result, molar heat capacities show some regularities which are hidden in specific heats.
Predicting molar specific heats

If we understand what happens when we put energy into a system, we should be able to predict molar specific heats. If we put energy into an ideal gas, the only form it can take is kinetic energy. But if we put energy into a solid, it can appear as both kinetic and potential energy. So to make the average kinetic energy of atoms in a solid increase, we’re going to have to add more energy per atom than we would in the gas; We’ll have to put some of it will go into potential energy too. As a result, we expect molar heat capacity of solids to be higher. How much higher?

The different forms of energy internal to a material are sometimes called “degrees of freedom”, and at a given temperature each atom in a material will have an average energy of 1/2k_BT in each degree of freedom. For an ideal gas, these degrees of freedom are motion in the three directions, x, y, and z. Each of these takes up 1/2k_BT, so the total energy per atom is 3/2k_BT, as we have seen before. What does this make the molar heat capacity?

To raise the temperature by an amount ΔT, we have to give each atom 3/2kBΔT of energy. If we do this for a mole of gas, we have to put in a total energy:

ΔQ = NA(3/2)kBΔT or (3/2)NAkB = ΔQ/ΔT

Remembering that NAkB = R, the universal gas constant, we see that the molar heat capacity of an ideal gas should be 3/2R, or about 12.5 J/mole*K. For the monatomic noble gases (He, Ne, Ar, etc.) this is just what you find. From this you can see that the molar heat capacity for a material will general be (1/2)R for each degree of freedom.

Diatomic gases

What about a diatomic gas like N₂ or H₂? A molecule like this is like a dumbbell, with the two atoms attached by a stiff spring. Here you can still put energy into translation, so you get 3/2R from that. There are other ways to store energy though. The dumbbell can rotate. Imagine it’s aligned along the x-axis. For this, you can put energy in by rotating around either the y or z axis. That’s two more degrees of freedom.

In addition, the two atoms can vibrate along the axis of the molecule (it gets longer and shorter). Because this motion involves oscillations around equilibrium, it includes both kinetic and potential energy, so that’s two more degrees of freedom. As a result, there should be seven total degrees of freedom (3 for translation, 2 for rotation, and 2 for vibration), and the molar heat capacity should be 7/2R.

At high temperatures, this is exactly right, and all diatomic gases have C = 7/2R when the temperature is high enough. But as you lower the temperature, the molar heat capacity drops; first to 5/2R, then to 3/2R. When this was discovered, it was a real mystery.

It appeared that, when the temperature was very low and the thermal energy small, it was impossible to put energy into the form of rotation or vibration, though it could still go into translation. So at low temperature this looked like the diatomic gas could only
translate, just like a monatomic gas. Then as you raised the temperature, it would gradually become possible for the molecule to vibrate, adding two more degrees of freedom and increase the molar heat capacity to 5/2R. Finally, it became possible for the molecule to rotate, and the molar heat capacity rose again to 7/2R.

This strange behavior was one of the earliest recognized signatures of quantum mechanics. The energies for vibration and rotation of these molecules are “quantized”. They can’t have just any values (from zero up) but can only take particular, well separated values. As a result, it is impossible for the molecule to absorb thermal energy into vibration until the typical amount of thermal energy around \( (k_B T) \) is comparable to the jump from one quantum state to another. Once it is, the molecule absorbs energy into this form as freely as any other.

**Solids**

What should molar heat capacities be for solids? Let’s just think about a simple solid, made of a regular lattice of just one kind of atom. Each atom in the solid can vibrate along x, y, or z. Each of these vibrations involves both kinetic and potential energy, so there are 2x3 = 6 degrees of freedom. This would lead us to expect \( C_{\text{solid}} = \frac{6}{2}R \), or around 25 J/mole*K, which is in fact what you usually find.

**More complex materials**

What would the molar heat capacity be for some more complex material, like ice, which is a solid made of molecules of water? For this, you might have the six modes seen for simple solids, in which the whole molecule oscillates back and forth in each direction, then also have modes for bending of the molecule, or oscillation of the molecule around different axes, or vibration along the bonds in the molecule. Each of these might, if it is accessible (doesn’t have quantum states that are too far apart) add to the specific heat. So we’d certainly expect \( C_{\text{ice}} \) to be bigger than \( C_{\text{simple solid}} \). And again, that’s just what we find. The molar heat capacity of ice is actually quite high; about 75 J/mole*K.

**Converting molar heat capacity to specific heat**

Molar heat capacity is the fundamental thing, as it tracks how much energy goes into each atom or molecule. But in ordinary life we don’t count the number of atoms in a sample, we weigh it. So it is useful to know how much heat is required to raise the temperature of a certain mass of material. This is the specific heat:

\[
\Delta Q = m c_{\text{specific}} \Delta T
\]

From this we can see that \( c_{\text{specific}} = \frac{C_{\text{molar}}}{M_{\text{molar}}} \). The fact that different materials have different molar masses means that, although they may have the same molar heat capacities, they will have quite different specific heats. Substances like the noble gases vary in molar mass by a factor of 30, so their specific heats will also vary by a factor of 30. Typical values for specific heats range from 500 to 15,000 J/kg*K. Usually they’re tabulated as J/gram*K, so the numbers you see would be 1000 times smaller.
In general substances with massive atoms will have low specific heats, though only because a sample of fixed mass is actually made of fewer atoms.

**Heat capacity and energy storage**

Once we know something about heat capacity, we can learn an important basic fact about thermal energy; we can quantify it. Let’s think about an example. Imagine you have a boiling hot cup of coffee. Eventually, it will cool off to room temperature, releasing some of its thermal energy into its surroundings. How much energy will this be?

Coffee is mostly water. We know that the specific heat of water is around 4200 J/kg*K. An 8 oz cup of water has a volume of about 240 cm³ or 2.4x10⁻⁴ m³ and a mass of about 0.24 kg. When it cools from boiling (373 K) to room temperature, it drops by about 75 K. The total energy released is 4200 J/kg*K * 0.24 kg * 75 K = 75,600 Joules. Is that a lot of energy?

If you could take this energy and convert it into your kinetic energy, you would end up traveling at:

\[
\frac{1}{2}mv^2 = \frac{1}{2}(80 \text{ kg})v^2 = 75,600 \text{ J} \quad \text{or} \quad v = 43.5 \text{ m/s} = 97.4 \text{ mph}
\]

So if you could somehow extract the excess energy in that cup of hot coffee and use it to get yourself going in the morning, you could in principle shoot out of the kitchen at 97 miles per hour. And that’s without even considering the caffeine! Not surprisingly, the proposed transformation (take all the excess heat and organize it into your motion) is impossible. We will see this more clearly when we talk about heat engines below.

Although this is kind of a joke example, but it does get at an essential point. There is a LOT of energy around in thermal form. This should not be a huge surprise, as we already know of the tendency for energy in any other form to quickly spread out into thermal energy. But it is remarkable just how much energy there is in something like a cup of coffee.

**Changes of state and heat of transformation**

There is one more thing to discuss when talking about how the addition of heat to materials causes changes of temperature; changes of state. When a material goes through a change of state from solid to liquid or liquid to gas, additional heat is required to break the bonds which are holding atoms together. This is illustrated in the figure below.

During the period while the change of state is occurring, heat which is added does not raise the temperature, but rather goes into breaking the bonds that hold the atoms together. The heat required to do this is quantified as the “latent heat of transformation”, and written \( \Delta Q = mL \). So to transform a mass \( m \) from solid to liquid, without raising the temperature, you get it to the melting point then add an amount of heat equal to the mass times the latent heat to break all the bonds.
These latent heats have a wide range of values, from $10^4 \text{ J/kg*K}$ to $2 \times 10^6 \text{ J/kg*K}$. So they are large, larger than specific heats. It takes more energy to break the bonds in a material than to raise the temperature of that same material by one degree.

Tables giving values for some specific heats, molar heat capacities, and latent heats of transformation are given on the next page.

There are a few features of these specific and latent heats which are obviously important for life. As always, water plays a special role. Its high specific heat helps it to play a role in temperature stability. If you’re a 60 kg person, made of mostly water, it takes around a quarter of a million joules to change your temperature by 1 K. Similarly the high latent heat for freezing water makes it difficult to freeze a pond to the bottom, making it easier for life to survive there.
<table>
<thead>
<tr>
<th>Substance</th>
<th>Phase</th>
<th>$c_p$ J g$^{-1}$ K$^{-1}$</th>
<th>$C_p$ J mol$^{-1}$ K$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air (Sea level, dry, 0 °C) gas</td>
<td>1.0035</td>
<td>29.07</td>
<td></td>
</tr>
<tr>
<td>Air (typical room conditions)</td>
<td>1.012</td>
<td>29.19</td>
<td></td>
</tr>
<tr>
<td>Aluminum</td>
<td>solid</td>
<td>0.897</td>
<td>24.2</td>
</tr>
<tr>
<td>Ammonia</td>
<td>liquid</td>
<td>4.700</td>
<td>80.08</td>
</tr>
<tr>
<td>Argon</td>
<td>gas</td>
<td>0.5203</td>
<td>20.7862</td>
</tr>
<tr>
<td>Copper</td>
<td>solid</td>
<td>0.385</td>
<td>24.47</td>
</tr>
<tr>
<td>Diamond</td>
<td>solid</td>
<td>0.5091</td>
<td>6.115</td>
</tr>
<tr>
<td>Gold</td>
<td>solid</td>
<td>0.1291</td>
<td>25.42</td>
</tr>
<tr>
<td>Helium</td>
<td>gas</td>
<td>5.1932</td>
<td>20.7862</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>gas</td>
<td>14.30</td>
<td>28.82</td>
</tr>
<tr>
<td>Iron</td>
<td>solid</td>
<td>0.450</td>
<td>25.1</td>
</tr>
<tr>
<td>Mercury</td>
<td>liquid</td>
<td>0.1395</td>
<td>27.98</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>gas</td>
<td>1.040</td>
<td>29.12</td>
</tr>
<tr>
<td>Oxygen</td>
<td>gas</td>
<td>0.918</td>
<td>29.38</td>
</tr>
<tr>
<td>Silica (fused)</td>
<td>solid</td>
<td>0.703</td>
<td>42.2</td>
</tr>
<tr>
<td>Uranium</td>
<td>solid</td>
<td>0.116</td>
<td>27.7</td>
</tr>
<tr>
<td></td>
<td>gas (100 °C)</td>
<td>2.080</td>
<td>37.47</td>
</tr>
<tr>
<td>Water</td>
<td>liquid (25 °C)</td>
<td>4.1813</td>
<td>75.327</td>
</tr>
<tr>
<td></td>
<td>solid (0 °C)</td>
<td>2.114</td>
<td>38.09</td>
</tr>
</tbody>
</table>

All measurements are at 25 °C unless otherwise noted. Notable minimums and maximums are shown in bold text.

<table>
<thead>
<tr>
<th>Substance</th>
<th>Latent Heat Melting J/g</th>
<th>Melting Temp °C</th>
<th>Latent Heat Vaporization J/g</th>
<th>Boiling Temp °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol, ethyl</td>
<td>108</td>
<td>-114</td>
<td>855</td>
<td>78.3</td>
</tr>
<tr>
<td>Ammonia</td>
<td>339</td>
<td>-75</td>
<td>1369</td>
<td>-33.34</td>
</tr>
<tr>
<td>Carbon Dioxide</td>
<td>184</td>
<td>-57</td>
<td>574</td>
<td>-78</td>
</tr>
<tr>
<td>Helium</td>
<td>21</td>
<td></td>
<td></td>
<td>-268.93</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>58</td>
<td>-259</td>
<td>455</td>
<td>-253</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>25.7</td>
<td>-210</td>
<td>200</td>
<td>-196</td>
</tr>
<tr>
<td>Oxygen</td>
<td>13.9</td>
<td>-219</td>
<td>213</td>
<td>-183</td>
</tr>
<tr>
<td>Water</td>
<td>335</td>
<td>0</td>
<td>2272</td>
<td>100</td>
</tr>
</tbody>
</table>
Physics for the Life Sciences: Fall 2008 Lecture #19

In our previous class, we talked a lot about adding heat to a system. Today we’ll start by discussing how it is that heat moves around. We know that if we put hot and cold things together energy will flow around until an equilibrium temperature is reached. There are three principle methods by which this heat will move

- Conduction
- Convection
- Radiation

We will discuss each in turn. All of these operate within the basic picture of statistical physics, which suggests that heat will tend to spread outward from places where it is localized.

**Conduction**

Conduction is the form of heat transfer which occurs through direct contact. For example, if you put one side of a slab in contact with something hot, and the other side in contact with something cold, heat will flow directly through the slab, as the atoms on the hot side rattle off their neighbors, which bang into their neighbors, and so on until the atoms on the cold side are heated up.

The flow of heat by conduction depends first on some obvious geometric factors. If the slab is thick, heat will flow more slowly. If the slab has a large area through which heat can flow, heat will flow more rapidly. Finally, the rate at which heat flows depends on the nature of the material itself, through a parameter called the *thermal conductivity* $k$ of the material. These are combined together in the following way:

\[
\frac{dQ}{dt} = kA \frac{dT}{dx} \quad \frac{dQ}{dt} = kA \frac{T_H - T_C}{L}
\]

The first equation shows that, at any point, the rate of heat flow will depend on the thermal conductivity, the area through which it can flow, and the temperature gradient. Notice that this depends on a gradient of temperature, in just the same way that diffusion depended on a density gradient. In fact, this equation is identical to the equation for Fick’s law. Heat flow is mathematically identical to diffusion. The second equation shows what this would be for a cylindrical slab like of area $A$ and length $L$, which has a hot end ($T_H$) and a cold end ($T_C$).

The thermal conductivity $k$ has units $J/m*s*K$, or since $1 \text{ J/s} = 1 \text{ Watt}$, you can write this as $W/m*K$. Here is a table of some thermal conductivities:
Thermal conductivity

\( W \cdot m^{-1} \cdot K^{-1} \)

Temperature

\( K \)

<table>
<thead>
<tr>
<th>Material</th>
<th>Thermal conductivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copper (Cu), pure</td>
<td>385 - 401</td>
</tr>
<tr>
<td>Gold (Au), pure</td>
<td>314 - 318</td>
</tr>
<tr>
<td>Aluminum (Al), pure</td>
<td>205 - 237 (220)</td>
</tr>
<tr>
<td>Carbon Steel (Fe+(1.5-0.5)%C)</td>
<td>36 - 54</td>
</tr>
<tr>
<td>Ice</td>
<td>1.6 - 2.2</td>
</tr>
<tr>
<td>Water</td>
<td>0.6</td>
</tr>
<tr>
<td>Wood</td>
<td>0.16 - 0.4</td>
</tr>
<tr>
<td>Snow (dry)</td>
<td>0.11</td>
</tr>
<tr>
<td>Air (1 atm)</td>
<td>0.024 - 0.0262</td>
</tr>
</tbody>
</table>

There are some things to notice about this. First, the metals at the top have very high thermal conductivity. Heat flows through these things very easily. This fact is related to their electrical conductivity. This high conductivity is why we use primarily metal pots and pans, rather than ceramic for instance. The other solids, like ice and wood, conduct heat hundreds of times more slowly. They make nice insulators. Better still is air, which conducts heat 15,000 times more slowly than copper. Air can make a great insulator.

Convection

A second way to move heat around is called convection. The idea of convection is simply that you can move heat by mixing hot and cold material. Mostly this happens in liquids and gases, where flow and the consequent mixing of material is easy. You cause heat flow like this when you stir cream into your coffee. This is sometimes called forced convection. A more important process is “free convection” which uses gravity to mix the hot and cold material, so that the process happens without intervention.

Since most materials expand when they are heated, hot material is less dense than cold. If the fluid is pulled on by gravity, this less dense material will rise due to the buoyant force, which we will explore in more detail later in the class. Any time you have a fluid which is hotter on the bottom than on the top, convection will work to mix it. One case you’re familiar with is heating a pot from the bottom.
Since the only life we know is on Earth, and it all lives under the influence of gravity, free convection like this is a very important process in nature. It drives atmospheric and oceanic currents, and is responsible for plate tectonics.

**Heat transfer through electromagnetic radiation**

The third process of heat transfer is due to electromagnetic radiation. The familiar form of electromagnetic radiation is visible light, but it appears in many guises. They are all really the same, and differ only wavelength. Other forms of electromagnetic radiation include radio waves, microwaves, infrared radiation, ultraviolet radiation, x-rays, and gamma-rays.

We have learned that the atoms which make up all materials are always in motion, shaking back and forth and bouncing off one another. It turns out that any time a charged particle (like an electron or a nucleus) is accelerated, it will emit electromagnetic radiation. The thermal motion of atoms always involves these accelerations, and they’re all made of charged particles, so all materials are always emitting electromagnetic radiation, and this radiation carries away energy.

The amount and nature of the radiation each object emits depends primarily just on temperature. The basic rule is encoded in Stefan’s Law:

\[
\frac{dQ}{ddtA} = R = e \sigma T^4
\]

The heat loss per unit area per unit time R depends on the temperature to the fourth power, a universal constant \( \sigma \), and a property of the surface emitting the radiation called the emissivity \( e \). The constant \( \sigma \), called Stefan’s constant, has a value of:

\[
\sigma = 5.7 \times 10^{-8} \text{ W/m}^2\text{K}^4
\]

The emissivity is a measure of how good a radiator this surface is. It depends on both the material the object is made of and the condition of the surface (is it rough or smooth). The emissivity takes on values from zero to one only. A perfect emitter would have \( e = 1 \) and a completely failed emitter would have \( e = 0 \).

Notice that the heat loss rate changes rapidly with the temperature, as the fourth power. If you double the temperature, you increase the heat loss rate by sixteen times. What’s the typical scale for this? Assume \( e = 1 \), and room temperature of 300 K. For this you get a heat loss rate \( R = 460 \text{ W/m}^2 \). Every object at room temperature emits approximately 400 Joules per second from every square meter of surface it has.

This is quite a loss rate. If everything is radiating away energy like this, why doesn’t everything rapidly cool off? The reason objects don’t cool wildly is because radiation goes both ways. A book on the table (for example) is surrounded by other objects at roughly the same temperature. Those other objects are emitting radiation which is then absorbed by the book. So while radiation is going out, it also comes in. Notice that this balancing process (out and in) will tend to make all the objects approach the same temperature.
There is another important point here. The emission of radiation and the absorption of radiation are inverse processes. That is to say, they’re really the same, just reversed in time. Because of this, any material which is a good emitter (ε ∼ 1) is also a good absorber, and any material which is a bad emitter (ε ∼ 0) is a bad absorber. This fact aids in allowing radiation to balance temperature. Anything which emits a lot will also absorb a lot, while anything which emits very little will absorb very little.

Think a bit about this. Which is a better emitter of visible light, a piece of black paper or a mirror?

**The nature of thermal electromagnetic radiation**

If everything is emitting radiation all the time, why can’t you see in the dark? Well, you can see some things in the dark, things which are hot enough to emit visible light!

The thermal radiation emitted by objects comes out in a broad distribution of wavelengths, but the peak of this distribution changes with temperature according to Wein’s law:

$$\lambda_{\text{peak}} = \frac{2.9 \times 10^{-3}}{T}$$

For a temperature of 300 K, this is about 10 μm. Radiation with this wavelength is in the infrared part of the spectrum, a part which your eyes cannot detect. As a result, you don’t see room temperature things through light which they emit. You only see them through visible light which they reflect. The figure below shows the “electromagnetic spectrum” and points out what part of the spectrum is emitted by objects of different temperatures. To emit significant amounts of visible light, an object must reach around 1000 K or so.
Size, thermoregulation, and life

We have seen that life uses random thermal motion to accomplish many things, including things like diffusion and osmosis. One strategy for ensuring the stability of these thermal effects is to maintain a constant temperature. At constant temperature all of these processes (diffusion, osmosis, etc.) run along at a fixed rate, and the organism can count on high performance at all times.

This constant temperature approach, taken by “homeotherms”, is effective, but costly in terms of energy usage. For many organisms (like us) it also requires constant vigilance. Many or our functions fail catastrophically if body temperature moves outside relatively narrow limits. The other major approach is taken by “ectotherms”, which generally take on the temperature of their surroundings. They have the advantage of paying the cost of maintaining their temperature, but they can only function very slowly at low temperature.

Size is a major factor in the difficulty of thermoregulation. Recall that surface area varies like size\(^2\), while volume varies like size\(^3\). The generation of heat in an organism depends on the number of cells, so basically the volume. Heat loss, by whatever means, depends on surface area. So the ratio of heat lost to heat generated changes as:

\[
\frac{\text{heat lost}}{\text{heat generated}} \propto \frac{\text{size}^2}{\text{size}^3} = \frac{1}{\text{size}}
\]

As organisms get large, heat loss is not a problem. In fact, for very large organisms, overheating is the problem, which is why the really large mammals are often hairless and include heat loss mechanisms like very large ears (which have very large surface area and little volume).

On the other end, small creatures lose the heat they produce more and more quickly. At this extreme lie creatures like shrews and hummingbirds. To replace the heat they lose so rapidly these tiny creatures must consume relatively enormous quantities of food, many times their own weight each day. Even doing this, these creatures live on the edge, and unlike many homeotherms they are able to survive dramatic reductions in body temperature. When faced with suddenly cold conditions they can enter a dormant state, allowing their body temperature to drop. Provided they can eventually use environmental heat to recover (sunlight for example) they survive this without harm. Many small homeotherms follow this strategy.

There are many other fascinating aspects to thermoregulation in organisms. It’s a major part of what life has to accomplish.
We’ve been learning about how life makes use of random thermal motion to get things done. The practical inevitability of the increase of entropy drives mechanisms like diffusion. We have seen that to make these processes happen reliably, many organisms maintain a fixed, generally elevated temperature.

Given the variety of mechanisms for heat loss, maintaining a constant temperature comes at significant cost. People, for example, must maintain their temperature in quite a narrow range around 37° C. If your temperature falls below 32° C you would suffer severe hypothermia and after a short time would die. If your temperature rises above above about 41° C you may suffer brain damage and above 45° C death is nearly certain.

Why do these dire consequences occur? At low temperatures, your normal cellular processes slow down, eventually become too slow to maintain the finely coordinated sequence of reactions that your body requires. At high temperatures everything cooks along, but the temperature becomes high enough to begin to destroy some of the very proteins you need for your cells to function. This narrow survival range is one of the clearest signs of how important these purely random thermal processes are for life.

Life and the formation of structures

One of the surprising things about life is its ability to construct very complex structures. On its face, this seems to fly in the face of the second law of thermodynamics, which tells us that the steady increase of entropy is inescapably likely. How can complex structures form unbidden in a universe governed by the steady increase of entropy?

The apparent conflict between the second law and the formation of structure is resolved by recalling that most of the systems we have studied were carefully arranged as closed systems. Typically neither matter nor energy is allowed to enter or leave. Such a system, like our box filled with gas atoms, will indeed rapidly approach a uniform, structureless, thermal equilibrium. Once it does so nothing significant will ever happen. Atoms will still move around, things change on the microscopic level, but the macroscopic state of the system will never change.

But few systems on Earth are ever so carefully isolated. Energy flows into and out of systems constantly. This changes the accounting completely, and allows surprisingly complex structures to form. To see how this works, leave life aside for a bit and consider how something as impressive as a snowflake forms.

It begins with some water vapor, spread more or less at random, containing some energy. Surround this vapor with a cold environment and the energy within it will, as usual, spread out as much as possible. This spreading could happen by just cooling the vapor, taking away the energy of each atom leaving them separate.
But it’s possible to do more. Water molecules can give up even more energy if they lock together in a solid form. To make this happen maximally, each molecule must lock into a very regular crystal structure with its neighbors. When it does this, even more of the energy originally localized in the water is allowed to spread out, an outcome which statistical mechanics suggests is inevitably likely.

The incredible structures you see in snowflakes are all examples of how such crystals can form when they grow. The great variety of structures you see come about because the conditions for crystal growth are each time just a bit different. In a laboratory it is possible to make the conditions very stable, and hence to repeatedly grow the same kinds of snowflakes. These amazing structures form completely by chance, through the operation of the same thermodynamic laws which guarantee that gas atoms released in a box will spread uniformly through it. They do so because, although the final structure of the atoms in these snowflakes is extremely ordered, the final distribution of the energy which they once possessed is vastly more disordered. Taken together, entropy has increased.

Structure formation in the non-living world is all around you, including the formation of galaxies, stars, and planets, all the way down to diamonds, raindrops, and clouds. These structures form because in doing so entropy increases.

The same essential mechanisms are used by life for the creation of complex structures. When your cells build a protein, they don’t grab each atom, carry it across the cell and carefully put it in place. Instead, they just have the right ingredients in the right conditions to allow random thermal motion to put the protein together. Making those conditions just right for all the interlocking constructs of biochemistry to appear in a coordinated way is what makes life seem so different from non-living matter. But in fact the same essential processes are occurring.

**Processes and cycles**

The formation of structures along the march toward increased entropy can be understood. But life is more than the formation of structure. Living things are active, they move around, grow, do things repeatedly rather than once. Living things have many cycles. How can cyclic processes like this occur if the inexorable march toward high entropy is behind everything that’s happening?

Again, the problem is in the accounting. Closed systems, left alone, can’t undergo cyclic behavior. Somewhere along the way this would have to violate the second law of thermodynamics. But the Earth is hardly a closed system. Energy flows into the Earth continuously from the Sun. It flows out continuously as well, spreading out into space. Take away the Sun and the Earth’s surface would rapidly cool to the nearly the
background temperature of space, about 3 K. Stop the Earth from radiating away heat and it would rapidly overheat. Of course that’s perhaps what we’re doing right now…

So the Earth is a system through which energy flows. When energy flows through a system, cyclic processes emerge very naturally. Think a bit about what the Earth would be like without life. Seen from space it would seem very much the same. Weather cycles would continue, cycling water through the atmosphere, lifting it across the continents where it could flow back toward the oceans. The annual weather cycles, driven by the changing orientation of the Earth relative to the sun would be unchanged. Thunderstorms, hurricanes, and tornados would continue to naturally concentrate enormous amounts of energy as a consequence of accidental conditions. Ocean currents would flow, annually carrying icebergs to lower latitude. Volcanoes would erupt, reforming the land, while within the Earth convection in the mantle would continue to drive continental drift, rearranging the continents every few hundred million years. The Earth, without life, remain a very lively place.

There are other places in the solar system where similarly interesting cyclic, grossly repeated behavior happens on its own. The Sun, site of such a huge energy flow, is perhaps most dramatic, with wild surface storms rising and falling in number through an ~11 year cycle. Similar weather affects the other planets too, perhaps most impressively on Jupiter, where the red spot (a 400 year old storm) was joined a few years ago by a new storm.

Two pictures of lively weather elsewhere in the solar system: On the left is a montage of images of the Sun taken as it progress from the peak of the solar cycle (on the left) to the minimum. On the right is a picture of Jupiter from April 2206, showing the new storm which has recently joined the famous “red spot” storm first observed by Giovanni Cassini in 1655.

Most of these phenomena are driven by convection. They “just happen” because they allow energy to spread more thoroughly, speeding its escape from these open systems.
What about life?

Is this what life is, a kind of natural thermodynamic process which happens because it speeds the increase of entropy? To think about this more, it’s useful to ask what life is, where we draw the line between the living and the non-living. There is no complete consensus about what marks this line. The Oxford English Dictionary waffles, defining life as:

“The property which constitutes the essential difference between a living animal or plant, or a living portion of organic tissue, and dead or non-living matter; the assemblage of the functional activities by which the presence of this property is manifested.4”

Not too satisfying. This definition says life is what makes the difference between living and non-living matter. Ernst Mayr, one of the 20th century’s leading biologists, expressed his exasperation with the problem this way: “Attempts have been made again and again to define 'life'. These endeavours are rather futile since it is now quite clear that there is no special substance, object, or force that can be identified with life.5”

Still it seems useful to try. Probably the narrowest, most widely accepted definition of life is something which reproduces itself with the possibility of modification. This lets in all the widely accepted living things. Some prefer a longer list of criteria, here’s the Wikipedia version6, which captures most of the properties usually raised:

1. **Homeostasis**: Regulation of the internal environment to maintain a constant state; for example, sweating to reduce temperature.
2. **Organization**: Being composed of one or more cells, which are the basic units of life.
3. **Metabolism**: Consumption of energy by converting nonliving material into cellular components (anabolism) and decomposing organic matter (catabolism). Living things require energy to maintain internal organization (homeostasis) and to produce the other phenomena associated with life.
4. **Growth**: Maintenance of a higher rate of synthesis than catalysis. A growing organism increases in size in all of its parts, rather than simply accumulating matter. The particular species begins to multiply and expand as the evolution continues to flourish.
5. **Adaptation**: The ability to change over a period of time in response to the environment. This ability is fundamental to the process of evolution and is determined by the organism's heredity as well as the composition of metabolized substances, and external factors present.
6. **Response to stimuli**: A response can take many forms, from the contraction of a unicellular organism when touched to complex reactions involving all the senses

---

4 Oxford English Dictionary online
of higher animals. A response is often expressed by motion, for example, the leaves of a plant turning toward the sun or an animal chasing its prey.

7. **Reproduction**: The ability to produce new organisms. Reproduction can be the division of one cell to form two new cells. Usually the term is applied to the production of a new individual (either asexually, from a single parent organism, or sexually, from at least two differing parent organisms), although strictly speaking it also describes the production of new cells in the process of growth.

Looking at this list, the feature most difficult to mimic with engineering is reproduction. Life is particularly good at this, and so perhaps this is remains the best dividing line between the living and the non-living.

A long discussion offering a variety of alternate definitions of life in the Encyclopedia Britannica concludes by pointing out the particular problem biology currently faces; we only know about one form of life, that on Earth, and all of the life on Earth seems completely related.

“The existence of diverse definitions of life surely means that life is something complicated. A fundamental understanding of biological systems has existed since the second half of the 19th century. But the number and diversity of definitions suggest something else as well. As detailed below, all the organisms on the Earth are extremely closely related, despite superficial differences. The fundamental ground pattern, both in form and in matter, of all life on Earth is essentially identical. As will emerge below, this identity probably implies that all organisms on Earth are evolved from a single instance of the origin of life. It is difficult to generalize from a single example, and in this respect the biologist is fundamentally handicapped as compared, say, to the chemist or physicist or geologist or meteorologist, who now can study aspects of his discipline beyond the Earth. If there is truly only one sort of life on Earth, then perspective is lacking in the most fundamental way.”

Among all this confusion there is a clear consensus that, whatever life is exactly, all living things exist in open thermodynamic circumstances in which they take in resources, use them to construct themselves and near replicas of themselves, then expend these resources, always at the expense of substantial increases in entropy.

We have seen that outcomes which increase entropy are statistically likely, so likely as to be inevitable. In this way of thinking, life may be made likely by the same mechanisms which cause other cyclic phenomena in open thermodynamic systems weather on the Earth. If this is so, life (whatever that is) should exist anywhere in the universe where conditions allow it. This prediction presents science with one of its most tantalizing challenges for the future. If we understand life, we really should find it in many places. There’s a real chance this will happen during your life, and perhaps one of you will find it.

---


Copyright: Regents of the University of Michigan, 2009
Physics for the Life Sciences: Fall 2008 Lecture #21

Thermodynamic cycles and engines

One way of better understanding how the complex engines of life might work is to consider simple, man-made thermodynamic cycles and the mechanical engines which they drive. Let’s define one first. A thermodynamic cycle is a system which starts in some state, then goes through a series of changes, returning eventually to the original state, while doing something along the way. This might be an engine going through a single power stroke, consuming some fuel along the way, then returning to its initial state. Or it might be a muscle cell, suddenly tugging itself shorter, using up some ATP along the way, then returning to its initial state.

To explore such thermodynamic cycles, we’ll consider a simple system: some quantity of ideal gas contained within a piston which begins at some initial temperature $T_i$, pressure $P_i$, and volume $V_i$. As it goes through a cycle, $T$, $P$, and $V$ will change in some way, eventually returning to the initial state. This ideal gas model system is useful to discuss partly because it is simple, but also because it actually underlies many of the engines we use to create lifelike motion. There are a semi-infinite variety of other possibilities we could consider, and it is important to remember that this little enclosed sample of ideal gas is only one of many.

Imagine you allow such a gas to expand a bit by pushing the piston outward through some distance $dx$. Since the gas is pushing on the piston with a force $F = PA$ (with $A$ the area of the piston), the gas must do work on the piston:

$$W_{\text{gas on piston}} = Fdx = PAdx = PdV$$

In the last step we noted that $Adx$ is just the change in volume of the gas $dV$. As we move the piston out, the pressure may change too, and we can track the total work done by the gas by adding up the quantity $PdV$ in each little step $dx$. 

![Diagram showing work done by ideal gas](image)
This is clarified by considering the whole ‘path’ of the system in the P, V plane. It starts at \((P_i, V_i)\), and moves step by step to \((P_f, V_f)\). For each little change in volume \(dV\), the gas does work \(PdV\). So the total work done by the gas is the area under this \((P, V)\) curve. If the gas is doing work positive work, it would be losing energy, but of course it’s possible that something is putting energy back into the gas. All that is required is for the gas to do negative work; then it must be absorbing energy. What we see here is how P and V change.

In a thermodynamic cycle, P and V start out somewhere, then change continuously, eventually returning to their starting point. The figure at right shows an example. To understand such thermodynamic cycles in detail, it is helpful to consider a few particular kinds of processes which the gas might undergo. We will then use these particular processes to construct thermodynamic cycles for use in machines.

**Isothermal processes**

One kind of process is an isothermal one, a process in which the temperature of the gas stays the same. Since the gas is doing work while expanding or contracting, it is either gaining or losing internal energy. To keep the temperature the same, something is going to have to either remove or add some heat while the gas expands or contracts. This might be accomplished by placing the gas in contact with a large heat reservoir. Then, if the gas gets a little cooler due to expansion, heat can flow in, replacing the energy lost due to expansion. If it gets hotter due to compression, heat can flow out, removing the energy gained in compression.

If the process is isothermal we can use the ideal gas law to tell us exactly what the path is in the \((P, V)\) plane.

\[ PV = nRT \quad \text{implies} \quad P = \frac{nRT}{V} \]

This figure shows both an isothermal expansion (which requires putting heat in to maintain the temperature) and an isothermal contraction (which requires removing heat to maintain the temperature). This particular path in the plane, with \(P \propto 1/V\), is the one for which \(T\) remains the same. Any path where \(P\) falls more rapidly than this is one in which the gas temperature drops. Any path where \(P\) falls slower than this is one in which the gas temperature rises.

For example, an “isochoric” path which maintains constant volume but changes pressure would be a line straight up or down in this plane. If you go straight down, pressure is falling more steeply than isothermal, and the temperature must be falling. If you go
straight up, pressure is falling more slowly than isothermal (in fact it’s increasing) and the temperature must be rising.

How much work is done in this isothermal expansion? We know that the work is given by the definition:

\[ W = \int P \, dV = \int (nRT/V) \, dV = nRT[\ln(V_f) - \ln(V_i)] \]

This is the amount of work done by the gas during an isothermal transformation. Notice that it depends on both the temperature at which it occurs (it would do more work if the temperature were higher) and on the change in volume \( V_f/V_i \) (it would do more work if the change in volume is large).

How much heat has to be put in to make this isothermal process happen? The amount of heat in has to exactly balance the work done, or else the temperature will change, so:

\[ Q_{\text{in}} = W_{\text{out}} = nRT\ln(V_f/V_i) \]

**Constructing a cycle, the Stirling Cycle**

Any closed path in this (P,V) plane would be a thermodynamic cycle, starting, moving around doing some work etc., and returning to the same initial condition. How much work would be done by a cycle? If the cycle goes clockwise, as shown, positive work is done during the expansion on the top, while negative work is done during the contraction on the bottom. The positive work is the area under the top curve, while the negative work is the area under the bottom, so the net work is the area in between the two.

To complete the cycle, we’d have to put a net amount of heat into the cycle which is just equal to the total work done. Otherwise we wouldn’t be able to return the gas to exactly the same point in the (P, V) plane it started at. From this, you can start to see what a thermodynamic cycle like this does. To complete it, you put thermal energy (heat) in, and you get mechanical energy (work, pushing on the piston) out. That’s what engines like steam engines do, and they are central to technology because they allowed people, for the first time, to harness thermal energy to accomplish mechanical work; tasks like moving things around, grinding wheat into flour, or operating machines.

To get a flavor for the details, we will consider in detail one particular cycle, though there are in fact infinitely many which are possible. This one is called the “Stirling cycle”. In this cycle you go through four steps:

1. Start out with an isothermal expansion at some high temperature \( T_H \) from volume \( V_1 \) to \( V_2 \). During this phase, the gas does positive work, pushing the piston
outward. You could use this part of the work to do something. The amount of positive work done is given by $W_H = nR\ln(V_2/V_1)$. To maintain the temperature you have to add some heat to balance this $Q_H = W_H$.

2. After the expansion, the gas undergoes an isochoric (constant volume) decrease in pressure. This is also a decrease in temperature to a new temperature $T_C$ (think of $PV=nRT$ with $V$ constant...), so heat must be removed in this step. How much heat is removed? How much heat depends on the molar heat capacity of the material, $\Delta Q_{\text{right}} = nC_{\text{gas}}(T_H-T_C)$.

3. Now the gas undergoes a new isothermal compression. During this compression, the gas does negative work (that is, work is done on the gas). To keep the temperature of the gas constant at $T_C$ while compressing it, heat must be removed. The amount of heat removed equals the work: $W_C = nR\ln(V_1/V_2)$

4. In the final step, you add back heat to take the gas back up to the hot temperature $T_H$. This requires you to add heat $\Delta Q_{\text{left}} = nC_{\text{gas}}(T_H-T_C)$. Note that in this step you’re just putting back exactly the same amount of heat you took out in step two.

What’s the total work done? $W_H - W_C = nR(T_H-T_C)\ln(V_2/V_1)$

This could be increased by either making the volume change larger, or by making the temperature difference between the hot and cold ends larger.

**Efficiency in thermodynamic cycles**

If you think of a thermodynamic cycle like this as something which converts random thermal motion into work (something like ordered motion) it’s worth asking how efficiently it does this. One way to measure this would be to compare the amount of work out to the amount of heat put in. For this Stirling cycle, the heat put in is $Q_H$, and the work is given by the equation above. So here we might have:

$\text{Efficiency} = \frac{\text{Work out}}{\text{Heat in}} = \frac{nR(T_H-T_C)\ln(V_2/V_1)}{nR\ln(V_2/V_1)}$

$\text{Efficiency} = \frac{(T_H-T_C)}{T_H} = 1 - \frac{T_C}{T_H}$

What does this mean? It says that, unless $T_C = 0$, the efficiency of this kind of cycle cannot be 100%. But it offers the possibility that, with $T_H$ very large and $T_C$ very low, a mechanism like this might allow you to turn most of the heat you put in into work. Remember too that this is the theoretical limit. Anything which happens during the cycle which creates any additional loss, like not being able to completely recycle the heat you took out on the right and put it back in on the left, will lower this further.

Let’s think about this practically though. Here on Earth, there are limits to how hot the hot end can be and how cold the cold end can be. Generally speaking, the “cold” end will be at the ambient surrounding temperature, while the hot end will be something like the
temperature of boiling water. So we might have $T_H = 373 \text{ K}$ and $T_C = 300 \text{ K}$. This would give an efficiency of:

$$\text{Efficiency} = 1 - \frac{300}{373} = 0.2 \quad \text{or} \quad 20\%$$

That’s the best you can do with those constraints. Only 20% of the heat you take from the boiling water will get turned into work by this Sterling cycle engine. What happens to the rest? The rest of the heat gets pulled out of the hot reservoir and dumped into the cold reservoir. That is, it gets spread out throughout the part of the world which is at the ambient temperature around 300K.

Why not just run the cycle again? Take that heat you remove back up to 373 and send it through again. The problem is, you can’t do that. Once it’s spread out at the low temperature, you can’t get it back up to the high temperature. So the heat you expel at low temperature, 80% of what you took out the boiling water, is wasted, lost in an unrecoverable way.

This is the key of thermodynamic cycles. They can create order, taking random motion and converting some of it into ordered motion, but they can only do this at the cost of more effectively spreading out some of the heat taken in.

Just like other thermodynamic processes which ‘just happen’, life does this too. It gets things done, creates order, moves around, reproduces. But all this happens only at the cost of speeding the collective increase in entropy.
Physics for the Life Sciences: Fall 2008 Lecture #22

A new subject: fluids

The Earth is a wet, airy place, distinguished from the other planets in the solar system by the presence of liquid water. Life exists in, and is largely made of, two different fluids; air and water. While the two share many properties as fluids, one is a gas, the other a liquid. We will see that this creates differences in both quantity (their densities differ by a factor of 1000) and quality (air is easily compressed, while water is as stiff as a solid).

The physical behavior of fluids includes a number of phenomena we haven’t discussed before. These unusual physical properties generate both opportunities and limits for life. As we have seen before, life exists within these limits, but evolution has allowed life to explore an extraordinary range of strategies for doing this.

The most important property of fluids is their ability to flow, to relatively freely alter the arrangements of their atoms. Life moves within fluids; using them to push against, slow down, and generate lift. It also pulls them in, pushes them out, and uses them in a form of forced convection to internally deliver nutrients and remove wastes. By understanding the physical properties of fluids, we’ll be able to understand better life on land, in the sea, in the air, and at the boundaries among these.

Fluids vs. Solids:

To start, let’s consider what makes a fluid a fluid, and different from a solid. We have already discussed how solids respond to the application of various static loads. We have seen that the basic response of a solid to a load is a deformation:

\[ \sigma_{\text{tensile}} = Ye_{\text{tensile}} \quad \text{for tensile or compressive loads} \]
\[ F/A = Y(\Delta L/L) \]
\[ \sigma_{\text{shear}} = Se_{\text{shear}} \quad \text{for shear loads} \]
\[ F/A = S(\Delta x/L) \]
\[ \sigma_{\text{bulk}} = Be_{\text{bulk}} \quad \text{for hydrostatic loads} \]
\[ F/A = B(\Delta V/V) \]

So although there are different kinds of loads, and different responses to them, the situation with solids is actually pretty simple. While many solids, especially biological materials, do not behave in precisely this linear ‘Hooke’s Law’ fashion, they still respond to loads by deforming. Apply a stress to a material and it will undergo some deformation until the forces balance, and then it will stop moving.

Fluids exhibit a fundamentally different behavior when a stress is applied to them. Their response is so different it’s somewhat difficult to imagine doing it. How do I apply a stress to a tank of water? You can’t simply grab it and pull. So we’ll start with a seemingly limited case. Consider a tank of water and imagine what happens if I apply a shear stress to it. We might do this in two ways.
First, if I could ‘grab’ the top of a tank of fluid and push it sideways, what would happen? Unlike a solid, the fluid doesn’t just distort a bit and push back on me. Instead it actually begins to flow, moving in the direction of the force without limit.

There is of another way to apply a shear to a fluid like this; tilt it sideways and let gravity apply the shear. Picture a cup filled with water. When it is straight up and down, gravity applies only a compressive stress to a layer of the water. The water can stand this. Like a solid is does squash inward a bit, until the pressure in it balances the downward force of gravity. Tip the cup on its side, however, and the stress becomes a combination of compressive stress into the fluid and a shear stress along the surface. This is illustrated in the picture below.

You can see that in a tilted object gravity exerts a shear on each layer. If I tilt a solid like this, the same shear is applied. But the solid just deforms a little, bending to the side. If I tilt a fluid like water, instead of distorting, it immediately begins to flow.

This is a key difference between a fluid and a solid. If you place a solid under a shear stress, it distorts by a fixed amount. If you place a fluid under a shear stress it flows instead of distorting. Schematically, this gets expressed in stress-strain equations like this.

For solids, we have the familiar:

\[
\text{Shear stress} \propto \text{Shear strain (or distortion)}
\]

For fluids we will have:

\[
\text{Shear stress} \propto \text{Shear rate (or flow rate)}
\]

How is it that a fluid can do this? In a fluid, the amount of energy required to move one atom in the material past another is relatively small. In particular, it is smaller than the typical thermal kinetic energy which the atoms possess. Since they already have enough energy to move past one another, they do so with no special effort. Give them the slightest urging, a little bit of shear stress and their motion past one another just happens.

Now consider a hydrostatic load. Start with a liquid. Fill a piston with water and push down. What happens? The water compresses a bit until it is pushing back up just as hard as we push down on it. This is just like the response of a solid, and liquids at least are almost as ‘stiff’ as solids, with bulk moduli in the \(10^9\) N/m\(^2\) range. It turns out that liquids also behave a lot like solids in hydrostatic tension too, though this is less commonly observed. We will talk about it in the context of siphons a bit later.
Gases in compression and tension are, as we’ve seen, different from liquids and solids. Exactly what happens when I compress a gas depends on whether I allow the temperature to change. The pressure the gas exerts back on me will generally change, but very slowly, approximately as the ideal gas law suggests:
\[ PV = nRT \]
So if, for example, T is held constant and the volume is decreased the pressure in the gas will increase, but slowly. To double the pressure you have to halve the volume. Changing the volume of a gas is very easy compared to changing the volume of a liquid or solid. Think about how hard you would have to push to halve the volume of a brick.

We will treat both kinds of fluids (gases and liquids) the same way most of the time, but it will occasionally be essential to remember the easy compressibility of gases.

Why do I have to put the fluid in a container before pushing down on it? I can take a block of solid and just push down on it, applying a normal compression. If I do that with a column of fluid, it immediately escapes my pressure by squirting out the sides. Fluids don’t support normal compressions and tension, but only ‘hydrostatic’ compression or tension which pushes in on every side equally. If you give the fluid any little avenue of escape, it will take it, immediately flowing out of whatever hole you provide.

Hydrostatic pressure like this is something which, in a sense, points in every direction. If the pressure at a point is \( P \), the fluid will push away from that point with a force per unit area \( F/A = P \) in every direction.

Q: Is hydrostatic pressure a vector quantity? What about mechanical pressure?

**Hydrostatics**

When we talk about fluids a key thing to understand is hydrostatic pressure. If you push down on the top of a fluid, it will flow in whatever direction it has to in order to escape this pressure. So if you place a fluid in a cylinder and push down on the top with a tightly sealed piston (so that it can't flow out) you will generate a "hydrostatic" pressure inside it. Pressure is measured in F per unit Area, or \( \text{N/m}^2 \). This unit, \( 1 \text{ N/m}^2 \), is also called a "Pascal".

What does this mean? If you push down on the top with a certain force per unit area, that water will push outward, in every direction, on every part of the cylinder, with exactly the same pressure. This is because the fluid is doing everything it can to escape this pressure. In other words, if I push on the top with a pressure of \( 20 \text{ N/m}^2 \), the pressure with which the fluid will push out against the walls will increase everywhere in it by \( 20 \text{ N/m}^2 \).

**Gravity and hydrostatic pressure**

You’re not the thing only which can create pressure in a fluid. Here on Earth at least, the fluid can do it on its own. Let's think about how the pressure in a fluid is affected by the weight of the fluid itself.
Consider a stable tank of fluid. Each layer is static, so $\sum F=0$. This means

$$F_{\text{down}} = F_{\text{up}}$$

$$P_{\text{top}} \times \text{Area} + \text{Weight} = P_{\text{bottom}} \times \text{Area}$$

$$\text{Weight} = \rho g A \Delta h$$

$$P_{\text{top}} + \rho g \Delta h = P_{\text{bottom}}$$

Where I have used the fact that $m=\rho A \Delta h$. Now I can do exactly the same thing we did when we considered how much force is required in a stack of bricks to support its own weight.

Consider a point a distance $H$ from the top; adding up the pressure from each layer above it we find:

$$P_H = P_{\text{top}} + \rho g H \quad (= \text{P}_{\text{atmosphere}} + \text{Gauge Pressure})$$

Where the first term is the external pressure (very often this is the atmospheric pressure) and the second term is the pressure due to the fluid itself, which is sometimes called the gauge pressure. It's called the gauge pressure because often this is the pressure which we measure; the pressure difference between the outside (atmospheric pressure) and the pressure at some depth. For the two fluids we're concerned with, the densities are roughly $\rho_{\text{air}} = 1.2 \text{ kg/m}^3$ and $\rho_{\text{water}} = 10^3 \text{ kg/m}^3$. The large difference between these means that pressure will increase about $1000x$ more rapidly with depth in water than in air.

It's important to stress that what causes the pressure increases here is gravity. In the absence of gravity there will be no gauge pressure in the fluid, no increase in pressure with depth.

Why is hydrostatic fluid pressure different from what happens with a stack of blocks? The difference is that the pressure in the stack of blocks is only up and down, it does not act in every direction. You don't have to push inward on the sides of a stack of blocks to keep them from squirting out sideways.

Does the change of pressure with depth depend on the shape of the vessel? No, you actually get the same change in pressure with depth independent of the shape of the vessel. The change in pressure with depth in these three vessels is exactly the same. This somewhat surprising fact is sometimes called the “hydrostatic paradox”, though of course it isn’t really a paradox, it’s just a surprising fact. We’ll discuss the explanation for this in class. For the meantime you can think it over.
Atmospheric Pressure

We have written the pressure as $P = P_{\text{atmosphere}} + \rho gh$. What is this atmospheric pressure? This is the hydrostatic pressure with which the atmosphere pushes in on every object. Every thing on the surface of the Earth is submerged in an ocean of air. That air squeezes inward, pressing in on every exposed surface of every object with this pressure $P_{\text{atm}}$.

How large is this pressure? At sea level (Ann Arbor is only 800 feet above this) the magnitude of atmospheric pressure is about $10^5$ N/m$^2$. This is a very large pressure; it means is that every square meter of exposed surface has a force of $10^5$ N pushing on it. A $10^5$ N force is equal to the weight of a 10,000 kilogram mass.

Take a smaller space, like the palm of your hand. This is maybe 8 cm x 8 cm, or 0.0064 m$^2$. The force on this area, on the palm of your hand, is about 640 N. So right now, there is a force pushing down on the palm of your hand equal to the weight of a 60 kg person. How come you don't feel this person standing on your hand? While there is a very large force pushing down on the top of your hand, there is an (almost exactly) equal force pushing up on the bottom of your hand.

Atmospheric pressure is hydrostatic; it surrounds you and pushes inward from every direction. It squeezes you inward, and if we were to remove it, you would expand outward a bit, with your atoms moving a little bit farther apart. How much would your volume change? To estimate this we could use:

$$\sigma_{\text{bulk}} = B \varepsilon_{\text{bulk}} \text{ or } \frac{F}{A} = B \left( \frac{\Delta V}{V} \right)$$

Putting in the change in pressure ($10^5$ N/m$^2$) and the bulk modulus of water (2.15 x $10^9$ N/m$^2$) we find: $\Delta V/V = 10^5$ N/m$^2$ / 2.15 x $10^9$ N/m$^2$ = 4.7 x $10^{-5}$ = 0.0047%. You'd expand a little, but hardly enough to notice.

Atmospheric pressure is generated by exactly the same thing which causes gauge pressure. The atmospheric pressure at sea level is just due to combined weight of the atmosphere piled up on top of us. We live at the bottom of an "ocean of air".

Why do you feel outside pressure in a different way when you're under the water? For example, if you dive to the bottom of a pool, you feel a dramatic water pressure pushing in on your ears. Recall the difference between gases and liquids. Gases are easily compressed, liquids are not. So as you go underwater, the pressure of the water increases. This pressure increase is more than enough to compress the gas inside you eardrums. So your eardrum bends inward, stretching it in a painful way.

One more thing to note; if you have ever gone snorkeling, you may have wondered why snorkels are always so short, barely long enough to reach above the surface. Imagine what would happen if you tried to a longer one, deeper in the water. Outside you the water pressure would be: $P_{\text{air}} + \rho gh$, and inside your lungs, which are directly connected to the air by the snorkel, the pressure would be $P_{\text{air}}$. 

Copyright: Regents of the University of Michigan, 2009
If you tried to do this at a depth of 10m, you would be squashed inward by the large pressure difference $\rho_{\text{water}}gH = 1000 \text{ kg/m}^3 \times 9.8 \text{ m/s}^2 \times 10 \text{ m} \sim 10^5 \text{ N/m}^2$. This kind of pressure would quickly squash your lungs forcing out all the air. In practice, the greatest depth the typical person can handle with a snorkel is only a few feet. This is because you can expand your chest against a pressure only a small fraction of atmospheric pressure. If you try to do it deeper than this your chest will be compressed and all the air squeezed out. This will leave you with the feeling of having the wind knocked out of you.

How do divers go deeper? Snorkelers do it by holding their breath, closing off the connection between the air at the surface and the air in their lungs. Then when they dive underwater the air in their lungs can be compressed without being forced out. As they return to the surface this air expands again, so that by the time they reconnect with the air through the snorkel it is again at normal atmospheric pressure.

Scuba divers use a fancier trick to allow them to go deeper still. The compressed air in their tanks is passed to their lungs through a “regulator” which matches the pressure of the air the diver receives to the pressure of the water which surrounds them. This generally works well, though there are important complications of breathing high pressure gas, especially the increased quantity of nitrogen which dissolves in the blood at high pressure.

What happens when the wind is knocked out of you? Usually you have air inside and outside your lungs at about the same pressure. So breathing in is just moving some air around. But if you empty your lung, then you have atmospheric pressure pushing in only, and it takes real effort to pull open your lungs again.

How does a suction cup work? Is it really "sucked onto" the surface? Imagine that I place an object on a surface and make it very flat, so that no air at all can be under it. Now the atmospheric pressure pushes down on the top, but there's no air under it to balance this, so it is pressed down very hard. That's what makes a suction cup work. It's not sucked onto the table, it is pressed in place by the atmospheric pressure. If you think about how you remove a suction cup this should be clear. You can’t pull it straight off, that’s much too hard. But if you peel up the edge just a bit, letting some air in underneath, it comes right off.

How does a straw work? Do you really “suck” the fluid up the straw? What happens with a straw is similar. You reduce a little the air pressure inside the straw, and the outside air pressure pushes the fluid up the straw.

It is worthwhile to stress that you can never "pull" something along using a gas. Gases don’t support tension at all. A vacuum cleaner cannot suck things up, and you can't be "sucked" out of a punctured airplane or a spaceship. You can be pushed out by the high pressure air inside your plane, racing to expand into the low pressure area outside the hole. Once the air which was inside the plane is gone, nothing further happens, and so long as you have something to breath, you can ride back to the surface in safety.
Pressure Measurement

Pressure is most often measured in a "differential" way. A simple device for doing this is the “U-tube manometer”. When a liquid is placed in this tube it flows until the liquid level is the same on both sides. If you now connect one side to some chamber containing gas at higher than atmospheric pressure, the fluid is pushed down on that side, and up on the other, until added pressure from the column of fluid equals the difference in pressure between the chamber and the outside.

The difference in pressure between inside and outside the vessel is then given by:

\[ P_{\text{in}} - P_{\text{out}} = \rho g \Delta h \]

This \( \Delta h \), which can be read off the scale, then measures the difference in pressure between the atmosphere and inside. This is why the term \( \rho g \Delta h \) is called the "gauge pressure"; it’s what you see on the gauge.

There are other kinds of pressure sensors, many of which are absolute (actually measuring \( F/A \)). Some of these are based on what are called "piezoelectric" materials. These are materials which generate a measurable voltage when you squeeze them. Measure the voltage and you get the pressure.

Pascal’s Principle and the hydraulic press:

As we said earlier, if you squeeze down on the top of the fluid, that pressure is transmitted equally to every point in the fluid. Any change in external pressure is propagated throughout the fluid.

Uses of this include the hydraulic press. If I apply a small force to a small area, that generates a certain pressure. This pressure, if applied to a large area, can then produce a large force.

\[ F_{\text{down}} / A_{\text{small}} = F_{\text{up}} / A_{\text{large}} \quad \text{or} \quad F_{\text{up}} = (A_{\text{large}} / A_{\text{small}}) F_{\text{down}} \]

This kind of mechanism is another kind of "force magnifier", like the block and tackle, the inclined plane, and lever. The trick here, as in all the other systems which magnify forces, is that to raise the big side a distance \( d_{\text{big}} \), you have to push down on the small side much farther, a distance given by \( d_{\text{small}} = d_{\text{big}} \times (A_{\text{large}}/A_{\text{small}}) \). So the work done is the same using Pascal's principle to lift with a small force as it would be simply lifting with a large
force. You’re just able to use a much smaller force to do the lifting than you otherwise would.

Hydraulic mechanisms like this are used in many machines which need to apply a really large force, but possess engines only capable of producing small ones. Most construction equipment, for example, uses hydraulic mechanisms extensively.

Archimedes principle and buoyancy

There is one very important consequence of the increase of pressure with depth in a fluid, used very extensively by life. We talked in the start of the class about how one of the greatest mechanical challenges for an organism is to support its own weight against the downward force of gravity. The increase of pressure with depth in a fluid can help with this.

If an organism is submerged in a fluid (we all are…) the fluid beneath the organism will always push upward on it with a somewhat larger force than the fluid above it pushes down. This difference generates a net upward force called buoyancy. As we’ll see, these buoyant forces are small in air (where the pressure changes slowly with depth), but large in water. In fact, they are large enough in water to make resisting gravity a nearly non-existent problem for creatures that spend their lives submerged.

To figure out exactly how large this buoyant force will be, imagine a thin cube filled with water, in the water. This could be made, perhaps, by submerging a very thin plastic cube filled with water. This cube would be stationary, in equilibrium, so we know that:

\[ \sum F_y = P_{\text{bottom}} \cdot \text{Area} - P_{\text{top}} \cdot \text{Area} - mg = 0 \]

or

\[ P_{\text{bottom}} \cdot \text{Area} - P_{\text{top}} \cdot \text{Area} = mg \]

In words, the body experiences a net upward force which is just equal to the weight of the water in the cube. This upward force, which is generated by the variation in pressure as you go deeper in a fluid, is called the Buoyant Force.

\[ \text{BF} = P_{\text{bottom}} \cdot A - P_{\text{top}} \cdot A = \text{weight of the water displaced} \]

Now replace the water filled plastic cube with a cube of wood in exactly the same shape and size. Does this change the pressure pushing on it? No, so the buoyant force is in this case exactly the same. It might seem that it does matter what's in this box, as this would affect the change in pressure from the bottom to the top. But this isn't right. The change in pressure in the fluid depends only on the distance below the surface (remember the hydrostatic paradox).

To recap the argument: we use the consideration of a cube of water to figure out how large the buoyant force is, and find that it’s exactly equal to the weight of the water which is inside the object. Then we make the point that the size of this buoyant force is the same
if I replace the cube of water with a cube of anything. The upward buoyant force on any object in a fluid is *always equal to the weight of the fluid which the object displaces.*

\[ \text{BF} = \text{Weight of fluid which was displaced} \]

**The shape of the object does not matter!**

We worked out the buoyant force by considering a cube submerged in the fluid. We did this because it was easy, with the pressure on the top and the pressure on the bottom taking on uniform values across the surfaces. We were also able to ignore the inward pressure on the sides. Using this, we got the simple answer that the buoyant force is equal to the weight of the fluid displaced.

But what’s the buoyant force is the submerged object has a different shape? What if it’s shaped like a fish, or a submarine, or a person? While it’s not obvious, it is nonetheless true that the buoyant force is always just the same as for the cube, it is always equal to the weight of the displaced fluid.

**Floating, sinking, and the buoyant force**

So now consider what happens to some object we submerge in the water. It is pressed upward with a force just equal to the weight of the fluid which it displaced.

\[ \text{BF} = m_{\text{displaced}} g = \rho_{\text{water}} V_{\text{object}} g \]

so, the total force on it is:

\[ \text{Net force} = \text{BF} - \text{Weight} = \rho_{\text{water}} V_{\text{object}} g - m_{\text{object}} g = \rho_{\text{water}} V_{\text{object}} g - \rho_{\text{object}} V_{\text{object}} g \]

\[ \text{Net force} = (\rho_{\text{water}} - \rho_{\text{object}}) V_{\text{object}} g \]

What does this mean? If \( \rho_{\text{water}} < \rho_{\text{object}} \), the net force is down, and the object sinks deeper in the water. If \( \rho_{\text{water}} > \rho_{\text{objects}} \), the net force is positive, and it rises to the top.

What happens when it reaches the top? Part of the object goes above the surface, until

\[ \rho_{\text{water}} V_{\text{in}} g = \rho_{\text{object}} V_{\text{object}} g \]

or \( \rho_{\text{object}} / \rho_{\text{water}} = V_{\text{in}} / V_{\text{object}} \)

In other words it floats with a fraction

\[ \rho_{\text{object}} / \rho_{\text{water}} \]

of its volume below the surface.

What happens if it sinks to the bottom? When it reaches the bottom, it is partly supported by the buoyant force, and partly by the bottom of the vessel. This is what happens with you when you stand in a swimming pool. Instead of the normally large normal force you feel between your feet and the floor on land, you feel a quite tiny little normal force between your feet and the bottom of the pool. The extra buoyant support you receive from the water supports most of your weight, and is part of what makes splashing about in the pool so delightful.
Some consequences of Archimedes principle and this idea

The impact of the buoyant force is completely different for organisms living in air and water. The reason is simple. Living things are made mostly of water, so to first order the weight of an organism is about \( \rho_{\text{water}} V_{\text{organism}} g \). In the air, the buoyant force is \( \rho_{\text{air}} V_{\text{organism}} g \), and since \( \rho_{\text{air}} \sim \rho_{\text{water}}/1000 \), the BF experienced by air dwelling creatures is negligible.

People have learned how to make hot air and Helium balloons which are supported in air by the buoyant force, but no organisms use the buoyant force of air to substantially support themselves. Many organisms live in the air, but their support has more to do with fluid flow and fluid friction (which we will discuss in a bit) than with buoyancy.

In water, by constrast, the buoyant force on an organism is quite naturally almost equal to its weight. The net force on an organism submerged in water is:

\[
F = (\rho_{\text{water}} - \rho_{\text{organism}}) V_{\text{organism}} g
\]

Since the density of most creatures is close to that of water, this net force is usually small, and may be either up or down. You probably know this from your own experience. When you swim, taking a deep breath (and hence decreasing your density) can make you positively buoyant, while expelling it (and increasing your density) can make you sink. Just that little adjustment is enough to change the balance.

The presence of this supportive buoyant force for organisms living submerged is the single largest difference between life on land and in the water. Things that live in water don’t have to support their weight, while those on land do. This fact was one of the several serious challenges evolution had to overcome when creatures first began to move from the sea onto land. It’s also why so many sea creatures seem utterly different from life on land. Enormous numbers of large invertebrates, supported completely by the buoyant force, exist there. Bring them on land and they become piles of goo, but back in the lovely, supportive ocean, they get along quite well. Many are aggressive, highly competitive carnivores, like the deep sea siphonophore shown in the picture.

The presence of the buoyant force in the water also eliminates the size restrictions which are so unavoidable on land. The blue whale, as far as we know the largest animal to ever live on Earth, can be so enormous only because it lives in the sea.

Buoyancy is also responsible for the phrase “the tip of the iceberg”. Unlike most materials, water has the property that ice, its solid form, is less dense than water, its liquid form. This has many important consequences, including some which have been emphasized in blockbuster disaster films. Approximate values for these densities of pure water are:

\[
\rho_{\text{ice}} = 0.914 \times 10^3 \text{ kg/m}^3 \\
\rho_{\text{water}} = 1.000 \times 10^3 \text{ kg/m}^3
\]
so about 0.914/1.000 or 91% of a floating iceberg is below the surface. Is this more or less in a real salt water ocean circumstance?
Liquids and their surfaces: life at the interface

Liquids are things which flow smoothly, like gases, but which hold together at their surfaces, like solids. They do not expand freely. So liquids and solids have something that a gas does not: a surface. In solids this is interesting, but not dynamic. Since atoms can’t move around in a solid, surfaces never just change, they have to be altered by external influences.

Liquids, by contrast, change their shape freely. Large quantities flow freely, squirting away from anything that tries to shear them. On large scales, in rivers and big pipes, that’s all you need to know. This is also about all large living things need to know about liquids.

But look more closely and a variety of really interesting things emerge. Small drops of fluid in free fall alter their own shape, pulling themselves together into drops as spherical as they can be. Drops of water on a waxed car “ball up” and roll off, while those on a warm frying pan spread out in a very thin layer. Water in a glass pulls itself up the edge a bit. These are all surface effects, and they become really important when you consider the behavior of fluids on small scales. As always, organisms living in this environment not only have to know about these properties, but actually take advantage of them in a wild variety of creative ways.

Surface energy

To understand the nature of this surface it is useful to consider what happens when an atom in a liquid moves around. Inside the liquid, the atom is attracted by its neighbors approximately equally in every direction, so it is relatively easy for it to move around. As it travels, it moves away from one atom, but towards another, so that on average it doesn’t require much energy to move from place to place.

At the surface the situation is different. Here the atom only has neighbors behind it, and is only pulled back into the liquid. So when the atom attempts to move outward, it is pulled back into the liquid. This attraction, in a liquid, is large enough to keep atoms with typical thermal motions (kinetic energy ~ k_B T) from escaping. Atoms continually move to the surface, but each time they’re pulled back, and remain in the liquid.

How is a gas different? In a gas the atoms are not connected to one another at all most of the time. An atom reaching the edge of a gas just keeps going and isn’t pulled back by anything. So gases don’t have surfaces and have no cohesion. This is also why you can’t place a gas in tension at all, you can’t stretch it. And it is why a vacuum cleaner can’t suck, it can only provide an empty space into which gas can push something.
What happens at the surface of a liquid is governed by random thermal motion, so to understand what happens we need to think about energy. As an atom moves from the bulk of the liquid to the surface, it moves against a restraining force, the same interatomic forces which hold the liquid together. Moving against this force means that work is being done. This work done by the rest of the liquid on the moving atom can be recovered if you let the atom turn around and go back into the liquid. So it makes sense to treat this “work done by the rest of the liquid” as a potential energy stored in the atom.

You can imagine an atom arriving at the surface with some thermal kinetic energy. It tries to keep going and exit the liquid, but the attraction of its neighbors pull it back inward, slowing it down, and stopping it (taking away its kinetic energy and replacing it with potential) and then reversing its motion, speeding it up, and sending it back into the liquid (reconverting that potential energy into kinetic energy).

Let’s think about what this energy picture implies. If I take a spherical drop and stretch it out into a thin sheet, I make its surface bigger, moving more atoms to the surface. This increases the amount of energy in this form of surface potential energy. The more surface I have, the more energy in the liquid is stored in the form of this “surface potential energy”. If I let this droplet go, that energy locked up in the surface will, as always, try to spread out. If the drop rearranges itself to be more and more spherical, that surface energy can be released, so it will be. Other things being equal, liquids will rearrange their shapes to minimize their surface areas.

Here are some typical values of the surface energies for three liquids: (in contact with air)

<table>
<thead>
<tr>
<th>Liquid</th>
<th>Surface Energy (J/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>0.073</td>
</tr>
<tr>
<td>Ethyl Alcohol</td>
<td>0.022</td>
</tr>
<tr>
<td>Mercury</td>
<td>0.49</td>
</tr>
</tbody>
</table>

For example, it costs 0.073 Joules to create one meter squared of surface area in water. That’s not a whole lot of energy, which is why surface effects are not very important for large organisms. Notice too that these surface energies vary a lot. The value for mercury is about 7 times bigger than that for water. Mercury really sticks to itself.

**Surface tension**

There is an alternate way of thinking about what happens at the surface of a liquid, though it is of course really the same. If I want to stretch the surface of a liquid (if I want to add more atoms to the surface) I must apply a force parallel to the surface. This is because all I really have to do is pull two atoms at the surface apart, and another atom from the bulk will then move up to fill the space. It will be jostled there by ordinary thermal motion. The resistance of a surface to this kind of stretching is called the “surface
tension” of the liquid, and it is measured in units of N/m, because the wider the surface is, the more force is needed to stretch it. The symbol $\gamma$ is usually used to represent this.

**Surface tension is the same thing as surface energy**

Consider a little rectangle of the surface of a liquid. I have to apply a force $F=\gamma L$ in order to stretch the surface.

The energy associated with this whole rectangle is then the work done by this force, which is $W=\int F\,dx = Fw = \gamma wL = \gamma A$ where $A$ is the surface area of the liquid. In fact it is often easier to think about the energy associated with a surface of a certain size. We said a minute ago that the surface tension was expressed in terms of N/m. But N/m = (kgm/s$^2$)/m = (kgm$^2$/s$^2$)/m$^2$ = J/m$^2$. So the same number which represents the surface tension of a liquid is also the surface energy per unit surface area, or the potential energy stored in the surface of a material.

The number which represents the surface energy of water, 0.073 J/m$^2$ is also the surface tension 0.073 N/m.

Is this a little or a lot on a macroscopic scale? The answer depends on scale, for the by now obvious reason that surface area varies like size$^2$ while volume varies like size$^3$. When you have really large amounts of water, like an 8 oz glass of water, this surface energy is usually smaller than other things like gravitational potential energy or bulk kinetic energy. For smaller amounts of water, like raindrops, surface tension (and surface energy) become more important. So as we have often seen, the importance of this feature of liquids is scale dependent. We will see that surface tension plays an important role in a lot of small scale biology.

**Some consequences of surface tension**

Liquids are systems which have internal friction which can dissipate energy. Any extra energy which is in the liquid will, because of the inexorable increase in entropy, be given up and spread out into the surroundings. So if you start out with a liquid that has excess energy in some form, for instance a larger surface than it needs, it will move to make the surface smaller, and as it moves, dissipate the extra energy in the form of heat. This means that any fluid will, pretty rapidly, settle down to an arrangement in which the surface area of the fluid is minimized.
This makes systems which are governed by surface tension some of natures most beautiful and elegant examples of minimization. We will look at a variety of different examples of how surface tension plays a role in the behavior of fluids, and you should keep in mind throughout this theme of optimization. Liquids change in ways which tend to spread out as much as possible their total energy. In part at least, this causes them to minimize their surface area. Surface tension is often what governs the shape of bodies of liquids.

Why do these liquids minimize their surface areas? What “forces” them to do so? Remember what we learned about statistical physics, about the idea of equipartition. Given a system with several ways of storing energy, you will find that random thermal rearrangement leads to the maximum sharing of energy among the different modes. How does this apply to a drop of liquid?

Imagine a drop of liquid is stretched out into a flat pancake. Such a drop would have a lot of energy in the form of this “surface energy”. If you could get this energy out of that form and spread it out into other forms like random thermal motion you would find your drop in a state with the energy more uniformly spread. The way to do this is to pull more of the atoms at the surface back into the liquid. When this happens, the bulk of the liquid will do positive work on the surface atom, reducing this surface potential energy and increasing the average kinetic energy of the atoms. So what happens is the drop pulls itself inward, doing its best to minimize its surface area.

This doesn’t happen without limit however. If it did, every drop you see would be a perfect sphere. But as the liquid pulls inward to reduce the surface area, it must raise the average height of the liquid. If the drop is large, the increase in gravitational energy associated with this rise is larger than decrease in surface energy you get from shrinking it. In this case the equilibrium point will be an elliptical drop, with a surface area larger than a sphere, but with an average height lower than it would have as a sphere. In this picture of droplets on a leaf you can see that small droplets are much more spherical than large ones.

**Three demonstrations of simple surface tension**

Surface tension is able to apply an extra force to objects at the surfaces of liquids, often supporting them when buoyant forces alone could not.

1: One famous example of this is a steel needle supported on the surface of water. The needle is much more dense than water, so it is not primarily buoyancy which supports the needle, but rather surface tension.
To is possible to do a crude calculation of the surface tension force which supports it:

\[ F_{\text{max}} = 2\sigma L = 2 \times 0.073 \text{N/m} \times 0.025 \text{m} = 0.004 \text{N} \]

How large a force is this? The weight of 1 gm of material is 0.001 kg \( \times 10 \text{m/s}^2 = 0.01 \text{N} \).

So this force would only hold up a needle if its mass is less than 0.4 gm. Remember though that there is also a buoyant force involved here. As the needle penetrates down into the fluid, it displaces water, creating an additional buoyant force. How large is this? If the thing displaces a volume almost equal to its own volume, this might be \( \rho_{\text{water}} V_{\text{needle}} g \). This provides an additional supporting force.

Various organisms use surface tension, some in just this way. The most familiar is probably the water strider. Robot versions of these have been made as well:

In addition to needles, it’s quite possible to “float” things on screens (like a window screen). In fact, a screen with sufficiently small holes will work just about as well as a solid object for keeping water out. This is why fabrics, at least non-wetting ones (see below), can be waterproof even though they have holes in them.

2: Air bubbles and water droplets. Surface tension tends to make the surface of a liquid as small as possible. This is true both for a droplet of liquid surrounded by gas and for a bubble of gas surrounded by liquid:
3: The dripping faucet: The shape which a droplet forms is governed by gravity and surface tension. Water is held up in the pipe until the energy gained by lowering the gravitational potential energy of the fluid is enough to make up for the increased surface energy caused by stretching the drop out. As it starts to get long and thin, it becomes unstable, and finds a lower energy solution by pinching off a drop.

**Interactions between liquids, solids, and gases**

To understand many biological phenomena it is necessary to consider the balance between cohesion (the forces which hold the liquid together) and adhesion (the forces which attract atoms in the liquid to the solid with which it comes into contact). One useful parameter to keep track of is the “contact angle” between a liquid and a solid in the presence of a gas.

If you put a drop of liquid on a surface, there will be some angle of contact at the point where the liquid first comes into contact with the solid. This “contact angle” is determined by a balance between the cohesion of the liquid and the adhesion between the liquid and the solid. If the atoms in the liquid are more attracted to other atoms in the liquid than to atoms in the solid, the liquid will bunch together (as in diagram A). If the reverse is true, and the atoms in the liquid are more strongly attracted to the atoms in the solid than to their neighbors in the liquid, the liquid will “wet” the surface, spreading out over it as in diagram B.
If the contact angle $\theta > 90$ the material does not "wet" and the drop is more attracted to itself than to the solid. If the contact angle $\theta < 90$ the liquid "wets" the material, and is more attracted to the solid than to itself. The value of this contact angle depends both on the kind of liquid and the composition and state of the solid surface, and even to a certain extent on the gas which is around. It is worth noting, however, that there is no adhesion between the liquid and the gas.

The interaction between water and solids can vary a lot. Some solids are hydrophilic (water loving), and attract water strongly. These solids will tend to wet easily. Others are hydrophobic (water fearing). On these hydrophobic materials water tends to ball up into droplets and run off. You have probably seen this on well waxed cars. It is evidenced in the phrase “like water off a duck’s back”.

Staying dry and getting wet are two very important activities for living things. So it’s not surprising that life has evolved some truly remarkable hydrophilic and hydrophobic materials. Many of them depend not just on the fundamental chemistry of the interaction between the solid and liquid atoms, but also on geometry. For example, a common tactic for making seriously hydrophobic surfaces is to start with a naturally hydrophobic material. Remember this is a material which bonds less strongly to water than water does to itself. To make a water molecule move from the bulk of the water into contact with this material costs energy. Now to make it really cost a lot of energy, construct a material with a lot of surface. This is generally done by making the surface very rough, or even hairy. This way each little bit of area (viewed head on) actually has much more surface area than you would guess. This very large actual area makes the cost of wetting it become enormous.

Not surprisingly, this approach has been mimicked in human technology recently. So that now there are both hard surfaces and fabrics that shed water incredibly well.

Adhesion is a surface phenomenon, very complex, and under extensive study because of its technological importance. It can be extremely dependent on the detailed state of the surface (clean or contaminated for example). This is one of the main reasons water birds spend so much time preening. Keeping the feathers in good order is essential for shedding water, and shedding water is essential to avoid drowning or freezing to death.
Liquid solid boundaries and capillary flow

The second important aspect of liquid-solid interfaces is capillary flow. When a fluid comes into contact with a vessel which it likes to adhere to, its contact angle will be less than 90°. This means that the adhesion of the liquid to the solid is stretching the fluid out, increasing its surface area. It can energetically afford to do this because the surface energy required to stretch the liquid is less than the energy gained by bonding more of the liquid to the solid.

Imagine what happens when you put such a liquid in contact with a vertical surface. For example, what happens when you place a hollow vertical cylinder down into the liquid? If you do this and the solid interacts well with the liquid, it will actually pull the liquid up the surface, suspending it from its surface tension.

This process is enabled by random thermal motion. When you place the vertical surface in contact with the fluid, the tiny random motions in the fluid enable atoms near the solid to reach up the surface slightly. When they do, they bond there, pulling the liquid behind them up as well. How rapidly this will happen depends on the temperature of the liquid as well as how strong the interaction between the liquid and the solid is.

How far can this adhesion pull a liquid up like this? Since the liquid is, in the end, hanging from its surface tension, the total upward component of the surface tension force is:

$$F_{up} = \gamma \times \text{length of contact} \times \cos \theta$$

Often, in large vessels, like a glass of water, this doesn't matter much. You can see the liquid pulled up at the edge, but it does little to the bulk. The situation is different when the tube you insert is very small. In this case you sometimes have enough force from this “capillary effect” to lift the fluid pretty far up the tube. Let’s work out how far.

The upward force, limited by the surface tension is:

$$F = 2\pi r \gamma \cos \theta$$

Pulling up on a column of liquid with weight

$$W = \pi r^2 \rho gh$$
This will lift the fluid until the forces balance:

$$\pi r^2 \rho gh = 2\pi \gamma \cos \theta$$

or

$$h = \frac{2\gamma \cos \theta}{\rho g r}$$

In the limited case of perfect wetting, where $\theta = 0$, $h = \frac{2\gamma}{\rho g r}$.

How does this behave as I change the size of the tube? If the tube is big the liquid won’t rise far at all. But if the tube is small, it can rise remarkably far. What sets the scale? The contact angle is determined by the fluid properties, the solid, and the conditions (temperature and pressure). The surface tension $\gamma$ and density $\rho$ are properties of the fluid.

Look at an example: a tube which has complete wetting (contact angle = 0) and the fluid is water ($\rho = 1000 \text{ kg/m}^3$ and $\gamma = 0.073 \text{ N/m}$). In this case:

$$h = 1.5 \times 10^{-5} \text{ m}^2 / r$$

What does this mean? If the radius of the tube $r$ is 1 cm, this rises by $1.5 \times 10^{-4}$ m, or 0.15 millimeters (not very much). But if the tube is thin, like $10^{-5}$ m or 10 $\mu$m, then the water might rise up the tube a distance of 1.5 m! It uses thermal motion, combined with the natural attraction between the solid and the liquid, to pump itself up more than a meter.

Why is this interesting? In part because it is the mechanism used by trees to raise water to their leaves. What travels from the roots of a tree to its leaves through a continuous, narrow channel in the “xylem”, hollow, dead, tubes in the trunk which are very effectively wetted by water. These tubes have typical radii of $2 \times 10^{-5}$ m. For such a tube we would calculate a capillary height of

$$h = 2 \times 0.073 / 1000 \times 9.8 \times 2 \times 10^{-5} = 0.8 \text{ m}$$

That’s not very far, at least not compared to the height of a tree. So how do trees manage to grow so tall? In fact this column of fluid is not being pulled up these relatively broad tubes. It is instead being held in place by capillary action in much thinner vessels separating the cells in the leaves. These vessels have typically radii of $5 \times 10^{-9}$ m. How tall a column would this support? We have to be careful here, because the column is broad in the xylem, and only narrows at the top…

Force available = $2\pi r_{pore} \gamma \ast \text{Number of pores}$

Force needed = $\pi r_{xylem}^2 \rho gh$

Setting these equal:

$$2\pi r_{pore} \gamma \ast \text{Number of pores} = \pi r_{xylem}^2 \rho gh$$

Or
What’s the number of pores? We might guess that the total area of pores equals the total area of the xylem tube, just as a first approximation. This would imply:

\[ N_{\text{pores}} \pi r_{\text{pore}}^2 = \pi r_{\text{xylem}}^2 \quad \text{or} \quad N_{\text{pores}} = \frac{r_{\text{xylem}}^2}{r_{\text{pore}}^2} \]

Putting this into the above we would find:

\[ h = 2r_{\text{pore}} \gamma \left( \frac{r_{\text{xylem}}^2}{r_{\text{pore}}^2} \right) / r_{\text{xylem}}^2 \rho g = 2\gamma / r_{\text{pore}} \rho g = 3000 \text{ m} \]

Wow. This suggests a tree could be 3 km tall, at least if our assumption about the number of pores is correct. Not surprisingly, other constraints intervene before this point. But this should make it clear that, despite the fact that trees have no pumps to do the job for them, the problem of moving water to the tops of trees is not a major limitation on tree size. Trees don't have to raise water. The water, taking advantage again of random thermal motion, does the work itself!

**Soap films**

Soap films and bubbles are the most familiar example of extreme surface tension effects, and especially of the beautiful minimization phenomena which it causes.

What is a soap film? Soap films are mostly made of water, and contain only a bit of soap. What makes it a soap, and why do you have to have this to make bubbles? Soaps are made of sodium salts of fatty acids. To the physicist this just means that they typically have big hydrophilic heads, and long hydrophobic tails. So what do they do when you put them in water? They try to find ways of arranging themselves so that their heads point into water, and their tails point out. This creates layers of material on the surface. This "difference" between surface and interior is what gives soap films their stability. What about the surface tension of soap films? Is this higher or lower than water? Can you blow a bubble with pure water?

Because the presence of the soap so dramatically lowers the surface tension, it’s possible to create a much larger surface and have it remain stable. This is what allows you to stretch out a small droplet of water into a large, very thin sheet.

Materials which have this kind of effect, which act at the surface of a liquid, are called “surfactants”. Because they naturally localize at the surface, adding just a small quantity can dramatically change the surface properties of a liquid. Living systems use kind of material constantly. For example, surfactants like this are used to reduce the surface tension of the fluid in your lungs. Without this surfactant, the surface tension of the pulmonary fluid would be large enough to collapse the alveoli, making respiration impossible.
Soap films and minimization

Because liquids with surface energy will flow until as much of that surface energy is released as possible, they will work to minimize their surface area. This minimization can be put to work. At the end of the 19th century a number of knotty mathematical problems about minimal surfaces were solved in part by reference to soap bubbles. These bubbles, because they sought their lowest surface energy states, often arranged themselves in truly minimal surfaces. Wouldn’t the minimal surface area always be a sphere? Yes, but by attaching the fluid to solid frames like this, or by putting it into a constrained geometry, you can trap the liquid in a local energy minimum, like a bubble. With a little energy input from the outside (say from a pinprick), you can help it to jump to a still lower energy state, like the little sphere.

Soap bubbles

What's going on in a bubble? Is the pressure higher inside or out? Consider a hemisphere of the bubble. The net force on it due to the inside/outside pressure difference is just the pressure difference between inside and outside (\(\Delta P\)) times the cross-sectional area of the hemisphere \(\pi r^2\). This is not completely obvious, because the internal pressure pushes straight out at all points, so it has both up and down components as well as leftward ones. But if you do the integral to add up all these components the up-and-down forces cancel, and only a net force to the left (in this picture) remains. Balancing this pressure to the left is a surface tension force to the right. This surface tension force it the usual \(F = \gamma L\), but we have to remember that there are two surfaces, both an inner and an outer surface. These two forces balance.

\[
\pi r^2 \Delta P = 2\pi r(2\gamma)
\]

or

\[
\Delta P = \frac{4\gamma}{r}
\]

The pressure inside a bubble is larger than the pressure outside. This is what pushes the soap film outward. Notice the dependence on both the surface tension and the radius. Larger surface tension squeezes the bubble inward, so the pressure difference has to be larger to maintain the same radius. Smaller radius makes it curve more, so again the pressure difference has to be more. Imagine a really big bubble, with large \(r\). The surface of such a bubble would be almost flat, and only a small pressure difference would be required to make it bend this tiny bit.
A big difference between surface tension and elastic tension

At first, it looks as if surface tension in a liquid is really very like the elastic tension which governs the stretching of a rubber sheet. But this is definitely not the case. Rubber sheets are elastic, which means that the force required to stretch them depends linearly on the amount by which they are stretched. This is just Hooke’s law: \( F = -kx \). So the farther you stretch them, the harder you have to pull.

This is not the case with surface tension. As there are essentially always more atoms to bring to the surface, stretching the surface another meter always requires the same force as was required for the first meter. The force required to stretch the surface never becomes larger.

Relation to solids and practical importance

There are many surface effects in solids which are analogous to the surface effects in liquids. The real difference of course, is that new atoms don’t move to the surface on their own: simply due to thermal motion. But very similar statements can be made about why solids don’t fall apart. The bonded material is energetically favored over more surface area, so to break it you have to apply energy. In order to split a diamond, you have to provide enough energy to account for the new surface energy of the two pieces you create.

Surface physics, the study of the details of what goes on right at the surface (primarily of solids) is a major field of study today. Why would the study of surfaces become more and more important as time goes on? In this age of minimization everything that we work with becomes more “surface-dominated” all the time. So like soap bubbles, the tiny structures used in the modern electronic industry are more and more dominated in their behavior by surface effects. The ratio of surface area to volume changes like \( 1/r \), so as the size of things gets small, the relative importance of surface phenomena increases.

When does a material become “mostly surface”? These are things called mesoscopic, they are between macroscopic and microscopic, and they have interesting properties. If an object is made up of little sphere (like atoms) with a radius \( r \), then it contains \( 4/3\pi R^3 / 4/3\pi r^3 \) little balls. (ignoring the loss due to packing). So there are about \( (R/r)^3 \) little balls in the big sphere. Meanwhile the number of atoms at the surface is something like \( 4\pi R^2 / \pi r^2 = 4(R^2/r^2) \). So the ratio of atoms at the surface to all atoms changes like:

\[
\frac{4(R^2/r^2)}{(R^3/r^3)} = 4(r/R)
\]

Once the radius of the atoms becomes a modest fraction of the total radius \( R \), the fraction of atoms which are at the surface becomes large, and surface effects become important.

Much effort now is going into “nanotechnology”, the manufacture of very tiny, subcellular technologies. Nanotech devices are very much in this surface dominated regime. That’s part of the reason why they offer new opportunities for engineering.
Physics for the Life Sciences: Fall 2008 Lecture #24

The key property of fluids, what makes them so interesting, is that they flow. The atoms which make up the fluid are able to move relative to one another, rearranging the structure of the fluid under even the most subtle influence. This ability to change, to mix and reshape, is why fluids are so important for life. In a solid, essentially nothing happens. All of its atoms remain fixed, never moving around, never reshaping themselves. Life is continuous change, cycling along on the flow of energy. This could never happen without taking advantage of the freedom of change which fluids provide.

Life lives in fluids, either air or water, and it is the usual thing for the organism and the fluid to move relative to one another. This may happen because of locomotion (a bird flying or fish swimming), because the fluid flows by (a tree in a storm), or as the fluid flows through the organism (a circulatory or respiratory system). Understanding how fluids flow is essential for understanding life.

We will approach this subject in steps. First, we’ll consider the flow of “ideal” fluids, because in this case some of the phenomena will be simpler. It turns out real fluids aren’t very ideal, but considering such things will nevertheless give us some useful ideas to apply to more realistic fluids. In the final two lectures we will consider some of the complications which make real fluids so much more interesting.

Fluid Flow, and the necessity for idealization

What is an ideal fluid? One which flows in such a way that the properties we described previously are exactly met. So, it should always flow to escape pressure (rather than be compressed for example). It should have no trouble flowing to escape pressure, which means both that it should flow without friction and it should not swirl around when it flows. These two constraints are much the same. Here are some technical ways of stating the what’s required for ideal fluid flow:

- Fluid flow is laminar and steady (no accelerations). This condition really implies that the shear stress on the fluid should be constant.
- The fluid is incompressible: its density doesn’t change
- The fluid is non-viscous: there is no friction in the fluid. This means no energy is lost in the continuous flow of the fluid.
- The fluid flow should be “irrotational”. It should have no “vorticity” and never swirl around, mixing one layer with another.

Now we often make assumptions like this in physics; we ignore air friction, talk about frictionless surfaces, pretend that objects are rigid and don’t deform etc. Each time we do this we make errors, and we need to be careful to understand what they might be.

The reason I raise this now is to stress that the assumption of ideal fluids, and ideal flow, is VERY rarely met. It is a much worse assumption than many we have made.
Nevertheless, assuming it will allow us to understand two simple principles which, while they may not be obeyed quantitatively, are very good qualitative descriptions of what happens in fluids. So beware, real fluid flow is very complicated. We’ll return to this and make our fluids a little more realistic before we’re done.

**Simplest property of fluid flow: continuity**

The first principle of fluid flow we will use is called the continuity equation. It concerns the flow of the fluid through a confined channel, and makes a simple assertion; what goes in must come out. The volume flow rate is a measure of the volume of fluid which flows past a point per unit time. This is related in a simple way to the cross-sectional area of the flow (A) and the speed of the flow (v):

\[
d\text{Volume}/dt = A(dx/dt) = Av
\]

If I make the channel wider while keeping the flow velocity constant, the flow rate increases. If I increase the velocity while keeping the area constant the flow rate increases.

The continuity equation just says that this flow rate must be the same at different locations in a confined flow;

\[A_1v_2 = A_2v_2\]

This is just saying that total amount of fluid passing each point along the flow is the same. If this weren’t true the fluid would have to pile up somewhere along the way.

Being a little more careful, the “area” used here should be perpendicular to the streamlines of fluid flow. This principle depends on the incompressibility of the fluid. If it is compressible, then it can change its volume, and so the volume rate of flow need not be conserved. For liquids at least, incompressibility is not such a bad approximation. For gases it is a very bad one indeed.

**A simple application of the continuity equation**

Here is a simple, and interesting, application of the continuity equation. Your circulatory system is a remarkable network, using fluid flow to deliver oxygen and nutrients to all the parts of your body. To do this, it must send all the blood through one channel in your heart, then split the flow out into many branches, sending blood off into ever narrowing channels, until finally at least one tiny capillary passes by every living cell in your body. It would be interesting to know just how many capillaries you have. But this is really hard to determine by counting.

Let’s use the continuity equation to get some idea of the number of capillaries you have. The aorta, the major blood vessel coming out of the heart, has a radius of about 1 cm, and blood flows through it at an average rate of about 30 cm/s. It’s not especially hard to observe the blood flow in a single capillary, and measure its rate. If blood flows through the capillaries at about 0.03 cm/s, what is the total area of capillaries in your body?
We can answer this using the continuity equation.

\[ A_{\text{aorta}} v_{\text{aorta}} = A_{\text{capillaries}} v_{\text{capillaries}} \quad \text{or} \quad A_{\text{capillaries}} = A_{\text{aorta}} \left( \frac{v_{\text{aorta}}}{v_{\text{capillaries}}} \right) \]

The area of the aorta is \( \pi r^2 = 3.1 \times 10^{-4} \text{ m}^2 \) and the ratio \( \frac{v_{\text{aorta}}}{v_{\text{capillaries}}} = 10^4 \), so the total area of the capillaries is \( A_{\text{capillaries}} = 3.1 \times 10^{-4} \text{ m}^2 \times 10^4 = 3.1 \text{ m}^2 \). Now that’s a very big cross-sectional area. A circular pipe with that large an area would have to have a radius of one meter!

Since capillaries have typical radii of \( 5 \times 10^{-6} \text{ m} \), the total number of capillaries separating your arteries and veins is about:

\[ N_{\text{capillaries}} = \frac{A_{\text{capillaries}} \text{ total}}{A_{\text{one capillary}}} = \frac{3.1 \text{ m}^2}{8 \times 10^{-11} \text{ m}^2} = 4 \times 10^{10} \]

The flow which starts in one large aorta splits out into tens of billions of capillaries. These capillaries service something like \( 5 \times 10^{14} \) (tens of trillions) cells in your body. So the very large number of capillaries is not so surprising.

**A somewhat more complex fluid flow relation: energy conservation**

In an ideal fluid, there is no loss of energy to friction, and the bulk mechanical energy in a flow of fluid is conserved. That is, there’s no conversion from bulk kinetic energy into random thermal energy. When a fluid flows along, it contains some amount of energy per unit mass. Just as when we throw a ball up it exchanges kinetic energy for potential, so too a fluid trades energy between different forms as it flows.

When water flows in a pipe, each bit of water has some energy. What are the forms this energy can take?

- Kinetic energy (bulk fluid flow)
- Gravitational potential energy (height of fluid)
- Elastic energy (how much is it squeezed, pressure)
- There would be thermal energy too, only we specified no friction, which means that energy does not flow from bulk mechanical forms to random thermal forms for such an ideal fluid…

Expressing each for as the energy **per unit volume** of fluid we find:

- \( \text{KE/Vol} = 1/2 \text{mv}^2 / \text{Vol} = 1/2 \rho v^2 \)
- \( \text{PE/Vol} = \text{mgh} / \text{Vol} = \rho gh \) (measured from some zero potential reference point)
- Elastic energy per unit volume in pressure is just \( \text{P} \). Why is this? If the fluid is at pressure \( \text{P} \), and undergoes an expansion \( dV \), it will do some work \( dW = \text{P}dV \). Since this work would represent energy taken from the fluid, we could say: (energy in the fluid) / (unit volume) = \( dW/dV = \text{P} \).

We expect to see the total amount of energy in these three forms conserved as the fluid flows from one place to another. Every little bit of fluid contains the same total amount of energy, though it may be kinetic energy in on part of the flow and potential energy in...
another. Comparing the mix of these three forms at two different locations in the flow we could write:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho gh_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho gh_2$$

For a simple proof: consider a tube of fluid flowing as shown:

Consider a bit of fluid in the lower part of the pipe (in the region labeled 1). This fluid has a total mass $m = \rho A v_1 \Delta t$. After it moves to the upper part of the pipe, it will have some new velocity $v_2$.

We know that $W = \Delta KE$, so

$$W_{\text{gravity}} + W_{\text{pressure}} = \frac{1}{2}m(v_2^2 - v_1^2)$$

Here

$$W_{\text{gravity}} = -mg(h_2 - h_1)$$

And

$$W_{\text{pressure}} = -(P_2 A_2 d_2 - P_1 A_1 d_1) = -(P_2 - P_1)V$$

So that:

$$-(P_2 - P_1)V - mg(h_2 - h_1) = \frac{1}{2}m(v_2^2 - v_1^2)$$

which can be rearranged as:

$$P_1 + \rho gh_1 + \frac{1}{2}\rho v_1^2 = P_2 + \rho gh_2 + \frac{1}{2}\rho v_2^2$$

This rule is called Bernoulli’s equation. It is just a statement of conservation of energy in non-viscous fluid-flow.

What the Bernoulli principle says is that for ideal fluids, energy is conserved. Any little piece of the fluid, as it moves along, will convert its energy from one form to another, but will neither gain nor lose energy. So, I can assume that the total energy per unit volume will be the same throughout the flow. Measure it at one place, and it will be the same someplace else. But it is most important that you just remember what this means, because
it will REALLY help you to apply this when the time comes. It means, find the energy/volume somewhere in the fluid, and it will be the same somewhere else.

**Examples of applying the Bernoulli equation**

Some simple examples:

1: Water pressure in your house. How does the water pressure change in your house? Think about how fast fluid comes out of the pipe on the first and second floors. In both cases the fluid coming out of the pipe is at atmospheric pressure. It must be because there's nothing else to prevent it from flowing. So:

\[
P_{\text{atm}} + \rho gh_1 + 1/2 \rho v_1^2 = P_{\text{atm}} + \rho gh_2 + 1/2 \rho v_2^2
\]

Or

\[
v_2^2 = v_1^2 - 2g(h_2 - h_1)
\]

The bigger the change in height, the lower the velocity with which it emerges. Too bad if you live on the fifth floor of a dorm. The water may well dribble out of your shower, even though it comes screaming out down on the first floor.

2: Flow in a pipe:

Here we again have:

\[
P_1 + \rho gh_1 + 1/2 \rho v_1^2 = P_2 + \rho gh_2 + 1/2 \rho v_2^2
\]

But here \( h_1 = h_2 \), so:

\[
P_1 - P_2 = 1/2 \rho (v_2^2 - v_1^2) \quad [ = 1/2 \rho ((A_1/A_2) - 1)v_1^2 ]
\]

How can we figure out the velocities? From the equation of continuity:

\[
A_1v_1 = A_2v_2 \quad \text{or} \quad v_2 = (A_1/A_2)v_1
\]

Now since \( A_1 > A_2 \), we know that \( v_2 > v_1 \). And from the equation we just wrote, we see that \( P_1 > P_2 \). When the fluid flows faster, the pressure is reduced.

Can that be true? Think of the flow in a pipe. If it is slow and I punch a hole, the fluid will leak out. As it gets faster, the fluid will leak out less, and if it is fast enough, it will leak out not at all, but instead will simply go zooming along in the pipe.

3: A final example: water leaking from a tank. If a tank, open to the air on top, has a leak 2m below the surface, how fast does the water come out?

At the top of the fluid: \( P_{\text{top}} = P_{\text{atm}} \), \( v_{\text{top}} = 0 \), and \( h_{\text{top}} = 2m \). At the hole \( h_{\text{hole}} = 0 \), \( v_{\text{hole}} \) is unknown, and \( P_{\text{hole}} = P_{\text{atm}} \). Why is this? When a bit of fluid is right at the hole, the only pressure holding it in is atmospheric pressure. So we have:

\[
P_{\text{atm}} + \rho gh_{\text{top}} + 1/2 \rho v_{\text{top}}^2 = P_{\text{atm}} + \rho g(0) + 1/2 \rho v_{\text{hole}}^2
\]

Or
\[ v_{\text{hole}} = \sqrt{2gh_{\text{top}}} = \sqrt{2 \times 9.8 \text{m/s}^2 \times 2 \text{m}} = 6.26 \text{m/s} \]

How good is the Bernoulli equation? What would happen if, during the flow, some of the energy was lost due to friction? This would change the Bernoulli equation into an inequality. Which way would this inequality point? Would you ever expect energy to flow from random thermal energy into the organized motion of a flow?

**Consequences of Bernoulli’s principle**

The basic notions of Bernoulli’s principle are seen in a wide range of circumstances relevant for life. The pressure differences generated by moving fluids are used by many organisms to induce flow. In many cases, organisms use the fact that fluid velocities increase with distance from a solid surface. For example, wind speed is low very close to the ground and much higher in the air. You may have experienced this at the beach. While lying on your towel, you get quite hot in the sun. But when you stand up you discover a nice cool breeze is blowing.

Sponges, an incredibly common and ancient though simple form of sea life, simplify their feeding by extending their tops up from the sea floor, into the more rapidly flowing fluid. This makes the pressure at the top of the sponge lower than at the base, driving a flow of water in through the sides of the sponge and up through its center. As the water passes through the sponge, tiny cilia strain it for food.

On land, prairie dogs in the American West use a similar strategy. Their very large networks of underground burrows need ventilation. The CO₂ which they exhale is denser than the O₂ they need to breathe in. If they didn’t replace the O₂, they would suffocate. The burrows are much too large for diffusion to replace the oxygen, so they need to use forced convection. To accomplish this they build a mound at one end of the system which sticks up from the ground. The mounds at the other end are lower and more rounded. This allows the pressure difference between the two ends to drive a flow of fresh air from the low mound through the high. Termite mounds, built up from the ground, perform a similar function.
Real fluids:

As we have mentioned several times, real fluids are more complex than the ideal fluids described by the continuity equation and Bernoulli’s equation. Fluid dynamics is one of the most complex subjects in physics, and is now addressed using the world’s largest supercomputers. John von Neumann, one of the great early physicist/computer scientists, once commented that assuming no friction in fluids is like “working with dry water”. Many of the important and interesting phenomena associated with fluids emerge from the “extra” properties of real fluids. As you might expect, these interesting properties of real fluids are regularly used by living things to accomplish their goals, so it is especially important that you should understand some aspects of them.

Today we will introduce real fluids, and especially get an idea of how we describe them. Along the way we’ll see how even a basic consideration of friction in fluids explains a lot of new phenomena. This is a rich and very current field, important because most of the matter in the universe is in fluid form. Even the Earth itself, the canonical solid object, is mostly a fluid, covered by a thin solid crust. In addition, all life we know moves in fluids, air or water, so the behavior of these real fluids places strong constraints on how living things evolved.

Returning to the definition of fluids and shear stress

Remember how shear stress works in solids:

\[(F/A) = S(\Delta x/L)\]

Where \(F/A\) is the shear stress (the force per unit area creating the shear) and \(\Delta x/L\) is the shear strain, how much the material distorts under the stress. See the picture below:

\[\Delta x\]

This is what happens when you apply a shear stress in a solid. What happens in a fluid? If I apply a shear stress, instead of just distorting and coming to a halt like the solid does, it begins to flow. There is no limit to the shear strain which can be produced in a fluid. Any shear stress can produce an infinite shear strain.

So, what’s the appropriate thing to associate with the varying properties of fluids? If I apply a shear stress to some water, and apply the same stress to tar, they both will flow, and both will flow as far as I like if I wait long enough. It’s clear that something is different about these two fluids. How should we quantify this difference?
Viscosity and time rate of change of fluid shear.

Fluids with low internal friction shear rapidly, those with high internal friction shear slowly. This can be expressed in the equation:

$$(F/A) = \eta (v/L)$$

Here the F/A is again the shear stress. Now instead of a strain ($\Delta x/L$), we have a time rate of change of a strain ($v/L$), and instead of the shear modulus $S$ we have the “viscosity” $\eta$.

Viscosity is a measure of the internal friction of fluids, how much they resist flow. The units of viscosity are:

$$\eta = \text{N/m}^2/[(\text{m/s})/\text{m}] = \text{Nsm}^{-2} = (\text{N/m}^2)\text{s} = \text{Pascal*second}$$

Unfortunately, this SI unit is not always used, and instead viscosity is usually tabulated in terms of “poise”, a unit named for the French scientist Poiseuille, and defined as:

$$1 \text{ poise} = 0.1 \text{ Nsm}^{-2}$$

In fact, viscosities are often given in tables in units of “centipoise”, or 100ths of a Poise. So you’ll have to be careful if you take viscosity values from the literature. Here are some example values for viscosity measured in the SI unit Pascal*second.

<table>
<thead>
<tr>
<th>Temp °C</th>
<th>Water</th>
<th>Air</th>
<th>Mercury Motor Oil</th>
<th>SAE 10 Motor Oil</th>
<th>SAE 30 Motor Oil</th>
<th>Glycerin</th>
<th>Honey</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.8x10^{-3}</td>
<td>1.7x10^{-5}</td>
<td>1.7x10^{-3}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.0x10^{-3}</td>
<td>1.8x10^{-5}</td>
<td>1.6x10^{-3}</td>
<td>7x10^{-2}</td>
<td>0.3</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>40</td>
<td>0.7x10^{-2}</td>
<td>1.9x10^{-5}</td>
<td>1.4x10^{-3}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.5x10^{-3}</td>
<td>2.0x10^{-5}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.4x10^{-3}</td>
<td>2.1x10^{-5}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.3x10^{-3}</td>
<td>2.2x10^{-5}</td>
<td>1.2x10^{-3}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note a few facts which are apparent from this table. Liquid viscosities decrease with temperature, while gas viscosities increase. This is because the increasing thermal energy available as temperature rises in liquids makes it easier for atoms to flow past one another. In a gas, like air, the increasing temperature implies more frequent collisions among the atoms, and hence greater viscosity. It’s also worth noting that fluid viscosities, even in this short table, can vary a lot. Honey is about 100,000 times as viscous as air.

Consequences of viscosity

How do fluids with viscosity flow differently from ideal fluids, which would have a viscosity of zero? Here are a few of the key consequences.

1: Loss of pressure in flow: Energy is lost as a fluid flows. This effectively adds another term to Bernoulli’s equation:
\[ P_1 + \rho gh_1 + \frac{1}{2}\rho v_1^2 = P_2 + \rho gh_2 + \frac{1}{2}\rho v_2^2 + \text{Losses of energy due to viscosity} \]

So even in constant velocity flow in a horizontal pipe (where \( h \) and \( v \) remain constant), pressure declines along a flow. This should not be too surprising. In the presence of friction, it costs something to keep a fluid flowing.

2: Flow layering and laminar flow: The velocity of flow at the wall of a pipe is zero. It rises to the center of the flow, and then falls back to zero at the opposite wall. Imagine water flowing down through a pipe or a river. The shear stress coming from gravity is the same everywhere, so you have:

\[ (F/A) = \eta (v/L) \]

where \( L \) is the distance from the wall. This means \( v = L(F/A\eta) \); the speed of the flow is proportional to the distance from the wall.

You can see this in many flows, such as when a river flows rapidly in the center and slowly at the edges. This layered, “laminar” flow is a widespread phenomenon for smooth flows, and plays an important role in many living systems.

3: Terminal velocity in falling bodies. The rate at which the fluid can move out of the way of the falling body is dependent on the viscosity of the fluid. Under a constant strain (the weight of the falling body) the fluid reaches a constant velocity, the value of which is governed by the viscosity \( \eta \).

This is actually a bit more complicated. Even motion through a fluid without friction will exhibit terminal velocity. Just accelerating the fluid to get it out of the way requires that a falling object exert a force on the fluid, implying a reciprocal force on the falling object. For ‘large’ objects falling at ‘large’ velocities, like you or I falling through air, this inertial effect is the most important. For small objects falling at low velocities, the frictional effects associated with viscosity are more important.

You can see that in the two fluid friction laws we discussed so much earlier in the class:

- \( F_f^\text{fluid} = 12\pi \eta Dv \) (small/slow)
- \( F_f^\text{fluid} = 1/2C\rho A v^2 \) (large/fast)

We will (finally) see how to decide what constitutes ‘large’ objects and velocities below.

Copyright: Regents of the University of Michigan, 2009
Limitations of this picture

This is a kind of first approximation, much better than assuming no friction. But it is limited. The assumption we have made that:

\[ (F/A) = \eta(v/L) \]

assumes that \( \eta \) is independent of velocity. Fluids which obey this are called Newtonian fluids (because he first stated this relationship).

There are many non-Newtonian fluids, including most biological fluids, blood for example. As the velocity of blood increases, cells align themselves with the flow, lowering the friction. Such a material is called “shear-thinning”, because it becomes less viscous once it starts to flow. This is common in suspensions and otherwise non-uniform liquids.

Many other common examples of non-Newtonian fluids exist, like egg white (albumen), a “visco-elastic” fluid, which is elastic for small shears and can spring back to its original shape, but will flow like any liquid for larger shears. Ketchup is another well known shear-thinning fluid. Once ketchup starts to flow, it does so very freely, blopping out all over your French fries. Then there is the famous corn starch and water concoction you may have played with in elementary school. This is a “shear-thickening” fluid. If you apply a small shear it will flow nicely, but if you apply a large shear it becomes very viscous indeed.

What we have said about simple works best with uniform, one component fluids. What this means for life is that if it wants to get around the limitations of simple fluids, it can often work up a kind of a cocktail mix of things which will behave the way it needs. A recently published example is the visco-elastic properties of the digestive juices of a carnivorous pitcher plant. This gooeyness greatly increases the chance that a fly landing in the fluid will be unable to escape.

Kinematic viscosity

This discussion of fluid flow is still really static; what stable state of flow does the fluid reach after we give it time to settle down. The dynamical response of a fluid to a stress depends on both the viscosity and the density of the fluid. This just reflects the difference in inertia between two fluids with the same viscosity. So it is often useful to talk about the kinematic viscosity of a fluid:

\[ \text{Kinematic viscosity} = \frac{\eta}{\rho} \]

A whirlpool of air stops spinning much faster than a whirlpool of water, because although the viscosity of air is 50 times less than that of water, the density is 1000 times less. So

---

although there is somewhat less viscous force available to stop the air, much less is needed, and it stops much more quickly. The difference in kinematic viscosity of these two expresses this different response to change much better than a simple comparison of their viscosities.
Physics for the Life Sciences: Fall 2008 Lecture #26

Real fluid flow and turbulence

We have been discussing fluid flow which is smooth and laminar, neatly layered and without mixing. But that’s not what we often see in fluid flow. Instead we see a swirling mixing of fluids, a much more complicated flow in which stability is not evident. What causes this turbulence, and what marks the transition between smooth flow and turbulence?

If the motion of a fluid is dominated by internal friction, by viscous drag, the flow will always be smooth. Any deviations from smoothness will be “damped out” by the friction before they have a chance to become large. This is what you usually see when you pour highly viscous liquids like honey or syrup. Just try mixing a jar of honey with a spoon and you’ll see what I mean.

On the other hand, if the friction is small and something disturbs the flow even slightly, the elements of the fluid which are disturbed will be able to travel far from where they would be in smooth flow before the viscous effects stop them. This in turn allows them to affect other parts of the fluid before they are stopped, leading to a kind of cascade of confusion. This is what you usually see in the flow of air, perhaps most clearly when you look at the smoke from a recently extinguished candle. Swirling around, it doesn’t take long for the smoke to be well mixed with the air around it.

Reynold’s number as the thing which characterizes fluid flow

Somehow the important thing is the balance between the inertia of a bit of the fluid, and the size of the viscous forces acting on it. What follows is a generic discussion of how these two things might vary with the parameters of the material.

First consider the inertia of the object: what size force does it take to stop it?

\[
Ma = (\rho L^3) a \approx (\rho L^3) \left( \frac{v}{\Delta t} \right) \approx (\rho L^3) \frac{v}{L/v} = \rho L^2 v^2
\]

Where I have assumed \( \Delta t = L/v \) is a reasonable approximation for how quickly I might make it stop. That is, I want it to stop before it would travel its own length.

Viscous forces: what size force is available to stop it?

\[
F/A = \eta (v/L) \text{ implies } F = \eta (v/L)L^2 = \eta v L
\]

The ratio of these two forces is a dimensionless number:

\[
R = \frac{\text{(inertial properties)}}{\text{(viscous forces)}} = \frac{\rho L^2 v^2}{\eta v L} = \rho L v / \eta
\]

This is the famous Reynolds’ number \( R \), names for British fluid mechanic Osborne Reynolds. Note that it is dimensionless, a pure number independent of what units we measure it in:
What does this dimensionless nature mean? It suggests that R is an important number, because it allows us to compare problems on all kinds of scales, independent of the units we use, so that they can be uniformly understood.

**Meaning of Reynolds’ number**

R is dimensionless, so it only has a value, not units. Remember that it’s the ratio of inertial properties to viscous forces. If R is small, then the viscous forces are dominant and the flow is smooth and steady. If R is large, then viscous forces are negligible and the flow rapidly becomes complex and turbulent.

What is “large” and “small”? Can we guess from the way we formulated it?

<table>
<thead>
<tr>
<th>R</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R &lt; 1/1000</td>
<td>Viscous forces completely dominate</td>
</tr>
<tr>
<td>1/10 &lt; R &lt; 10</td>
<td>Mixed behavior</td>
</tr>
<tr>
<td>1000 &lt; R &lt; 3000</td>
<td>Drag negligible, but still influences large scale flow patterns</td>
</tr>
<tr>
<td>3000 &lt; R</td>
<td>Flow is “fully” turbulent</td>
</tr>
</tbody>
</table>

We brushed over a problem though, what is this length scale L we have used in determining R? L is the thing which sets the scale for the problem. Could be the size of an object moving through a fluid, diameter of a pipe through which it flows, etc. It’s the scale on which we want to examine the fluid.

We can ask about different length scales within the same problem even. Imagine the flow of water in a pipe of radius r\text{pipe}. If we wish to know about the flow in the pipe overall, we might use this length scale. But imagine that we want to know what the flow is like around little bumps on the wall. If these are smaller, with size r\text{bump} << r\text{pipe}, the flow around them may have a different nature. For example, the Reynolds’ number for the pipe as a whole might be large (both r\text{pipe} and v are large for the whole pipe) while R for the bumps may be small (both r\text{bump} and v\text{wall} are small). This is very typical of flows; turbulence on large scales accompanied by smooth laminar flow on small scales.

**Implications of scale dependence**

The scale dependence (the dependence on L) in the Reynold’s number shows that what we will see depends on what we look at. Let’s consider some large and small things.

Consider a falling dust particle: \( \rho = 1.3 \text{kg/m}^3, v = 0.001 \text{m/s}, L = 0.0001 \text{m}, \eta = 1.8 \times 10^{-5} \text{Nsm}^{-2} \)  
So R = 1.3(0.001)(0.0001)/1.8x10^{-5} = 0.007. This means a falling dust grain is in the viscous regime and will always experience the small-slow kind of friction.

Redoing this for a falling raindrop:  
\( \rho = 1.3 \text{kg/m}^3, v = 0.1 \text{m/s}, L = 0.005 \text{m}, \eta = 1.8 \times 10^{-5} \text{Nsm}^{-2} \)  
we get
Here the motion is dominated by the inertia of the drop, but affected still by the friction which is involved in moving the air out of the way.

Redoing this for a falling person
\[ \rho = 1.3 \text{kg/m}^3, \, v = 30 \text{m/s}, \, L = 2 \text{m}, \, \eta = 1.8 \times 10^{-5} \text{Nsm}^{-2} \]
we get
\[ R = 4.3 \times 10^6 \]
Here the motion is completely dominated by the inertia of the fluid. The fact that there is viscosity, that there is fluid friction, is really not important.

R describes the quality of the flow around an object. So if you want to model a flow, by placing a model in a wind tunnel for example, you have to do it with the same R as you would have in the real system. This means that if you use a 1/10 scale model (L becomes L/10) you have to somehow compensate for this in your test, perhaps by going 10x faster. This presents real problems for the testing of aircraft, as going 10x faster would usually require traveling faster than sound. Other problems occur when you try to blast air over a model at speed greater than the speed of sound.

**Life at large and small Reynolds’ number**

People live in a world where R is large; \( v \) is big, \( \eta \) is small, \( L \) is large, etc. So we usually are not dominated by viscosity. As a result, we also live in a world where turbulent mixing is easy to achieve. If we want to mix cream into our coffee uniformly we just do it. With no special effort we can make the uniformly smooth liquid we so enjoy.

Bacteria, on the other hand, live in a world totally dominated by viscosity. Their speed \( v \) is small, and size \( L \) is small, so they’re always in the viscous dominated, low Reynolds’ number regime. Every time such a creature stops pushing for forward motion, it immediately stops moving. This is a world in which \( F \propto v \), rather than \( F \propto a \). It’s an Aristotelian world in which you have to have a forward force constantly to have motion. As soon as it disappears you stop.

In addition, the fluid in the world of low Reynolds’ number does not mix. Since it flows only laminarly, it never gets the kind of turbulent mixing that allows cream to mix into your coffee. This is very important if you're a bacteria hoping that a little food will come your way; basically it won't.

**Reynold's number and terminal velocity**

Earlier in the term we learned that for a large object moving through a fluid the frictional resistance is given by:
\[
\text{Drag} = \frac{1}{2} \rho C_d v^2
\]
Where $C$ is a property of the shape of the object, and $A$ is its area. This is approximately true for most large objects, moving in low viscosity air. This is basically saying that the Reynolds’ number for the motion must be large.

What happens in the small Reynold's number case? In this case drag is given by Stokes law:

$$\text{Drag} = 12\pi \eta D v$$

This is quite different. It depends on the diameter of the object $D$, not on its area $A$, and on $v$, not on $v^2$. This is because this drag is dominated by fluid friction, rather than by fluid inertia. Note too that viscosity shows up here, because it matters.