Saint Anselm College

Calculus-Based Physics

Jeffrey W. Schnick

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

This text was compiled on 04/15/2025

TABLE OF CONTENTS

Licensing

Why Write the Book, and, Why Release it for Free?

Volume A: Kinetics, Statics, and Thermodynamics

- 1A: Mathematical Prelude
- 2A: Conservation of Mechanical Energy I: Kinetic Energy & Gravitational Potential Energy
- 3A: Conservation of Mechanical Energy II: Springs, Rotational Kinetic Energy
- 4A: Conservation of Momentum
- 5A: Conservation of Angular Momentum
- 6A: One-Dimensional Motion (Motion Along a Line): Definitions and Mathematics
- 7A: One-Dimensional Motion: The Constant Acceleration Equations
- 8A: One-Dimensional Motion: Collision Type II
- 9A: One-Dimensional Motion Graphs
- 10A: Constant Acceleration Problems in Two Dimensions
- 11A: Relative Velocity
- 12A: Gravitational Force Near the Surface of the Earth, First Brush with Newton's 2nd Law
- 13A: Freefall, a.k.a. Projectile Motion
- 14A: Newton's Laws #1: Using Free Body Diagrams
- 15A: Newton's Laws #2: Kinds of Forces, Creating Free Body Diagrams
- 16A: Newton's Laws #3: Components, Friction, Ramps, Pulleys, and Strings
- 17A: The Universal Law of Gravitation
- 18A: Circular Motion - Centripetal Acceleration
- 19A: Rotational Motion Variables, Tangential Acceleration, Constant Angular Acceleration
- 20A: Torque & Circular Motion
- 21A: Vectors - The Cross Product & Torque
- 22A: Center of Mass, Moment of Inertia
- 23A: Statics
- 24A: Work and Energy
- 25A: Potential Energy, Conservation of Energy, Power
- 26A: Impulse and Momentum
- 27A: Oscillations: Introduction, Mass on a Spring
- 28A: Oscillations: The Simple Pendulum, Energy in Simple Harmonic Motion
- 29A: Waves: Characteristics, Types, Energy
- 30A: Wave Function, Interference, Standing Waves
- 31A: Strings, Air Columns
- 32A: Beats and the Doppler Effect
- 33A: Fluids: Pressure, Density, Archimedes' Principle
- 34A: Pascal's Principle, the Continuity Equation, and Bernoulli's Principle
- 35A: Temperature, Internal Energy, Heat and Specific Heat Capacity
- 36A: Heat: Phase Changes
- 37A: The First Law of Thermodynamics

Volume B: Electricity, Magnetism, and Optics

- B1: Charge & Coulomb's Law
- B2: The Electric Field - Description and Effect
- B3: The Electric Field Due to one or more Point Charges

- B4: Conductors and the Electric Field
- B5: Work Done by the Electric Field and the Electric Potential
- B6: The Electric Potential Due to One or More Point Charges
- B7: Equipotential Surfaces, Conductors, and Voltage
- B8: Capacitors, Dielectrics, and Energy in Capacitors
- 9B: Electric Current, EMF, and Ohm's Law
- B10: Resistors in Series and Parallel; Measuring I & V
- B11: Resistivity and Power
- B12: Kirchhoff's Rules, Terminal Voltage
- B13: RC Circuit
- B14: Capacitors in Series & Parallel
- B15: Magnetic Field Introduction - Effects
- B16: Magnetic Field - More Effects
- B17: Magnetic Field: Causes
- B18: Faraday's Law and Lenz's Law
- B19: Induction, Transformers, and Generators
- B20: Faraday's Law and Maxwell's Extension to Ampere's Law
- B21: The Nature of Electromagnetic Waves
- B22: Huygens's Principle and 2-Slit Interference
- B23: Single-Slit Diffraction
- B24: Thin Film Interference
- B25: Polarization
- B26: Geometric Optics, Reflection
- B27: Refraction, Dispersion, Internal Reflection
- B28: Thin Lenses - Ray Tracing
- B29: Thin Lenses - Lens Equation, Optical Power
- B30: The Electric Field Due to a Continuous Distribution of Charge on a Line
- B31: The Electric Potential due to a Continuous Charge Distribution
- B32: Calculating the Electric Field from the Electric Potential
- B33: Gauss's Law
- B34: Gauss's Law Example
- B35: Gauss's Law for the Magnetic Field and Ampere's Law Revisited
- B36: The Biot-Savart Law
- B37: Maxwell's Equations

Index

Glossary

Detailed Licensing

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing****.*

Why Write the Book, and, Why Release it for Free?

While investigating how best to implement active learning in the classroom, I found that among the various active learning methods I looked into, the critical commonality was the reading of the textbook by the students, prior to the classroom session on the material of the reading assignment. So whichever active learning method I decided on (I chose Eric Mazur's Peer Tutoring method), the question of paramount importance was, "How do I maximize the timely readership of the textbook?" Ideally, each student would read each reading assignment prior to the classroom session on the material of that reading assignment. So another way of posing the question would be, "What must I do (aside from starting each classroom session with a quiz on the reading assignment) to encourage the students to most closely approach that ideal?"

In attempting to answer the question at hand, I took advantage of the fact that it didn't seem like it had been all that long since I was a student. Here are the characteristics I think each reading assignment would have to have in order to maximize the likelihood that, if I were the student, I would do the reading assignment for a class, just prior to the class, for each and every classroom session:

1. "It must be short. Above all else it must be short. I've got problems to solve and laboratory reports to write up and besides that I'm taking four or five other classes. If you expect me to carry out three separate reading assignments each week you better make each reading assignment as short as possible. I don't need you to start off each reading assignment with a story to get me interested in it. I'm already motivated to find out what I need to know. Just tell me what I need to know. Oh, and don't try to entertain me with pictures that are at best peripherally related to what you are writing about. Include only pictures that promote the understanding of what you are supposed to be helping me understand. Also, don't tell me what you are going to write about, and then write about it, and then summarize what you wrote about—just write about it once. I'll find out what you were going to write about by reading it, and I'll summarize it myself. If you keep it short enough in the first place, there is no reason for all that. Don't give me all the banal learning objectives. I know I want to be able to apply the concepts to answer thoughtprovoking questions and to solve physics problems. You don't have to keep writing that over and over again. If it's not going to be on the test, don't make me read about it. I hate it when I struggle mightily to understand some concept in the reading assignment, and, failing to do so, go to you for help only to be told that I don't have to worry about it because it won't be on the test. If it won't be on the test, that means that you don't consider it important, so why make me read it? Keep it short! And listen, you don't have to put each and every thing you want me to learn in the book. Save some of it for the classroom and the laboratory. I can only read so much in one week. How about if you ask me to read the essential facts and explanations of concepts in the textbook and save some of the applications, elaborations, and caveats for class? If you try to do everything with the textbook, it will make it too long. I really want you to keep it short!"
2. "Make it understandable. Be aware that jargon that is as familiar to you as your own name is brand new to me. Don't bandy it about as if it were old hat for me. Write in clear plain English. If you have to use jargon, spell it out clearly when you first use it. And don't be afraid to repeat a definition if it's been a while since you last used it. No, this is not a contradiction to my 'keep it short' mantra. If I have to look up the definition it will take me longer to do the reading assignment than it would have if you had reminded me of the meaning. When I say, 'Keep it short!' I mean, write it so that it takes me as little time as possible to read and understand it."
3. "Don't make me keep searching for equations and figures appearing earlier in the text. Each time you do that, I have to take the time to find the equation or figure to which you are referring and then, after I read the equation or look at the figure, I have to find my place in the textbook again. It not only wastes my time but it reduces my understanding of what you are writing about because it interrupts the flow of the text. Would it kill you to just copy it down again for me right there at the point in the text where you want me to look at it? Once again, I need you to understand that when I say, 'Keep it short!' I don't mean, minimize the total number of pages; rather; I mean, write it so that it takes me as little time as possible to gain the understanding that you were hoping I would gain by reading what you wrote."

Okay, that's enough of me trying to remember how I felt about physics reading assignments when I was a student. I am going to put my professor's cap back on. In trying to create the kind of reading assignments (in textbooks in existence before I wrote mine) that I think that I the student would have been most likely to actually carry out on time, I kept running into a conflict with one of the most important goals of the course as a whole, namely, "to maximize the students' understanding of a selected set of physics concepts and the capability of the students to apply those concepts to answer thoughtprovoking questions and to solve physics problems." These goals ("maximize on-time reading" and "maximize learning") should be in concert with each other. Let me give you an example of one such conflict:

Based on my experience in learning and teaching physics, I have found that it is important to order the topics from familiar to abstract. As such, the traditional ordering of the first semester (starting with kinematics followed by Newton's second law) is appropriate but if adhered to strictly results in the application, by students, of Newton's 2nd Law followed by the application of one or more of the constant acceleration equations, in the solution of problems for which the acceleration is not constant (e.g. problems involving springs, problems involving the universal law of gravitation, and problems involving the application of Coulomb's law). I have found that if one precedes the kinematics with some brief experiences for the students in applying conservation laws, then students tend to use conservation laws more often when they are called for. If one starts the course with the conservation laws, then the most appropriate reading assignments in a textbook with a conventional ordering of topics will involve terms and concepts, assumed known to the student from earlier chapters in the textbook. Textbooks using a "momentum and energy first" ordering tend to go way beyond the "brief initial experience with conservation laws" that I am striving for. Traditional customizable books do not solve the problem—moving chapters around is not sufficient or advisable in the case of physics because of the way later material builds on earlier material. My order is consistent with a spiral approach to learning as students study the conservation laws in more depth later in the course, but you don't need to agree with my conclusions on the best ordering of material to see my point. The point is that a conventional textbook does not allow for the flexibility needed to achieve both "maximization of on-time reading" and "maximization of learning."

So what I needed was a fairly conventional physics textbook that I could easily edit to make the "maximization of on-time reading" support the goal of "maximization of learning." I searched for one on the internet. I didn't find what I was looking for so I wrote one myself. I wanted it to make it so that the next person who was looking for an introductory physics textbook that would be easy to edit would find what they were looking for so I released the book for free.

One of the titles I considered for the book was "Starting Point Physics." The title was to have three different equally valid interpretations: The obvious interpretation is that it would be the book for the first college physics course taken by a student and as such the starting point for the student's study of physics. The second interpretation that I had in mind is that experienced teachers of physics would use it as the starting point for their own physics textbook. In this interpretation, a physics teacher would rearrange the chapters, add a chapter here, remove a chapter there, change a variable name here, add a new equation, there, thoroughly revise an explanation here, delete a long derivation there, etc. until the book was perfectly suited to the course taught by the teacher the way the teacher liked to teach it. The third interpretation involves the new teacher of physics. The selection, ordering, and depth of coverage of topics, as well as the clarity of explanation are, in my opinion, of course (otherwise I would have made them different) excellent. As such, the book would, in my biased opinion, be an excellent choice as the starting point textbook for a new teacher to use the first time that new teacher teaches a calculus-based physics course. With experience, the teacher could customize the book to take better advantage of the teacher's own strengths and the academic background of the students enrolling in the teacher's course.

It might be a pipe dream, but I envision Calculus-Based Physics being edited and used by physicists. After doing the hard work to make the book perfect for one's own course, I would imagine that a physicist would like to share one's own version of the book with the world by posting it to the internet. I am hoping that after several years, a physics professor who decides to adopt the book for her or his course will have several options from which to choose.

Another reason behind the free release of Calculus-Based Physics is the capacity it gives me for immediate correction of mistakes. At one time I was surprised by the prevalence of errors in physics textbooks. Now I recognize it as a natural byproduct of the demands of the marketplace for frequent new editions. I see it as a result of the release of a newer version of a book before there has been time to work out all the bugs in the older version. My plan is to make a correction to the on-line version of the book as soon as possible after I have been made aware of an erratum. The intension is to have Calculus-Based Physics progress toward perfection.

I cannot claim the rising costs of textbooks to have been a major reason behind the writing and free release of Calculus-Based Physics. I consider the no to low cost of Calculus-Based Physics to be a side benefit but I am not convinced that the textbook publishers are out to get the consumer. I do understand the concern with the fact that the folks footing the bill for a textbook are not the ones who are making the decision on what will be purchased but I think that this aspect of the textbook market is just the nature of the beast and that student advocacy groups (do an internet search on "Make Textbooks Affordable" to see what I'm talking about here) are having an impact. I think that market forces are responsible for things like the blossoming of color photographs, new editions, encyclopedic coverage, and multitudes of problems, exercises and examples in physics textbooks today. Authors and publishers put them in there, at substantial cost to the publisher, because it makes the textbooks sell. I think that authors and textbook publishers are doing an excellent job of responding to research into how humans learn physics and creating and promoting textbooks and ancillary materials that incorporate the results of that research. I think it would be a mistake for a physics professor to adopt a poor textbook over a good textbook simply on the basis of cost. Despite the fact that a savings of a hundred dollars to an in-debt student could mean a substantially greater savings over the lifetime of the loans taken out to meet college expenses, the cost of a poor preparation in the foundations of physics, due to the selection of the wrong textbook by a professor, to a person's science or engineering career could be much greater. But I do think it is important to evaluate textbooks and, if they are found to be equal or nearly equal in terms of the expected impact on a student's educational experience, to make cost a serious consideration in the professor's selection of the textbook for the course. A reduction in the cost of the textbook can make additional costs such as web-based homework assignment systems or audience response transceiver rentals (yes, these items can be obtained independently—you don't have to adopt the book with which they are sometimes packaged to take advantage of them) more bearable to the student and thus more viable for the course taught by a professor who believes they will increase learning. I want physics professors to adopt Calculus-Based Physics, not because of the savings to the student, but because, in each case, the physics professor considers Calculus-Based Physics to be the best textbook for the course, the one that will enable the student taking the course to gain the most physics-related knowledge and skill.

The irony about releasing a book for free (Calculus-Based Physics is free in electronic form, either as a pdf file or a zipped Microsoft WordTM Document, but one must pay the cost of production, \$9.98 for volume I and \$11.48 for volume II at the time of this writing, and shipping to get it in black and white paperback form) is that it might not reach as wide an audience as it would if it were released at a substantially higher cost. There is the assumption that low or no cost means low quality. I think folks that read the book will discover the fallacy of that argument.

The bottom line is that I wanted a book that would be ideally suited to my own calculusbased physics course and I wanted to make it available to physics professors in a way that would allow them either to either use it as is or to easily edit it to create a book ideally suited, in the case of each professor, to their own physics course, a book that they could, if desired, release to the public to provide another starting point for the cycle. That's why I wrote it, and that's why I released it for free. I hope you find it useful.

CHAPTER OVERVIEW

Volume A: Kinetics, Statics, and Thermodynamics

1A: Mathematical Prelude
2A: Conservation of Mechanical Energy I: Kinetic Energy & Gravitational Potential Energy
3A: Conservation of Mechanical Energy II: Springs, Rotational Kinetic Energy
4A: Conservation of Momentum
5A: Conservation of Angular Momentum
6A: One-Dimensional Motion (Motion Along a Line): Definitions and Mathematics
7A: One-Dimensional Motion: The Constant Acceleration Equations
8A: One-Dimensional Motion: Collision Type II
9A: One-Dimensional Motion Graphs
10A: Constant Acceleration Problems in Two Dimensions
11A: Relative Velocity
12A: Gravitational Force Near the Surface of the Earth, First Brush with Newton's 2nd Law
13A: Freefall, a.k.a. Projectile Motion
14A: Newton's Laws #1: Using Free Body Diagrams
15A: Newton's Laws #2: Kinds of Forces, Creating Free Body Diagrams
16A: Newton's Laws #3: Components, Friction, Ramps, Pulleys, and Strings
17A: The Universal Law of Gravitation
18A: Circular Motion - Centripetal Acceleration
19A: Rotational Motion Variables, Tangential Acceleration, Constant Angular Acceleration
20A: Torque & Circular Motion
21A: Vectors - The Cross Product & Torque
22A: Center of Mass, Moment of Inertia
23A: Statics
24A: Work and Energy
25A: Potential Energy, Conservation of Energy, Power
26A: Impulse and Momentum
27A: Oscillations: Introduction, Mass on a Spring
28A: Oscillations: The Simple Pendulum, Energy in Simple Harmonic Motion
29A: Waves: Characteristics, Types, Energy
30A: Wave Function, Interference, Standing Waves
31A: Strings, Air Columns
32A: Beats and the Doppler Effect
33A: Fluids: Pressure, Density, Archimedes' Principle
34A: Pascal's Principle, the Continuity Equation, and Bernoulli's Principle
35A: Temperature, Internal Energy, Heat and Specific Heat Capacity
36A: Heat: Phase Changes
37A: The First Law of Thermodynamics

Thumbnail: Roller coaster "Blue Fire" at Europa Park. (CC SA 3.0; Coaster J).

1A: Mathematical Prelude

Just below the title of each chapter is a tip on what I perceive to be the most common mistake made by students in applying material from the chapter. I include these tips so that you can avoid making the mistakes. Here's the first one:

The reciprocal of $\frac{1}{x}+\frac{1}{y}$ is not $x+y$. Try it in the case of some simple numbers. Suppose $x=2$ and $y=4$. Then $\frac{1}{x}+\frac{1}{y}=\frac{1}{2}+\frac{1}{4}=\frac{2}{4}+\frac{1}{4}=\frac{3}{4}$, and the reciprocal of $\frac{4}{3}$is $\frac{3}{4}$ which is clearly not 6 (which is what you obtain if you take the reciprocal of $\frac{1}{2}+\frac{1}{4}$ to be $2+4$). So what is the reciprocal of $\frac{1}{x}+\frac{1}{y}$? The reciprocal of $\frac{1}{x}+\frac{1}{y}$ is $\frac{1}{\frac{1}{x}+\frac{1}{y}}$.

Note

This book is a physics book, not a mathematics book. One of your goals in taking a physics course is to become more proficient at solving physics problems, both conceptual problems involving little to no math, and problems involving some mathematics. In a typical physics problem you are given a description about something that is taking place in the universe and you are supposed to figure out and write something very specific about what happens as a result of what is taking place. More importantly, you are supposed to communicate clearly, completely, and effectively, how, based on the description and basic principles of physics, you arrived at your conclusion. To solve a typical physics problem you have to: (1) form a picture based on the given description, quite often a moving picture, in your mind, (2) concoct an appropriate mathematical problem based on the picture, (3) solve the mathematical problem, and (4) interpret the solution of the mathematical problem. The physics occurs in steps 1, 2, and 4. The mathematics occurs in step 3. It only represents about 25% of the solution to a typical physics problem.

You might well wonder why we start off a physics book with a chapter on mathematics. The thing is, the mathematics covered in this chapter is mathematics you are supposed to already know. The problem is that you might be a little bit rusty with it. We don't want that rust to get in the way of your learning of the physics. So, we try to knock the rust off of the mathematics that you are supposed to already know, so that you can concentrate on the physics. As much as we emphasize that this is a physics course rather than a mathematics course, there is no doubt that you will advance your mathematical knowledge if you take this course seriously. You will use mathematics as a tool, and as with any tool, the more you use it the better you get at using it. Some of the mathematics in this book is expected to be new to you. The mathematics that is expected to be new to you will be introduced in recitation on an as-needed basis. It is anticipated that you will learn and use some calculus in this course before you ever see it in a mathematics course. (This book is addressed most specifically to students who have never had a physics course before and have never had a calculus course before but are currently enrolled in a calculus course. If you have already taken calculus, physics, or both, then you have a wellearned advantage.) Two points of emphasis regarding the mathematical component of your solutions to physics problems that have a mathematical component are in order:

1. You are required to present a clear and complete analytical solution to each problem. This means that you will be manipulating symbols (letters) rather than numbers.
2. For any physical quantity, you are required to use the symbol which is conventionally used by physicists, and/or a symbol chosen to add clarity to your solution. In other words, it is not okay to use the symbol x to represent every unknown.

Aside from the calculus, here are some of the kinds of mathematical problems you have to be able to solve:

Problems Involving Percent Change

A cart is traveling along a track. As it passes through a **photogate**[1] its speed is measured to be $3.40 m/s$. Later, at a second photogate, the speed of the cart is measured to be $3.52 m/s$. Find the percent change in the speed of the cart.

> The percent change in anything is the change divided by the original, all times 100%. (I've emphasized the word "original" because the most common mistake in these kinds of problems is dividing the change by the wrong thing.) The change in a quantity is the new value minus the original value. (The most common mistake here is reversing the order. If you forget which way it goes, think of a simple problem for which you know the answer and see how you must arrange the new and original values to make it come out right. For instance, suppose you gained 2 kg over the summer. You know that the change in your mass is +2 kg. You can calculate the difference both ways—we're talking trial and error with at most two trials. You'll quickly find out that it is "the new value minus the original value" a.k.a. "final minus initial" that yields the correct value for the change.)

Okay, now let's solve the given problem

$$\%Change = \frac{change}{original} 100\%$$

Recalling that the change is the new value minus the original value we have

$$\%Change = \frac{new - original}{original} 100\%$$

While it's certainly okay to memorize this by accident because of familiarity with it, you should concentrate on being able to derive it using common sense (rather than working at memorizing it). Substituting the given values for the case at hand we obtain

$$\%Change = \frac{3.52\frac{m}{s} - 3.40\frac{m}{s}}{3.40\frac{m}{s}} 100\%$$

$$\%Change = 3.5\%$$

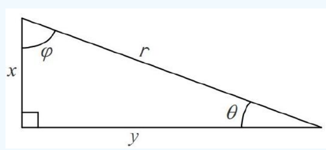
Problems Involving Right Triangles

Example 1A.1:

The length of the shorter side of a right triangle is x and the length of the hypotenuse is r. Find the length of the longer side and find both of the angles, aside from the right angle, in the triangle.

Solution

Draw the triangle such that it is obvious

which side is the shorter side → 

Pythagorean Theorem → $r^2 = x^2 + y^2$

Subtract \x^2 from both sides of the equation → $r^2 - x^2 = y^2$

Swap sides → $y^2 = r^2 - x^2$

Take the square root of both

sides of the equation → \(y=\sqrt{r^2-x^2}\)

By definition, the sine of θ is the side

opposite θ divided by the hypotenuse → $\sin\theta = \frac{x}{r}$

Take the arcsine of both sides of the

equation in order to get θ by itself → $\theta = \sin^{-1}\frac{x}{r}$

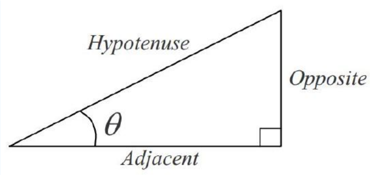
By definition, the cosine of ϕ is the side

adjacent to ϕ divided by the hypotenuse → $\cos\phi = \frac{x}{r}$

Take the arccosine of both sides of the

equation in order to get ϕ by itself → $\phi = \cos^{1}\frac{x}{r}$

To solve a problem like the one above, you need to memorize the relations between the sides and the angles of a right triangle. A convenient mnemonic[2] for doing so is "SOHCAHTOA "



pronounced as a single word.

Referring to the diagram above right:

$$\text{SOH reminds us that:: } \sin\theta = \frac{Opposite}{Hypotenuse}$$

$$\text{CAH reminds us that:: } \cos\theta = \frac{Adjacent}{Hypotenuse}$$

$$\text{TOA reminds us that:: } \tan\theta = \frac{Opposite}{Adhacent}$$

Points to remember:
1. The angle θ is never the 90 degree angle.
2. The words "opposite" and "adjacent" designate sides relative to the angle. For instance, the cosine of θ is the length of the side adjacent to θ divided by the length of the hypotenuse.

You also need to know about the arcsine and the arccosine functions to solve the example problem above. The arcsine function is the inverse of the sine function. The answer to the question, "What is the arcsine of 0.44?" is, "that angle whose sine is 0.44 ." There is an arcsine button on your calculator. It is typically labeled sin-1, to be read, "arcsine." To use it you probably have to hit the inverse button or the second function button on your calculator first.

The inverse function of a function undoes what the function does. Thus:

$$\sin^{-1}\sin\theta = \theta$$

Furthermore, the sine function is the inverse function to the arcsine function and the cosine function is the inverse function to the arccosine function. For the former, this means that:
$\sin(\sin^{-1}x)=x$

Problems Involving the Quadratic Formula

First comes the quadratic equation, then comes the quadratic formula. The quadratic formula is the solution to the **quadratic equation:**

$$ax^2+bx+c=0$$

in which

x is the variable whose value is sought, and a, b, and c are constants

The goal is to find the value of x that makes the left side 0. That value is given by the **quadratic formula:**

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

to be read/said:

'x' equals minus 'b', plus-or-minus the square root of 'b' squared minus four 'a' 'c', all over two 'a'.

So, how do you know when you have to use the quadratic formula? There is a good chance that you need it when the square of the variable for which you are solving, appears in the equation you are solving. When that is the case, carry out the algebraic steps needed to arrange the terms as they are arranged in equation 1-8 above. If this is impossible, then the quadratic formula is not to be used. Note that in the quadratic equation you have a term with the variable to the second power, a term with the variable to the first power, and a term with the variable to the zeroth power (the constant term). If additional powers also appear, such as the one-half power (the square root), or the third power, then the quadratic formula does not apply. If the equation includes additional terms in which the variable whose value is sought appears as the argument of a special function such as the sine function or the exponential function, then the quadratic formula does not apply. Now suppose that there is a square term and you can get the equation that you are solving in the form of equation 1-8 above but that either b or c is zero. In such a case, you can use the quadratic formula, but it is overkill. If b in equation 1-8 above is zero then the equation reduces to:

$$ax^2+bx=0$$

The easy way to solve this problem is to recognize that there is at least one x in each term, and to factor the x out. This yields:

$$(ax+b)x=0$$

Then you have to realize that a product of two multiplicands is equal to zero if either multiplicand is equal to zero. Thus, setting either multiplicand equal to zero and solving for x yields a solution. We have two multiplicands involving x, so, there are two solutions to the equation. The second multiplicand in the expression $(ax+b)x=0$ is x itself, so

$$x=0$$

is a solution to the equation. Setting the first term equal to zero gives:

$$ax+b=0$$

$$ax=-b$$

$$x=-\frac{b}{a}$$

Now suppose the b in the quadratic equation $ax^2+bx+c=0$, equation 1-8, is zero. In that case, the quadratic equation reduces to:

$$ax^2+c=0$$

which can easily be solved without the quadratic formula as follows:

$$ax^2=-c$$

$$x^2=-\frac{c}{a}$$

$$x=\pm\sqrt{-\frac{c}{a}}$$

where we have emphasized the fact that there are two square roots to every value by placing a plus-or-minus sign in front of the radical.

Now, if upon arranging the given equation in the form of the quadratic equation (equation 1-8):

$$ax^2+bx+c=0$$

you find that a, b, and c are all non-zero, then you should use the quadratic formula. Here we present an example of a problem whose solution involves the quadratic formula:

Example $1A.1$: Quadratic Formula Example Problem

Given

$$3+x=\frac{24}{x+1}$$

find x.

Solution

At first glance, this one doesn't look like a quadratic equation, but as we begin isolating x, as we always strive to do in solving for x, (hey, once we have x all by itself on the left side of the equation, with no x on the right side of the equation, we have indeed solved for x—that's what it means to solve for x) we quickly find that it is a quadratic equation. Whenever we have the unknown in the denominator of a fraction, the first step in isolating that unknown is to multiply both sides of the equation by the denominator. In the case at hand, this yields:

$$(x+1)(x+3)=24$$

Multiplying through on the left we find

$$3x+3+x^2+x=24$$

At this point it is pretty clear that we are dealing with a quadratic equation so our goal becomes getting it into the standard form of the quadratic equation, the form of equation 1-8, namely: $ax^2+bx+c=0$. Combining the terms involving x on the left and rearranging we obtain

$$x^2+4x+3=24$$

Subtracting 24 from both sides yields:

$$x^2+4x-21=0$$

which is indeed in the standard quadratic equation form. Now we just have to use inspection to identify which values in our given equation are the a, b, and c that appear in the standard quadratic equation (equation 1-8) $ax^2+bx+c=0$. Although it is not written, the constant multiplying the x^2, in the case at hand, is just 1. So we have a = 1, b = 4, and c = −21.

Substituting these values into the quadratic formula (equation 1-9):

$$x=\frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

yields

$$x=\frac{-4 \pm \sqrt{4^2-4(1)(21)}}{2(1)}$$

which results in x=3,x=-7

as the solutions to the problem. As a quick check we substitute each of these values back into the original equation:

$$3+x=\frac{24}{x+1}$$

and find that each substitution leads to an identity. (An identity is an equation whose validity is trivially obvious, such as 6 = 6.)

This chapter does not cover all the non-calculus mathematics you will encounter in this course. I've kept the chapter short so that you will have time to read it all. If you master the concepts in this chapter (or re-master them if you already mastered them in high school) you will be on your way to mastering all the non-calculus mathematics you need for this course. Regarding reading it all: By the time you complete your physics course, you are supposed to have read this book from cover to cover. Reading physics material that is new to you is supposed to be slow going. By the word reading in this context, we really mean reading with understanding. Reading a physics text involves not only reading but taking the time to make sense of diagrams, taking the time to make sense of mathematical developments, and taking the time to make sense of the words themselves. It involves rereading. The method I use is to push my way through a chapter once, all the way through at a novel-reading pace, picking up as much as I can on the way but not allowing myself to slow down. Then, I really read it. On the second time through I pause and ponder, study diagrams, and ponder over phrases, looking up words in the dictionary and working through examples with pencil and paper as I go. I try not to go on to the next paragraph until I really understand what is being said in the paragraph at hand. That first read, while of little value all by itself, is of great benefit in answering the question, "Where is the author going with this?", while I am carrying out the second read.

2A: Conservation of Mechanical Energy I: Kinetic Energy & Gravitational Potential Energy

Physics professors often assign conservation of energy problems that, in terms of mathematical complexity, are very easy, to make sure that students can demonstrate that they know what is going on and can reason through the problem in a correct manner, without having to spend much time on the mathematics. A good before-and-after-picture correctly depicting the configuration and state of motion at each of two well-chosen instants in time is crucial in showing the appropriate understanding. A presentation of the remainder of the conceptual plus-mathematical solution of the problem starting with a statement in equation form that the energy in the before picture is equal to the energy in the after picture, continuing through to an analytical solution and, if numerical values are provided, only after the analytical solution has been arrived at, substituting values with units, evaluating, and recording the result is almost as important as the picture. The problem is that, at this stage of the course, students often think that it is the final answer that matters rather than the communication of the reasoning that leads to the answer. Furthermore, the chosen problems are often so easy that students can arrive at the correct final answer without fully understanding or communicating the reasoning that leads to it. Students are unpleasantly surprised to find that correct final answers earn little to no credit in the absence of a good correct before andafter picture and a well-written remainder of the solution that starts from first principles, is consistent with the before and after picture, and leads logically, with no steps omitted, to the correct answer. Note that students who focus on correctly communicating the entire solution, on their own, on every homework problem they do, stand a much better chance of successfully doing so on a test than those that "just try to get the right numerical answer" on homework problems.

Mechanical Energy

Energy is a transferable physical quantity that an object can be said to have. If one transfers energy to a material particle that is initially at rest, the speed of that particle changes to a value which is an indicator of how much energy was transferred. Energy has units of joules, abbreviated J. Energy can't be measured directly but when energy is transferred to or from an object, some measurable characteristic (or characteristics) of that object changes (change) such that, measured values of that characteristic or those characteristics (in combination with one or more characteristics such as mass that do not change by any measurable amount) can be used to determine how much energy was transferred. Energy is often categorized according to which measurable characteristic changes when energy is transferred. In other words, we categorize energy in accord with the way it reveals itself to us. For instance, when the measurable characteristic is temperature, we call the energy thermal energy; when the measurable quantity is speed, we call the energy kinetic energy. While it can be argued that there is only one form or kind of energy, in the jargon of physics we call the energy that reveals itself one way one kind or form of energy (such as thermal energy) and the energy that reveals itself another way another kind or form of energy (such as kinetic energy). In physical processes it often occurs that the way in which energy is revealing itself changes. When that happens we say that energy is transformed from one kind of energy to another.

Kinetic Energy is energy of motion. An object at rest has no motion; hence, it has no kinetic energy. The kinetic energy K of a non-rotating rigid object in motion depends on the mass m and speed v of the object **as follows**:

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

The mass m of an object is a measure of the object's inertia, the object's inherent tendency to maintain a constant velocity. The inertia of an object is what makes it hard to get that object moving. The words "mass" and "inertia" both mean the same thing. Physicists typically use the word "inertia" when talking about the property in general conceptual terms, and the word "mass" when they are assigning a value to it, or using it in an equation. Mass has units of kilograms, abbreviated kg. The speed v has units of meters per second, abbreviated m/s. Check out the units in equation 2-1:

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

On the left we have the kinetic energy which has units of joules. On the right we have the product of a mass and the square of a velocity. Thus the units on the right are $kg\frac{m^2}{s^2}$, and we can deduce that a joule is a $kg\frac{m^2}{s^2}$.

Potential Energy is energy that depends on the arrangement of matter. Here, we consider one type of potential energy:

The Gravitational Potential Energy **of an object** near the surface of the earth is the energy (relative to the gravitational potential energy that the object has when it is at the reference level about to me mentioned) that the object has because it is "up high" above a reference level such as the ground, the floor, or a table top. In characterizing the relative gravitational potential energy of an object it is important to specify what you are using for a reference level. In using the concept of near-earth gravitational potential energy to solve a physics problem, although you are free to choose whatever you want to as a reference level, it is important to stick with one and the same reference level throughout the problem. The relative gravitational potential energy U_g of an object near the surface of the earth depends on the object's height y above the chosen reference level, the object's mass m, and the magnitude g of the earth's gravitational field, which to a good approximation has the same value $g = 9.80\dfrac{N}{Kg}$ everywhere near the surface of the earth, as follows: \[U_g=mgy \nonumber \]

The N in $g = 9.80\dfrac{N}{Kg}$ stands for newtons, the unit of force. (Force is an ongoing push or pull.) Since it is an energy, the units of U_g are joules, and the units on the right side of equation 2-2, with the height y being in meters, work out to be newtons times meters. Thus a joule must be a newton meter, and indeed it is. Just above we showed that a joule is a $kg\dfrac{m^2}{s^2}$. If a joule is also a newton meter then a newton must be a $kg\dfrac{m^2}{s^2}$.

A Special Case of the Conservation of Mechanical Energy

Energy is very useful for making predictions about physical processes because it is never created or destroyed. To borrow expressions from economics, that means we can use simple bookkeeping or accounting to make predictions about physical processes. For instance, suppose we create, for purposes of making such a prediction, an imaginary boundary enclosing part of the universe. Then any change in the total amount of energy inside the boundary will correspond exactly to energy transfer through the boundary. If the total energy inside the boundary increases by ΔE, then exactly that same amount of energy ΔE must have been transferred through the boundary into the region enclosed by the boundary from outside that region. And if the total energy inside the boundary decreases by ΔE, then exactly that amount of energy ΔE must have been transferred through the boundary out of the region enclosed by the boundary from inside that region. Oddly enough, in keeping book on the energy in such an enclosed part of the universe, we rarely if ever know or care what the overall total amount of energy is. It is sufficient to keep track of changes. What can make the accounting difficult is that there are so many different ways in which energy can manifest itself (what we call the different “forms” of energy), and there is no simple energy meter that tells us how much energy there is in our enclosed region. Still, there are processes for which the energy accounting is relatively simple. For instance, it is relatively simple when there is no (or negligible) transfer of energy into or out of the part of the universe that is of interest to us, and when there are few forms of energy for which the amount of energy changes. The two kinds of energy discussed above (the kinetic energy of a rigid non-rotating object and gravitational potential energy) are both examples of mechanical energy, to be contrasted with, for example, thermal energy. Under certain conditions the total mechanical energy of a system of objects does not change even though the configuration of the objects does. This represents a special case of the more general principle of the conservation energy. The conditions under which the total mechanical energy of a system doesn’t change are:

1. No energy is transferred to or from the surroundings.
2. No energy is converted to or from other forms of energy (such as thermal energy).

Consider a couple of processes in which the total mechanical energy of a system does not remain the same:

Case #1
A rock is dropped from shoulder height. It hits the ground and comes to a complete stop.

The "system of objects" in this case is just the rock. As the rock falls, the gravitational potential energy is continually decreasing. As such, the kinetic energy of the rock must be continually increasing in order for the total energy to be staying the same. On the collision with the ground, some of the kinetic energy gained by the rock as it falls through space is transferred to the ground and the rest is converted to thermal energy and the energy associated with sound. Neither condition (no transfer and no transformation of energy) required for the total mechanical energy of the system to remain the same is met; hence, it would be incorrect to write an equation setting the initial mechanical energy of the rock (upon release) equal to the final mechanical energy of the rock (after landing).

Can the idea of an unchanging total amount of mechanical energy be used in the case of a falling object? The answer is yes. The difficulties associated with the previous process occurred upon collision with the ground. You can use the idea of an unchanging total amount of mechanical energy to say something about the rock if you end your consideration of the rock before it hits the ground. For instance, given the height from which it is dropped, you can use the idea of an unchanging total amount of mechanical energy to determine the speed of the rock at the last instant before it strikes the ground. The "last instant before" it hits the ground corresponds to the situation in which the rock has not yet touched the ground but will touch the ground in an amount of time that is too small to measure and hence can be neglected. It is so close to the ground that the distance between it and the ground is too small to measure and hence can be neglected. It is so close to the ground that the additional speed that it would pick up in continuing to fall to the ground is too small to be measured and hence can be neglected. The total amount of mechanical energy does not change during this process. It would be correct to write an equation setting the initial mechanical energy of the rock (upon release) equal to the final mechanical energy of the rock (at the last instant before collision).

Case #2
A block, in contact with nothing but a sidewalk, slides across the sidewalk.

The total amount of mechanical energy does not remain the same because there is friction between the block and the sidewalk. In any case involving friction, mechanical energy is converted into thermal energy; hence, the total amount of mechanical energy after the sliding, is not equal to the total amount of mechanical energy prior to the sliding.

Applying the Principle of the Conservation of Energy for the Special Case in which the Mechanical Energy of a System does not Change

In applying the principle of conservation of mechanical energy for the special case in which the mechanical energy of a system does not change, you write an equation which sets the total mechanical energy of an object or system objects at one instant in time equal to the total mechanical energy at another instant in time. Success hangs on the appropriate choice of the two instants. The principal applies to all pairs of instants of the time interval during which energy is neither transferred into or out of the system nor transformed into non-mechanical forms. You characterize the conditions at the first instant by means of a "Before Picture" and the conditions at the second instant by means of an "After Picture.” In applying the principle of conservation of mechanical energy for the special case in which the mechanical energy of a system does not change, you write an equation which sets the total mechanical energy in the Before Picture equal to the total mechanical energy in the After Picture. (In both cases, the “total” mechanical energy in question is the amount the system has relative to the mechanical energy it would have if all objects were at rest at the reference level.) To do so effectively, it is necessary to sketch a Before Picture and a separate After Picture. After doing so, the first line in one's solution to a problem involving an unchanging total of mechanical energy always reads

$$\text{energy before} = \text{energy after}$$

We can write this first line more symbolically in several different manners:

$$E_1 = E_2$$

or

$$E_i = E_f$$

or

$$E = E'$$

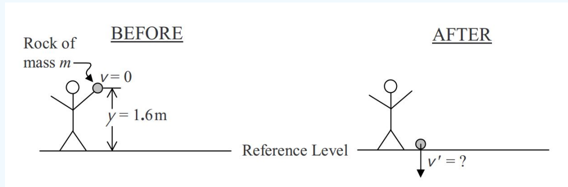
The first two versions use subscripts to distinguish between "before picture" and "after picture" energies and are to be read "E-sub-one equals E-sub-two" and "E-sub-i equals E-sub-f." In the latter case the symbols i and f stand for initial and final. In the final version, the prime symbol is added to the E to distinguish "after picture" energy from "before picture" energy. The last equation is to be read "E equals E-prime." (The prime symbol is sometimes used in mathematics to distinguish one variable from another and it is sometimes used in mathematics to signify the derivative with respect to x. It is never used it to signify the derivative in this book.) The unprimed/prime notation is the notation that will be used in the following example:

Example 2.1

A rock is dropped from a height of 1.6 meters. How fast is the rock falling just before it hits the ground?

Solution

Choose the "before picture" to correspond to the instant at which the rock is released, since the conditions at this instant are specified ("dropped" indicates that the rock was released from rest—its speed is initially zero, the initial height of the rock is given). Choose the "after picture" to correspond to the last instant before the rock makes contact with the ground since the question pertains to a condition (speed) at this instant.



$$E = E'$$

0 (since at rest) 0 (since at ground level)

$$\cancel{K} + U = K' + \cancel{U'}$$

Note that we have omitted the subscript g (for "gravitational") from both U and U'. When you are dealing with only one kind of potential energy, you don't need to use a subscript to distinguish it from other kinds.

$$mgy = \frac{1}{2}mv'^2$$

$$v'^2 = 2gy$$

$$v' = \sqrt{2gy}$$

$$v' = \sqrt{2(9.80\,\mathrm{m/s^2})1.6\,\mathrm{m}}$$

$$\boxed{v' = 5.6\,\frac{\mathrm{m}}{\mathrm{s}}}$$

Note that the unit, 1 newton, abbreviated as 1 N, is $1\dfrac{kg*m}{s^2}$. Hence, the magnitude of the earth's near-surface gravitational field $g = 9.80\dfrac{N}{kg}$ can also be expressed as $g = 9.80\dfrac{m}{s^2}$ as we have done in the example for purposes of working out the units.

The solution presented in the example provides you with an example of what is required of students in solving physics problems. In cases where student work is evaluated, it is the solution which is evaluated, not just the final answer. In the following list, general requirements for solutions are discussed, with reference to the solution of the example problem:

1. Sketch (the before and after pictures in the example).
Start each solution with a sketch or sketches appropriate to the problem at hand. Use the sketch to define symbols and, as appropriate, to assign values to symbols. The sketch aids you in solving the problem and is important in communicating your solution to the reader. Note that each sketch depicts a configuration at a particular instant in time rather than a process which extends over a time interval.

2. Write the "Concept Equation" ($E = E'$ in the example).

3. Replace quantities in the "Concept Equation" with more specific representations of the same quantities. Repeat as appropriate.

In the example given, the symbol E representing total mechanical energy in the before picture is replaced with "what it is," namely, the sum of the kinetic energy and the potential energy $K+U$ of the rock in the before picture. On the same line E' has been replaced with what it is, namely, the sum of the kinetic energy and the potential energy $K'+U'$ in the after picture. Quantities that are obviously zero have slashes drawn through them and are omitted from subsequent steps.

This step is repeated in the next line ($mgy = \frac{1}{2}mV'^2$) in which the gravitational potential energy in the before picture, U, has been replaced with what it is, namely mgy, and on the right, the kinetic energy in the after picture has been replaced with what it is, namely, $\frac{1}{2}mV'^2$. The symbol m that appears in this step is defined in the diagram.

4. Solve the problem algebraically. The student is required to solve the problem by algebraically manipulating the symbols rather than substituting values and simultaneously evaluating and manipulating them. The reasons that physics teachers require students taking college level physics courses to solve the problems algebraically in terms of the symbols rather than working with the numbers are:

(a) College physics teachers are expected to provide the student with experience in "the next level" in abstract reasoning beyond working with the numbers. To gain this experience, the students must solve the problems algebraically in terms of symbols.

(b) Students are expected to be able to solve the more general problem in which, whereas certain quantities are to be treated as if they are known, no actual values are given. Solutions to such problems are often used in computer programs which enable the user to obtain results for many different values of the "known quantities.” Actual values are assigned to the known quantities only after the user of the program provides them to the program as input—long after the algebraic problem is solved.

(c) Many problems more complicated than the given example can more easily be solved algebraically in terms of the symbols. Experience has shown that students accustomed to substituting numerical values for symbols at the earliest possible stage in a problem are unable to solve the more complicated problems.

In the example, the algebraic solution begins with the line $\frac{1}{2}mV'^2$. The m's appearing on both sides of the equation have been canceled out (this is the algebraic step) in the solution provided. Note that in the example, had the m's not canceled out, a numerical answer to the problem could not have been determined since no value for m was given. The next two lines represent the additional steps necessary in solving algebraically for the final speed v'. The final line in the algebraic solution ($v' = \sqrt{2gy}$ in the example) always has the quantity being solved for all by itself on the *left* side of the equation being set equal to an expression involving only known quantities on the right side of the equation. The algebraic solution is not complete if unknown quantities (especially the quantity sought) appear in the expression on the right hand side. Writing the final line of the algebraic solution in the reverse order, e.g.$\sqrt{2gy} = v'$, is unconventional and hence unacceptable. If your algebraic solution naturally leads to that, you should write one more line with the algebraic answer written in the correct order.

5. Replace symbols with numerical values with units, $v' = \sqrt{2(9.80\frac{m}{s^2})1.6m}$ in the example; the units are the units of measurement: $\frac{m}{s^2}$ and m in the example).

No computations should be carried out at this stage. Just copy down the algebraic solution but with symbols representing known quantities replaced with numerical values with units. Use parentheses and brackets as necessary for clarity.

6. Write the final answer with units ($v' = 5.6\frac{m}{s}$ in the example).

Numerical evaluations are to be carried out directly on the calculator and/or on scratch paper. It is unacceptable to clutter the solution with arithmetic and intermediate numerical answers between the previous step and this step. Units should be worked out and provided with the final answer. It is good to show some steps in working out the units but for simple cases units (not algebraic solutions) may be worked out in your head. In the example provided, it is easy to see that upon taking the square root of the product of $\frac{m}{s^2}$ and m, one obtains $\frac{m}{s}$ hence no additional steps were depicted.

3A: Conservation of Mechanical Energy II: Springs, Rotational Kinetic Energy

Note

A common mistake involving springs is using the length of a stretched spring when the amount of stretch is called for. Given the length of a stretched spring, you have to subtract off the length of that same spring when it is neither stretched nor compressed to get the amount of stretch.

Spring Potential Energy is the potential energy stored in a spring that is compressed or stretched. The spring energy depends on how stiff the spring is and how much it is stretched or compressed. The stiffness of the spring is characterized by the force constant of the spring, k. k is also referred to as the spring constant for the spring. The stiffer the spring, the bigger its value of k is. The symbol x is typically used to characterize the amount by which a spring is compressed or stretched. It is important to note that x is not the length of the stretched or compressed spring. Instead, it is the difference between the length of the stretched or compressed spring and the length of the spring when it is neither stretched nor compressed. The amount of energy U_S stored in a spring with a force constant (spring constant) k that has either been stretched by an amount x or compressed by an amount x is:

$$U=\frac{1}{2}kx^2 \tag{3A.1}$$

Rotational Kinetic Energy is the energy that a spinning object has because it is spinning. When an object is spinning, every bit of matter making up the object is moving in a circle (except for those bits on the axis of rotation). Thus, every bit of matter making up the object has some kinetic energy $\frac{1}{2}mV^2$ where V is the speed of the bit of matter in question and m is its mass. The thing is, in the case of an object that is just spinning, the object itself is not going anywhere, so it has no speed, and the different bits of mass making up the object have different speeds, so there is no one speed V that we can use for the speed of the object in our old expression for kinetic energy $K=\frac{1}{2}mV^2$. The amount of kinetic energy that an object has because it is spinning can be expressed as:

$$K=\frac{1}{2}I\omega^2 \tag{3A.2}$$

where the Greek letter omega ω (please don't call it double-u) is used to represent the magnitude of the angular velocity of the object and the symbol I is used to represent the moment of inertia, a.k.a. rotational inertia, of the object. The magnitude of the angular velocity of the object is how fast the object is spinning and the moment of inertia of the object is a measure of the object's natural tendency to spin at a constant rate. The greater the moment of inertia of an object, the harder it is to change how fast that object is spinning.

The magnitude of the angular velocity, the spin rate ω, is measured in units of radians per second where the radian is a unit of angle. An angle is a fraction of a rotation and hence a unit of angle is a fraction of a rotation. If we divide a rotation up into 360 parts then each part is $\frac{1}{360}$ of a rotation and we call each part a degree. In the case of radian measure, we divide the rotation up into 2π parts and call each part a radian. Thus a radian is $\frac{1}{2\pi}$ of a rotation. The fact that an angle is a fraction of a rotation means that an angle is really a pure number and the word "radian" abbreviated rad, is a reminder about how many parts the rotation has been divided up into, rather than a true unit. In working out the units in cases involving radians, one can simply erase the word radian. This is not the case for actual units such as meters or joules.

The moment of inertia I has units of $kg\cdot m^2$. The units of the right hand side of equation 3-2, $K=\frac{1}{2}I\omega^2$, thus work out to be $kg\cdot m^2\frac{rad^2}{s^2}$. Taking advantage of the fact that a radian is not a true unit, we can simply erase the units rad^2 leaving us with units of $kg\cdot\frac{m^2}{s^2}$, a combination that we recognize as a joule which it must be since the quantity on the left side of the equation $K=\frac{1}{2}I\omega^2$ (equation 3-2) is an energy.

Energy of Rolling

An object which is rolling is both moving through space and spinning so it has both kinds of kinetic energy, the $\frac{1}{2}mV^2$ and the $\frac{1}{2}I\omega^2$. The movement of an object through space is called translation. To contrast it with rotational kinetic energy, the ordinary kinetic energy $K=\frac{1}{2}mV^2$ is referred to as translational kinetic energy. So, the total kinetic energy of an object that is rolling can be expressed as

$$K_{Rolling}=K_{translation}+K_{rotation} \tag{3A.3}$$

$$K_{Rolling}=\frac{1}{2}mV^2+\frac{1}{2}I\omega^2 \tag{3A.4}$$

Now you probably recognize that an object that is rolling without slipping is spinning at a rate that depends on how fast it is going forward. That is to say that the value of ω depends on the value of v. Let's see how. When an object that is rolling without slipping completes one rotation, it moves a distance equal to its circumference which is 2π times the radius of that part of the object on which the object is rolling.

$$\text{Distance traveled in one rotation}=2\pi r \tag{3A.5}$$

Now if we divide both sides of this equation by the amount of time that it takes for the object to complete one rotation we obtain on the left, the speed of the object and, on the right, we can interpret the 2π as 2π radians and, since 2π radians is one rotation the 2π radians divided by the time it takes for the object to complete one rotation is just the magnitude of the angular velocity ω. Hence we arrive at

$$v=\omega r$$

which is typically written:

$$v=r\omega \tag{3A.6}$$

4A: Conservation of Momentum

> A common mistake involving conservation of momentum crops up in the case of totally inelastic collisions of two objects, the kind of collision in which the two colliding objects stick together and move off as one. The mistake is to use conservation of mechanical energy rather than conservation of momentum. One way to recognize that some mechanical energy is converted to other forms is to imagine a spring to be in between the two colliding objects such that the objects compress the spring. Then imagine that, just when the spring is at maximum compression, the two objects become latched together. The two objects move off together as one as in the case of a typical totally inelastic collision. After the collision, there is energy stored in the compressed spring so it is clear that the total kinetic energy of the latched pair is less than the total kinetic energy of the pair prior to the collision. There is no spring in a typical inelastic collision. The mechanical energy that would be stored in the spring, if there was one, results in permanent deformation and a temperature increase of the objects involved in the collision.

The momentum of an object is a measure of how hard it is to stop that object. The momentum of an object depends on both its mass and its velocity. Consider two objects of the same mass, e.g. two baseballs. One of them is coming at you at 10 mph, and the other at 100 mph. Which one has the greater momentum? Answer: The faster baseball is, of course, harder to stop, so it has the greater momentum. Now consider two objects of different mass with the same velocity, e.g. a Ping-Pong ball and a cannon ball, both coming at you at 25 mph. Which one has the greater momentum? The cannon ball is, of course, harder to stop, so it has the greater momentum.

The momentum p of an object is equal to the product of the object's mass m and velocity V:

$$p = mV$$

Momentum has direction. Its direction is the same as that of the velocity. In this chapter we will limit ourselves to motion along a line (motion in one dimension). Then there are only two directions, forward and backward. An object moving forward has a positive velocity/momentum and one moving backward has a negative velocity/momentum. In solving physics problems, the decision as to which way is forward is typically left to the problem solver. Once the problem solver decides which direction is the positive direction, she must state what her choice is (this statement, often made by means of notation in a sketch, is an important part of the solution), and stick with it throughout the problem.

The concept of momentum is important in physics because the total momentum of any system remains constant unless there is a net transfer of momentum to that system, and if there is an ongoing momentum transfer, the rate of change of the momentum of the system is equal to the rate at which momentum is being transferred into the system. As in the case of energy, this means that one can make predictions regarding the outcome of physical processes by means of simple accounting (bookkeeping) procedures. The case of momentum is complicated by the fact that momentum has direction, but in this initial encounter with the conservation of momentum you will deal with cases involving motion along a straight line. When all the motion is along one and the same line, there are only two possible directions for the momentum and we can use algebraic signs (plus and minus) to distinguish between the two. The principle of Conservation of Momentum applies in general. At this stage in the course however, we will consider only the special case in which there is no net transfer of momentum to (or from) the system from outside the system.

> *Conservation of Momentum in One Dimension for the Special Case in which there is No Transfer of Momentum to or from the System from Outside the System.*

In any process involving a system of objects which all move along one and the same line, as long as none of the objects are pushed or pulled along the line by anything outside the system of objects (it's okay if they push and pull on each other), the total momentum before, during, and after the process remains the same.

The total momentum of a system of objects is just the algebraic sum of the momenta of the individual objects. That adjective "algebraic" means you have to pay careful attention to the plus and minus signs. If you define "to the right" as your positive direction and your system of objects consists of two objects, one moving to the right with a momentum of $12 kg \cdot m/s$ and the other moving to the left with momentum $5 kg \cdot m/s$, then the total momentum is (+12 kg·m/s) + (−5 kg·m/s) which is +7 kg·m/s. The plus sign in the final answer means that the total momentum is directed to the right.

Upon reading this selection you'll be expected to be able to apply conservation of momentum to two different kinds of processes. In each of these two classes of processes, the system of objects will consist of only two objects. In one class, called collisions, the two objects bump into each other. In the other class, anti-collisions the two objects start out together, and spring apart. Some further breakdown of the collisions class is pertinent before we get into examples. The two extreme types of collisions are the completely **inelastic** collision, and the completely **elastic** collision.

Upon a completely inelastic collision, the two objects stick together and move off as one. This is the easy case since there is only one final velocity (because they are stuck together, the two objects obviously move off at one and the same velocity). Some mechanical energy is converted to other forms in the case of a completely inelastic collision. It would be a big mistake to apply the principle of conservation of mechanical energy to a completely inelastic collision. Mechanical energy is **not** conserved. The words "completely inelastic" tell you that both objects have the same velocity (as each other) after the collision.

In a completely **elastic** collision (often referred to simply as an **elastic** collision), the objects bounce off each other in such a manner that no mechanical energy is converted into other forms in the collision. Since the two objects move off independently after the collision there are two final velocities. If the masses and the initial velocities are given, conservation of momentum yields one equation with two unknowns—namely, the two final velocities. Such an equation cannot be solved by itself. In such a case, one must apply the principle of conservation of mechanical energy. It does apply here. The expression "completely elastic" tells you that conservation of mechanical energy does apply.

In applying conservation of momentum one first sketches a before and an after picture in which one defines symbols by labeling objects and arrows (indicating velocity), and defines which direction is chosen as the positive direction. The first line in the solution is always a statement that the total momentum in the before picture is the same as the total momentum in the after picture. This is typically written by means an equation of the form:

$$\sum p \rightarrow = \sum p' \rightarrow$$

The Σ in this expression is the upper case Greek letter "sigma" and is to be read "the sum of." Hence the equation reads: "The sum of the momenta to the right in the before picture is equal to the sum of the momenta to the right in the after picture." In doing the sum, a leftward momentum counts as a negative rightward momentum. The arrow subscript is being used to define the positive direction.

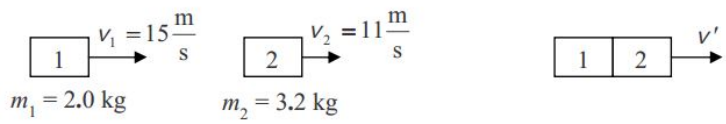
We'll use the examples to clarify what is meant by collisions and anti-collisions; to introduce one more concept, namely, relative velocity (sometimes referred to as muzzle velocity); and of course, to show the reader how to apply conservation of momentum.

✔ Example $4A.1$: A Collision Problem

Two objects move on a horizontal frictionless surface along the same line in the same direction which we shall refer to as the forward direction. The trailing object of mass $2.0 kg$ has a velocity of $15 m/s$ forward. The leading object of mass $3.2 kg$ has a velocity of $11 m/s$ forward. The trailing object catches up with the leading object and the two objects experience a completely inelastic collision. What is the final velocity of each of the two objects?

Solution

BEFORE AFTER



$$\Sigma p_{\rightarrow} = \Sigma p'_{\rightarrow}$$

$$p_1 + p_2 = p'_{12}$$

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2)v'$$

$$v' = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2}$$

$$v' = \frac{2.0\,\text{kg}(15\,\text{m/s}) + 3.2\,\text{kg}(11\,\text{m/s})}{2.0\,\text{kg} + 3.2\,\text{kg}}$$

$$v' = 12.54\,\frac{\text{m}}{\text{s}}$$

$$v' = 13\,\frac{\text{m}}{\text{s}}$$

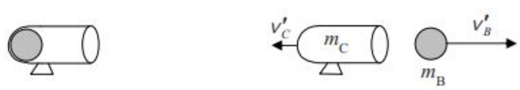
The final velocity of each of the objects is $\boxed{13\,\frac{\text{m}}{\text{s}} \text{ forward.}}$

✓ Example $4A.2$: An Anti-Collision Problem

A cannon of mass m_C, resting on a frictionless surface, fires a ball of mass m_B. The ball is fired horizontally. The muzzle velocity is v_M . Find the velocity of the ball and the recoil velocity of the cannon.

Solution

BEFORE AFTER



$$\Sigma p_{\rightarrow} = \Sigma p'_{\rightarrow}$$

$$0 = -m_C v'_C + m_B v'_B \qquad (1)$$

Also, from the definition of muzzle velocity:

$$v_M = v'_B + v'_C$$

$$v'_C = v_M - v'_B \qquad (2)$$

Substituting this result into equation (1) yields:

$$0 = -m_C(v_M - v'_B) + m_B v'_B$$

$$0 = -m_C v_M + m_C v'_B + m_B v'_B$$

$$m_C v'_B + m_B v'_B = m_C v_M$$

$$(m_C + m_B)v'_B = m_C v_M$$

$$\boxed{v'_B = \frac{m_C}{m_C + m_B} v_M}$$

Now substitute this result into equation (2) above. This yields:

$$v'_C = v_M - \frac{m_C}{m_C + m_B} v_M$$

$$v'_C = \frac{(m_C + m_B)v_M - m_C v_M}{m_C + m_B}$$

$$v'_C = \frac{m_C v_M + m_B v_M - m_C v_M}{m_C + m_B}$$

$$\boxed{v'_C = \frac{m_B}{m_C + m_B} v_M}$$

This is an example of an anti-collision problem. It also involves the concept of relative velocity. The muzzle velocity is the relative velocity between the ball and the cannon. It is the velocity at which the two separate. If the velocity of the ball relative to the ground is v_B' to the right, and the velocity of the cannon relative to the ground is v_c' to the left, then the velocity of the ball relative to the cannon, also known as the muzzle velocity of the ball, is $v_M = v_B' + v_c'$. In cases not involving guns or cannons one typically uses the notation v_{rel} for "relative velocity" or, relating to the example at hand, v_{BC} for "velocity of the ball relative to the cannon."

5A: Conservation of Angular Momentum

Much as in the case of linear momentum, the mistake that tends to be made in the case of angular momentum is not using the principle of conservation of angular momentum when it should be used, that is, applying conservation of mechanical energy in a case in which mechanical energy is not conserved but angular momentum is. Consider the case, for instance, in which one drops a disk (from a negligible height) that is not spinning, onto a disk that is spinning, and after the drop, the two disks spin together as one. The "together as one" part tips you off that this is a completely inelastic (rotational) collision. Some mechanical energy is converted into thermal energy (and other forms not accounted for) in the collision. It's easy to see that mechanical energy is converted into thermal energy if the two disks are CD's and the bottom one is initially spinning quite fast (but is not being driven). When you drop the top one onto the bottom one, there will be quite a bit of slipping before the top disk gets up to speed and the two disks spin as one. During the slipping, it is friction that increases the spin rate of the top CD and slows the bottom one. Friction converts mechanical energy into thermal energy. Hence, the mechanical energy prior to the drop is less than the mechanical energy after the drop

The angular momentum of an object is a measure of how difficult it is to stop that object from spinning. For an object rotating about a fixed axis, the angular momentum depends on how fast the object is spinning, and on the object's rotational inertia (also known as moment of inertia) with respect to that axis.

Rotational Inertia (a.k.a. Moment of Inertia)

The rotational inertia of an object with respect to a given rotation axis is a measure of the object's tendency to resist a change in its angular velocity about that axis. The rotational inertia depends on the mass of the object and how that mass is distributed. You have probably noticed that it is easier to start a merry-go-round spinning when it has no children on it. When the kids climb on, the mass of what you are trying to spin is greater, and this means the rotational inertia of the object you are trying to spin is greater. Have you also noticed that if the kids move in toward the center of the merry-go-round it is easier to start it spinning than it is when they all sit on the outer edge of the merry-go-round? It is. The farther, on the average, the mass of an object is distributed away from the axis of rotation, the greater the object's moment of inertia with respect to that axis of rotation. The rotational inertia of an object is represented by the symbol I. During this initial coverage of angular momentum, you will not be required to calculate I from the shape and mass of the object. You will either be given I or expected to calculate it by applying conservation of angular momentum (discussed below).

Angular Velocity

The angular velocity of an object is a measure of how fast it is spinning. It is represented by the Greek letter omega, written w, (not to be confused with the letter w which, unlike omega, is pointed on the bottom). The most convenient measure of angle in discussing rotational motion is the radian. Like the degree, a radian is a fraction of a revolution. But, while one degree is $\frac{1}{360}$ of a revolution, one radian is $\frac{1}{2\pi}$ of a revolution. The units of angular velocity are then *radians per second* or, in notational form, $\frac{rad}{s}$. Angular velocity has direction or sense of rotation associated with it. If one defines a rotation which is clockwise when viewed from above as a positive rotation, then an object which is rotating counterclockwise as viewed from above is said to have a negative angular velocity. In any problem involving angular velocity, one is free to choose the positive sense of rotation, but then one must stick with that choice throughout the problem.

Angular Momentum

The angular momentum L of an object is given by:

$$L = I\omega$$

Note that this is consistent with our original definition of angular momentum as a measure of the degree of the object's tendency to keep on spinning, once it is spinning. The greater the rotational inertia of the object, the more difficult it is to stop the object from spinning, and the greater the angular velocity of the object, the more difficult it is to stop the object from spinning.

The direction of angular momentum is the same as the direction of the corresponding angular velocity.

Torque

We define torque by analogy with force which is an ongoing push or pull on an object. When there is a single force acting on a particle, the momentum of that particle is changing. A torque is what you are exerting on the lid of a jar when you are trying to remove the lid. When there is a single torque acting on a rigid object, the angular momentum of that object is changing.

Conservation of Angular Momentum

Angular Momentum is an important concept because, if there is no angular momentum transferred to or from a system, the total angular momentum of that system does not change, and if there is angular momentum being transferred to a system, the rate of change of the angular momentum of the system is equal to the rate at which angular momentum is being transferred to the system. As in the case of energy and momentum, this means we can use simple accounting (bookkeeping) procedures for making predictions on the outcomes of physical processes. In this chapter we focus on the special case in which there are no external torques which means that no angular momentum is transferred to or from the system.

Definition: Conservation of Angular Momentum

Conservation of Angular Momentum for the Special Case in which no Angular Momentum is Transferred to or from the System from Outside the System

In any physical process involving an object or a system of objects free to rotate about an axis, as long as there are no external torques exerted on the system of objects, the total angular momentum of that system of objects remains the same throughout the process.

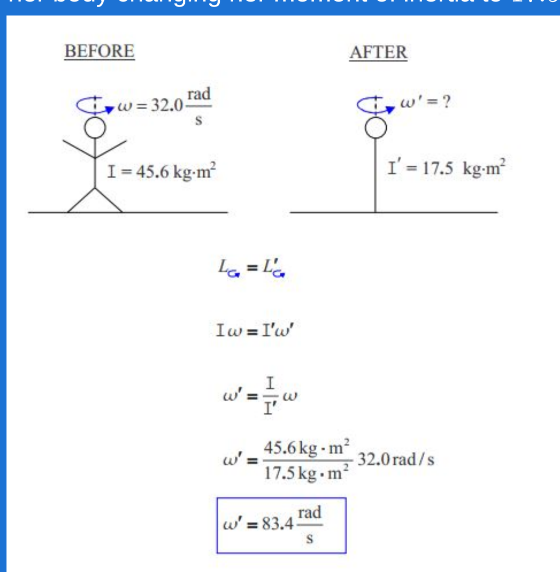
The application of the conservation of angular momentum in solving physics problems for cases involving no transfer of angular momentum to or from the system from outside the system (no external torque) is very similar to the application of the conservation of energy and to the application of the conservation of momentum. One selects two instants in time, defines the earlier one as the before instant and the later one as the after instant, and makes corresponding sketches of the object or objects in the system. Then one writes

$$L = L'$$

meaning "the angular momentum in the before picture equals the angular momentum in the after picture." Next, one replaces each L with what it is in terms of the moments of inertia and angular velocities in the problem and solves the resulting algebraic equation for whatever is sought.

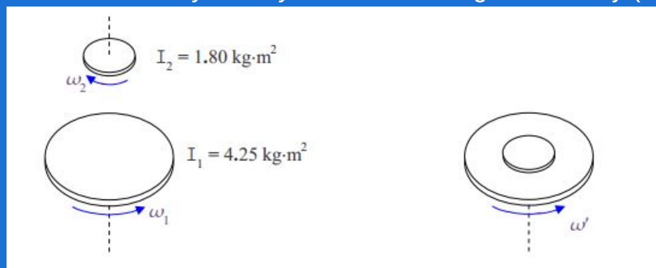
A skater is spinning at $32.0 rad/s$ with her arms and legs extended outward. In this position her moment of inertia with respect to the vertical axis about which she is spinning is $45.6 kg \cdot m^2$. She pulls her arms and legs in close to her body changing her moment of inertia to $17.5 kg \cdot m^2$. What is her new angular velocity?

Solution



A horizontal disk of rotational inertia $4.25 kg \cdot m^2$ with respect to its axis of symmetry is spinning counterclockwise about its axis of symmetry, as viewed from above, at 15.5 revolutions per second on a frictionless massless bearing. A second disk, of rotational inertia $1.80 kg \cdot m^2$ with respect to its axis of symmetry, spinning clockwise as viewed from above about the same axis (which is also its axis of symmetry) at 14.2 revolutions per second, is dropped on top of the first disk. The two disks stick together and rotate as one about their common axis of symmetry at what new angular velocity (in units of radians per second)?

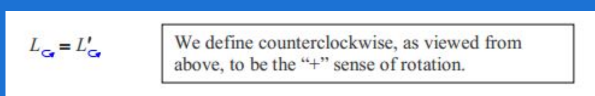
Solution



Some preliminary work (expressing the given angular velocities in units of rad/s):

$$\omega 1 = 15.5 \frac{rev}{s} \left(\frac{2\pi * rad}{rev}\right) = 97.39 \frac{rad}{s} \qquad \omega 2 = 14.2 \frac{rev}{s} \left(\frac{2\pi * rad}{rev}\right) = 89.22 \frac{rad}{s}$$

Now we apply the principle of conservation of angular momentum for the special case in which there is no transfer of angular momentum to or from the system from outside the system. Referring to the diagram:



$$\omega' = \frac{I1\omega 1 - I2\omega 2}{I1 + I2}$$

$$I1\omega 1 - I2\omega 2 = (I1 + I2)\omega'$$

$$\omega' = \frac{(4.25 kg \cdot m^2) 97.39 rad/s - (1.80 kg \cdot m^2) 89.22 rad/s}{4.25 kg \cdot m^2 + 1.80 kg \cdot m^2}$$

$\omega' = 41.9 \frac{rad}{s}$ (Counterclockwise as viewed from above.)

6A: One-Dimensional Motion (Motion Along a Line): Definitions and Mathematics

> A mistake that is often made in linear motion problems involving acceleration, is using the velocity at the end of a time interval as if it was valid for the entire time interval. The mistake crops up in constant acceleration problems when folks try to use the definition of average velocity $\bar{v} = \dfrac{\triangle x}{\triangle t}$ in the solution. Unless you are asked specifically about average velocity, you will never need to use this equation to solve a physics problem. Avoid using this equation—it will only get you into trouble. For constant acceleration problems, use the set of constant acceleration equations provided you.

Here we consider the motion of a particle along a straight line. The particle can speed up and slow down and it can move forward or backward but it does not leave the line. While the discussion is about a particle (a fictitious object which at any instant in time is at a point in space but has no extent in space—no width, height, length, or diameter) it also applies to a rigid body that moves along a straight line path without rotating, because in such a case, every particle of the body undergoes one and the same motion. This means that we can pick one particle on the body and when we have determined the motion of that particle, we have determined the motion of the entire rigid body.

So how do we characterize the motion of a particle? Let's start by defining some variables:

- t: How much time t has elapsed since some initial time. The initial time is often referred to as "the start of observations" and even more often assigned the value 0. We will refer to the amount of time t that has elapsed since time zero as the stopwatch reading. A time interval $\triangle t$ (to be read "delta t") can then be referred to as the difference between two stopwatch readings.
- x : Where the object is along the straight line. To specify the position of an object on a line, one has to define a reference position (the start line) and a forward direction. Having defined a forward direction, the backward direction is understood to be the opposite direction. It is conventional to use the symbol x to represent the position of a particle. The values that x can have, have units of length. The SI unit of length is the meter. (SI stands for "Systeme International," the international system of units.) The symbol for the meter is m. The physical quantity x can be positive or negative where it is understood that a particle which is said to be minus five meters forward of the start line (more concisely stated as $x = -5m$) is actually five meters behind the start line.
- v: How fast and which way the particle is going—the velocity of the object. Because we are considering an object that is moving only along a line, the "which way" part is either forward or backward. Since there are only two choices, we can use an algebraic sign ("+" or "−") to characterize the direction of the velocity. By convention, a positive value of velocity is used for an object that is moving forward, and a negative value is used for an object that is moving backward. Velocity has both magnitude and direction. The magnitude of a physical quantity that has direction is how big that quantity is, regardless of its direction. So the magnitude of the velocity of an object is how fast that object is going, regardless of which way it is going. Consider an object that has a velocity of $5m/s$. The magnitude of the velocity of that object is $5m/s$. Now consider an object that has a velocity of $-5m/s$. (It is going backward at $5m/s$.) The magnitude of its velocity is also $5m/s$. Another name for the magnitude of the velocity is the speed. In both of the cases just considered, the speed of the object is $5m/s$ despite the fact that in one case the velocity was $-5m/s$. To understand the "how fast" part, just imagine that the object whose motion is under study has a built-in speedometer. The magnitude of the velocity, a.k.a. the speed of the object, is simply the speedometer reading.
- a: Next we have the question of how fast and which way the velocity of the object is changing. We call this the acceleration of the object. Instrumentally, the acceleration of a car is indicated by how fast and which way the tip of the speedometer needle is moving. In a car, it is determined by how far down the gas pedal is pressed or, in the case of car that is slowing down, how hard the driver is pressing on the brake pedal. In the case of an object that is moving along a straight line, if the object has some acceleration, then the speed of the object is changing.

Okay, we've got the quantities used to characterize motion. Soon, we're going to develop some useful relations between those variables. While we're doing that, I want you to keep these four things in mind:

1. We're talking about an object moving along a line.
2. Being in motion means having your position change with time.
3. You already have an intuitive understanding of what instantaneous velocity is because you have ridden in a car. You know the difference between going 65 mph and 15 mph and you know very well that you neither have to go 65 miles nor travel for an hour to be going 65 mph. In fact, it is entirely possible for you to have a speed of 65 mph for just an instant (no time interval at all —it's how fast you are going (what your speedometer reading is) at that instant. To be sure, the speedometer needle may be

just “swinging through” that reading, perhaps because you are in the process of speeding up to 75 mph from some speed below 65 mph, but the 65 mph speed still has meaning and still applies to that instant when the speedometer reading is 65 mph. Take this speed concept with which you are so familiar, tack on some directional information, which for motion on a line just means, specify “forward” or “backward; and you have what is known as the instantaneous velocity of the object whose motion is under consideration.

A lot of people say that the speed of an object is how far that object travels in a certain amount of time. No! That’s a distance. Speed is a rate. Speed is never how far, it is how fast. So if you want to relate it to a distance you might say something like, “Speed is what you multiply by a certain amount of time to determine how far an object would go in that amount of time if the speed stayed the same for that entire amount of time.” For instance, for a car with a speed of 25 mph, you could say that 25 mph is what you multiply by an hour to determine how far that car would go in an hour if it maintained a constant speed of 25 mph for the entire hour. But why explain it in terms of position? It is a rate. It is how fast the position of the object is changing. If you are standing on a street corner and a car passes you going 35 mph, I bet that if I asked you to estimate the speed of the car that you would get it right within 5 mph one way or the other. But if we were looking over a landscape on a day with unlimited visibility and I asked you to judge the distance to a mountain that was 35 miles away just by looking at it, I think the odds would be very much against you getting it right to within 5 miles. In a case like that, you have a better feel for “how fast” than you do for “how far.” So why define speed in terms of distance when you can just say that the speed of an object is how fast it is going?

4. You already have an intuitive understanding of what acceleration is. You have been in a car when it was speeding up. You know what it feels like to speed up gradually (small acceleration) and you know what it feels like to speed up rapidly (big, “pedalto the-metal,” acceleration).

All right, here comes the analysis. We have a start line ($x = 0$) and a positive direction (meaning the other way is the negative direction).

Consider a moving particle that is at position x_1 when the clock reads t_1 and at position x_2 when the clock reads t_2.

The displacement of the particle is, by definition, the change in position $\Delta x = x_2 - x_1$ of the particle. The average velocity $barv$ is, by definition,

$$\bar{v} = \frac{\triangle x}{\triangle t}$$

where $\Delta t = t_2 - t_1$ is the change in clock reading. Now the average velocity is not something that one would expect you to have an intuitive understanding for, as you do in the case of instantaneous velocity. The average velocity is not something that you can read off the speedometer, and frankly, it’s typically not as interesting as the actual (instantaneous) velocity, but it is easy to calculate and we can assign a meaning to it (albeit a hypothetical meaning). It is the constant velocity at which the particle would have to travel if it was to undergo the same displacement $\Delta x = x_2 - x_1$ in the same time $\Delta t = t_2 - t_1$ at constant velocity. The importance of the average velocity in this discussion lies in the fact that it facilitates the calculation of the instantaneous velocity.

Calculating the instantaneous velocity in the case of a constant velocity is easy. Looking at what we mean by average velocity, it is obvious that if the velocity isn’t changing, the instantaneous velocity is the average velocity. So, in the case of a constant velocity, to calculate the instantaneous velocity, all we have to do is calculate the average velocity, using any displacement with its corresponding time interval, that we want. Suppose we have position vs. time data on, for instance, a car traveling a straight path at 24 m/s.

Here’s some idealized fictitious data for just such a case.

Data Reading Number	Times [second]	Position [meter]
0	0.00	0.0

Data Reading Number	Times [second]	Position [meter]
1	0.100	2.30
2	1.0	23.0
3	10.0	230
4	100.0	2300

Remember, the speedometer of the car is always reading 24 m/s. (It should be clear that the car was already moving as it crossed the start line at time zero—think of time zero as the instant a stopwatch was started and the times in the table as stopwatch readings.) The position is the distance forward of the start line.

Note that for this special case of constant velocity, you get the same average velocity, the known value of constant speed, no matter what time interval you choose. For instance, if you choose the time interval from 1.00 seconds to 10.0 seconds:

$$\bar{v}=\frac{\triangle x}{\triangle t}\text{ (\textbf{Average velocity})}$$

$$\bar{v}=\frac{x_3-x_2}{t_3-t_2}$$

$$\bar{v}=\frac{230m-23.0m}{10.0s-1.0s}$$

$$\bar{v}=23.0\frac{m}{s}$$

and if you choose the time interval 0.100 seconds to 100.0 seconds:

$$\bar{v}=\frac{\triangle x}{\triangle t}$$

$$\bar{v}=\frac{x_4-x_1}{t_4-t_1}$$

$$\bar{v}=\frac{2300m-2.30m}{100s-0.100s}$$

$$\bar{v}=23.0\frac{m}{s}$$

The points that need emphasizing here are that, if the velocity is constant then the calculation of the average speed yields the instantaneous speed (the speedometer reading, the speed we have an intuitive feel for), and when the velocity is constant, it doesn't matter what time interval you use to calculate the average velocity; in particular, a small time interval works just as well as a big time interval.

So how do we calculate the instantaneous velocity of an object at some instant when the instantaneous velocity is continually changing? Let's consider a case in which the velocity is continually increasing. Here we show some idealized fictitious data (consistent with the way an object really moves) for just such a case.

Data Reading Number	Time since object was at start time [s]	Position (distance ahead of start line) [m]	Velocity (This is what we are trying to calculate. Here are correct answers.) [m/s]
0	0	0	10
1	1	14	18
2	1.01	14.1804	18.08
3	1.1	15.84	18.8
4	2	36	26

Data Reading Number	Time since object was at start time [s]	Position (distance ahead of start line) [m]	Velocity (This is what we are trying to calculate. Here are correct answers.) [m/s]
5	5	150	50

What I want to do with this fictitious data is to calculate an average velocity during a time interval that begins with t = 1s and compare the result with the actual velocity at time t = 1s. The plan is to do this repeatedly, with each time interval used being smaller than the previous one.

Average velocity from t = 1s to t = 5s:

$$\bar{v} = \frac{\triangle x}{\triangle t}$$

$$\bar{v} = \frac{x_5 - x_1}{t_5 - t_1}$$

$$\bar{v} = \frac{150m - 14m}{5s - 1s}$$

$$\bar{v} = 34.0\frac{m}{s}$$

Note that this value is quite a bit larger than the correct value of the instantaneous velocity at t =1s (namely 18 m/s). It does fall between the instantaneous velocity of 18 m/s at t = 1s and the instantaneous velocity of 50 m/s at t = 5 seconds. That makes sense since, during the time interval, the velocity takes on various values which for 1s < t < 5s are all greater than 18 m/s but less than 50 m/s.

For the next two time intervals in decreasing time interval order (calculations not shown):

- Average velocity from t=1 to t =2s: 22 m/s
- Average velocity from t=1 to t =1.1s: 18.4 m/s

And for the last time interval, we do show the calculation

$$\bar{v} = \frac{\triangle x}{\triangle t}$$

$$\bar{v} = \frac{x_2 - x_1}{t_2 - t_1}$$

$$\bar{v} = \frac{14.1804m - 14m}{1.01s - 1s}$$

$$\bar{v} = 18.04\frac{m}{s}$$

Here I copy all the results so that you can see the trend:

- Average velocity from t = 1 to t = 5 s: 34 m/s
- Average velocity from t = 1 to t = 2 s: 22 m/s
- Average velocity from t = 1 to t = 1.1 s: 18.4 m/s
- Average velocity from t = 1 to t = 1.01 s: 18.04 m/s

Every answer is bigger than the instantaneous velocity at t =1s (namely 18 m/s). Why? Because the distance traveled in the time interval under consideration is greater than it would have been if the object moved with a constant velocity of 18 m/s. Why? Because the object is speeding up, so, for most of the time interval the object is moving faster than 18 m/s, so, the average value during the time interval must be greater than 18 m/s. But notice that as the time interval (that starts at t = 1s) gets smaller and smaller, the average velocity over the time interval gets closer and closer to the actual instantaneous velocity at t = 1s. By induction, we conclude that if we were to use even smaller time intervals, as the time interval we chose to use was made smaller and smaller, the average velocity over that tiny time interval would get closer and closer to the instantaneous velocity, so that when the time interval got to be so small as to be virtually indistinguishable from zero, the value of the average velocity would get to be indistinguishable from the value of the instantaneous velocity. We write that:

$$v=\lim_{\triangle x\to 0}\frac{\triangle x}{\triangle t}$$

(Note the absence of the bar over the v. This v is the instantaneous velocity.) This expression for v is, by definition, the derivative of x with respect to t. The derivative of x with respect to t is $\frac{dx}{dt}$ written as which means that

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

Note that, as mentioned, $\frac{dx}{dt}$ is the derivative of x with respect to t. It is not some variable d times x all divided by d times t. It is to be read “dee ex by dee tee” or, better yet, “the derivative of x with respect to t.” Conceptually what it means is, starting at that value of time t at which you wish to find the velocity, let t change by a very small amount. Find the (also very small) amount by which x changes as a result of the change in t and divide the tiny change in x by the tiny change in t. Fortunately, given a function that provides the position x for any time t, we don’t have to go through all of that to get v, because the branch of mathematics known as differential calculus gives us a much easier way of determining the derivative of a function that can be expressed in equation form. A function, in this context, is an equation involving two variables, one of which is completely alone on the left side of the equation, the other of which, is in a mathematical expression on the right. The variable on the left is said to be a function of the variable on the right. Since we are currently dealing with how the position of a particle depends on time, we use x and t as the variables in the functions discussed in the remainder of this chapter. In the example of a function that follows, we use the symbols x_o, v_o, and a to represent constants:

$$x=x_o+v_ot+\frac{1}{2}at^2$$

The symbol t represents the reading of a running stopwatch. That reading changes so t is a variable. For each different value of t, we have a different value of x, so x is also a variable. Some folks think that any symbol whose value is not specified is a variable. Not so. If you know that the value of a symbol is fixed, then that symbol is a constant. You don’t have to know the value of the symbol for it to be a constant; you just have to know that it is fixed. This is the case for x_o, v_o, and a in equation above.

Acceleration

At this point you know how to calculate the rate of change of something. Let’s apply that knowledge to acceleration. Acceleration is the rate of change of velocity. If you are speeding up, then your acceleration is how fast you are speeding up. To get an average value of acceleration over a time interval $\triangle t$, we determine how much the velocity changes during that time interval and divide the change in velocity by the change in stopwatch reading. Calling the velocity change $\triangle v$, we have

$$\bar{a}=\frac{\triangle v}{\triangle t}$$

To get the acceleration at a particular time t we start the time interval at that time t and make it an infinitesimal time interval. That is:

$$a=\lim_{\triangle t\to 0}\frac{\triangle v}{\triangle t}$$

The right side is, of course, just the derivative ofv with respect to t:

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

7A: One-Dimensional Motion: The Constant Acceleration Equations

> **Note**
>
> The constant acceleration equations presented in this chapter are only applicable to situations in which the acceleration is constant. The most common mistake involving the constant acceleration equations is using them when the acceleration is changing.

In chapter 6 we established that, by definition of accelaration

$$a=\frac{dv}{dt} \tag{7A.1}$$

where a is the acceleration of an object moving along a straight line path, v is the velocity of the object and t, which stands for time, represents the reading of a stopwatch.

This equation is called a differential equation because that is the name that we give to equations involving derivatives. It's true for any function that gives a value of a for each value of t. An important special case is the case in which a is simply a constant. Here we derive some relations between the variables of motion for just that special case, the case in which a is constant.

$a=\frac{dv}{dt}$, with a stipulated to be a constant, can be considered to be a relationship between v and t. Solving it is equivalent to finding an expression for the function that gives the value of v for each value of t. So our goal is to find the function whose derivative $\frac{dv}{dt}$ is a constant. The derivative, with respect to t, of a constant times t is just the constant. Recalling that we want that constant to be a, let's try:

$$v=at$$

We'll call this our trial solution. Let's plug it into Equation 7A.1, and see if it works. The equation can be written:

$$a=\frac{d}{dt}v$$

and when we plug our trial solution $v=at$ into it we get:

$$a=\frac{d}{dt}(at)$$

$$a=a\frac{d}{dt}t$$

$$a=a\cdot 1$$

$$a=a$$

That is, our trial solution $v=at$ leads to an identity. Thus, our trial solution is indeed a solution to the Equation 7A.1. Let's see how this solution fits in with the linear motion situation under study.

In that situation, we have an object moving along a straight line and we have defined a one-dimensional coordinate system which can be depicted as

and consists of nothing more than an origin and a positive direction for the position variable x. We imagine that someone starts a stopwatch at a time that we define to be "time zero," $t=0$, a time that we also refer to as "the start of observations." Rather than limit ourselves to the special case of an object that is at rest at the origin at time zero, we assume that it could be moving with any velocity and be at any position on the line at time zero and define the constant x_0 to be the position of the object at time zero and the constant v_0 to be the velocity of the object at time zero.

Now the solution $v=at$ to the differential equation $a=\frac{dv}{dt}$ yields the value $v=0$ when $t=0$ (just plug $t=0$ into $v=at$ to see this). So, while $v=at$ does solve $a=\frac{dv}{dt}$, it does not meet the conditions at time zero, namely that $v=v_0$ at time zero. We can fix the initial condition problem easily enough by simply adding v_0 to the original solution yielding

$$v=v_0+at$$

This certainly makes it so that v evaluates to v_0 when $t=0$. But is it still a solution to $a=\frac{dv}{dt}$?

Let's try it. If $v=v_0+at$, then

$$a=\frac{dv}{dt}=\frac{d}{dt}(v_0+at)=\frac{d}{dt}v_0+\frac{d}{dt}(at)=0+a\frac{d}{dt}t=a$$

$v=v_0+at$, when substituted into Equation 7A.1 leads to an identity so $v=v_0+at$ is a solution to Equation 7A.1. What we have done is to take advantage of the fact that the derivative of a constant is zero, so if you add a constant to a function, you do not change the derivative of that function. The solution $v=v_0+at$ is not only a solution to the equation $a=\frac{dv}{dt}$ (with a stipulated to be a constant) but it is a solution to the whole problem since it also meets the initial value condition that $v=v_0$ at time zero. The solution:

$$v=v_0+at$$

is the first of a set of four constant acceleration equations to be developed in this chapter.

The other definition provided in the last section was:

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

which in words can be read as: The velocity of an object is the rate of change of the position of the object (since the derivative of the position with respect to time is the rate of change of the position). Substituting our recently-found expression for velocity yields

$$v_0+at=\frac{dx}{dt}$$

which can be written as:

$$\frac{dx}{dt}=v_0+at$$

We seek a function that gives a value of x for every value of t, whose derivative $\frac{dx}{dt}$ is the sum of terms v_0+at . Given the fact that the derivative of a sum will yield a sum of terms, namely the sum of the derivatives, let's try a function represented by the expression $x=x_1+x_2$. This works if $\frac{dx_1}{dt}$ is v_0 and $\frac{dx_2}{dt}$ is at. Let's focus on x_1 first. Recall that v_0 is a constant. Further recall that the derivative-with-respect-to-t of a constant times t, yields that constant. So check out $x_1=v_0t$. Sure enough, the derivative of v_0t with respect to t is v_0, the first term in equation above. So far we have

$$x=v_0t+x_2$$

Now let's work on x_2. We need $\frac{dx_2}{dt}$ to be at. Knowing that when we take the derivative of something with t^2 in it we get something with t in it we try $x_2=constant\cdot t^2$. The derivative of that is $2\cdot constant\cdot t$ which is equal to at if we choose $\frac{1}{2}a$ for the constant. If the constant is $\frac{1}{2}a$ then our trial solution for x_2 is $x_2=\frac{1}{2}at^2$. Plugging this in for x_2 in equation $x=v_0t+x_2$, yields:

$$x=v_0t+\frac{1}{2}at^2$$

Now we are in a situation similar to the one we were in with our first expression for $v(t)$. This expression for x does solve

$$\frac{dx}{dt} = v_0 + at$$

but it does not give x_0 when you plug 0 in for t. Again, we take advantage of the fact that you can add a constant to a function without changing the derivative of that function. This time we add the constant x_0 so

$$x = x_0 + v_0 t + \frac{1}{2} a t^2$$

This meets both our criteria: It solves equation $\frac{dx}{dt} = v_0 + at$, and it evaluates to x_0 when $t = 0$. We have arrived at the second equation in our set of four constant acceleration equations. The two that we have so far are:

$$x = x_0 + v_0 t + \frac{1}{2} a t^2$$

and

$$v = v_0 + at$$

These two are enough, but to simplify the solution of constant acceleration problems, we use algebra to come up with two more constant acceleration equations. Solving $v = v_0 + at$, for a yields $a = \frac{v - v_0}{t}$ and if you substitute that into $x = x_0 + v_0 t + \frac{1}{2} a t^2$ you quickly arrive at the third constant acceleration equation:

$$x = x_0 + \frac{V_0 + V}{2} t$$

Solving $v = v_0 + at$ for t yields $t = \frac{v - v_0}{a}$ and if you substitute that into $x = x_0 + v_0 t + \frac{1}{2} a t^2$ you quickly arrive at the final constant acceleration equation:

$$V^2 = V_0^2 + 2a(x - x_0)$$

For your convenience, we copy down the entire set of constant acceleration equations that you are expected to use in your solutions to problems involving constant acceleration:

$$x = x_0 + v_0 t + \frac{1}{2} a t^2$$

$$x = x_0 + \frac{V_0 + V}{2} t$$

$$v = v_0 + at$$

$$V^2 = V_0^2 + 2a(x - x_0)$$

8A: One-Dimensional Motion: Collision Type II

A common mistake one often sees in incorrect solutions to collision type two problems is using a different coordinate system for each of the two objects. It is tempting to use the position of object 1 at time 0 as the origin for the coordinate system for object 1 and the position of object 2 at time 0 as the origin for the coordinate system for object 2. This is a mistake. One should choose a single origin and use it for both particles. (One should also choose a single positive direction.)

We define a **Collision Type II problem** to be one in which two objects are moving along one and the same straight line and the questions are, "When and where are the two objects at one and the same position?" In some problems in this class of problems, the word "collision" can be taken literally, but the objects don't have to actually crash into each other for the problem to fall into the "Collision Type II" category. Furthermore, the restriction that both objects travel along one and the same line can be relaxed to cover for instance, a case in which two cars are traveling in adjacent lanes of a straight flat highway. The easiest way to make it clear what we mean here is to give you an example of a Collision Type II problem.

A car traveling along a straight flat highway is moving along at $41.0 m/s$ when it passes a police car standing on the side of the highway. $3.00 s$ after the speeder passes it, the police car begins to accelerate at a steady $5.00 m/s^2$. The speeder continues to travel at a steady $41.0 m/s$. (a) How long does it take for the police car to catch up with the speeder? (b) How far does the police car have to travel to catch up with the speeder? (c) How fast is the police car going when it catches up with the speeder?

Solution

The first step in any "Collision Type II" problem is to establish one and the same coordinate system for both objects. Since we are talking about one-dimensional motion, the coordinate system is just a single axis, so what we are really saying is that we have to establish a start line (the zero value for the position variable x) and a positive direction, and we have to use the same start line and positive direction for both objects.

A convenient start line in the case at hand is the initial position of the police car. Since both cars go in the same direction, the obvious choice for the positive direction is the direction in which both cars go.

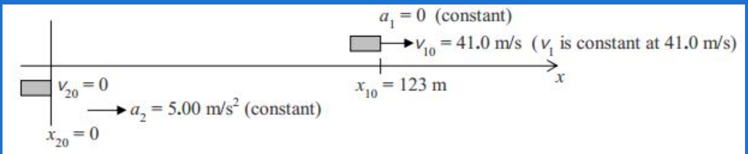
Next, we establish one and the same time variable t for both objects. More specifically, we establish what we mean by time zero, a time zero that applies to both objects. To choose time zero wisely, we actually have to think ahead to the next step in the problem, a step in which we use the constant acceleration equations to write an expression for the position of each object in terms of the time t. We want to choose a time zero, $t=0$, such that for all positive values of t, that is for all future times, the acceleration of each object is indeed constant. In the case at hand, the first choice that suggests itself to me is the instant at which the speeder first passes the police car. But if we "start the stopwatch" at that instant, we find that as time passes, the acceleration of the police car is not constant; rather, the police car has an acceleration of zero for three seconds and then, from then on, it has an acceleration of $5.00 m/s^2$. So we wouldn't be able to use a single constant acceleration equation to write down an expression for the position of the police car that would be valid for all times $t \geq 0$. Now the next instant that suggests itself to me as a candidate for time zero is the instant at which the police car starts accelerating. This turns out to be the right choice. From that instant on, both cars have constant acceleration (which is 0 in the case of the speeder and $5.00 m/s^2$ in the case of the police car). Furthermore, we have information on the conditions at that instant. For instance, based on our start line, we know that the position of the police car is zero, the velocity of the police car is zero, and the acceleration of the police car is $5.00 m/s^2$ at that instant. These become our "initial values" when we choose time zero to be the instant at which the police car starts accelerating. The one thing we don't know at that instant is the position of the speeder. But we do have enough information to determine the position of the speeder at the instant that we choose to call time zero. Our choice of time zero actually causes the given problem to break up into two problems: (1) Find the position of the speeder at time 0, and (2) Solve the "Collision Type II" problem.

The solution of the preliminary problem, finding the position of the speeder at time 0, is quite easy in this case because the speed of the speeder is constant. Thus the distance traveled is just the speed times the time.

$$d = v_s t' \quad d = (41.0 \frac{m}{s})(3.00 s) \quad d = 123 m$$

I used the symbol t' here to distinguish this time from the time t that we will use in the "Collision Type II" part of the problem. We can think of the problem as one that requires two stopwatches: One stopwatch, we start at the instant the speeder passes the police car. This one is used for the preliminary problem and we use the symbol t' to represent the value of its reading. The second one is used for the "Collision Type II" problem. It is started at the instant the police car starts accelerating and we will use the symbol t to represent the value of its reading. Note that $d = 123 m$ is the position of the speeder, relative to our established start line, at $t = 0$.

Now we are in a position to solve the "Collision Type II" problem. We begin by making a sketch of the situation. The sketch is a critical part of our solution. Sketches are used to define constants and variables. The required sketch for a "Collision Type II" problem is one that depicts the initial conditions.



We have defined the speeder's car to be car 1 and the police car to be car 2. From the constant acceleration equation (the one that gives the position of an object as a function of time) we have for the speeder:

$$x_1 = x_{10} + v_{10}t + \frac{1}{2}\overset{0}{\cancel{a_1}}t^2$$

$$x_1 = x_{10} + v_{10}t \qquad (8\text{-}1)$$

where we have incorporated the fact that a1 is zero. For the police car:

$$x_2 = \overset{0}{\cancel{x_{20}}} + \overset{0}{\cancel{v_{20}}}t + \frac{1}{2}a_2t^2$$

$$x_2 = \frac{1}{2}a_2t^2 \qquad (8\text{-}2)$$

where we have incorporated the fact that $x_20 = 0$ and the fact that $v_20 = 0$. Note that both equations have the same time variable t. The expression for x_1, gives the position of the speeder's car for any time t. You tell me the time t, and I can tell you where the speeder's car is at that time t just by plugging it into equation 8-1. Similarly, equation 8-2 for x_2 gives the position of the police car for any time t. Now there is one special time t, let's call it t^* when both cars are at the same position. The essential part of solving a "Collision Type II" problem is finding that that special time t^* which we refer to as the "collision time." Okay, now here comes the big central point for the "Collision Type II" problem. At the special time t^*,

$x_1 = x_2$

This small simple equation is the key to solving every "Collision Type II" problem. Substituting our expressions for x_1 and x_2 in equations 1 and 2 above, and designating the time as the collision time t^* we have

$$x_{10} + v_{10}t^* = \frac{1}{2}a_2t^{*2}$$

This yields a single equation in a single unknown, namely, the collision time t^*. We note that t^* appears to the second power. This means that the equation is a quadratic equation so we will probably (and in this case it turns out that we do) need the quadratic formula to solve it. Thus, we need to rearrange the terms as necessary to get the equation in the form of the standard quadratic equation $ax^2 + bx + c = 0$ (recognizing that our variable is t^* rather than x). Subtracting $x_{10} + v_{10}t^*$ from both sides, swapping sides, and reordering the terms yields

$$\frac{1}{2}a_2t^{*2} - v_{10}t^* - x_{10} = 0$$

which is the standard form for the quadratic equation. The quadratic formula $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ then yields

$$t^* = \frac{-(-v_{10}) \pm \sqrt{(-v_{10})^2 - 4(\frac{1}{2}a_2)(-x_{10})}}{2(\frac{1}{2}a_2)}$$

which simplifies ever so slightly to

$$t^* = \frac{v_{10} \pm \sqrt{v_{10}^2 + 2a_2x_{10}}}{a_2}$$

Substituting values with units yields:

$$t^* = \frac{41.0\frac{m}{s} \pm \sqrt{(41.0\frac{m}{s})^2 + 2(5.00\frac{m}{s^2})123m}}{5.00\frac{m}{s^2}}$$

Evaluation gives two results for t^*, namely $t^* = 19.0s$ and $t^* = -2.59s$. While the negative value is a valid solution to the mathematical equation, it corresponds to a time in the past and our expressions for the physical positions of the cars were written to be valid from time 0 on. Prior to time 0, the police car had a different acceleration than the $5.00\frac{m}{s^2}$ that we used in the expression for the position of the police car. Because we know that our equation is not valid for times earlier that $t = 0$ we must discard the negative solution. We are left with $t^* = 19.0s$ for the time when the police car catches up with the speeder. Once you find the "collision" time in a "Collision Type II" problem, the rest is easy. Referring back to the problem statement, we note that the collision time itself $t^* = 19.0s$ is the answer to part a, "How long does it take for the police car to catch up with the speeder?" Part b asks, "How far must the police car travel to catch up with the speeder?" At this point, to answer that, all we have to do is to substitute the collision time t^* into equation 8-2, the equation that gives the position of the police car at any time:

$$x_2 = \frac{1}{2}a_2 t^{*2}$$

$$x_2 = \frac{1}{2}(5.00\frac{m}{s^2})(19.0s)^2$$

$$x_2 = 902m$$

Finally, in part c of the problem statement we are asked to find the velocity of the police car when it catches up with the speeder. First we turn to the constant acceleration equations to get an expression for the velocity of the police car at as a function of time:

$$v_2 = v_{02} + a_2 t$$

The velocity of the police car at time zero is 0 yielding:

$$v_2 = a_2 t$$

To get the velocity of the police car at the "collision" time, we just have to evaluate this at $t = t^* = 19.0s$. This yields:

$$v_2 = (5.00\frac{m}{s^2})19.0s$$

$$v_2 = 95.0\frac{m}{s}$$

for the velocity of the police car when it catches up with the speeder.

9A: One-Dimensional Motion Graphs

Consider an object undergoing motion along a straight-line path, where the motion is characterized by a few consecutive time intervals during each of which the acceleration is constant but typically at a different constant value than it is for the adjacent specified time intervals. The acceleration undergoes abrupt changes in value at the end of each specified time interval. The abrupt change leads to a jump discontinuity in the Acceleration vs. Time Graph and a discontinuity in the slope (but not in the value) of the Velocity vs. Time Graph (thus, there is a "corner" or a "kink" in the trace of the Velocity vs. Time graph). The thing is, the trace of the Position vs. Time graph extends smoothly through those instants of time at which the acceleration changes. Even folks that get quite proficient at generating the graphs have a tendency to erroneously include a kink in the Position vs. Time graph at a point on the graph corresponding to an instant when the acceleration undergoes an abrupt change.

Your goals here all pertain to the motion of an object that moves along a straight line path at a constant acceleration during each of several time intervals but with an abrupt change in the value of the acceleration at the end of each time interval (except for the last one) to the new value of acceleration that pertains to the next time interval. Your goals for such motion are:

1. Given a description (in words) of the motion of the object; produce a graph of position vs. time, a graph of velocity vs. time, and a graph of acceleration vs. time, for that motion.
2. Given a graph of velocity vs. time, and the initial position of the object; produce a description of the motion, produce a graph of position vs. time, and produce a graph of acceleration vs. time.
3. Given a graph of acceleration vs. time, the initial position of the object, and the initial velocity of the object; produce a description of the motion, produce a graph of position vs. time, and produce a graph of velocity vs. time.

The following example is provided to more clearly communicate what is expected of you and what you have to do to meet those expectations:

A car moves along a straight stretch of road upon which a start line has been painted. At the start of observations, the car is already $225m$ ahead of the start line and is moving forward at a steady $15m/s$. The car continues to move forward at $15m/s$ for 5.0 seconds. Then it begins to speed up. It speeds up steadily, obtaining a speed of $35m/s$ after another 5.0 seconds. As soon as its speed gets up to $35m/s$, the car begins to slow down. It slows steadily, coming to rest after another 10.0 seconds. Sketch the graphs of position vs. time, velocity vs. time, and acceleration vs. time pertaining to the motion of the car during the period of time addressed in the description of the motion. Label the key values on your graphs of velocity vs. time and acceleration vs. time.

Solution

Okay, we are asked to draw three graphs, each of which has the time, the same "stopwatch readings" plotted along the horizontal axis. The first thing I do is to ask myself whether the plotted lines/curves are going to extend both above and below the time axis. This helps to determine how long to draw the axes. Reading the description of motion in the case at hand, it is evident that:

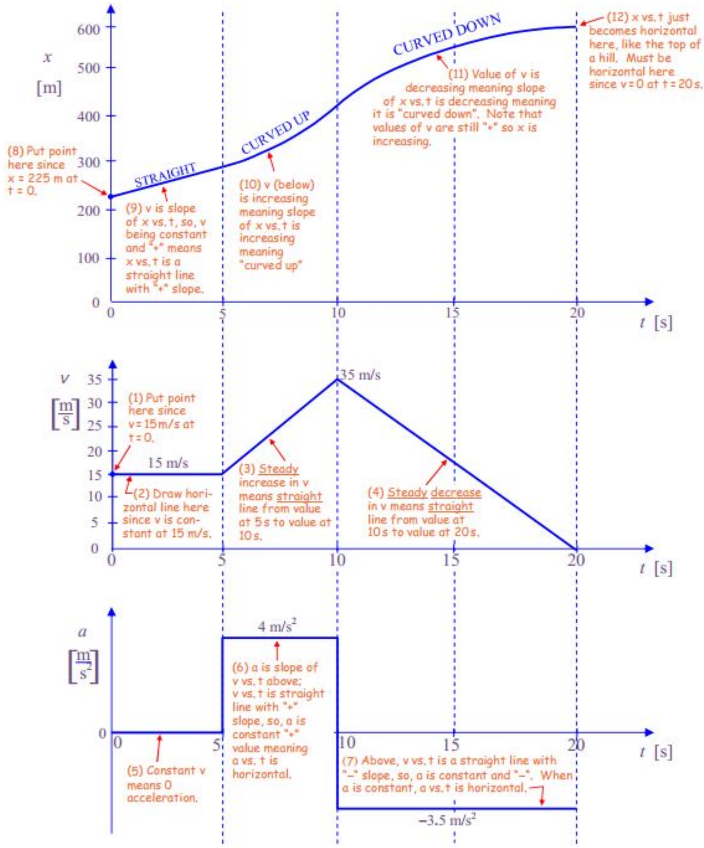
1. The car goes forward of the start line but it never goes behind the start line. So, the x vs. t graph will extend above the time axis (positive values of x) but not below it (negative values of x).
2. The car does take on positive values of velocity, but it never backs up, that is, it never takes on negative values of velocity. So, the v vs. t graph will extend above the time axis but not below it.
3. The car speeds up while it is moving forward (positive acceleration), and it slows down while it is moving forward (negative acceleration). So, the a vs. t graph will extend both above and below the time axis.

Next, I draw the axes, first for x vs. t, then directly below that set of axes, the axes for v vs. t, and finally, directly below that, the axes for a vs. t. Then I label the axes, both with the symbol used to represent the physical quantity being plotted along the axis and, in brackets, the units for that quantity.

Now I need to put some tick marks on the time axis. To do so, I have to go back to the question to find the relevant time intervals. I've already read the question twice and I'm getting tired of reading it over and over again. This time I'll take some notes:

At t = 0:	x = 225 m
	v = 15 m/s
0-5 s:	v = 15 m/s (constant)
5-10 s:	v increases steadily from 15 m/s to 35 m/s
10-20 s:	v decreases steadily from 35 m/s to 0 m/s

From my notes it is evident that the times run from 0 to 20 seconds and that labeling every 5 seconds would be convenient. So I put four tick marks on the time axis of x vs. t. I label the origin 0, 0 and label the tick marks on the time axis 5, 10, 15, and 20 respectively. Then I draw vertical dotted lines, extending my time axis tick marks up and down the page through all the graphs. They all share the same times and this helps me ensure that the graphs relate properly to each other. In the following diagram we have the axes and the graph. Except for the labeling of key values I have described my work in a series of notes. To follow my work, please read the numbered notes, in order, from 1 to 10.



The key values on the v vs. t graph are givens so the only "mystery," about the diagram above, that remains is, "How were the key values on a vs. t obtained?" Here are the answers:

On the time interval from $t = 5$ seconds to $t = 10$ seconds, the velocity changes from $15\frac{m}{s}$ to $35\frac{m}{s}$. Thus, on that time interval the acceleration is given by:

$$a = \frac{\triangle v}{\triangle t} = \frac{v_f - v_i}{t_f - t_i} = \frac{35\frac{m}{s} - 15\frac{m}{s}}{10s - 5s} = 4\frac{m}{s^2}$$

On the time interval from $t = 10$ seconds to $t = 20$ seconds, the velocity changes from $35\frac{m}{s}$ to $0\frac{m}{s}$. Thus, on that time interval the acceleration is given by:

$$a = \frac{\triangle v}{\triangle t} = \frac{v_f - v_i}{t_f - t_i} = \frac{0\frac{m}{s} - 35\frac{m}{s}}{20s - 10s} = -3.5\frac{m}{s^2}$$

10A: Constant Acceleration Problems in Two Dimensions

In solving problems involving constant acceleration in two dimensions, the most common mistake is probably mixing the x and y motion. One should do an analysis of the x motion and a separate analysis of the y motion. The only variable common to both the x and the y motion is the time. Note that if the initial velocity is in a direction that is along neither axis, one must first break up the initial velocity into its components.

In the last few chapters we have considered the motion of a particle that moves along a straight line with constant acceleration. In such a case, the velocity and the acceleration are always directed along one and the same line, the line on which the particle moves. Here we continue to restrict ourselves to cases involving constant acceleration (constant in both magnitude and direction) but lift the restriction that the velocity and the acceleration be directed along one and the same line. If the velocity of the particle at time zero is not collinear with the acceleration, then the velocity will never be collinear with the acceleration and the particle will move along a curved path. The curved path will be confined to the plane that contains both the initial velocity vector and the acceleration vector, and in that plane, the trajectory will be a parabola. (The trajectory is just the path of the particle.)

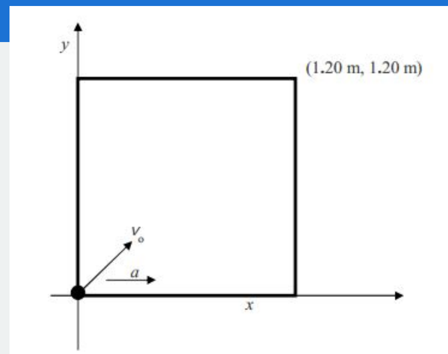
You are going to be responsible for dealing with two classes of problems involving constant acceleration in two dimensions:

1. Problems involving the motion of a single particle.
2. Collision Type II problems in two dimensions

We use sample problems to illustrate the concepts that you must understand in order to solve two-dimensional constant acceleration problems.

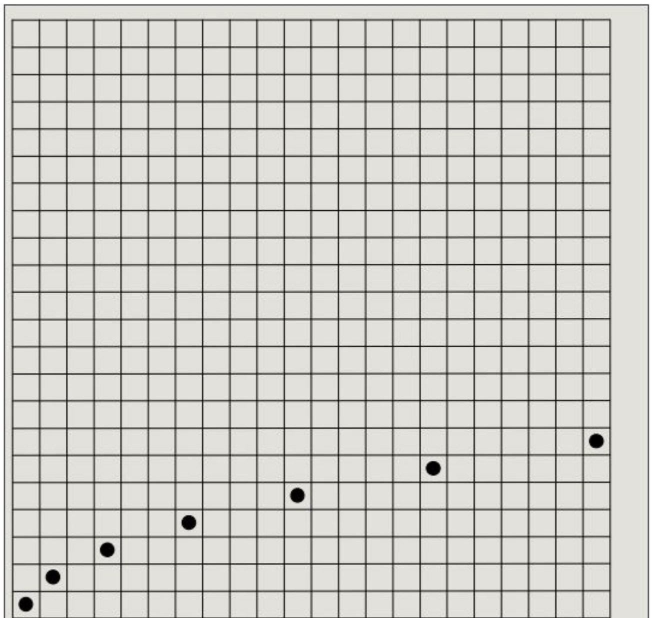
A horizontal square of edge length 1.20 m is situated on a Cartesian coordinate system such that one corner of the square is at the origin and the corner opposite that corner is at (1.20 m, 1.20 m). A particle is at the origin. The particle has an initial velocity of 2.20 m/s directed toward the corner of the square at (1.20 m, 1.20 m) and has a constant acceleration of $4.87 m/s^2$ in the $+x$ direction. Where does the particle hit the perimeter of the square?

Solution Let's start with a diagram.



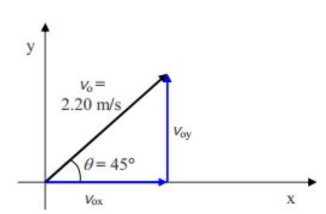
Now let's make some conceptual observations on the motion of the particle. Recall that the square is horizontal so we are looking down on it from above. It is clear that the particle hits the right side of the square because: It starts out with a velocity directed toward the far right corner. That initial velocity has an x component and a y component. The y component never changes because there is no acceleration in the y direction. The x component, however, continually increases. The particle is going rightward faster and faster. Thus, it will take less time to get to the right side of the square then it would without the acceleration and the particle will get to the right side of the square before it has time to get to the far side.

An important aside on the trajectory (path) of the particle: Consider an ordinary checker on a huge square checkerboard with squares of ordinary size (just a lot more of them then you find on a standard checkerboard). Suppose you start with the checker on the extreme left square of the end of the board nearest you (square 1) and every second, you move the checker right one square and forward one square. This would correspond to the checker moving toward the far right corner at constant velocity. Indeed you would be moving the checker along the diagonal. Now let's throw in some acceleration. Return the checker to square 1 and start moving it again. This time, each time you move the checker forward, you move it rightward one more square than you did on the previous move. So first you move it forward one square and rightward one square. Then you move it forward another square but rightward two more squares. Then forward one square and rightward three squares. And so on. With each passing second, the rightward move gets bigger. (That's what we mean when we say the rightward velocity is continually increasing.) So what would the path of the checker look like? Let's draw a picture.



As you can see, the checker moves on a curved path. Similarly, the path of the particle in the problem at hand is curved.

Now back to the problem at hand. The way to attack these two-dimensional constant acceleration problems is to treat the x motion and the y motion separately. The difficulty with that, in the case at hand, is that the initial velocity is neither along x nor along y but is indeed a mixture of both x motion and y motion. What we have to do is to separate it out into its x and y components. Let's proceed with that. Note that, by inspection, the angle that the velocity vector makes with the x axis is 45.0°.



$\cos\theta = \frac{v_{ox}}{v_o}$ $v_{ox} = v_o \cos\theta$ $v_{ox} = 2.20\frac{\text{m}}{\text{s}}\cos 45.0°$ $v_{ox} = 1.556\frac{\text{m}}{\text{s}}$	By inspection (because the angle is 45.0°): $v_{oy} = v_{ox}$ So: $v_{oy} = 1.556\frac{\text{m}}{\text{s}}$

Now we are ready to attack the x motion and the y motion separately. Before we do, let's consider our plan of attack. We have established, by means of conceptual reasoning, that the particle will hit the right side of the square. This means that we already have the answer to half of the question "Where does the particle hit the perimeter of the square?" It hits it at $x = 1.20m$ and y = ? . All we have to do is to find out the value of y. We have established that it is the x motion that determines the time it takes for the particle to hit the perimeter of the square. It hits the perimeter of the square at that instant in time when x achieves the value of $1.20m$. So our plan of attack is to use one or more of the x-motion constant acceleration equations to determine the time at which the particle hits the perimeter of the square and to plug that time into the appropriate y-motion constant acceleration equation to get the value of y at which the particle hits the side of the square. Let's go for it.

x motion

We start with the equation that relates position and time:

$$x = \overset{0}{\cancel{x_o}} + v_{ox}t + \frac{1}{2}a_x t^2 \quad \text{(We need to find the time that makes } x = 1.20 \text{ m.)}$$

The x component of the acceleration is the total acceleration, that is $a_x = a$. Thus,

$$x = v_{0x}t + \frac{1}{2}at^2$$

Recognizing that we are dealing with a quadratic equation we get it in the standard form of the quadratic equation.

$$\frac{1}{2}at^2 + v_{0x}t - x = 0$$

Now we apply the quadratic formula:

$$t = \frac{-v_{0x} \pm \sqrt{v_{0x}^2 - 4(\frac{1}{2}a)(-x)}}{2(\frac{1}{2}a_x)}$$

$$t = \frac{-v_{0x} \pm \sqrt{v_{0x}^2 + 2ax}}{a_x}$$

substituting values with units (and, in this step, doing no evaluation) we obtain:

$$t = \frac{-1.556\frac{m}{s} \pm \sqrt{(1.556\frac{m}{s})^2 + 2(4.87\frac{m}{s^2})1.20m}}{4.87\frac{m}{s^2}}$$

Evaluating this expression yields:

$$t = 0.4518s$$

and

$$t = -1.091s.$$

We are solving for a future time so we eliminate the negative result on the grounds that it is a time in the past. We have found that the particle arrives at the right side of the square at time $t = 0.4518s$. Now the question is, "What is the value of y at that time?"

y motion

Again we turn to the constant acceleration equation relating position to time, this time writing it in terms of the y variables:

$$y = \overset{0}{\cancel{y_o}} + v_{oy}t + \frac{1}{2}\overset{0}{\cancel{a_y}}t^2$$

We note that y_0 is zero because the particle is at the origin at time 0 and ay is zero because the acceleration is in the $+x$ direction meaning it has no y component. Rewriting this:

$$y = V_{0y}t$$

Substituting values with units,

$$y = 1.556\frac{m}{s}(0.4518s)$$

evaluating, and rounding to three significant figures yields:

$$y = 0.703m$$

Thus, the particle hits the perimeter of the square at

$$(1.20m, 0.703m)$$

Next, let's consider a 2-D Collision Type II problem. Solving a typical 2-D Collision Type II problem involves finding the trajectory of one of the particles, finding when the other particle crosses that trajectory, and establishing where the first particle is when the second particle crosses that trajectory. If the first particle is at the point on its own trajectory where the second particle crosses that trajectory then there is a collision. In the case of objects rather than particles, one often has to do some further reasoning to solve a 2-D Collision Type II problem. Such reasoning is illustrated in the following example involving a rocket.

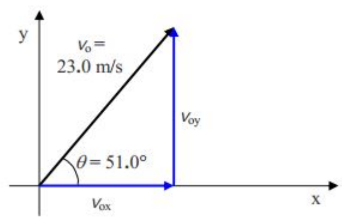
The positions of a particle and a thin (treat it as being as thin as a line) rocket of length $0.280m$ are specified by means of Cartesian coordinates. At time 0 the particle is at the origin and is moving on a horizontal surface at $23.0m/s$ at 51.0°. It has a constant acceleration of $2.43m/s^2$ in the $+y$ direction. At time 0 the rocket is at rest and it extends from $(-.280m, 50.0m)$ to $(0, 50.0m)$, but it has a constant acceleration in the $+x$ direction. What must the acceleration of the rocket be in order for the particle to hit the rocket?

Solution

Based on the description of the motion, the rocket travels on the horizontal surface along the line $y = 50.0m$. Let's figure out where and when the particle crosses this line. Then we'll calculate the acceleration that the rocket must have in order for the nose of the rocket to be at that point at that time and repeat for the tail of the rocket. Finally, we'll quote our answer as being any acceleration in between those two values.

When and where does the particle cross the line $y = 50.0m$?

We need to treat the particle's x motion and the y motion separately. Let's start by breaking up the initial velocity of the particle into its x and y components.



$\cos\theta = \frac{V_{ox}}{V_o}$	$\sin\theta = \frac{V_{oy}}{V_o}$
$V_{ox} = V_o \cos\theta$	$V_{oy} = V_o \sin\theta$
$V_{ox} = 23.0\frac{\text{m}}{\text{s}}\cos 51.0°$	$V_{oy} = 23.0\frac{\text{m}}{\text{s}}\sin 51.0°$
$V_{ox} = 14.47\frac{\text{m}}{\text{s}}$	$V_{oy} = 17.87\frac{\text{m}}{\text{s}}$

Now in this case, it is the y motion that determines when the particle crosses the trajectory of the rocket because it does so when $y = 50.0m$. So let's address the y motion first.

y motion of the particle

$$y = \overset{0}{\cancel{y_o}} + V_{oy}\,t + \frac{1}{2}a_y t^2$$

Note that we can't just assume that we can cross out yo but in this case the time zero position of the particle was given as $(0, 0)$ meaning that y_o is indeed zero for the case at hand. Now we solve for t:

$$y = V_{0y}t + \frac{1}{2}a_y t^2 \quad \frac{1}{2}a_y t^2 + V_{0y}t - y = 0 \quad t = \frac{-V_{0y} \pm \sqrt{V_{0y}^2 - 4(\frac{1}{2}a_y)(-y)}}{2(\frac{1}{2}a_y)}$$

$$t = \frac{-V_{0y} \pm \sqrt{V_{0y}^2 + 2a_y y}}{a_y} \quad t = \frac{-17.87\frac{m}{s} \pm \sqrt{(17.87\frac{m}{s})^2 + 2(2.43\frac{m}{s^2})50.0m}}{2.43\frac{m}{s^2}}$$

$t = 2.405s$ and $t = -17.11s$

Again, we throw out the negative solution because it represents an instant in the past and we want a future instant. Now we turn to the x motion to determine where the particle crosses the trajectory of the rocket.

x motion of the particle

Again we turn to the constant acceleration equation relating position to time, this time writing it in terms of the x variables:

$$x = \overset{0 \text{ (because the particle starts at the origin)}}{\cancel{x_o}} + v_{ox} t + \frac{1}{2} \overset{0 \text{ (because the acceleration is in the y direction)}}{\cancel{a_x}} t^2$$

$$x = v_{ox} t$$

$$x = 14.47 \frac{\text{m}}{\text{s}} (2.405 \text{ s})$$

$$\text{x} = 34.80 \text{ m.}$$

So the particle crosses the rocket's path at $(34.80m, 50.0m)$ at time $t = 2.450s$. Let's calculate the acceleration that the rocket would have to have in order for the nose of the rocket to be there at that instant. The rocket has x motion only. It is always on the line $y = 50.0m$.

Motion of the Nose of the Rocket

$$x'_n = \overset{0}{\cancel{x'_{on}}} + \overset{0}{\cancel{v'_{oxn}}} t + \frac{1}{2} a'_n t^2$$

where we use the subscript n for "nose" and a prime to indicate "rocket." We have crossed out x'_{on} because the nose of the rocket is at $(0, 50.0m)$ at time zero, and we have crossed out v'_{oxn} because the rocket is at rest at time zero.

$$x'_n = \frac{1}{2} a'_n t^2$$

Solving for a'_n yields:

$$a'_n = \frac{2x'_n}{t^2}$$

Now we just have to evaluate this expression at $t = 2.405s$, the instant when the particle crosses the trajectory of the rocket, and at $x'_n = x = 34.80m$, the value of x at which the particle crosses the trajectory of the rocket.

$$a'_n = \frac{2(34.80m)}{(2.405s)^2} \quad a'_n = 12.0 \frac{m}{s^2}$$

It should be emphasized that the n for "nose" is not there to imply that the nose of the rocket has a different acceleration than the tail; rather; the whole rocket must have the acceleration $a'_n = 12.0 \frac{m}{s^2}$ in order for the particle to hit the rocket in the nose. Now let's find the acceleration at'that the entire rocket must have in order for the particle to hit the rocket in the tail.

Motion of the Tail of the Rocket

$$x'_t = x'_{ot} + \overset{0}{\cancel{v'_{oxt}}} t + \frac{1}{2} a'_t t^2$$

where we use the subscript t for "tail" and a prime to indicate "rocket." We have crossed out v'_{oxt} because the rocket is at rest at time zero, but x'_{ot} is not zero because the tail of the rocket is at $(-.280, 50.0m)$ at time zero.

$$x'_t = x'_{ot} + \frac{1}{2} a'_t t^2$$

Solving for a'_t yields:

$$a'_t = \frac{2(x'_t - x'_{ot})}{t^2}$$

Evaluating at $t = 2.405s$ and $x'_t = x = 34.80m$ yields

$$a'_t = \frac{2(34.80m - (-0.280m))}{(2.405s)^2} \quad a'_t = 12.1 \frac{m}{s^2}$$

as the acceleration that the rocket must have in order for the particle to hit the tail of the rocket.

Thus: The acceleration of the rocket must be somewhere between $12.0 \frac{m}{s^2}$ and $12.1 \frac{m}{s^2}$, inclusive, in order for the rocket to be hit by the particle.

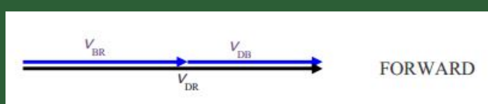
11A: Relative Velocity

Vectors add like vectors, not like numbers. Except in that very special case in which the vectors you are adding lie along one and the same line, you can't just add the magnitudes of the vectors.

Imagine that you have a dart gun with a **muzzle velocity** of $45 mph$. Further imagine that you are on a bus traveling along a straight highway at $55 mph$ and that you point the gun so that the barrel is level and pointing directly forward, toward the front of the bus. Assuming no recoil, as it leaves the muzzle of the gun, how fast is the dart traveling relative to the road? That's right! $100 mph$. The dart is already traveling forward at $55 mph$ relative to the road just because it is on a bus that is moving at $55 mph$ relative to the road. Add to that the velocity of $45 mph$ that it acquires as a result of the firing of the gun and you get the total velocity of the dart relative to the road. This problem is an example of a class of vector addition problems that come under the heading of "Relative Velocity." It is a particularly easy vector addition problem because both velocity vectors are in the same direction. The only challenge is the vector addition diagram, since the resultant is right on top of the other two. We displace it to one side a little bit in the diagram below so that you can see all the vectors. Defining

$\vec{V}_{BR}$ to be the velocity of the bus relative to the road,

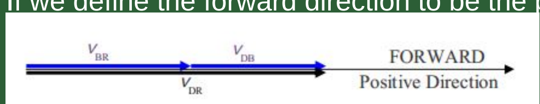
- $\vec{V}_{DB}$ to be the velocity of the dart relative to the bus, and
- $\vec{V}_{DR}$ to be the velocity of the dart relative to the road; we have



The vector addition problem this illustrates is

$$\vec{V}_{DR} = \vec{V}_{BR} + \vec{V}_{DB}$$

If we define the forward direction to be the positive direction,



then, because the vectors we are adding are both in the same direction, we are indeed dealing with that very special case in which the magnitude of the resultant is just the sum of the magnitudes of the vectors we are adding:

$$\vec{V}_{DR} = \vec{V}_{BR} + \vec{V}_{DB} \quad V_{DR} = V_{BR} + V_{DB} \quad V_{DR} = 55 mph + 45 mph \quad V_{DR} = 100 mph$$

$\vec{V}_{DR} = 100 mph$ in the direction in which the bus is traveling

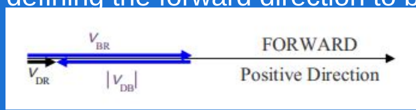
You already know all the concepts you need to know to solve relative velocity problems (you know what velocity is and you know how to do vector addition) so the best we can do here is to provide you with some more worked examples. We've just addressed the easiest kind of relative velocity problem, the kind in which all the velocities are in one and the same direction. The second easiest kind is the kind in which the two velocities to be added are in opposite directions.

A bus is traveling along a straight highway at a constant $55mph$. A person sitting at rest on the bus fires a dart gun that has a muzzle velocity of $45mph$ straight backward, (toward the back of the bus). Find the velocity of the dart, relative to the road, as it leaves the gun.

Again defining:

- $\vec{V}_{BR}$ to be the velocity of the bus relative to the road,
- $\vec{V}_{DB}$ to be the velocity of the dart relative to the bus, and
- $\vec{V}_{DR}$ to be the velocity of the dart relative to the road, and

defining the forward direction to be the positive direction; we have



$$\vec{V}_{DR} = \vec{V}_{BR} + \vec{V}_{DB}$$

$$V_{DR} = V_{BR} - |V_{DB}|$$

$V_{DR} = 55mph - 45mph$ $\quad V_{DR} = 10mph$ $\quad \vec{V}_{DR} = 10mph$ in the direction in which the bus is traveling

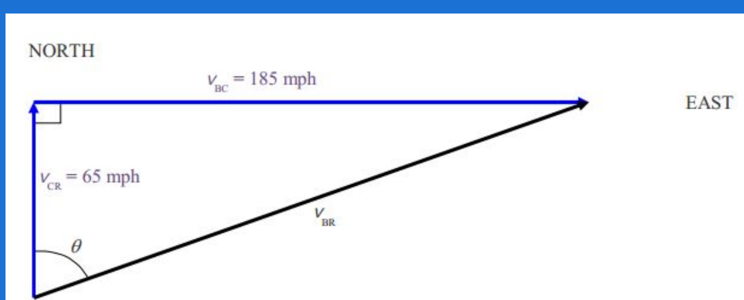
It would be odd looking at that dart from the side of the road. Relative to you it would still be moving in the direction that the bus is traveling, tail first, at $10mph$.

The next easiest kind of vector addition problem is the kind in which the vectors to be added are at right angles to each other. Let's consider a relative velocity problem involving that kind of vector addition problem.

A boy sitting in a car that is traveling due north at $65mph$ aims a BB gun (a gun which uses a compressed gas to fire a small metal or plastic ball called a BB), with a muzzle velocity of $185mph$, due east, and pulls the trigger. Recoil (the backward movement of the gun resulting from the firing of the gun) is negligible. In what compass direction does the BB go?

Defining

- $\vec{V}_{CR}$ to be the velocity of the car relative to the road,
- $\vec{V}_{BC}$ to be the velocity of the BB relative to the car, and
- $\vec{V}_{BR}$ to be the velocity of the BB relative to the road; we have



$$tan\theta = \frac{V_{BC}}{V_{CR}} \quad \theta = tan^{-1}\frac{V_{BC}}{V_{CR}}$$

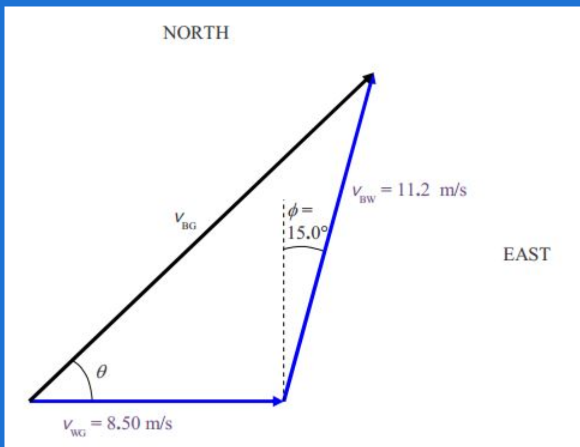
$$\theta = tan^{-1}\frac{185mph}{65mph} \quad \theta = 70.6^\circ$$

The BB travels in the direction for which the compass heading is 70.6°.

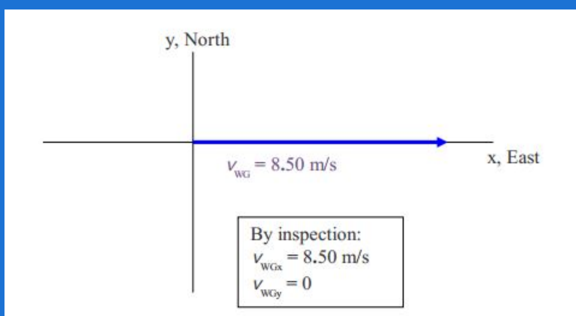
A boat is traveling across a river that flows due east at $8.50 m/s$. The compass heading of the boat is 15.0°. Relative to the water, the boat is traveling straight forward (in the direction in which the boat is pointing) at $11.2 m/s$. How fast and which way is the boat moving relative to the banks of the river?

Okay, here we have a situation in which the boat is being carried downstream by the movement of the water at the same time that it is moving relative to the water. Note the given information means that if the water was dead still, the boat would be going 11.2 m/s at 15.0° East of North. The water, however, is not still. Defining

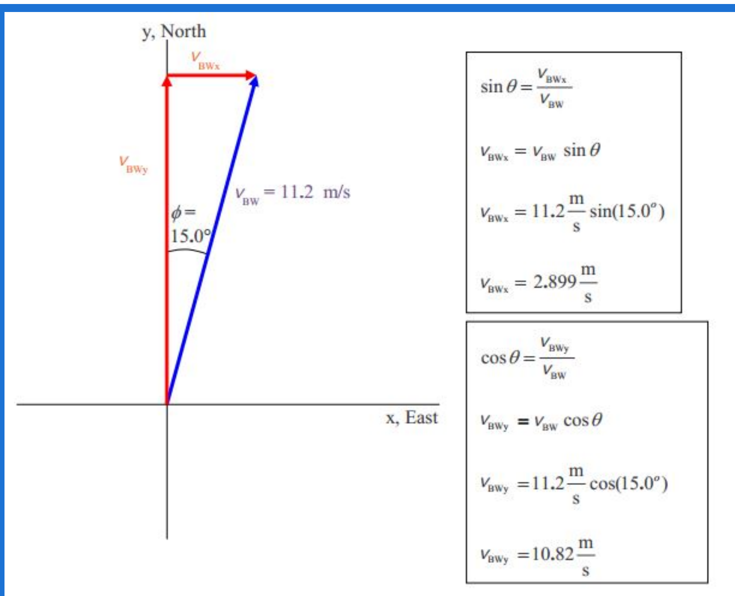
- $\vec{V}_{WG}$ to be the velocity of the water relative to the ground,
- $\vec{V}_{BW}$ to be the velocity of the boat relative to the water, and
- $\vec{V}_{BG}$ to be the velocity of the boat relative to the ground; we have



Solving this problem is just a matter of following the vector addition recipe. First we define $+x$ to be eastward and $+y$ to be northward. Then we draw the vector addition diagram for $\vec{V}_{WG}$. Breaking it up into components is trivial since it lies along the x-axis:



Breaking $\vec{V}_{BW}$ does involve a little bit of work:



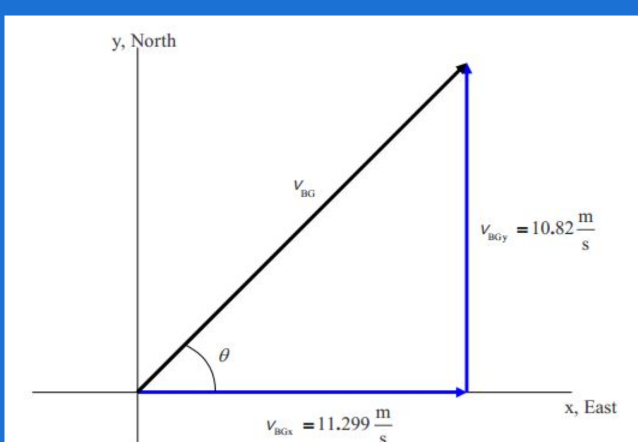
Now we add the x components to get the x-component of the resultant $V_{BGx} = V_{WGx} + V_{BWx}$

$$V_{BGx} = 8.50 \frac{m}{s} + 2.899 \frac{m}{s} \quad V_{BGx} = 11.299 \frac{m}{s}$$

and we add the y components to get the y-component of the resultant: $V_{BGy} = V_{WGy} + V_{BWy}$

$$V_{BGy} = 0 \frac{m}{s} + 10.82 \frac{m}{s} \quad V_{BGy} = 10.82 \frac{m}{s}$$

Now we have both components of the velocity of the boat relative to the ground. We need to draw the vector component diagram for $\vec{V}_{BG}$ to determine the direction and magnitude of the velocity of the boat relative to the ground.



We then use the Pythagorean Theorem to get the magnitude of the velocity of the boat relative to the ground,

$$\vec{V}_{BG} = \sqrt{V_{BGx}^2 + V_{BGy}^2} \quad \vec{V}_{BG} = \sqrt{(11.299 m/s)^2 + (10.82 m/s)^2} \quad \vec{V}_{BG} = 15.64 m/s$$

and the definition of the tangent to determine the direction of $\vec{V}_{BG}$: $tan\theta = \frac{V_{BGy}}{V_{BGx}} \quad \theta = tan^{-1}\frac{V_{BGy}}{V_{BGx}}$

$\theta = tan^{-1}\frac{10.82m/s}{11.299m/s} \quad \theta = 43.8^{\circ}$ Hence, $\vec{V}_{BG} = 15.64m/s$ at 43.8° North of East.

12A: Gravitational Force Near the Surface of the Earth, First Brush with Newton's 2nd Law

Some folks think that every object near the surface of the earth has an acceleration of $9.8m/s^2$ downward relative to the surface of the earth. That just isn't so. In fact, as I look around the room in which I write this sentence, all the objects I see have zero acceleration relative to the surface of the earth. Only when it is in freefall, that is, only when nothing is touching or pushing or pulling on the object except for the gravitational field of the earth, will an object experience an acceleration of $9.8m/s^2$ downward relative to the surface of the earth.

13A: Freefall, a.k.a. Projectile Motion

> **Note**
>
> The constant acceleration equations apply from the first instant in time after the projectile leaves the launcher to the last instant in time before the projectile hits something, such as the ground. Once the projectile makes contact with the ground, the ground exerts a huge force on the projectile causing a drastic change in the acceleration of the projectile over a very short period of time until, in the case of a projectile that doesn't bounce, both the acceleration and the velocity become zero. To take this zero value of velocity and plug it into constant acceleration equations that are devoid of post-ground-contact acceleration information is a big mistake. In fact, at that last instant in time during which the constant acceleration equations still apply, when the projectile is at ground level but has not yet made contact with the ground, (assuming that ground level is the lowest elevation achieved by the projectile) the magnitude of the velocity of the projectile is at its biggest value, as far from zero as it ever gets!

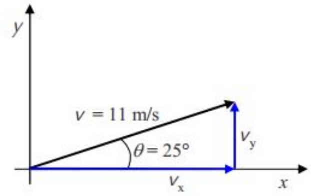
Consider an object in freefall with a non-zero initial velocity directed either horizontally forward; or both forward and vertically (either upward or downward). The object will move forward, and upward or downward—perhaps upward and then downward—while continuing to move forward. In all cases of freefall, the motion of the object (typically referred to as the projectile when freefall is under consideration) all takes place within a single vertical plane. We can define that plane to be the x-y plane by defining the forward direction to be the x direction and the upward direction to be the y direction.

One of the interesting things about projectile motion is that the horizontal motion is independent of the vertical motion. Recall that in freefall, an object continually experiences a downward acceleration of $9.80\frac{m}{s^2}$ but has no horizontal acceleration. This means that if you fire a projectile so that it is approaching a wall at a certain speed, it will continue to get closer to the wall at that speed, independently of whether it is also moving upward and/or downward as it approaches the wall. An interesting consequence of the independence of the vertical and horizontal motion is the fact that, neglecting air resistance, if you fire a bullet horizontally from, say, shoulder height, over flat level ground, and at the instant the bullet emerges from the gun, you drop a second bullet from the same height, the two bullets will hit the ground at the same time. The forward motion of the fired bullet has no effect on its vertical motion.

> The most common mistake that folks make in solving projectile motion problems is combining the x and y motion in one standard constant-acceleration equation. Don't do that. Treat the x-motion and the y-motion separately.

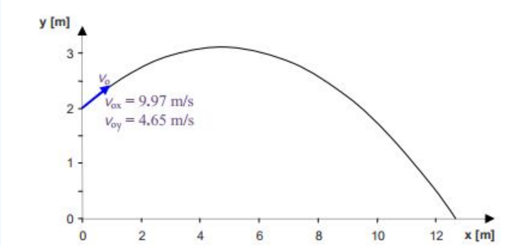
In solving projectile motion problems, we take advantage of the independence of the horizontal (x) motion and the vertical (y) motion by treating them separately. The one thing that is common to both the x motion and the y motion is the time. The key to the solution of many projectile motion problems is finding the total time of "flight." For example, consider the following example.

> **Example $13A.1$**
>
> A projectile is launched with a velocity of $11m/s$ at an angle of 28° above the horizontal over flat level ground from a height of $2.0m$ above ground level. How far forward does it go before hitting the ground? (Assume that air resistance is negligible.)
>
> **Solution**
>
> Before getting started, we better clearly establish what we are being asked to find. We define the forward direction as the x direction so what we are looking for is a value of x. More specifically, we are looking for the distance, measured along the ground, from that point on the ground directly below the point at which the projectile leaves the launcher, to the point on the ground where the projectile hits. This distance is known as the range of the projectile. It is also known as the range of the launcher for the given angle of launch and the downrange distance traveled by the projectile.
>
> Okay, now that we know what we're solving for, let's get started. An initial velocity of $11m/s$ at 28° above the horizontal, eh? Uh oh! We've got a dilemma. The key to solving projectile motion problems is to treat the x motion and the y motion separately. But we are given an initial velocity vo which is a mix of the two of them. We have no choice but to break up the initial velocity into its x and y components.



$\frac{v_x}{v} = \cos\theta$	$\frac{v_y}{v} = \sin\theta$
$v_x = v\cos\theta$	$v_y = v\sin\theta$
$v_x = 11\frac{\text{m}}{\text{s}}\cos 25°$	$v_y = 11\frac{\text{m}}{\text{s}}\sin 25°$
$v_x = 9.97\frac{\text{m}}{\text{s}}$	$v_y = 4.65\frac{\text{m}}{\text{s}}$

Now we're ready to get started. We'll begin with a sketch which defines our coordinate system, thus establishing the origin and the positive directions for x and y.



Recall that in projectile motion problems, we treat the x and y motion separately. Let's start with the x motion. It is the easier part because there is no acceleration.

x motion

$$x = \cancelto{0}{x_o} + v_{ox}\, t + \frac{1}{2}\cancelto{0}{a_x} t^2$$

$$x = v_{ox}t \tag{13A.1}$$

Note that for the x-motion, we start with the constant acceleration equation that gives the position as a function of time. (Imagine having started a stopwatch at the instant the projectile lost contact with the launcher. The time variable t represents the stopwatch reading.) As you can see, because the acceleration in the x direction is zero, the equation quickly simplifies to $x = V_{0x}t$. We are "stuck" here because we have two unknowns, x and t, and only one equation. It's time to turn to the y motion.

It should be evident that it is the y motion that yields the time, the projectile starts off at a known elevation $(y = 2.0m)$ and the projectile motion ends when the projectile reaches another known elevation, namely, $y = 0$.

y-motion

$$y = y_0 + V_{0y}t + \frac{1}{2}a_y t^2 \tag{13A.2}$$

This equation tells us that the y value at any time t is the initial y value plus some other terms that depend on t. It's valid for any time t, starting at the launch time $t = 0$, while the object is in projectile motion. In particular, it is applicable to that special time t, the last instant before the object makes contact with the ground, that instant that we are most interested in, the time when $y = 0$. What we can do, is to plug 0 in for y, and solve for that special time t that, when plugged into Equation 13A.2, makes y be 0. When we rewrite Equation 13A.2 with y set to 0, the symbol t takes on a new meaning. Instead of being a variable, it becomes a special time, the time that makes the y in the actual Equation 13A.2 ($y = y = y_0 + V_{0y}t + \frac{1}{2}a_y t^2$) zero.

$$0 = y_0 + V_{0y}t_* + \frac{1}{2}a_y t_*^2 \tag{13A.3}$$

To emphasize that the time in Equation 13A.3 is a particular instant in time rather than the variable time since launch, I have written it as t_* to be read "t star." Everything in Equation 13A.3 is a given except t_* so we can solve Equation 13A.3 for t_*. Recognizing that Equation 13A.3 is a quadratic equation in t_* we first rewrite it in the form of the standard quadratic equation $ax + bx^2 + c = 0$. This yields:

$$\frac{1}{2}a_y t_*^2 + V_{0y}t_* + y_0 = 0$$

Then we use the quadratic formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ which for the case at hand appears as:

$$t_* = \frac{-V_{0y} \pm \sqrt{V_{0y}^2 - 4(\frac{1}{2}a_y)y_0}}{2(\frac{1}{2}a_y)}$$

which simplifies to

$$t_* = \frac{-V_{0y} \pm \sqrt{V_{0y}^2 - 2a_y y_0}}{a_y}$$

Substituting values with units yields:

$$t_* = \frac{-4.65\frac{m}{s} \pm \sqrt{(-4.65\frac{m}{s})^2 - 2(-9.80\frac{m}{s^2})2.0m}}{-9.80\frac{m}{s^2}}$$

which evaluates to
$t_* = -0.321s$ and $t_* = 1.27s$

We discard the negative answer because we know that the projectile hits the ground after the launch, not before the launch.

Recall that t_* is the stopwatch reading when the projectile hits the ground. Note that the whole time it has been moving up and down, the projectile has been moving forward in accord with Equation 13A.1, $x = V_{0x}t$. At this point, all we have to do is plug $t_* = 1.27s$ into Equation 13A.1 and evaluate:

$$\begin{aligned} x &= V_{0x}t_* \\ &= 9.97\frac{m}{s}(1.27s) \\ &= 13m \end{aligned}$$

This is the answer. The projectile travels $13m$ forward before it hits the ground.

14A: Newton's Laws #1: Using Free Body Diagrams

> *If you throw a rock upward in the presence of another person, and you ask that other person what keeps the rock going upward, after it leaves your hand but before it reaches its greatest height, that person may incorrectly tell you that the force of the person's hand keeps it going. This illustrates the common misconception that force is something that is given to the rock by the hand and that the rock "has" while it is in the air. It is not. A force is all about something that is being done to an object. We have defined a force to be an ongoing push or a pull. It is something that an object can be a victim to, it is never something that an object has. While the force is acting on the object, the motion of the object is consistent with the fact that the force is acting on the object. Once the force is no longer acting on the object, there is no such force, and the motion of the object is consistent with the fact that the force is absent. (As revealed in this chapter, the correct answer to the question about what keeps the rock going upward, is, "Nothing." Continuing to go upward is what it does all by itself if it is already going upward. You don't need anything to make it keep doing that. In fact, the only reason the rock does not continue to go upward forever is because there is a downward force on it. When there is a downward force and only a downward force on an object, that object is experiencing a downward acceleration. This means that the upward-moving rock slows down, then reverses its direction of motion and moves downward ever faster.)*

Imagine that the stars are fixed in space so that the distance between one star and another never changes. (They are not fixed. The stars are moving relative to each other.) Now imagine that you create a Cartesian coordinate system; a set of three mutually orthogonal axes that you label x, y, and z. Your Cartesian coordinate system is a reference frame. Now as long as your reference frame is not rotating and is either fixed or moving at a constant velocity relative to the (fictitious) fixed stars, then your reference frame is an inertial reference frame. Note that velocity has both magnitude and direction and when we stipulate that the velocity of your reference frame must be constant in order for it to be an inertial reference frame, we aren't just saying that the magnitude has to be constant but that the direction has to be constant as well. The magnitude of the velocity is the speed. So, for the magnitude of the velocity to be constant, the speed must be constant. For the direction to be constant, the reference frame must move along a straight line path. So an inertial reference frame is one that is either fixed or moving at a constant speed along a straight line path, relative to the (fictitious) fixed stars.

The concept of an inertial reference frame is important in the study of physics because it is in inertial reference frames that the laws of motion known as Newton's Laws of Motion apply. Here are Newton's three laws of motion, observed to be adhered to by any particle of matter in an inertial reference frame:

I. **If there is no net force acting on a particle, then the velocity of that particle is not changing.**
II. **If there is a net force on a particle, then that particle is experiencing an acceleration that is directly proportional to the force, with the constant of proportionality being the reciprocal of the mass of the particle.**
III. **Anytime one object is exerting a force on a second object, the second object is exerting an equal but opposite force back on the first object.**

Discussion of Newton's 1st Law

Despite the name, it was actually Galileo that came up with the first law. He let a ball roll down a ramp with another ramp facing the other way in front of it so that, after it rolled down one ramp, the ball would roll up the other. He noted that the ball rolled up the second ramp, slowing steadily until it reached the same elevation as the one from which the ball was originally released from\ rest. He then repeatedly reduced the angle that the second ramp made with the horizontal and released the ball from rest from the original position for each new inclination of the second ramp. The smaller the angle, the more slowly the speed of the ball was reduced on the way up the second ramp and the farther it had to travel along the surface of the second ramp before arriving at its starting elevation. When he finally set the angle to zero, the ball did not appear to slow down at all on the second ramp. He didn't have an infinitely long ramp, but he induced that if he did, with the second ramp horizontal, the ball would keep on rolling forever, never slowing down because no matter how far it rolled, it would never gain any elevation, so it would never get up to the starting elevation. His conclusion was that if an object was moving, then if nothing interfered with its motion it would keep on moving at the same speed in the same direction. So what keeps it going? The answer is "nothing." That is the whole point. An object doesn't need anything to keep it going. If it is already moving, going at a constant velocity is what it does as long as there is no net force acting on it. In fact, it takes a force to change the velocity of an object.

It's not hard to see why it took a huge chunk of human history for someone to realize that if there is no net force on a moving object, it will keep moving at a constant velocity, because the thing is, where we live, on the surface of the Earth, there is inevitably a net force on a moving object. You throw something up and the Earth pulls downward on it the whole time the object is in flight. It's not going to keep traveling in a straight line upward, not with the Earth pulling on it. Even if you try sliding something across the smooth surface of a frozen pond where the downward pull of the Earth's gravitational field is cancelled by the ice pressing up on the object, you find that the object slows down because of a frictional force pushing on the object in the direction opposite that of the object's velocity and indeed a force of air resistance doing the same thing. In the presence of these ubiquitous forces, it took humankind a long time to realize that if there were no forces, an object in motion would stay in motion along a straight line path, at constant speed, and that an object at rest would stay at rest.

Discussion of Newton's 2nd Law

Galileo induced something else of interest from his ball-on-the-ramp experiments by focusing his attention on the first ramp discussed above. Observation of a ball released from rest revealed to him that the ball steadily sped up on the way down the ramp. Try it. As long as you don't make the ramp too steep, you can see that the ball doesn't just roll down the ramp at some fixed speed, it accelerates the whole way down. Galileo further noted that the steeper the ramp was, the faster the ball would speed up on the way down. He did trial after trial, starting with a slightly inclined plane and gradually making it steeper and steeper. Each time he made it steeper, the ball would, on the way down the ramp, speed up faster than it did before, until the ramp got so steep that he could no longer see that it was speeding up on the way down the ramp—it was simply happening too fast to be observed. But Galileo induced that, as he continued to make the ramp steeper, the same thing was happening. That is that the ball's speed was still increasing on its way down the ramp and the greater the angle, the faster the ball would speed up. In fact, he induced that if he increased the steepness to the ultimate angle, 90°, that the ball would speed up the whole way down the ramp faster than it would at any smaller angle but that it would still speed up on the way down. Now, when the ramp is tilted at 90°, the ball is actually falling as opposed to rolling down the ramp, so Galileo's conclusion was that when you drop an object (for which air resistance is negligible), what happens is that the object speeds up the whole way down, until it hits the Earth.

Galileo thus did quite a bit to set the stage for Sir Isaac Newton, who was born the same year that Galileo died. It was Newton who recognized the relationship between force and motion. He is the one that realized that the link was between force and acceleration, more specifically, that whenever an object is experiencing a net force, that object is experiencing an acceleration in the same direction as the force. Now, some objects are more sensitive to force than other objects—we can say that every object comes with its own sensitivity factor such that the greater the sensitivity factor, the greater the acceleration of the object for a given force. The sensitivity factor is the reciprocal of the mass of the object, so we can write that

$$\vec{a} = \frac{1}{m} \sum \vec{F} \tag{14A.2}$$

where $\vec{a}$ is the acceleration of the object, m is the mass of the object, and $\Sigma \vec{F}$ is the vector sum of all the forces acting on the object, that is to say that $\sum \vec{F}$ is the net force acting on the object.

Discussion of Newton's Third Law

In realizing that whenever one object is in the act of exerting a force on a second object, the second object is always in the act of exerting an equal and opposite force back on the first object, Newton was recognizing an aspect of nature that, on the surface, seems quite simple and straightforward, but quickly leads to conclusions that, however correct they may be, and indeed they are correct, are quite counterintuitive. Newton's 3rd law is a statement of the fact that any force whatsoever is just one half of an interaction where an interaction in this sense is the mutual pushing or pulling that quite often occurs when one object is in the vicinity of another.

In some cases, where the effect is obvious, the validity of Newton's third law is fairly evident. For instance if two people who have the same mass are on roller skates and are facing each other and one pushes the other, we see that both skaters go rolling backward, away from each other. It might at first be hard to accept the fact that the second skater is pushing back on the hands of the first skater, but we can tell that the skater that we think of as the pusher, must also be a "pushee," because we can see that she experiences a backward acceleration. In fact, while the pushing is taking place, the force exerted on her must be just as great as the force she exerts on the other skater because we see that her final backward speed is just as great as that of the other (samemass) skater.

But how about those cases where the effect of at least one of the forces in the interaction pair is not at all evident? Suppose for instance that you have a broom leaning up against a slippery wall. Aside from our knowledge of Newton's laws, how can we convince ourselves that the broom is pressing against the wall, that is, that the broom is continually exerting a force on the wall; and; how can we convince ourselves that the wall is exerting a force back on the broom? One way to convince yourself is to let your hand play the role of the wall. Move the broom and put your hand in the place of the wall so that the broom is leaning against the palm of your hand at the same angle that it was against the wall with the palm of your hand facing directly toward the tip of the handle. You can feel the tip of the handle pressing against the palm of your hand. In fact, you can see the indentation that the tip of the broom handle makes in your hand. You can feel the force of the broom handle on your hand and you can induce that when the wall is where your hand is, relative to the broom, the broom handle must be pressing on the wall with the same force.

How about this business of the wall exerting a force on (pushing on) the tip of the broom handle? Again, with your hand playing the role of the wall, quickly move your hand out of the way. The broom, of course, falls down. Before moving your hand, you must have been applying a force on the broom or else the broom would have fallen down then. You might argue that your hand wasn't necessarily applying a force but rather that your hand was just "in the way." Well I'm here to tell you that "being in the way" is all about applying a force. When the broom is leaning up against the wall, the fact that the broom does not fall over means that the wall is exerting a force on the broom that cancels the other forces so that they don't make the broom fall over. In fact, if the wall was not strong enough to exert such a force, the wall would break. Still, it would be nice to get a visceral sense of the force exerted on the broom by the wall. Let your hand play the role of the wall, but this time, let the broom lean against your pinky, near the tip of your finger. To keep the broom in the same orientation as it was when it was leaning against the wall, you can feel that you have to exert a force on the tip of the broom handle. In fact, if you increase this force a little bit, the broom handle tilts more upward, and if you decrease it, it tilts more downward. Again, you can feel that you are pushing on the tip of the broom handle when you are causing the broom handle to remain stationary at the same orientation it had when it was leaning against the wall, and you can induce that when the wall is where your hand is, relative to the broom, the wall must be pressing on the broom handle with the same force. Note that the direction in which the wall is pushing on the broom is away from the wall at right angles to the wall. Such a force is exerted on any object that is in contact with a solid surface. This contact force exerted by a solid surface on an object in contact with that surface is called a "normal force" because the force is perpendicular to the surface and the word "normal" means perpendicular.

Using Free Body Diagrams

The key to the successful solution of a Newton's 2$^{\text{nd}}$ Law problem is to draw a good free body diagram of the object whose motion is under study and then to use that free body diagram to expand Newton's 2$^{\text{nd}}$ Law, that is, to replace the $\Sigma\vec{F}$ with an the actual term-by-term sum of the forces. Note that Newton's 2$^{\text{nd}}$ Law

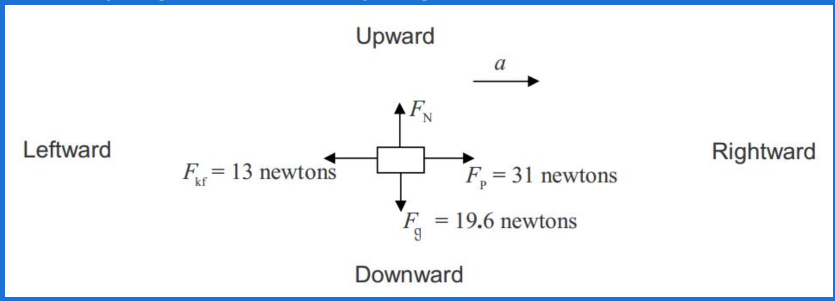
$$\vec{a} = \frac{1}{m}\sum\vec{F}$$

is a vector equation and hence, in the most general case (3 dimensions) is actually three scalar equations in one, one for each of the three possible mutually orthogonal directions in space. (A scalar is a number. Something that has magnitude only, as opposed to a vector which has magnitude and direction.) In your physics course, you will typically be dealing with forces that all lie in the same plane, and hence, you will typically get two equations from

$$\vec{a} = \frac{1}{m}\Sigma\vec{F}.$$

Regarding the Free Body Diagrams: The hard part is creating them from a description of the physical process under consideration; the easy part is using them. In what little remains of this chapter, we will focus on the easy part: Given a Free Body Diagram, use it to find an unknown force or unknown forces, and/or use it to find the acceleration of the object.

For example, given the free body diagram



for an object of mass $2.00kg$, find the magnitude of the normal force F_N and find the magnitude of the acceleration a. (Note that we define the symbols that we use to represent the components of forces and the component of the acceleration, in the free body diagram. We do this by drawing an arrow whose shaft represents a line along which the force lies, and whose arrowhead we define to be the positive direction for that force component, and then labeling the arrow with our chosen symbol. A negative value for a symbol thus defined, simply means that the corresponding force or acceleration is in the direction opposite to the direction in which the arrow is pointing.

Solution

Note that the acceleration and all of the forces lie along one or the other of two imaginary lines (one of which is horizontal and the other of which is vertical) that are perpendicular to each other. The acceleration along one line is independent of any forces perpendicular to that line so we can consider one line at a time. Let's deal with the horizontal line first. We write Newton's 2nd Law for the horizontal line as

a_{\rightarrow}=\dfrac{1}{m} \Sigma F_{\rightarrow} \label{14-2}

in which the shafts of the arrows indicate the line along which we are summing forces (the shafts in equation 14-2 are horizontal so we must be summing forces along the horizontal) and the arrowhead indicates which direction we consider to be the positive direction (any force in the opposite direction enters the sum with a minus sign).

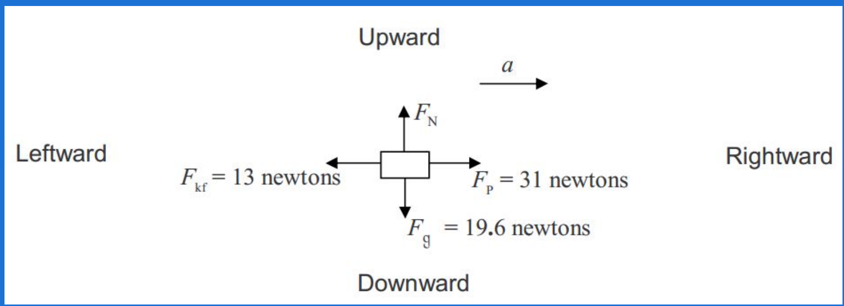
The next step is to replace $a_{\rightarrow}$ with the symbol that we have used in the diagram to represent the rightward acceleration and the $\Sigma F_{\rightarrow}$ with an actual term-by-term sum of the forces which includes only horizontal forces and in which rightward forces enter with a "$+$" and leftward forces enter with a "$-$". This yields:

$$a=\frac{1}{m}(F_p-F_{kf})$$

Substituting values with units and evaluating gives:

$$a=\frac{1}{2.00kg}(31N-13N)=9.0\frac{m}{s^2}$$

Now we turn our attention to the vertical direction. For your convenience, the free body diagram is replicated here:



Again we start with Newton's 2nd Law, this time written for the vertical direction:

$$a_{\downarrow}=\frac{1}{m}\Sigma F_{\downarrow}$$

We replace $a_\downarrow$ with what it is and we replace $\Sigma F_\downarrow$ with the term-by-term sum of the forces with a "$+$" for downward forces and a "$-$" for upward forces. Note that the only a in the free body diagram is horizontal. Whoever came up with that free body diagram is telling us that there is no acceleration in the vertical direction, that is, that $a_\downarrow = 0$. Thus:

$$0 = \frac{1}{m}(F_g - F_N)$$

Solving this for F_N yields:

$$F_N = F_g$$

Substituting values with units results in a final answer of:

$$F_N = 19.6\ N$$

15A: Newton's Laws #2: Kinds of Forces, Creating Free Body Diagrams

> *There is no "force of motion" acting on an object. Once you have the force or forces exerted on the object by everything (including any force-per-mass field at the location of the object) that is touching the object, you have all the forces. Do not add a bogus "force of motion" to your free body diagram. It is especially tempting to add a bogus force when there are no actual forces in the direction in which an object is going. Keep in mind, however, that an object does not need a force on it to keep going in the direction in which it is going; moving along at a constant velocity is what an object does when there is no net force on it.*

Now that you've had some practice using free body diagrams it is time to discuss how to create them. As you draw a free body diagram, there are a couple of things you need to keep in mind:

1. Include only those forces acting ON the object whose free body diagram you are drawing. Any force exerted BY the object on some other object belongs on the free body diagram of the other object.
2. All forces are contact forces and every force has an agent. The agent is "that which is exerting the force." In other words, the agent is the life form or thing that is doing the pushing or pulling on the object. No agent can exert a force on an object without being in contact with the object. (We are using the field point of view, rather than the action-at-adistance point of view for the fundamental forces of nature. Thus, for instance, it is the earth's gravitational field at the location of the object, rather than the material earth itself, that exerts the gravitational force on an object.)

We are going to introduce the various kinds of forces by means of examples. Here is the first example:

A rock is thrown up into the air by a person. Draw the free body diagram of the rock while it is up in the air. (Your free body diagram is applicable for any time after the rock leaves the thrower's hand, until the last instant before the rock makes contact with whatever it is destined to hit.) Neglect any forces that might be exerted on the rock by the air.

Solution

If you see the rock flying through the air, it may very well look to you like there is nothing touching the rock. But the earth's gravitational field is everywhere in the vicinity of the earth. It can't be blocked. It can't be shielded. It is in the air, in the water, even in the dirt. It is in direct contact with everything in the vicinity of the earth. It exerts a force on every object near the surface of the earth. We call that force the gravitational force. You have already studied the gravitational force. We give a brief synopsis of it here.

The Gravitational Force Exerted on Objects Near the Surface of the Earth.

Because it has mass, the earth has a gravitational field. The gravitational field is a force-per-mass field. It is invisible. It is not matter. It is an infinite set of force-per-mass vectors, one at every point in space in the vicinity of the surface of the earth. Each force-per-mass vector is directed downward, toward the center of the earth and, near the surface of the earth, has a magnitude of $9.80\dfrac{N}{kg}$. The symbol used to represent the earth's gravitational field vector at any point where it exists is $\vec{g}$. Thus, $\vec{g} = 9.80\dfrac{N}{kg}$ Downward. The effect of the earth's gravitational field is to exert a force on any object that is in the earth's gravitational field. The force is called the gravitational force and is equal to the product of the mass of the object and the earth's gravitational field vector: $F_g = m\vec{g}$. The magnitude of the gravitational force is given by

$$F_g = mg \tag{15A.1}$$

where $g = 9.80\dfrac{N}{kg}$ is the magnitude of the earth's gravitational field vector. The direction of the near-earth's-surface gravitational force is downward, toward the center of the earth.

Here is the free body diagram and the corresponding table of forces for Example 15A.1:

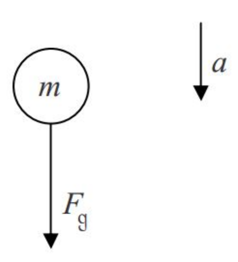


Table of Forces			
Symbol=?	**Name**	**Agent**	**Victim**
$F_g = mg$	Gravitational Force	The Earth's Gravitational Field	The Rock

Note:
1) The only thing touching the object while it is up in the air (neglecting the air itself) is the earth's gravitational field. So there is only one force on the object, namely the gravitational force. The arrow representing the force vector is drawn so that the tail of the arrow is touching the object, and the arrow extends away from the object in the direction of the force.

2) Unless otherwise stipulated and labeled on the diagram, upward is toward the top of the page and downward is toward the bottom of the page.

3) The arrow representing the acceleration must be near but not touching the object. (If it is touching the object, one might mistake it for a force.)

4) There is no velocity information on a free body diagram.

5) There is no force of the hand acting on the object because, at the instant in question, the hand is no longer touching the object. When you draw a free body diagram, only forces that are acting on the object at the instant depicted in the diagram are included. The acceleration of the object depends only on the currently-acting forces on the object. The force of the hand is of historical interest only.

6) Regarding the table of forces:

a) Make sure that for any free body diagram you draw, you are capable of making a complete table of forces. You are not required to provide a table of forces with every free body diagram you draw, but you should expect to be called upon to create a table of forces more than once.

b) In the table of forces, the agent is the life form or thing that is exerting the force and the victim is the object on which the force is being exerted. Make sure that, in every case, the victim is the object for which the free body diagram is being drawn.

c) In the case at hand, there is only one force so there is only one entry in the table of forces.

d) For any object near the surface of the earth, the agent of the gravitational force is the earth's gravitational field. It is okay to abbreviate that to "Earth" because the gravitational field of the earth can be considered to be an invisible part of the earth, but it is NOT okay to call it "gravity." Gravity is a subject heading corresponding to the kind of force the gravitational force is, gravity is not an agent.

A ball of mass m hangs at rest, suspended by a string. Draw the free body diagram for the ball, and create the corresponding table of forces.

Solution

To do this problem, you need the following information about strings:

The Force Exerted by a Taut String on an Object to Which it is Affixed

(This also applies to ropes, cables, chains, and the like.)

The force exerted by a string, on an object to which it is attached, is always directed away from the object, along the length of the string.

Note that the force in question is exerted by the string, not for instance, by some person pulling on the other end of the string.

The force exerted by a string on an object is referred to as a “tension force” and its magnitude is conventionally represented by the symbol F_T.

Note: There is no formula to tell you what the tension force is. If it is not given, the only way to get it is to use Newton’s 2$^{\text{nd}}$ Law.

Here is the free body diagram of the ball, and the corresponding table of forces for Example 15-2:

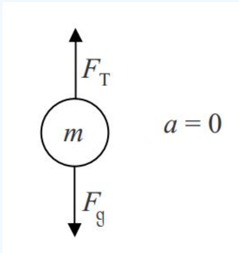


Table of Forces

Symbol=?	Name	Agent	Victim
F_T	Tension Force	The String	The Ball
$F_g = mg$	Gravitational Force	The Earth's Gravitational Field	The Ball

A sled of mass m is being pulled forward over a horizontal frictionless surface by means of a horizontal rope attached to the front of the sled. Draw the free body diagram of the sled and provide the corresponding table of forces.

Solution

Aside from the rope and the earth’s gravitational field, the sled is in contact with a solid surface. The surface exerts a kind of force that we need to know about in order to create the free body diagram for this example.

The Normal Force

When an object is in contact with a surface, that surface exerts a force on the object. The surface presses on the object. The force on the object is away from the surface, and it is perpendicular to the surface. The force is called the normal force because “normal” means perpendicular, and as mentioned, the force is perpendicular to the surface. We use the symbol F_N to represent the magnitude of the normal force.

Note: There is no formula to tell you what the normal force is. If it is not given, the only way to get it is to use Newton’s 2$^{\text{nd}}$ Law.

Here is the free body diagram of the sled as well as the corresponding table of forces.

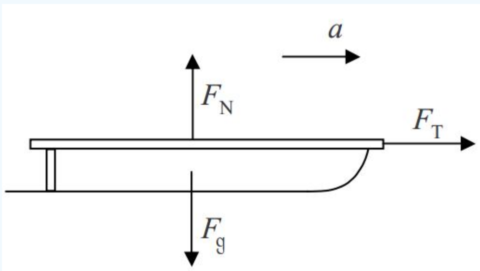


Table of Forces

	Name	Agent	Victim
F_N	Normal Force	The Horizontal Surface	The Sled
F_T	Tension Force	The Rope	The Sled
$F_g = mg$	Gravitational Force	The Earth's Gravitational Field	The Sled

Note: The word "Free" in "Free Body Diagram" refers to the fact that the object is drawn free of its surroundings. Do not include the surroundings (such as the horizontal surface on which the sled is sliding in the case at hand) in your Free Body Diagram.

A block of mass m rests on a frictionless horizontal surface. The block is due west of a west-facing wall. The block is attached to the wall by an ideal massless uncompressed/unstretched spring whose force constant is k. The spring is perpendicular to the wall. A person pulls the block a distance x directly away from the wall and releases it from rest. Draw the free body diagram of the block appropriate for the first instant after release. Provide the corresponding table of forces.

Solution

Now, for the first time, we have a spring exerting a force on the object for which we are drawing the free body diagram. So, we need to know about the force exerted by a spring.

The Force Exerted by a Spring

The force exerted by an ideal massless spring on an object in contact with one end of the spring is directed along the length of the spring, and

- away from the object if the spring is stretched,
- toward the object if the spring is compressed.

For the spring to exert a force on the object in the stretched-spring case, the object must be attached to the end of the spring. Not so in the compressed-spring case. The spring can push on an object whether or not the spring is attached to the object.

The force depends on the amount $|x|$ by which the spring is stretched or compressed, and on a measure of the stiffness of the spring known as the force constant of the spring a.k.a. the spring constant and represented by the symbol k. The magnitude of the spring force is typically represented by the symbol F_s. The spring force is directly proportional to the amount of stretch $|x|$. The spring constant k is the constant of proportionality. Thus,

$$F_s = k|x| \tag{15A.2}$$

Here is the free body diagram of the block, and the corresponding table of forces for Example 15-4:

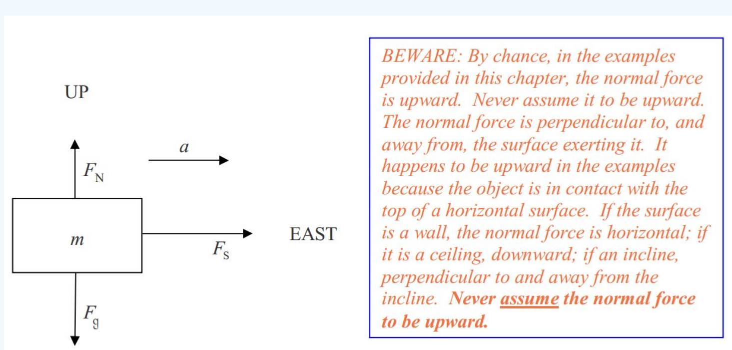


Table of Forces

	Name	Agent	Victim
F_N	Normal Force	The Horizontal Surface	The Block
$F_s = k\|x\|$	Spring Force	The Spring	The Block
$F_g = mg$	Gravitational Force	The Earth's Gravitational Field	The Block

From your vantage point, a crate of mass m is sliding rightward on a flat level concrete floor. Nothing solid is in contact with the crate except for the floor. Draw the free body diagram of the crate. Provide the corresponding table of forces.

Solution

From our experience with objects sliding on concrete floors, we know that the crate is slowing down at the instant under consideration. It is slowing because of kinetic friction.

Kinetic Friction

A surface, upon which an object is sliding, exerts (in addition to the normal force) a retarding force on that object. The retarding force is in the direction opposite that of the velocity of the object. In the case of an object sliding on a dry surface of a solid body (such as a floor) we call the retarding force a kinetic frictional force. Kinetic means motion and we include the adjective kinetic to make it clear that we are dealing with an object that is in motion.

The kinetic frictional formula given below is an empirical result. This means that it is derived directly from experimental results. It works only in the case of objects sliding on dry surfaces. It does not apply, for instance, to the case of an object sliding on a greased surface.

We use the symbol F_{kf} for the kinetic friction force. The kinetic frictional formula reads

$$F_{kf} = \mu_K F_N \tag{15A.3}$$

F_N is the magnitude of the normal force. Its presence in the formula indicates that the more strongly the surface is pressing on the object, the greater the frictional force.

μ_K (mu-sub-K) is called the coefficient of kinetic friction. Its value depends on the materials of which both the object and the surface are made as well as the smoothness of the two contact surfaces. It has no units. It is just a number. The magnitude of the kinetic frictional force is some fraction of the magnitude of the normal force; μ_K is that fraction. Values of μ_K for various pairs of materials can be found in handbooks. They tend to fall between 0 and 1. The actual value for a given pair of materials depends on the smoothness of the surface and is typically quoted with but a single significant digit.

IMPORTANT: μ_K is a coefficient (with no units) used in calculating the frictional force. It is not a force itself.

Here is the free body diagram and the table of forces for the case at hand. The crate is moving rightward and slowing down—it has a leftward acceleration.

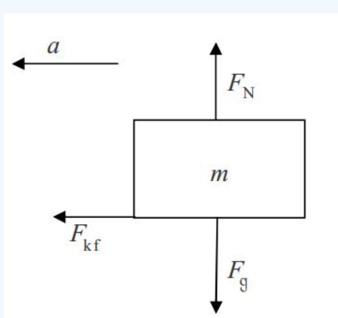


Table of Forces

Symbol=?	Name	Agent	Victim
F_N	Normal Force	The Concrete Floor	The Crate
$F_{kf}=\mu_K F_N$	Kinetic Friction Force	The Concrete Floor	The Crate
$F_g=mg$	Gravitational Force	The Earth's Gravitational Field	The Crate

A person has pushed a brick along a tile floor toward an eastward-facing wall trapping a spring of unstretched length L_0 and force constant k between the wall and the end of the brick that is facing the wall. That end of the brick is a distance d from the wall. The person has released the brick, but the spring is unable to budge it— the brick remains exactly where it was when the person released it. Draw the free body diagram for the brick and provide the corresponding table of forces.

Solution

A frictional force is acting on an object at rest. Typically, an object at rest clings more strongly to the surface with which it is in contact than the same object does when it is sliding across the same surface. What we have here is a case of static friction.

Static Friction Force

A surface that is not frictionless can exert a static friction force on an object that is in contact with that surface. The force of static friction is parallel to the surface. It is in the direction opposite the direction of impending motion of the stationary object. The direction of impending motion is the direction in which the object would accelerate if there was no static friction.

In general, there is no formula for calculating static friction—to solve for the force of static friction, you use Newton's 2nd Law. The force of static friction is whatever it has to be to make the net parallel-to-the-surface force zero.

We use the symbol F_{sf} to represent the magnitude of the static friction force.

SPECIAL CASE: Imagine trying to push a refrigerator across the floor. Imagine that you push horizontally, and that you gradually increase the force with which you are pushing. Initially, the harder you push, the bigger the force of static friction. But it can't grow forever. There is a maximum possible static friction force magnitude for any such case. Once the magnitude of your force exceeds that, the refrigerator will start sliding. The maximum possible force of static friction is given by:

$$F_{sf}^{max}=\mu_s F_N \tag{15A.4}$$

The unitless quantity μ_s is the coefficient of static friction specific to the type of surface the object is sliding on and the nature of the surface of the object. Values of μ_s tend to fall between 0 and 1. For a particular pair of surfaces, μ_s is at least as large as, and typically larger than, μ_K.

Clearly, this formula $(F_{sf}^{\text{max possible}}=\mu_s F_N)$ is only applicable when the question is about the maximum possible force of static friction. You can use this formula if the object is said to be on the verge of slipping, or if the question is about how hard one must push to budge an object. It also comes in handy when you want to know whether or not an object will stay put. In such a case you would use Newton's 2nd to find out the magnitude of the force of static friction needed to keep the object from accelerating. Then you would compare that magnitude with the maximum possible magnitude of the force of static friction.

Here is the free body diagram of the brick and the table of forces for Example 15-6:

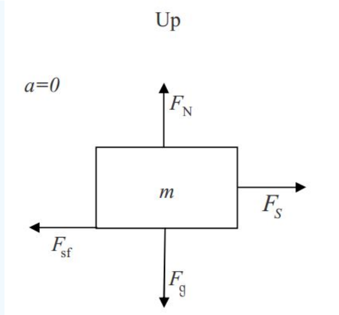


Table of Forces

Symbol=?	Name	Agent	Victim
F_N	Normal Force	The Tile Floor	The Brick
F_{sf}	Static Friction Force	The Tile Floor	The Brick
$F_g = mg$	Gravitational Force	The Earth's Gravitational Field	The Brick
$F_s = k\lvert x\rvert$	Spring Force	The Spring	The Brick

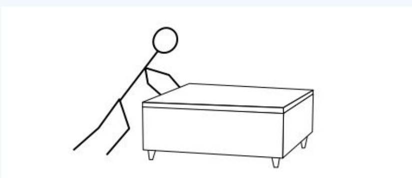
16A: Newton’s Laws #3: Components, Friction, Ramps, Pulleys, and Strings

> *Having learned how to use free body diagrams, and then having learned how to create them, you are in a pretty good position to solve a huge number of Newton’s 2nd Law problems. An understanding of the considerations in this chapter will enable to you to solve an even larger class of problems. Again, we use examples to convey the desired information.*

A professor is pushing on a desk with a force of magnitude F at an acute angle θ below the horizontal. The desk is on a flat, horizontal tile floor and it is not moving. For the desk, draw the free body diagram that facilitates the direct and straightforward application of Newton’s 2nd Law of motion. Give the table of forces.

Solution

While not a required part of the solution, a sketch often makes it easier to come up with the correct free body diagram. Just make sure you don’t combine the sketch and the free body diagram. In this problem, a sketch helps clarify what is meant by “at an acute angle θ below the horizontal.”



Pushing with a force that is directed at some acute angle below the horizontal is pushing horizontally and downward at the same time.

Here is the initial free body diagram and the corresponding table of forces

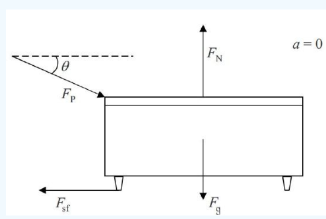


Table of Forces

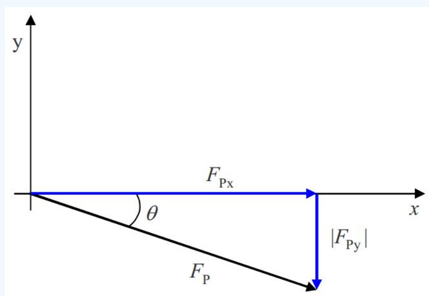
Symbol=?	Name	Agent	Victim
F_N	Normal Force	The Floor	Desk
F_{sf}	Static Friction Force	The Floor	Desk
$F_g = mg$	Gravitational Force	The Earth's Gravitational Field	Desk
F_p	Force of Professor	Hands of Professor	Desk

Note that there are no two mutually perpendicular lines to which all of the forces are parallel. The best choice of mutually perpendicular lines would be a vertical and a horizontal line. Three of the four forces lie along one or the other of such lines. But the force of the professor does not. We cannot use this free body diagram directly. We are dealing with a case which requires a second free body diagram.

Cases Requiring a Second Free Body Diagram in Which One of More of the Forces that was in the First Free Body Diagram is Replaced With its Components

Establish a pair of mutually perpendicular lines such that most of the vectors lie along one or the other of the two lines. After having done so, break up each of the other vectors, the ones that lie along neither of the lines, (let's call these the rogue vectors) into components along the two lines. (Breaking up vectors into their components involves drawing a vector component diagram.) Draw a second free body diagram, identical to the first, except with rogue vectors replaced by their component vectors. In the new free body diagram, draw the component vectors in the direction in which they actually point and label them with their magnitudes (no minus signs).

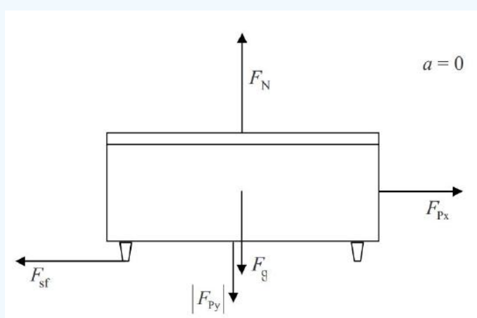
For the case at hand, our rogue force is the force of the professor. We break it up into components as follows:



$$\frac{F_{Px}}{F_P} = \cos\theta \quad \frac{|F_{Py}|}{F_P} = \sin\theta$$

$$F_{Px} = F_P\cos\theta \quad |F_{Py}| = F_P\sin\theta$$

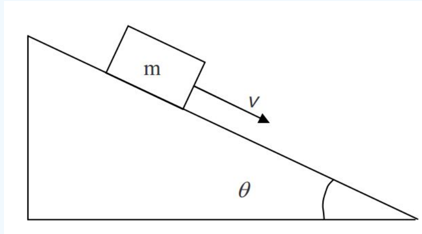
Then we draw a second free body diagram, the same as the first, except with F_P replaced by its component vectors:



A wooden block of mass m is sliding down a flat metal incline (a flat metal ramp) that makes an acute angle θ with the horizontal. The block is slowing down. Draw the directly-usable free body diagram of the block. Provide a table of forces.

Solution

We choose to start the solution to this problem with a sketch. The sketch facilitates the creation of the free body diagram but in no way replaces it.



Since the block is sliding in the down-the-incline direction, the frictional force must be in the up-the-incline direction. Since the block's velocity is in the down-the-incline direction and decreasing, the acceleration must be in the up-the-incline direction.

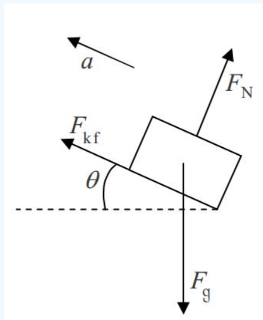


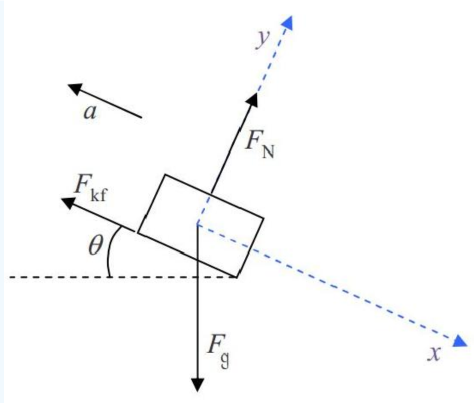
Table of Forces

Symbol=?	Name	Agent	Victim
F_N	Normal Force	The Ramp	The Block
$F_{kf} = \mu_K F_N$	Kinetic Friction Force	The Ramp	The Block
$F_g = mg$	Gravitation Force	The Earth's Gravitational Field	The Block

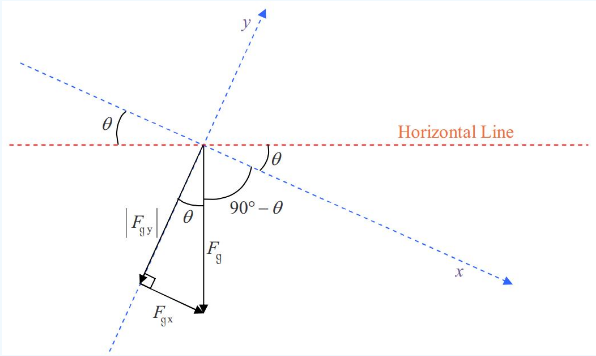
No matter what we choose for a pair of coordinate axes, we cannot make it so that all the vectors in the free body diagram are parallel to one or the other of the two coordinate axes lines. At best, the pair of lines, one line parallel to the frictional force and the other perpendicular to the ramp, leaves one rogue vector, namely the gravitational force vector. Such a coordinate system is tilted on the page.

> **Cases Involving Tilted Coordinate Systems**
>
> *For effective communication purposes, students drawing diagrams depicting phenomena occurring near the surface of the earth are required to use either the convention that downward is toward the bottom of the page (corresponding to a side view) or the convention that downward is into the page (corresponding to a top view). If one wants to depict a coordinate system for a case in which the direction "downward" is parallel to neither coordinate axis line, the coordinate system must be drawn so that it appears tilted on the page.*

In the case of a tilted-coordinate system problem requiring a second free body diagram of the same object, it is a good idea to define the coordinate system on the first free body diagram. Use dashed lines so that the coordinate axes do not look like force vectors. Here we redraw the first free body diagram. (When you get to this stage in a problem, just add the coordinate axes to your existing diagram.)



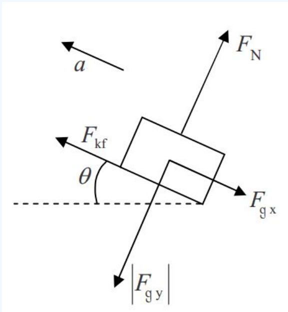
Now we break up F_g into its x, y component vectors. This calls for a vector component diagram.



$$\frac{F_{gx}}{F_g} = \sin\theta \quad \frac{|F_{gy}|}{F_g} = \cos\theta$$

$$F_{gx} = F_g \sin\theta \quad |F_{gy}| = F_g \cos\theta$$

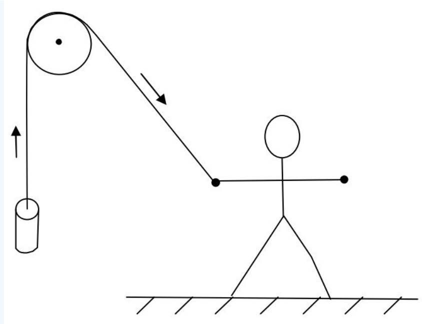
Next, we redraw the free body diagram with the gravitational force vector replaced by its component vectors.



A solid brass cylinder of mass m is suspended by a massless string which is attached to the top end of the cylinder. From there, the string extends straight upward to a massless ideal pulley. The string passes over the pulley and extends downward, at an acute angle theta to the vertical, to the hand of a person who is pulling on the string with force F_T. The pulley and the entire string segment, from the cylinder to hand, lie in one and the same plane. The cylinder is accelerating upward. Provide both a free body diagram and a table of forces for the cylinder.

Solution

A sketch comes in handy for this one:



To proceed with this one, we need some information on the effect of an ideal massless pulley on a string that passes over the pulley.

Effect of an Ideal Massless Pulley

> *The effect of an ideal massless pulley on a string that passes over the pulley is to change the direction in which the string extends, without changing the tension in the string.*

By pulling on the end of the string with a force of magnitude F_T, the person causes there to be a tension F_T in the string. (The force applied to the string by the hand of the person, and the tension force of the string pulling on the hand of the person, are a Newton's-3rd-law interaction pair of forces. They are equal in magnitude and opposite in direction. We choose to use one and the same symbol F_T for the magnitude of both of these forces. The directions are opposite each other.) The tension is the same throughout the string, so, where the string is attached to the brass cylinder, the string exerts a force of magnitude F_T directed away from the cylinder along the length of the string. Here is the free body diagram and the table of forces for the cylinder:

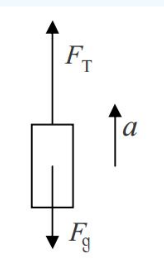


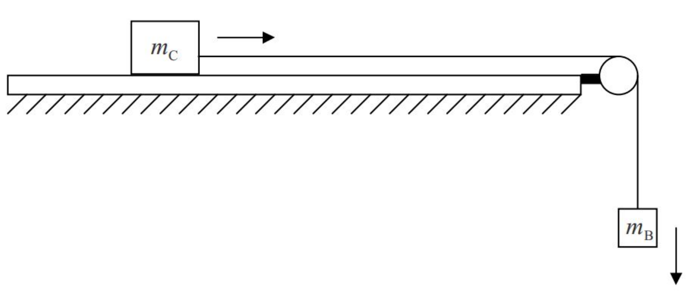
Table of Forces

Symbol=?	Name	Agent	Victim
F_T	Tension Force	The String	The Cylinder
$F_g = mg$	Gravitational Force	The Earth's Gravitational Field	The Cylinder

A cart of mass m_c is on a horizontal frictionless track. One end of an ideal massless string segment is attached to the middle of the front end of the cart. From there the string extends horizontally, forward and away from the cart, parallel to the centerline of the track, to a vertical pulley. The string passes over the pulley and extends downward to a solid metal block of mass m_B. The string is attached to the block. A person was holding the cart in place. The block was suspended at rest, well above the floor, by the string. The person has released the cart. The cart is now accelerating forward and the block is accelerating downward. Draw a free body diagram for each object.

Solution

A sketch will help us to arrive at the correct answer to this problem.



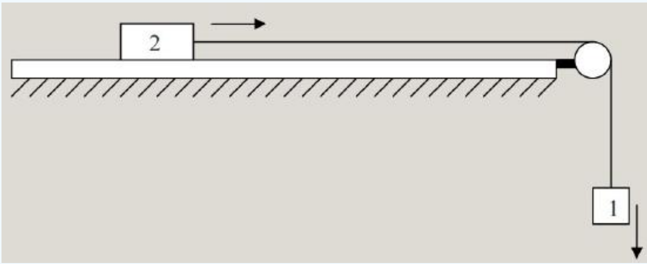
Recall from the last example that there is only one tension in the string. Call it F_T. Based on our knowledge of the force exerted on an object by a string, viewed so that the apparatus appears as it does in the sketch, the string exerts a rightward force F_T on the cart, and an upward force of magnitude FT on the block.

There is a relationship between each of several variables of motion of one object attached by a taut string, which remains taut throughout the motion of the object, and the corresponding variables of motion of the second object. The relationships are so simple that you might consider them to be trivial, but they are critical to the solution of problems involving objects connected by a taut spring.

The Relationships Among the Variables of Motion For Two Objects, One at One End and the Other at the Other

End, of an always-taut, Unstretchable String

Consider the following diagram.



Because they are connected together by a string of fixed length, if object 1 *goes downward* $5cm$*, then object* 2 *must go rightward* $5cm$*. So if object* 1 *goes downward at* $5cm/s$ *then object* 2 *must go rightward at* $5cm/s$*. In fact, no matter how fast object* 1 *goes downward, object* 2 *must go rightward at the exact same speed (as long as the string does not break, stretch, or go slack). In order for the speeds to always be the same, the accelerations have to be the same as each other. So if object* 1 *is picking up speed in the downward direction at, for instance,* $5cm/s^2$*, then object* 2 *must be picking up speed in the rightward direction at* $5cm/s^2$*. The magnitudes of the accelerations are identical. The way to deal with this is to use one and the same symbol for the magnitude of the acceleration of each of the objects. The ideas relevant to this simple example apply to any case involving two objects, one on each end of an inextensible string, as long as each object moves only along a line collinear with the string segment to which the object is attached.*

Let's return to the example problem involving the cart and the block. The two free body diagrams follow:

Note that the use of the same symbol F_T in both diagrams is important, as is the use of the same symbol a in both diagrams.

17A: The Universal Law of Gravitation

> *Consider an object released from rest an entire moon's diameter above the surface of the moon. Suppose you are asked to calculate the speed with which the object hits the moon. This problem typifies the kind of problem in which students use the universal law of gravitation to get the force exerted on the object by the gravitational field of the moon, and then mistakenly use one or more of the constant acceleration equations to get the final velocity. The problem is: the acceleration is not constant. The closer the object gets to the moon, the greater the gravitational force, and hence, the greater the acceleration of the object. The mistake lies not in using Newton's second law to determine the acceleration of the object at a particular point in space; the mistake lies in using that one value of acceleration, good for one object-to-moon distance, as if it were valid on the entire path of the object. The way to go on a problem like this, is to use conservation of energy.*

Back in chapter 12, where we discussed the near-surface gravitational field of the earth, we talked about the fact that any object that has mass creates an invisible force-per-mass field in the region of space around it. We called it a gravitational field. Here we talk about it in more detail. Recall that when we say that an object causes a gravitational field to exist, we mean that it creates an invisible force-per-mass vector at every point in the region of space around itself. The effect of the gravitational field on any particle (call it the victim) that finds itself in the region of space where the gravitational field exists, is to exert a force, on the victim, equal to the force-per mass vector at the victim's location, times the mass of the victim.

Now we provide a quantitative discussion of the gravitational field. (Quantitative means, involving formulas, calculations, and perhaps numbers. Contrast this with qualitative which means descriptive/conceptual.) We start with the idealized notion of a point particle of matter. Being matter, it has mass. Since it has mass it has a gravitational field in the region around it. The direction of a particle's gravitational field at point P, a distance r away from the particle, is toward the particle and the magnitude of the gravitational field is given by

$$g=G\frac{m}{r^2} \tag{17A.1}$$

where:

> G is the universal gravitational constant: $G=6.67\times10^{-11}\,\frac{N\cdot m^2}{kg^2}$
>
> m is the mass of the particle, and:
>
> r is the distance that P is from the particle.

In that point P can be any empty (or occupied) point in space whatsoever, this formula gives the magnitude of the gravitational field of the particle at all points in space. Equation 17A.1 is the equation form of Newton's Universal Law of Gravitation.

The Total Gravitational Field Vector at an Empty Point in Space

Suppose that you have two particles. Each has its own gravitational field at all points in space. Let's consider a single empty point in space. Each of the two particles has its own gravitational field vector at that empty point in space. We can say that each particle makes its own vector contribution to the total gravitational field at the empty point in space in question. So how do you determine the total gravitational field at the empty point in space? You guessed it! Just add the individual contributions. And because the contribution due to each particle is a vector, the two contributions add like vectors.

What the Gravitational Field does to a Particle that is in the Gravitational Field

Now suppose that you have the magnitude and direction of the gravitational field vector $\vec{g}$ at a particular point in space. The gravitational field exerts a force on any "victim" particle that happens to find itself at that location in space. Suppose, for instance, that a particle of mass m finds itself at a point in space where the gravitational field (of some other particle or particles) is $\vec{g}$. Then the particle of mass m is subject to a force

$$\vec{F}_G = m\vec{g} \tag{17A.2}$$

The Gravitational Effect of one Particle on Another

Let's put the two preceding ideas together. Particle 1 of mass m_1 has, among its infinite set of gravitational field vectors, a gravitational field vector at a location a distance r_{12} away from itself, a point in space that happens to be occupied by another particle, particle 2, of mass m_2. The gravitational field of particle 1 exerts a force on particle 2. The question is, how big and which way is the force?

Let's start by identifying the location of particle 2 as point P. Point P is a distance r_{12} away from particle 1. Thus, the magnitude of the gravitational field vector (of particle 1) at point P is:

$$g = G\frac{m_1}{r_{12}^2} \tag{17A.3}$$

Now the force exerted on particle 2 by the gravitational field of particle 1 is given by equation 17A.2, $\vec{F}_G = m_2\,\vec{g}$. Using $\vec{F}_{12}$ for $\vec{F}_G$ to emphasize the fact that we are talking about the force of 1 on 2, and writing the equation relating the magnitudes of the vectors we have

$$F_{12} = m_2\, g$$

Replacing the g with the expression we just found for the magnitude of the gravitational field due to particle number 1 we have

$$F_{12} = m_2\, G\frac{m_1}{r_{12}^2}$$

which, with some minor reordering can be written as

$$F_{12} = G\frac{m_1\, m_2}{r_{12}^2} \tag{17A.4}$$

This equation gives the force of the gravitational field of particle 1 on particle 2. Neglecting the "middleman" (the gravitational field) we can think of this as the force of particle 1 on particle 2. You can go through the whole argument again, with the roles of the particles reversed, to find that the same expression applies to the force of particle 2 on particle 1, or you can simply invoke Newton's 3rd Law to arrive at the same conclusion.

Objects, Rather than Point Particles

The vector sum of all the gravitational field vectors due to a spherically symmetric distribution of point particles (for instance, a spherically symmetric solid object), at a point outside the distribution (e.g. outside the object), is the same as the gravitational field vector due to a single particle, at the center of the distribution, whose mass is equal to the sum of the masses of all the particles. Also, for purposes of calculating the force exerted by the gravitational field of a point object on a spherically symmetric victim, one can treat the victim as a point object at the center of the victim. Finally, regarding either object in a calculation of the gravitational force exerted on a rigid object by the another object: if the separation of the objects is very large compared to the dimensions of the object, one can treat the object as a point particle located at the center of mass of the object and having the same mass as the object. This goes for the gravitational potential energy, discussed below, as well.

How Does this Fit in with g= 9.80 N/kg?

When we talked about the earth's near-surface gravitational field before, we used a single value for its magnitude, namely 9.80 N/kg and said that it always directed downward, toward the center of the earth. 9.80 N/kg is actually an average sea level value. g varies within about a half a percent of that value over the surface of the earth and the handbook value includes a tiny centrifugal pseudo force-per-mass field contribution (affecting both magnitude and direction) stemming from the rotation of the earth. So how does the value 9.80 N/kg for the magnitude of the gravitational field near the surface of the earth relate to Newton's Universal Law of Gravitation?

Certainly the direction is consistent with our understanding of what it should be: The earth is roughly spherically symmetric so for purposes of calculating the gravitational field outside of the earth we can treat the earth as a point particle located at the center of the earth. The direction of the gravitational field of a point particle is toward that point particle, so, anywhere outside the earth, including at any point just outside the earth (near the surface of the earth), the gravitational field, according to the Universal Law of Gravitation, must be directed toward the center of the earth, a direction we earthlings call downward.

But how about the magnitude? Shouldn't it vary with elevation according to the Universal Law of Gravitation? First off, how does the magnitude calculated using $g = G\frac{m}{r^2}$ compare with 9.80 N/kg at, for instance, sea level. A point at sea level is a distance $r = 6.37 \times 10^6$ m from the center of the earth. The mass of the earth is $\mathrm{m} = 5.98 \times 10^{21} \mathrm{kg}$. Substituting these values into our expression for g (equation 17A.4 which reads $g = G\frac{\mathrm{m}}{r_{12}^2}$) we find:

$$g = 6.67 \times 10^{-11} \frac{N \cdot m^2}{kg^2} \frac{5.98 \times 10^{24} kg}{(6.37 \times 10^6 m)^2}$$

$$g = 9.83 \frac{N}{kg}$$

which does agree with the value of 9.80 N/kg to within about 0.3 percent. (The difference is due in part to the centrifugal pseudo force-per-mass field associated with the earth's rotation.) We can actually see, just from the way that the value of the radius of the earth, 6.37×10^6 m, is written, that increasing our elevation, even by a mile 1.61 km, is not going to change the value of g to three significant figures. We'd be increasing r from 6.37×10^6 m to 6.37161×10^6 m which, to three significant figures is still 6.37×10^6 m. This brings up the question, "How high above the surface do you have to go before $g = 9.80$ N/kg is no longer within a certain percentage of the value obtained by means of the Universal Law of Gravitation?" Let's investigate that question for the case of a 1 percent difference. In other words, how high above sea level would we have to go to make g = 99% of g at sea level, that is $g = (.99)(9.80 m/s^2) = 9.70$ m/s^2. Letting $r = r_E + h$ where r_E is the radius of the earth, and using g = 9.70 N/kg so that we can find the h that makes 9.70 N/kg we have:

$$g = \frac{Gm_E}{(r_E + h)^2}$$

Solving this for h yields

$$h = \sqrt{\frac{Gm_E}{g}} - r_E$$

$$h = \sqrt{\frac{6.67 \times 10^{-11} \frac{N \cdot m^2}{kg^2} 5.98 \times 10^{24} kg}{9.70 \frac{N}{kg}}} - 6.37 \times 10^6 m$$

$$h = .04 \times 10^6 m$$

That is to say that at any altitude above 40 km above the surface of the earth, you should use Newton's Universal Law of Gravitation directly in order for your results to be good to within 1%.

In summary, g=9.80 N/kg for the near-earth-surface gravitational field magnitude is an approximation to the Universal Law of Gravitation good to within about 1% anywhere within about 40 km of the surface of the earth. In that region, the value is approximately a constant because changes in elevation represent such a tiny fraction of the total earth's-center-to-surface distance as to be negligible.

The Universal Gravitational Potential Energy

So far in this course you have become familiar with two kinds of potential energy, the near-earth's-surface gravitational potential energy $U_g = mgy$ and the spring potential energy $U_s = \frac{1}{2}kx^2$. Here we introduce another expression for gravitational potential energy. This one is pertinent to situations for which the Universal Law of Gravitation is appropriate.

$$U_G = -\frac{Gm_1\, m_2}{r_{12}} \tag{17A.5}$$

This is the gravitational potential energy of a pair of particles, one of mass m_1 and the other of mass m_2, which are separated by a distance of r_{12}. Note that for a given pair of particles, the gravitational potential energy can take on values from negative infinity up to zero. Zero is the highest possible value and it is the value of the gravitational potential energy at infinite separation. That is to say that $U_G \xrightarrow{\lim r_{12}\to\infty} 0$. The lowest conceivable value is negative infinity and it would be the value of the gravitational potential energy of the pair of particles if one could put them so close together that they were both at the same point in space. In mathematical notation: $U_G \xrightarrow{\lim r_{12}\to 0} -\infty$.

Conservation of Mechanical Energy Problems Involving Universal Gravitational Potential Energy for the Special Case in which the Total Amount of Mechanical Energy does not Change

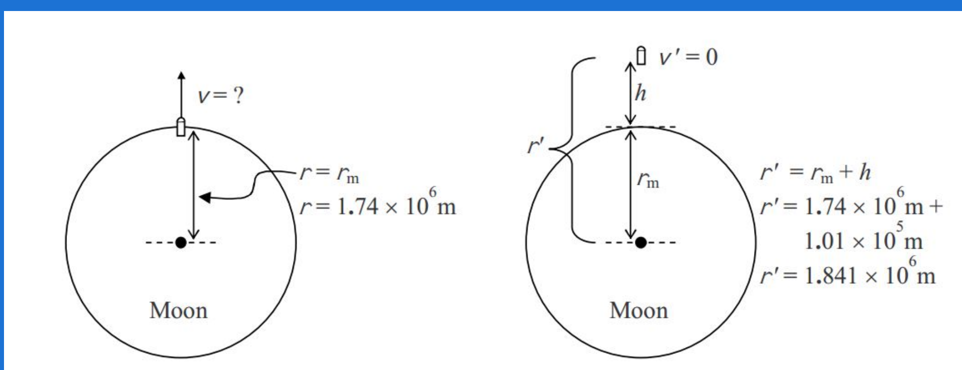
You solve fixed amount of mechanical energy problems involving universal gravitational potential energy just as you solved fixed amount of mechanical energy problems involving other forms of potential energy back in chapters 2 and 3. Draw a good before and after picture, write an equation stating that the energy in the before picture is equal to the energy in the after picture, and proceed from there. To review these procedures, check out the example problem on the next page:

How great would the muzzle velocity of a gun on the surface of the moon have to be in order to shoot a bullet to a altitude of 101 km?

Solution: We'll need the following lunar data:

Mass of the moon: $m_m = 7.35\times 10^{22} kg$; Radius of the moon: $r_m = 1.74\times 10^6 m$;

$E =$



$$K + U = K' + U'$$

$$\frac{1}{2}mv^2 + -\frac{Gm_m m}{r} = -\frac{Gm_m m}{r'} \quad m_m \text{is the mass of the moon and m is the mass of the bul}$$

$$\frac{1}{2}v^2 = \frac{G_m}{r} - \frac{G_m}{r'} \qquad \frac{1}{2}v^2 = G_m(\frac{1}{r} - \frac{1}{r'}) \qquad v = \sqrt{2Gm_m(\frac{1}{r} - \frac{1}{r'})}$$

$$v = \sqrt{2(6.67\times 10^{-11}\frac{N\cdot m^2}{kg^2})7.35\times 10^{22} kg(\frac{1}{1.74\times 10^6 m} - \frac{1}{1.841\times 10^6 m})}$$

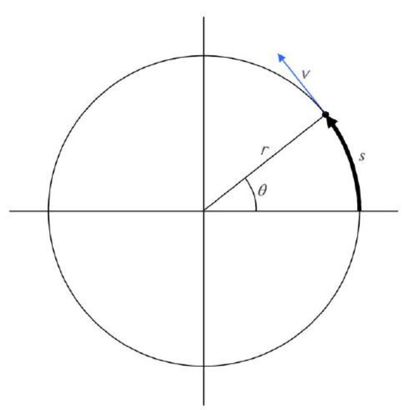
$$v = 556\frac{m}{s}$$

18A: Circular Motion - Centripetal Acceleration

> *There is a tendency to believe that if an object is moving at constant speed then it has no acceleration. This is indeed true in the case of an object moving along a straight line path. On the other hand, a particle moving on a curved path is accelerating whether the speed is changing or not. Velocity has both magnitude and direction. In the case of a particle moving on a curved path, the direction of the velocity is continually changing, and thus the particle has acceleration.*

We now turn our attention to the case of an object moving in a circle. We'll start with the simplest case of circular motion, the case in which the speed of the object is a constant, a case referred to as uniform circular motion. For the moment, let's have you be the object. Imagine that you are in a car that is traveling counterclockwise, at say 40 mph, as viewed from above, around a fairly small circular track. You are traveling in a circle. Your velocity is not constant. The magnitude of your velocity is not changing (constant speed), but the direction of your velocity is continually changing, you keep turning left! Now if you are continually turning left then you must be continually acquiring some leftward velocity. In fact, your acceleration has to be exactly leftward, at right angles to your velocity because, if your speed is not changing, but your velocity is continually changing, meaning you have some acceleration $\vec{a} = \dfrac{d\vec{v}}{dt}$, then for every infinitesimal change in clock reading dt, the change in velocity $d\vec{v}$ that occurs during that infinitesimal time interval must be perpendicular to the velocity itself. (If it wasn't perpendicular, then the speed would be increasing or decreasing.) So no matter where you are in the circle (around which you are traveling counterclockwise as viewed from above) you have an acceleration directed exactly leftward, perpendicular to the direction of your velocity. Now what is always directly leftward of you if you are traveling counterclockwise around a circle? Precisely! The center of the circle is always directly leftward of you. Your acceleration is thus, always, center directed. We call the center-directed acceleration associated with circular motion centripetal acceleration because the word "centripetal" means "center-directed." Note that if you are traveling around the circle clockwise as viewed from above, you are continually turning right and your acceleration is directed rightward, straight toward the center of the circle. These considerations apply to any object—an object moving in a circle has centripetal (center-directed) acceleration.

We have a couple of ways of characterizing the motion of a particle that is moving in a circle. First, we characterize it in terms of how far the particle has traveled along the circle. If we need a position variable, we establish a start point on the circle and a positive direction. For instance, for a circle centered on the origin of an x-y plane we can define the point where the circle intersects the positive x axis as the start point, and define the direction in which the particle must move to go counterclockwise around the circle as the positive direction. The name given to this position variable is s. The position s is the total distance, measured along the circle, that the particle has traveled. The speed of the particle is then the rate of change of s, $\dfrac{ds}{dt}$ and the direction of the velocity is tangent to the circle. The circle itself is defined by its radius. The second method of characterizing the motion of a particle is to describe it in terms of an imaginary line segment extending from the center of a circle to the particle. To use this method, one also needs to define a reference line segment—the positive x axis is the conventional choice for the case of a circle centered on the origin of an x-y coordinate system. Then, as long as you know the radius r of the circle, the angle $theta$ that the line to the particle makes with the reference line completely specifies the location of the particle.



In geometry, the position variable s, defines an arc length on the circle. Recall that, by definition, the angle θ in radians is the ratio of the arc length to the radius:

$$\theta=\frac{s}{r}$$

Solving for s we have:

$$s=r\theta \tag{18A.1}$$

in which we interpret the s to be the position-on-the-circle of the particle and the θ to be the angle that an imaginary line segment, from the center of the circle to the particle, makes with a reference line segment, such as the positive x-axis. Clearly, the faster the particle is moving, the faster the angle theta is changing, and indeed we can get a relation between the speed of the particle and the rate of change of θ just by taking the time derivative of both sides of Equation 18A.1. Let's do that.

We start by taking the derivative of both sides of Equation 18A.1 with respect to time:

$$\frac{ds}{dt}=r\frac{d\theta}{dt}$$

and then rewrite the result as:

$$\dot{s}=r\dot{\theta}$$

just to get the reader used to the idea that we represent the time derivative of a variable, that is the rate of change of that variable, by the writing the symbol for the variable with a dot over it. Then we rewrite the result as

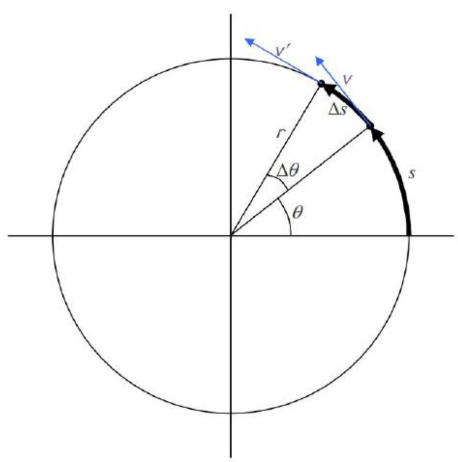
$$v=r\dot{\theta} \tag{18A.2}$$

to emphasize the fact that the rate of change of the position-on-the-circle is the speed of the particle (the magnitude of the velocity of the particle). Finally, we define the variable ω ("omega") to be the rate of change of the angle, meaning that ω is $\frac{d\theta}{dt}$ and ω is $\dot{\theta}$. It should be clear that ω is the spin rate for the imaginary line from the center of the circle to the particle. We call that spin rate the magnitude of the angular velocity of the line segment. (The expression angular velocity, ω, is more commonly used to characterize how fast and which way a rigid body, rather than an imaginary line, is spinning.) Rewriting $v=r\dot{\theta}$ with $\dot{\theta}$ replaced by ω yields:

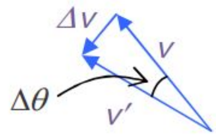
$$v=r\omega \tag{18A.3}$$

How the Centripetal Acceleration Depends on the Speed of the Particle and the Size of the Circle

We are now in a position to derive an expression for that center-directed (centripetal) acceleration we were talking about at the start of this chapter. Consider a short time interval Δt. (We will take the limit as Δt goes to zero before the end of this chapter.) During that short time interval, the particle travels a distance Δs along the circle and the angle that the line, from the center of the circle to the particle, makes with the reference line changes by an amount $\Delta\theta$.



Furthermore, in that time Δt, the velocity of the particle changes from $\vec{v}$ to $\vec{v}'$, a change $\Delta\vec{v}$ defined by $vecv' = \vec{v} + \Delta\vec{v}$ depicted in the following vector diagram (in which the arrows representing the vectors $\vec{v}$ and $\vec{v}'$ have been copied from above with no change in orientation or length). Note that the small angle $\Delta\theta$ appearing in the vector addition diagram is the same $\Delta\theta$ that appears in the diagram above.



While $\vec{v}'$ is a new vector, different from $\vec{v}$, we have stipulated that the speed of the particle is a constant, so the vector $\vec{v}'$ has the same magnitude as the vector $\vec{v}$. That is, $\vec{v}' = \vec{v}$. We redraw the vector addition diagram labeling both velocity vectors with the same symbol v.

The magnitude of the centripetal acceleration, by definition, can be expressed as

$$a_c = \lim_{\Delta t \to 0} \frac{\Delta V}{\Delta t}$$

Look at the triangle in the vector addition diagram above. It is an isosceles triangle. The two unlabeled angles in the triangle are equal to each other. Furthermore, in the limit as Δt approaches 0, $\Delta\theta$ approaches 0, and as $\Delta\theta$ approaches 0, the other two angles must each approach 90° in order for the sum of the angles to remain 180°, as it must, because the sum of the interior angles for any triangle is 180°. Thus in the limit as Δt approaches 0, the triangle is a right triangle and in that limit we can write:

$$\frac{\Delta v}{v} = tan(\Delta\theta)$$

$$\Delta v = v\tan(\Delta\theta)$$

Substituting this into our expression for a_c we have:

$$a_c = \lim_{\Delta t \to 0} \frac{vtan(\Delta\theta)}{\Delta t} \tag{18A.4}$$

Now we invoke the small angle approximation from the mathematics of plane geometry, an approximation which becomes an actual equation in the limit as $\Delta\theta$ approaches zero.

The Small Angle Approximation

For any angle that is very small compared to π radians (the smaller the angle the better the approximation), the tangent of the angle is approximately equal to the angle itself, expressed in radians; and the sine of the angle is approximately equal to the angle itself, expressed in radians. In fact,

$$tan(\Delta\theta) \underset{\Delta\theta \to 0}{\rightarrow} \Delta\theta$$

and

$$sin(\Delta\theta) \underset{\Delta\theta \to 0}{\rightarrow} \Delta\theta$$

where $\Delta\theta$ is in radians.

The small angle approximation allows us to write

$$a_c = \lim_{\Delta t \to 0} \frac{v\Delta\theta}{\Delta t}$$

[where we have replaced the $tan(\Delta\theta)$ in Equation 18A.4 above with $\Delta\theta$].

The constant v can be taken outside the limit yielding $a_c = \lim_{\Delta t \to 0} \dfrac{\Delta\theta}{\Delta t}$. But the $\lim_{\Delta t \to 0} \dfrac{\Delta\theta}{\Delta t}$ is the rate of change of the angle θ, which is, by definition, the angular velocity ω. Thus

$$a_c = v\omega$$

According to Equation 18A.3, $v = r\omega$. Solving that for ω we find that $\omega = \dfrac{v}{r}$. Substituting this into our expression for a_c yields

$$a_c = \frac{v^2}{r} \tag{18A.5}$$

Please sound the drum roll! This is the result we have been seeking. Note that by substituting $r\omega$ for v, we can also write our result as

$$a_c = r\omega^2 \tag{18A.6}$$

It should be pointed out that, despite the fact that we have been focusing our attention on the case in which the particle moving around the circle is moving at constant speed, the particle has centripetal acceleration whether the speed is changing or not. If the speed of the particle is changing, the centripetal acceleration at any instant is (still) given by Equation 18A.5 with the v being the speed of the particle at that instant (and in addition to the centripetal acceleration, the particle also has some along-the-circular-path acceleration known as tangential acceleration). The case that we have investigated is, however the remarkable case. Even if the speed of the particle is constant, the particle has some acceleration just because the direction of its velocity is continually changing. What's more, the centripetal acceleration is not a constant acceleration because its direction is continually changing. Visualize it. If you are driving counterclockwise (as viewed from above) around a circular track, the direction in which you see the center of the circle is continually changing (and that direction is the direction of the centripetal acceleration). When you are on the easternmost point of the circle the center is to the west of you. When you are at the northernmost point of the circle the center is to the south of you. When you are at the westernmost point of the circle, the center is to the east of you. And when you are at the southernmost point of the circle, the center is to the north of you.

19A: Rotational Motion Variables, Tangential Acceleration, Constant Angular Acceleration

> *Because so much of the effort that we devote to dealing with angles involves acute angles, when we go to the opposite extreme, e.g. to angles of thousands of degrees, as we often do in the case of objects spinning with a constant angular acceleration, one of the most common mistakes we humans tend to make is simply not to recognize that when someone asks us; starting from time zero, how many revolutions, or equivalently how many turns or rotations an object makes; that someone is asking for the value of the angular displacement $\Delta\theta$. To be sure, we typically calculate $\Delta\theta$ in radians, so we have to convert the result to revolutions before reporting the final answer, but the number of revolutions is simply the value of $\Delta\theta$*

In the last chapter we found that a particle in uniform circular motion has centripetal acceleration given by equations ??? and ???:

$$a_c = \frac{v^2}{r} \quad a_c = r\omega^2$$

It is important to note that any particle undergoing circular motion has centripetal acceleration, not just those in uniform (constant speed) circular motion. If the speed of the particle (the value of v in $a_c = \frac{v^2}{r}$) is changing, then the value of the centripetal acceleration is clearly changing. One can still calculate it at any instant at which one knows the speed of the particle.

If, besides the acceleration that the particle has just because it is moving in a circle, the speed of the particle is changing, then the particle also has some acceleration directed along (or in the exact opposite direction to) the velocity of the particle. Since the velocity is always tangent to the circle on which the particle is moving, this component of the acceleration is referred to as the tangential acceleration of the particle. The magnitude of the tangential acceleration of a particle in circular motion is simply the absolute value of the rate of change of the speed of the particle $a_t = \left|\frac{dv}{dt}\right|$. The direction of the tangential acceleration is the same as that of the velocity if the particle is speeding up, and in the direction opposite that of the velocity if the particle is slowing down. Recall that, starting with our equation relating the position s of the particle along the circle to the angular position θ of a particle, $s = r\theta$, we took the derivative with respect to time to get the relation $v = r\omega$. If we take a second derivative with respect to time we get

$$\frac{dv}{dt} = r\frac{d\omega}{dt}$$

On the left we have the tangential acceleration a_t of the particle. The $\frac{d\omega}{dt}$ on the right is the time rate of change of the angular velocity of the object. The angular velocity is the spin rate, so a non-zero value of $\frac{d\omega}{dt}$ means that the imaginary line segment that extends from the center of the circle to the particle is spinning faster or slower as time goes by. In fact, $\frac{d\omega}{dt}$ is the rate at which the spin rate is changing. We call it the angular acceleration and use the symbol $\propto$ (the Greek letter alpha) to represent it. Thus, the relation $\frac{dv}{dt} = r\frac{d\omega}{dt}$ can be expressed as

$$a_t = r\propto \tag{19A.2}$$

A Rotating Rigid Body

The characterization of the motion of a rotating rigid body has a lot in common with that of a particle traveling on a circle. In fact, every particle making up a rotating rigid body is

undergoing circular motion. But different particles making up the rigid body move on circles of different radii and hence have speeds and accelerations that differ from each other. For instance, each time the object goes around once, every particle of the object goes all the way around its circle once, but a particle far from the axis of rotation goes all the way around circle that is bigger than the one that a particle that is close to the axis of rotation goes around. To do that, the particle far from the axis of rotation must be moving faster. But in one rotation of the object, the line from the center of the circle that any particle of the object is on, to the particle, turns through exactly one rotation. In fact, the angular motion variables that we have been using to characterize the motion of a line extending from the center of a circle to a particle that is moving on that circle can be used to characterize the motion of a spinning rigid body as a whole. There is only one spin rate for the whole object, the angular velocity ω, and if that spin rate is changing, there is only one rate of change of the spin rate, the angular acceleration $\propto$. To specify the angular position of a rotating rigid body, we need to establish a reference line on the rigid body, extending away from a point on the axis of rotation in a direction perpendicular to the axis of rotation. This reference line rotates with the object. Its motion is the angular motion of the object. We also need a reference line segment that is fixed in space, extending from the same point on the axis, and away from the axis in a direction perpendicular to the axis. This one does not rotate with the object. Imagining the two lines to have at one time been collinear, the net angle through which the first line on the rigid body has turned relative to the fixed line is the angular position θ of the object.

The Constant Angular Acceleration Equations

While physically, there is a huge difference, mathematically, the rotational motion of a rigid body is identical to motion of a particle that only moves along a straight line. As in the case of linear motion, we have to define a positive direction. We are free to define the positive direction whichever way we want for a given problem, but we have to stick with that definition throughout the problem. Here, we establish a viewpoint some distance away from the rotating rigid body,but on the axis of rotation, and state that, from that viewpoint, counterclockwise is the positive sense of rotation, or alternatively, that clockwise is the positive sense of rotation. Whichever way we pick as positive, will be the positive sense of rotation for angular displacement (change in angular position), angular velocity, angular acceleration, and angular position relative to the reference line that is fixed in space. Next, we establish a zero for the time variable; we imagine a stopwatch to have been started at some instant that we define to be time zero. We call values of angular position and angular velocity, at that instant, the initial values of those quantities.

Given these criteria, we have the following table of corresponding quantities. Note that a rotational motion quantity is in no way equal to its linear motion counterpart, it simply plays a role in rotational motion that is mathematically similar to the role played by its counterpart in linear motion.

Linear Motion Quantity	Corresponding Angular Motion Quantity
x	θ
v	ω
a	$\propto$

The one variable that the two different kinds of motion do have in common is the stopwatch reading t.

Recall that, by definition,

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

$$\text{and} \quad \propto=\frac{d\omega}{dt}$$

While it is certainly possible for $\propto$ to be a variable, many cases arise in which $\propto$ is a constant. Such a case is a special case. The following set of constant angular acceleration equations apply in the special case of constant angular acceleration: (The derivation of these equations is mathematically equivalent to the derivation of the constant linear acceleration equations. Rather than derive them again, we simply present the results.)

$$\theta=\theta_0+\omega_0 t+\frac{1}{2}\propto t^2 \tag{19A.3}$$

$$\theta=\theta_0+\frac{\omega_0+\omega}{2}t \tag{19A.4}$$

$$\omega=\omega_0+\propto t \tag{19A.5}$$

$$\omega^2=\omega_0^2+2\propto\Delta\theta \tag{19A.6}$$

The rate at which a sprinkler head spins about a vertical axis increases steadily for the first 2.00 seconds of its operation such that, starting from rest, the sprinkler completes 15.0 revolutions clockwise (as viewed from above) during that first 2.00 seconds of operation. A nozzle, on the sprinkler head, at a distance of 11.0 cm from the axis of rotation of the sprinkler head, is initially due west of the axis of rotation. Find the direction and magnitude of the acceleration of the nozzle at the instant the sprinkler head completes its second (good to three significant figures) rotation.

Solution

We're told that the sprinkler head spin rate increases steadily, meaning that we are dealing with a constant angular acceleration problem, so, we can use the constant angular acceleration equations. The fact that there is a non-zero angular acceleration means that the nozzle will have some tangential acceleration $\vec{a_t}$. Also, the sprinkler head is spinning at the instant in question so the nozzle will have some centripetal acceleration $\vec{a_c}$. We'll have to find both $\vec{a_t}$ and $\vec{a_c}$ and add them like vectors to get the total acceleration of the nozzle. Let's get started by finding the angular acceleration $\propto$. We start with the first constant angular acceleration equation (equation 19A.3):

$$\theta = 0 + 0 \cdot t + \frac{1}{2} \propto t^2$$

The initial angular velocity ω_0 is given as zero. We have defined the initial angular position to be zero. This means that, at time $t = 2.00s$, the angular position θ is $15.0\,\text{rev} = 15.0\,\text{rev}\,\frac{2\pi\,\text{rad}}{\text{rev}} = 94.25\text{rad}$.

Solving equation 19A.3 above for $\propto$ yields:

$$\propto = \frac{2\theta}{t^2} \quad \propto = \frac{2(94.25\text{rad})}{(2.00s)^2} \quad \propto = 47.12\frac{\text{rad}}{s^2}$$

Substituting this result into equation 19A.2: $a_t = r \propto$ gives us $a_t = (.110m)47.12\text{rad}/s^2$

which evaluates to

$$a_t = 5.18\frac{m}{s^2}$$

Now we need to find the angular velocity of the sprinkler head at the instant it completes 2.00 revolutions. The angular acceleration $\propto$ that we found is constant for the first fifteen revolutions, so the value we found is certainly good for the first two turns. We can use it in the fourth constant angular acceleration equation (equation 19A.6):

$$\omega^2 = 0 + 2 \propto \Delta\theta$$ where $\Delta\theta = 2\text{ rev} = 2.00\text{ rev}\,\frac{2\pi\,\text{rad}}{\text{rev}} = 4.00\pi\text{ rad}$ $\omega = \sqrt{2 \propto \Delta\theta}$

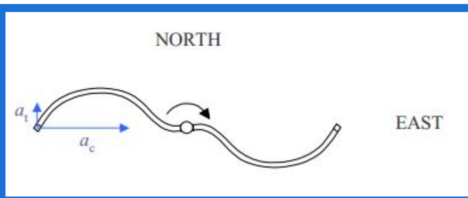
$$\omega = \sqrt{2(94.25\text{rad}/s^2)4.00\pi\text{rad}}$$ \omega=48.67 \mbox{rad}/s\label{19-6}

(at that instant when the sprinkler head completes its 2nd turn)

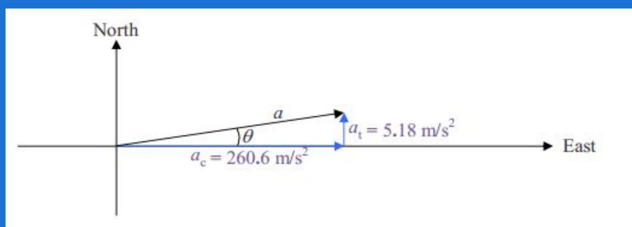
Now that we have the angular velocity, to get the centripetal acceleration we can use equation ???:

$$a_c = r\omega^2 \quad a_c = .110m(48.67\text{rad}/s)^2 \quad a_c = 260.6\frac{m}{s^2}$$

Given that the nozzle is initially at a point due west of the axis of rotation, at the end of 2.00 revolutions it will again be at that same point.



Now we just have to add the tangential acceleration and the centripetal acceleration vectorially to get the total acceleration. This is one of the easier kinds of vector addition problems since the vectors to be added are at right angles to each other.



From Pythagorean's theorem we have

$$a = \sqrt{a_c^2 + a_t^2}$$

$$a = \sqrt{(260.6 m/s^2)^2 + (5.18 m/s^2)^2}$$

$$a = 261 m/s^2$$

From the definition of the tangent of an angle as the opposite over the adjacent:

$$tan\theta = \frac{a_t}{a_c} \quad \theta = tan^{-1}\frac{5.18 m/s^2}{260.6 m/s^2} \quad \theta = 1.14^\circ$$

Thus,

$$a = 261 m/s^2 \quad \text{at } 1.14^\circ \text{ North of East}$$

When the Angular Acceleration is not Constant

The angular position of a rotating body undergoing constant angular acceleration is given, as a function of time, by our first constant angular acceleration equation, equation 19A.3:

$$\theta = \theta_0 + \omega_0 t + \frac{1}{2} \propto t^2$$

If we take the 2nd derivative of this with respect to time, we get the constant $\propto$. (Recall that the first derivative yields the angular velocity ω and that $\propto = \frac{d\omega}{dt}$.) The expression on the right side of $\theta = \theta_0 + \omega_0 t + \frac{1}{2} \propto t^2$ contains three terms: a constant, a term with t to the first power, and a term with t to the 2nd power. If you are given θ in terms of t, and it cannot be rearranged so that it appears as one of these terms or as a sum of two or all three such terms; then; $\propto$ is not a constant and you cannot use the constant angular acceleration equations. Indeed, if you are being asked to find the angular velocity at a particular instant in time, then you'll want to take the derivative $\frac{d\theta}{dt}$ and evaluate the result at the given stopwatch reading. Alternatively, if you are being asked to find the angular acceleration at a particular instant in time, then you'll want to take the second derivative $\frac{d^2\theta}{dt^2}$ and evaluate the result at the given stopwatch reading. Corresponding arguments can be made for the case of ω. If you are given ω as a function of t and the expression cannot be made to "look like" the constant angular acceleration equation $\omega = \omega_0 + \propto t$ then you are not dealing with a constant angular acceleration situation and you should not use the constant angular acceleration equations.

20A: Torque & Circular Motion

> The mistake that crops up in the application of Newton's 2nd Law for Rotational Motion involves the replacement of the sum of the torques about some particular axis, $\sum \tau_{o\circlearrowleft}$, with a sum of terms that are not all torques. Oftentimes, the errant sum will include forces with no moment arms (a force times a moment arm is a torque, but a force by itself is not a torque) and in other cases the errant sum will include a term consisting of a torque times a moment arm (a torque is already a torque, multiplying it by a moment arm yields something that is not a torque). Folks that are in the habit of checking units will catch their mistake as soon as they plug in values with units and evaluate.

We have studied the motion of spinning objects without any discussion of torque. It is time to address the link between torque and rotational motion. First, let's review the link between force and translational motion. (Translational motion has to do with the motion of a particle through space. This is the ordinary motion that you've worked with quite a bit. Until we started talking about rotational motion we called translational motion "motion." Now, to distinguish it from rotational motion, we call it translational motion.) The real answer to the question of what causes motion to persist, is nothing—a moving particle with no force on it keeps on moving at constant velocity. However, whenever the velocity of the particle is changing, there is a force. The direct link between force and motion is a relation between force and acceleration. The relation is known as Newton's 2^{nd} Law of Motion which we have written as equation ???:

$$\vec{a} = \frac{1}{m} \sum \vec{F}$$

in which,

- $\vec{a}$ is the acceleration of the object, how fast and which way its velocity is changing
- m is the mass, a.k.a. inertia, of the object. $\frac{1}{m}$ can be viewed as a sluggishness factor, the bigger the mass m, the smaller the value of $\frac{1}{m}$ and hence the smaller the acceleration of the object for a given net force. ("Net" in this context just means "total".)
- $\sum \vec{F}$ is the vector sum of the forces acting on the object, the net force.

We find a completely analogous situation in the case of rotational motion. The link in the case of rotational motion is between the angular acceleration of a rigid body and the torque being exerted on that rigid body.

$$\vec{\alpha} = \frac{1}{I} \sum \vec{\tau} \tag{20A.1}$$

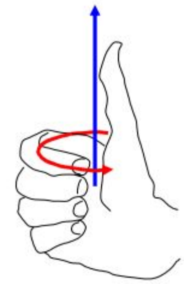
in which,

- $\vec{\alpha}$ is the angular acceleration of the rigid body, how fast and which way the angular velocity is changing
- I is the moment of inertia, a.k.a. the rotational inertia (but not just plain old inertia, which is mass) of the rigid body. It is the rigid body's inherent resistance to a change in how fast it (the rigid body) is spinning. ("Inherent" means "of itself", "part of its own being.") $\frac{1}{I}$ can be viewed as a sluggishness factor, the bigger the rotational inertia I, the smaller the value of $\frac{1}{I}$ and hence the smaller the angular acceleration of the object for a given net torque.
- $\sum \vec{\tau}$ is the net torque acting on the object. (A torque is what you are applying to a bottle cap or jar lid when you are trying to unscrew it.)

The Vector Nature of Torque and Angular Velocity

You've surely noticed the arrows over the letters used to represent torque, angular acceleration, and angular velocity; and as you know, the arrows mean that the quantities in question are vector quantities. That means that they have both magnitude and direction. Some explanation about the direction part is in order. Let's start with the torque. As mentioned, it is a twisting action such as that which you apply to bottle cap to loosen or tighten the bottle cap. There are two ways to specify the direction associated with torque. One way is to identify the axis of rotation about which the torque is being applied, then to establish a viewpoint, a position on the axis, at a location that is in one direction or the other direction away from the object. Then either state that the torque is clockwise, or state that it is counterclockwise, as viewed from the specified viewpoint. Note that it is not sufficient to identify the axis and state "clockwise" or "counterclockwise" without giving the viewpoint—a torque which is clockwise from one of the two viewpoints is counterclockwise from the other. The second method of specifying the direction is to give the torque vector direction. The convention for the torque vector is that the axis of rotation is the line on which the torque vector lies, and the direction is in accord with "the right hand rule for something curly something straight."

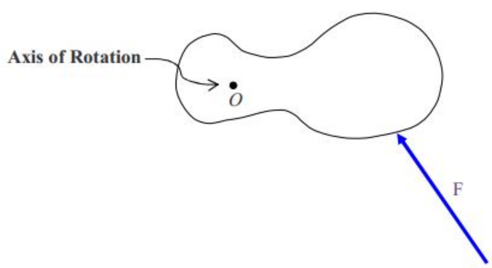
By the “the right hand rule for something curly something straight”, if you point the thumb of your cupped right hand in the direction of the torque vector, the fingers will be curled around in that direction which corresponds to the sense of rotation (counterclockwise as viewed from the head of the torque vector looking back along the shaft) of the torque.



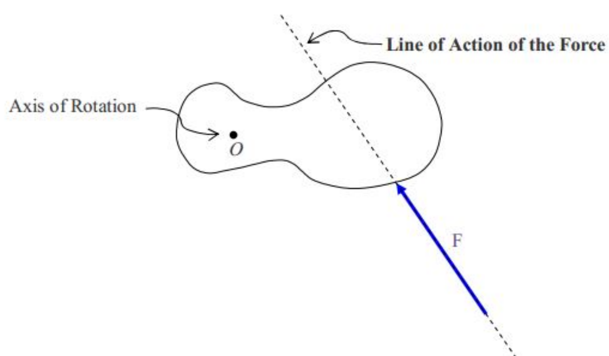
The angular acceleration vector $\vec{\propto}$ and the angular velocity vector $\vec{\omega}$ obey the same convention. These vectors, which point along the axis about which the rotation they represent occurs, are referred to as axial vectors.

The Torque Due to a Force

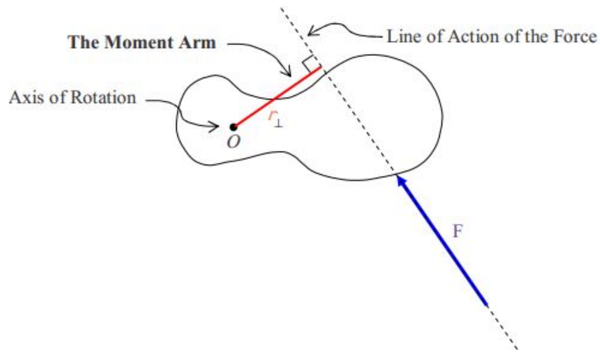
When you apply a force to a rigid body, you are typically applying a torque to that rigid body at the same time. Consider an object that is free to rotate about a fixed axis. We depict the object as viewed from a position on the axis, some distance away from the object. From that viewpoint, the axis looks like a dot. We give the name “point O” to the position at which the axis of rotation appears in the diagram and label it “O” in the diagram to make it easier to refer to later in this discussion. There is a force $\vec{F}$ acting on the object.



The magnitude of the torque due to a force is the magnitude of the force times the moment arm $r_{\perp}$ (read “r perp”) of the force. The moment arm (r_{\bot}\) is the perpendicular distance from the axis of rotation to the line of action of the force. The line of action of the force is a line that contains the force vector. Here we redraw the given diagram with the line of action of the force drawn in.



Next we extend a line segment from the axis of rotation to the line of action of the force, in such a manner that it meets the line of action of the force at right angles. The length of this line segment is the moment arm $r_{\perp}$.



The magnitude of the torque about the specified axis of rotation is just the product of the moment arm and the force.

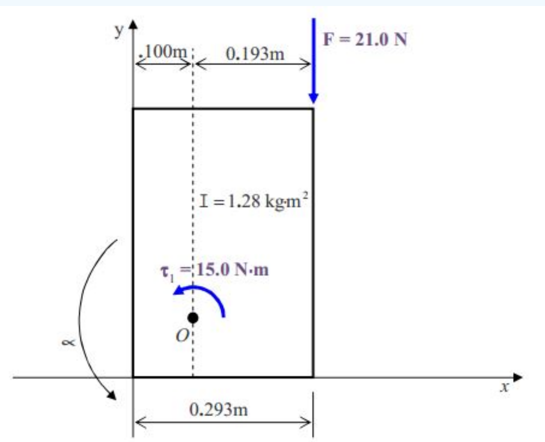
$$\tau = r_\perp F \tag{20A.2}$$

Applying Newton's Second Law for Rotational Motion in Cases Involving a Fixed Axis

Starting on the next page, we tell you what steps are required (and what diagram is required) in the solution of a fixed-axis "Newton's 2nd Law for Rotational Motion" problem by means of an example

A flat metal rectangular 293 mm $\times$ 452 mm plate lies on a flat horizontal frictionless surface with (at the instant in question) one corner at the origin of an x-y coordinate system and the opposite corner at point P which is at (293 mm, 452 mm). The plate is **pin connected to the horizontal surface** at (10.0 cm, 10.0 cm). A counterclockwise (as viewed from above) torque, with respect to the pin, of 15.0 N·m, is being applied to the plate and a force of 21.0 N in the –y direction is applied to the corner of the plate at point P. The moment of inertia of the plate, with respect to the pin, is 1.28 $kg \cdot m^2$. Find the angular acceleration of the plate at the instant for which the specified conditions prevail.

We start by drawing a pseudo free body diagram of the object as viewed from above (downward is "into the page"):

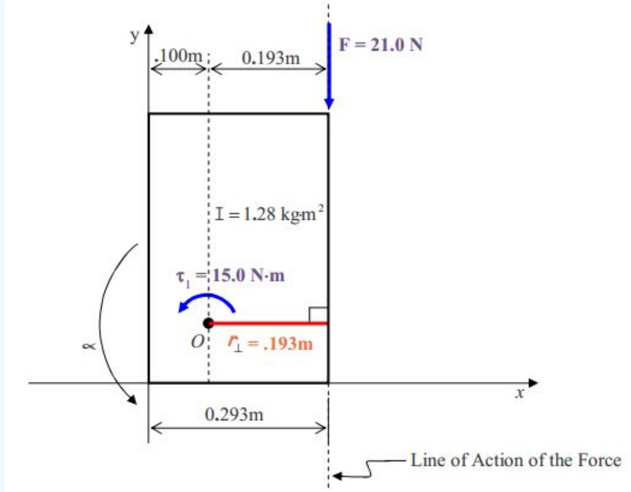


We refer to the diagram as a pseudo free body diagram rather than a free body diagram because:

a. We omit forces that are parallel to the axis of rotation (because they do not affect the rotation of the object about the axis of rotation). In the case at hand, we have omitted the force exerted on the plate by the gravitational field of the earth (which would be "into the page" in the diagram) as well as the normal force exerted by the frictionless surface on the plate ("out of the page").

b. We ignore forces exerted on the plate by the pin. (Such forces have no moment arm and hence do not affect the rotation of the plate about the axis of rotation. Note, a pin can, however, exert a frictional torque—assume it to be zero unless otherwise specified.)

Next we annotate the pseudo free body diagram to facilitate the calculation of the torque due to the force F:



Now we go ahead and apply Newton's 2nd Law for Rotational Motion, equation 20A.1:

$$\vec{\propto} = \frac{1}{I} \sum \vec{\tau}$$

As in the case of Newton's 2nd Law (for translational motion) this equation is three scalar equations in one, one equation for each of three mutually perpendicular axes about which rotation, under the most general circumstances, could occur. In the case at hand, the object is constrained to allow rotation about a single axis. In our solution, we need to indicate that we are summing torques about that axis, and we need to indicate which of the two possible rotational senses we are defining to be positive. We do that by means of the subscript $\circ_{\circlearrowleft}$ to be read "counterclockwise about point O." Newton's 2nd Law for rotational motion about the vertical axis (perpendicular to the page and represented by the dot labeled "O" in the diagram) reads:

$$\propto_{O_{\circlearrowleft}} = \frac{1}{I} \sum \tau_{O_{\circlearrowleft}} \tag{20A.3}$$

Now, when we replace the expression $\frac{1}{I} \sum \tau_{O_{\circlearrowleft}}$ with the actual term-by-term sum of the torques, we note that τ_1 is indeed counterclockwise as viewed from above (and hence positive) but that the force $\vec{F}$, where it is applied, would tend to cause a clockwise rotation of the plate, meaning
that the torque associated with force $\vec{F}$ is clockwise and hence, must enter the sum with a minus sign.

$$\propto = \frac{1}{I} (\tau_1 - r_\perp F$$

Substituting values with units yields:

$$\propto = \frac{1}{1.28 kg \cdot m^2} \left[15.0 N \cdot m - 0.193 m (21.0 N) \right]$$

Evaluating and rounding the answer to three significant figures gives us the final answer:

$$\propto = 8.55 \frac{\text{rad}}{s^2} (\text{counterclockwise as viewed from above})$$

Regarding the units we have:

$$\frac{1}{kg \cdot m^2} N \cdot m = \frac{1}{kg \cdot m} \frac{kg \cdot m}{s^2} = \frac{1}{s^2} = \frac{\text{rad}}{s^2}$$

where we have taken advantage of the fact that a newton is a $\frac{kg \cdot m}{s^2}$ and the fact that the radian is not a true unit but rather a marker that can be inserted as needed.

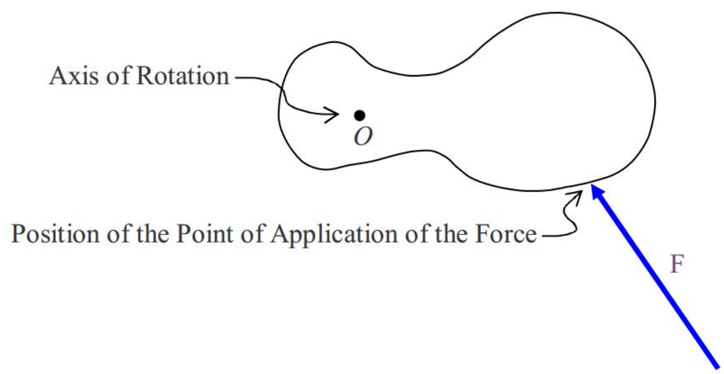
21A: Vectors - The Cross Product & Torque

> *Do not use your left hand when applying either the right-hand rule for the cross product of two vectors (discussed in this chapter) or the right-hand rule for "something curly something straight" discussed in the preceding chapter.*

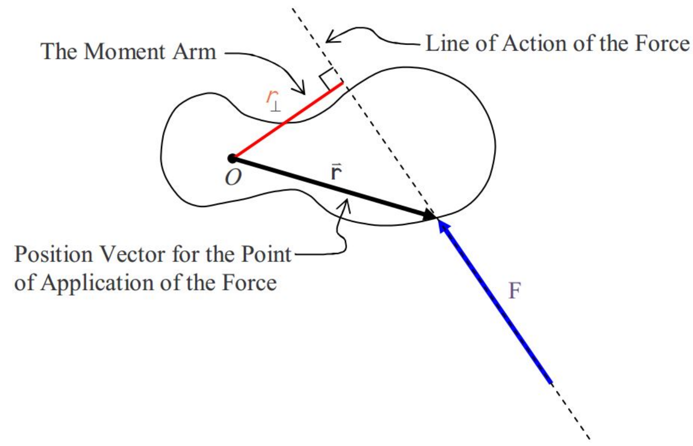
There is a **relational operator** for vectors that allows us to bypass the calculation of the moment arm. The relational operator is called the cross product. It is represented by the symbol "×" read "cross." The torque $\vec{\tau}$ can be expressed as the cross product of the position vector $\vec{r}$ for the point of application of the force, and the force vector $\vec{F}$ itself:

$$\vec{\tau} = \vec{r} \times \vec{F} \tag{21A.1}$$

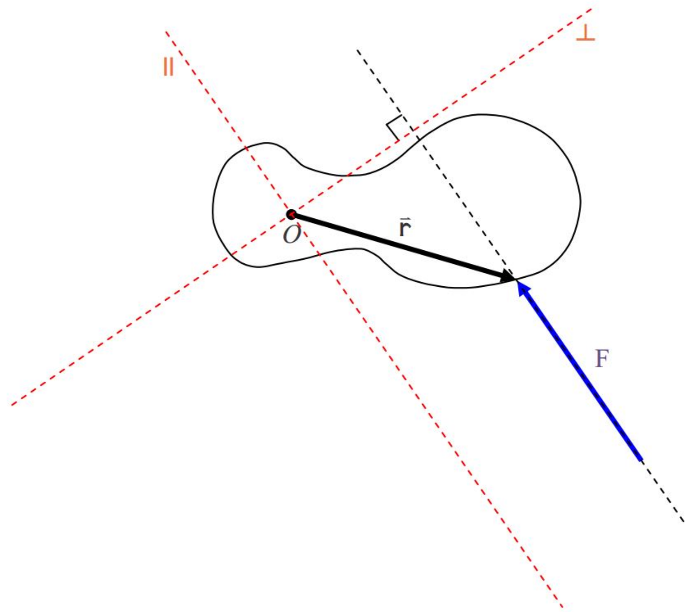
Before we begin our mathematical discussion of what we mean by the cross product, a few words about the vector $\vec{r}$ are in order. It is important for you to be able to distinguish between the position vector $\vec{r}$ for the force, and the moment arm, so we present them below in one and the same diagram. We use the same example that we have used before:



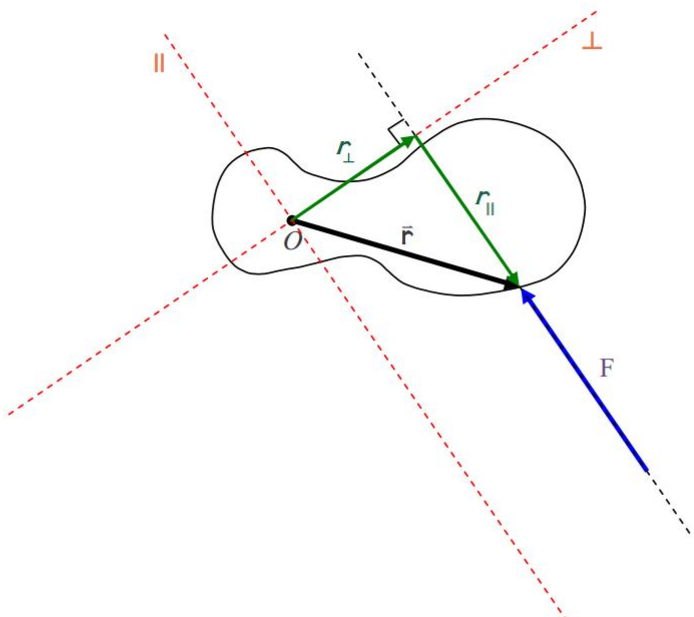
in which we are looking directly along the axis of rotation (so it looks like a dot) and the force lies in a plane perpendicular to that axis of rotation. We use the diagramatic convention that, the point at which the force is applied to the rigid body is the point at which one end of the arrow in the diagram touches the rigid body. Now we add the line of action of the force and the moment arm $r_{\perp}$ to the diagram, as well as the position vector $\vec{r}$ of the point of application of the force.



The moment arm can actually be defined in terms of the position vector for the point of application of the force. Consider a tilted x-y coordinate system, having an origin on the axis of rotation, with one axis parallel to the line of action of the force and one axis perpendicular to the line of action of the force. We label the x axis $\perp$ for "perpendicular" and the y axis $\|$ for "parallel."

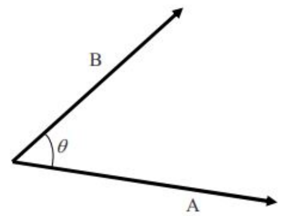


Now we break up the position vector $\vec{r}$ into its component vectors along the $\perp$ and $||$ axes.



From the diagram it is clear that the moment arm $r_\perp$ is just the magnitude of the component vector, in the perpendicular-to-the-force direction, of the position vector of the point of application of the force.

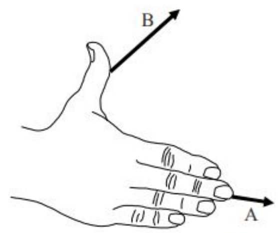
Now let's discuss the cross product in general terms. Consider two vectors, $\vec{A}$ and (\vec{B}\) that are neither parallel nor **anti-parallel** to each other. Two such vectors define a plane. Let that plane be the plane of the page and define θ to be the smaller of the two angles between the two vectors when the vectors are drawn tail to tail.



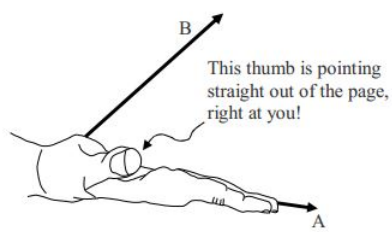
The magnitude of the cross product vector $\vec{A} \times \vec{B}$ is given by

$$|\vec{A} \times \vec{B}| = ABsin\theta \tag{21A.2}$$

Keeping your fingers aligned with your forearm, point your fingers in the direction of the first vector (the one that appears before the "$\times$" in the mathematical expression for the cross product; e.g. the $\vec{A}$ in $vecA \times \vec{B}$).



Now rotate your hand, as necessary, about an imaginary axis extending along your forearm and along your middle finger, until your hand is oriented such that, if you were to close your fingers, they would point in the direction of the second vector.



Your thumb is now pointing in the direction of the cross product vector. $\vec{C} = \vec{A} \times \vec{B}$. The cross product vector $\vec{C}$ is always perpendicular to both of the vectors that are in the cross product (the $\vec{A}$ and the $\vec{B}$ in the case at hand). Hence, if you draw them so that both of the vectors that are in the cross product are in the plane of the page, the cross product vector will always be perpendicular to the page, either straight into the page, or straight out of the page. In the case at hand, it is straight out of the page.

When we use the cross product to calculate the torque due to a force $\vec{F}$ whose point of application has a position vector $\vec{r}$, relative to the point about which we are calculating the torque, we get an axial torque vector $\vec{\tau}$. To determine the sense of rotation that such a torque vector would correspond to, about the axis defined by the torque vector itself, we use The Right Hand Rule For Something Curly Something Straight. Note that we are calculating the torque with respect to a point rather than an axis—the axis about which the torque acts, comes out in the answer.

Calculating the Cross Product of Vectors that are Given in $\hat{i}$, $\hat{j}$, $\hat{k}$ Notation

Unit vectors allow for a straightforward calculation of the cross product of two vectors under even the most general circumstances, e.g. circumstances in which each of the vectors is pointing in an arbitrary direction in a three-dimensional space. To take advantage of the method, we need to know the cross product of the Cartesian coordinate axis unit vectors $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$ with each other.

First off, we should note that any vector crossed into itself gives zero. This is evident from Equation 21A.2:

$$|\vec{A} \times \vec{B}| = ABsin\theta$$

because if A and B are in the same direction, then $\theta = 0^\circ$, and since $sin 0^\circ = 0$, we have $|\vec{A} \times \vec{B}| = 0$. Regarding the unit vectors, this means that:

$$\hat{\imath} \times \hat{\imath} = 0$$

$$\hat{\jmath} \times \hat{\jmath} = 0$$

$$\hat{k} \times \hat{k} = 0$$

Next we note that the magnitude of the cross product of two vectors that are perpendicular to each other is just the ordinary product of the magnitudes of the vectors. This is also evident from equation 21A.2:

$$|\vec{A} \times \vec{B}| = AB \sin\theta$$

because if $\vec{A}$ is perpendicular to $\vec{B}$ then $\theta = 90^\circ$ and $sin 90^\circ = 1$ so

$$|\vec{A} \times \vec{B}| = AB$$

Now if $\vec{A}$ and $\vec{B}$ are unit vectors, then their magnitudes are both 1, so, the product of their magnitudes is also 1. Furthermore, the unit vectors $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$ are all perpendicular to each other so the magnitude of the cross product of any one of them with any other one of them is the product of the two magnitudes, that is, 1.

Now how about the direction? Let's use the right hand rule to get the direction of $\hat{\imath} \times \hat{\jmath}$:

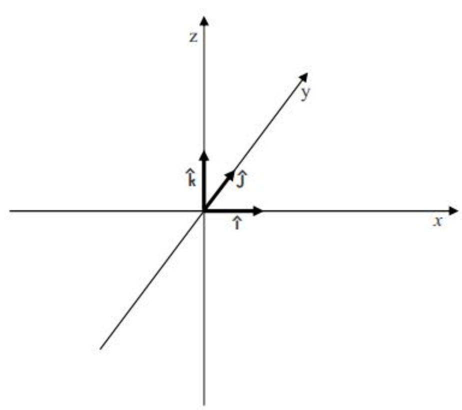


Figure 1

With the fingers of the right hand pointing directly away from the right elbow, and in the same direction as $hati$, (the first vector in '$\hat{\imath} \times \hat{\jmath}$') to make it so that if one were to close the fingers, they would point in the same direction as $\hat{\jmath}$, the palm must be facing in the +y direction. That being the case, the extended thumb must be pointing in the +z direction. Putting the magnitude (the magnitude of each unit vector is 1) and direction (+z) information together we see that $\hat{\imath} \times \hat{\jmath} = \hat{k}$. Similarly: $\hat{\jmath} \times \hat{k} = \hat{\imath}$, $\hat{k} \times \hat{\imath} = \hat{\jmath}$, $\hat{\jmath} \times \hat{\imath} = -\hat{k}$, $\hat{k} \times \hat{\jmath} = -\hat{\imath}$, and $\hat{\imath} \times \hat{k} = -\hat{\jmath}$. One way of remembering this is to write $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$ twice in succession:

$$\hat{\imath}, \hat{\jmath}, \hat{k}, \hat{\imath}, \hat{\jmath}, \hat{k}$$

Then, crossing any one of the first three vectors into the vector immediately to its right yields the next vector to the right. But crossing any one of the last three vectors into the vector immediately to its left yields the negative of the next vector to the left (left-to-right "+", but right-to-left "-").

Now we're ready to look at the general case. Any vector $\vec{A}$ can be expressed in terms of unit vectors:

$$\vec{A} = A_x \hat{\imath} + A_y \hat{\jmath} + A_z \hat{k}$$

Doing the same for a vector $\vec{B}$ then allows us to write the cross product as:

$$\vec{A} \times \vec{B} = (A_x \hat{\imath} + A_y \hat{\jmath} + A_z \hat{k}) \times (B_x \hat{\imath} + B_y \hat{\jmath} + B_z \hat{k})$$

Using the distributive rule for multiplication we can write this as:

$$\vec{A}\times\vec{B}=A_x\hat{i}\times(B_x\hat{i}+B_y\hat{j}+B_z\hat{k})+A_y\hat{j}\times(B_x\hat{i}+B_y\hat{j}+B_z\hat{k})+A_z\hat{k}\times(B_x\hat{i}+B_y\hat{j}+B_z\hat{k})$$

$$\vec{A}\times\vec{B}=A_x\hat{i}\times B_x\hat{i}+A_x\hat{i}\times B_y\hat{j}+A_x\hat{i}\times B_z\hat{k}+A_y\hat{j}\times B_x\hat{i}+A_y\hat{j}\times B_y\hat{j}+A_y\hat{j}\times B_z\hat{k}+A_z\hat{k}\times B_x\hat{i}+A_z\hat{k}\times B_y\hat{j}+A_z\hat{k}\times B_z\hat{k}$$

Using, in each term, the commutative rule and the associative rule for multiplication we can write this as:

$$\vec{A}\times\vec{B}=A_xB_x(\hat{i}\times\hat{i})+A_xB_y(\hat{i}\times\hat{j})+A_xB_z(\hat{i}\times\hat{k})+A_yB_x(\hat{j}\times\hat{i})+A_yB_y(\hat{j}\times\hat{j})+A_yB_z(\hat{j}\times\hat{k})+A_zB_x(\hat{k}\times\hat{i})+A_zB_y(\hat{k}\times\hat{j})+A_zB_z(\hat{k}\times\hat{k})$$

Now we evaluate the cross product that appears in each term:

$$\vec{A}\times\vec{B}=A_xB_x(0)+A_xB_y(\hat{k})+A_xB_z(-\hat{j})+A_yB_x(-\hat{k})+A_yB_y(0)+A_yB_z(\hat{i})+A_zB_x(\hat{j})+A_zB_y(-\hat{i})+A_zB_z(0)$$

Eliminating the zero terms and grouping the terms with $\hat{i}$ together, the terms with $\hat{j}$ together, and the terms with $\hat{k}$ together yields:

$$\vec{A}\times\vec{B}=A_yB_z(\hat{i})+A_zB_y(-\hat{i})+A_zB_x(\hat{j})+A_xB_z(-\hat{j})+A_xB_y(\hat{k})+A_yB_x(-\hat{k})$$

Factoring out the unit vectors yields:

$$\vec{A}\times\vec{B}=(A_yB_z-A_zB_y)\hat{i}+(A_zB_x-A_xB_z)\hat{j}+(A_xB_y-A_yB_x)\hat{k}$$

which can be written on one line as:

$$\vec{A}\times\vec{B}=(A_yB_z-A_zB_y)\hat{i}+(A_zB_x-A_xB_z)\hat{j}+(A_xB_y-A_yB_x)\hat{k} \tag{21A.3}$$

This is our end result. We can arrive at this result much more quickly if we borrow a tool from that branch of mathematics known as linear algebra (the mathematics of matrices).

We form the 3×3 matrix

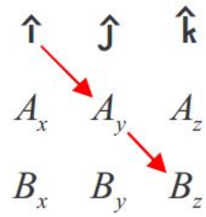
$$\begin{bmatrix}\hat{i} & \hat{j} & \hat{k}\\ A_x & A_y & A_z\\ B_x & B_y & B_z\end{bmatrix}$$

by writing \hat{i}, \hat{j}, \hat{k} as the first row, then the components of the first vector that appears in the cross product as the second row, and finally the components of the second vector that appears in the cross product as the last row. It turns out that the cross product is equal to the determinant of that matrix. We use absolute value signs on the entire matrix to signify "the determinant of the matrix." So we have:

$$\vec{A}\times\vec{B}=\begin{vmatrix}\hat{i} & \hat{j} & \hat{k}\\ A_x & A_y & A_z\\ B_x & B_y & B_z\end{vmatrix} \tag{21A.4}$$

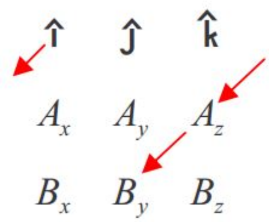
To take the determinant of a 3×3 matrix you work your way across the top row. For each element in that row you take the product of the elements along the diagonal that extends down and to the right, minus the product of the elements down and to the left; and you add the three results (one result for each element in the top row) together. If there are no elements down and to the appropriate side, you move over to the other side of the matrix (see below) to complete the diagonal.

For the first element of the first row, the $\hat{i}$, take the product down and to the right,



(this yields $\hat{\imath} A_y B_z$)

minus the product down and to the left



(the product down-and-to-the-left is $\hat{\imath} A_z B_y$)

For the first element in the first row, we thus have: $\hat{\imath} A_y B_z - \hat{\imath} A_z B_y$ which can be written as: $(A_y B_z - A_z B_y)\hat{\imath}$. Repeating the process for the second and third elements in the first row (the $\hat{\jmath}$ and the $\hat{k}$) we get $(A_z B_x - A_x B_z)\hat{\jmath}$ and $(A_x B_y - A_y B_x)\hat{k}$ respectively. Adding the three results, to form the determinant of the matrix results in:

$$\vec{A} \times \vec{B} = (A_y B_z - A_z B_y)\hat{\imath} + (A_z B_x - A_x B_z)\hat{\jmath} + (A_x B_y - A_y B_x)\hat{k}$$

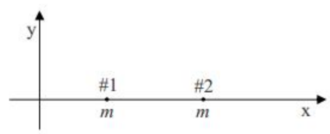
as we found before, “the hard way.”

22A: Center of Mass, Moment of Inertia

> *A mistake that crops up in the calculation of moments of inertia, involves the Parallel Axis Theorem. The mistake is to interchange the moment of inertia of the axis through the center of mass, with the one parallel to that, when applying the Parallel Axis Theorem. Recognizing that the subscript "CM" in the parallel axis theorem stands for "center of mass" will help one avoid this mistake. Also, a check on the answer, to make sure that the value of the moment of inertia with respect to the axis through the center of mass is smaller than the other moment of inertia, will catch the mistake.*

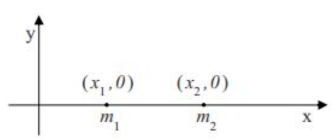
Center of Mass

Consider two particles, having one and the same mass m, each of which is at a different position on the x axis of a Cartesian coordinate system.



Common sense tells you that the average position of the material making up the two particles is midway between the two particles. Common sense is right. We give the name "center of mass" to the average position of the material making up a distribution, and the center of mass of a pair of same-mass particles is indeed midway between the two particles.

How about if one of the particles is more massive than the other? One would expect the center of mass to be closer to the more massive particle, and again, one would be right. To determine the position of the center of mass of the distribution of matter in such a case, we compute a weighted sum of the positions of the particles in the distribution, where the weighting factor for a given particle is that fraction, of the total mass, that the particle's own mass is. Thus, for two particles on the x axis, one of mass m_1, at x_1, and the other of mass m_2, at x_2,



the position $\bar{x}$ of the center of mass is given by

$$\bar{x}=\frac{m_1}{m_1+m_2}x_1+\frac{m_2}{m_1+m_2}x_2 \tag{22A.5}$$

Note that each weighting factor is a proper fraction and that the sum of the weighting factors is always 1. Also note that if, for instance, m_1 is greater than m_2, then the position x_1 of particle 1 will count more in the sum, thus ensuring that the center of mass is found to be closer to the more massive particle (as we know it must be). Further note that if $m_1=m_2$, each weighting factor is $\frac{1}{2}$, as is evident when we substitute m for both m_1 and m_2 in Equation 22A.5:

$$\bar{x}=\frac{m}{m+m}x_1+\frac{m}{m+m}x_2$$

$$\bar{x}=\frac{1}{2}x_1+\frac{1}{2}x_2$$

$$\bar{x}=\frac{x_1+x_2}{2}$$

The center of mass is found to be midway between the two particles, right where common sense tells us it has to be.

The Center of Mass of a Thin Rod

Quite often, when the finding of the position of the center of mass of a distribution of particles is called for, the distribution of particles is the set of particles making up a rigid body. The easiest rigid body for which to calculate the center of mass is the thin rod because it extends in only one dimension. (Here, we discuss an ideal thin rod. A physical thin rod must have some nonzero diameter. The ideal thin rod, however, is a good approximation to the physical thin rod as long as the diameter of the rod is small compared to its length.)

In the simplest case, the calculation of the position of the center of mass is trivial. The simplest case involves a uniform thin rod. A uniform thin rod is one for which the linear mass density μ, the mass-per-length of the rod, has one and the same value at all points on the rod. The center of mass of a uniform rod is at the center of the rod. So, for instance, the center of mass of a uniform rod that extends along the x axis from $x=0$ to $x=L$ is at (L/2, 0).

The linear mass density μ, typically called linear density when the context is clear, is a measure of how closely packed the elementary particles making up the rod are. Where the linear density is high, the particles are close together.

To picture what is meant by a non-uniform rod, a rod whose linear density is a function of position, imagine a thin rod made of an alloy consisting of lead and aluminum. Further imagine that the percentage of lead in the rod varies smoothly from 0% at one end of the rod to 100% at the other. The linear density of such a rod would be a function of the position along the length of the rod. A one-millimeter segment of the rod at one position would have a different mass than that of a one-millimeter segment of the rod at a different position.

People with some exposure to calculus have an easier time understanding what linear density is than calculus-deprived individuals do because linear density is just the ratio of the amount of mass in a rod segment to the length of the segment, in the limit as the length of the segment goes to zero. Consider a rod that extends from 0 to L along the x axis. Now suppose that $m_s(x)$ is the mass of that segment of the rod extending from 0 to x where $x \geq 0$ but $x < L$. Then, the linear density of the rod at any point x along the rod, is just $\frac{dm_s}{dx}$ evaluated at the value of x in question.

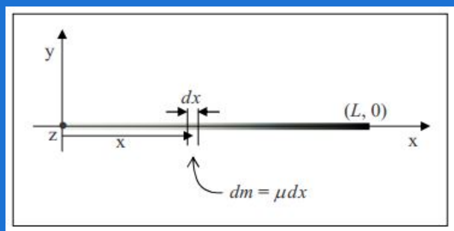
Now that you have a good idea of what we mean by linear mass density, we are going to illustrate how one determines the position of the center of mass of a non-uniform thin rod by means of an example.

Solution Find the position of the center of mass of a thin rod that extends from 0 to $.890\text{m}$ along the x axis of a Cartesian coordinate system and has a linear density given by $\mu(x)=0.650\frac{kg}{m^3}x^2$.

In order to be able to determine the position of the center of mass of a rod with a given length and a given linear density as a function of position, you first need to be able to find the mass of such a rod. To do that, one might be tempted to use a method that works only for the special case of a uniform rod, namely, to try using $m=\mu L$ with L being the length of the rod. The problem with this is, that μ varies along the entire length of the rod. What value would one use for μ? One might be tempted to evaluate the given μ at $x=L$ and use that, but that would be acting as if the linear density were constant at $\mu=\mu(L)$. It is not. In fact, in the case at hand, $\mu(L)$ is the maximum linear density of the rod, it only has that value at one point on the rod.

What we can do is to say that the infinitesimal amount of mass dm in a segment dx of the rod is μdx. Here we are saying that at some position x on the rod, the amount of mass in the infinitesimal length dx of the rod is the value of μ at that x value, times the infinitesimal length dx. Here we don't have to worry about the fact that μ changes with position since the segment dx is infinitesimally long, meaning, essentially, that it has zero length, so the whole segment is essentially at one position x and hence the value of μ at that x is good for the whole segment dx.

dm=\mu (x) dx \label{22-2}



Now this is true for any value of x, but it just covers an infinitesimal segment of the rod at x. To get the mass of the whole rod, we need to add up all such contributions to the mass. Of course, since each dm corresponds to an infinitesimal length of the rod, we will have an infinite number of terms in the sum of all the dm's. An infinite sum of infinitesimal terms, is an integral.

\int dm=\int_{0}{L} \mu (x) dx \label{22-3}

where the values of x have to run from 0 to L to cover the length of the rod, hence the limits on the right. Now the mathematicians have provided us with a rich set of algorithms for evaluating integrals, and indeed we will have to reach into that toolbox to evaluate the integral on the right, but to evaluate the integral on the left, we cannot, should not, and will not turn to such an algorithm. Instead, we use common sense and our conceptual understanding of what the integral on the left means. In the context of the problem at hand, $\int dm$ means "the sum of all the infinitesimal bits of mass making up the rod." Now, if you add up all the infinitesimal bits of mass making up the rod, you get the mass of the rod. So $\int dm$ is just the mass of the rod, which we will call m. Equation $22A.2$ then becomes

m=\int_{0}{L} \mu(x) dx \label{22-4}

Replacing $\mu(x)$ with the given expression for the linear density $\mu=0.650\frac{kg}{m^3}x^2$ which I choose to write as $\mu=bx^2$ with b being defined by $b=0.650\frac{kg}{m^3}$ we obtain

$$m=\int_0 Lbx^2dx$$

Factoring out the constant yields

$$m=b\int_0 Lx^2dx$$

When integrating the variable of integration raised to a power all we have to do is increase the power by one and divide by the new power. This gives

$$m=b\frac{x^3}{3}\Big|_0^L$$

Evaluating this at the lower and upper limits yields

$$m=b(\frac{L^3}{3}-\frac{0^3}{3})\quad m=\frac{bL^3}{3}$$

The value of L is given as $0.890m$ and we defined b to be the constant $0.650\frac{kg}{m^3}$ in the given expression for μ, $\mu = 0.650\frac{kg}{m^3}x^2$, so

$$m = \frac{0.650\frac{kg}{m^3}(0.890m)^3}{3} \quad m = 0.1527kg$$

That's a value that will come in handy when we calculate the position of the center of mass. Now, when we calculated the center of mass of a set of discrete particles (where a discrete particle is one that is by itself, as opposed, for instance, to being part of a rigid body) we just carried out a weighted sum in which each term was the position of a particle times its weighting factor and the weighting factor was that fraction, of the total mass, represented by the mass of the particle. We carry out a similar procedure for a continuous distribution of mass such as that which makes up the rod in question. Let's start by writing one single term of the sum. We'll consider an infinitesimal length dx of the rod at a position x along the length of the rod. The position, as just stated, is x, and the weighting factor is that fraction of the total mass m of the rod that the mass dm of the infinitesimal length dx represents. That means the weighting factor is $\frac{dm}{m}$, so, a term in our weighted sum of positions looks like:

$$\frac{dm}{m}x$$

Now, dm can be expressed as μdx so our expression for the term in the weighted sum can be written as

$$\frac{\mu dx}{m}x$$

That's one term in the weighted sum of positions, the sum that yields the position of the center of mass. The thing is, because the value of x is unspecified, that one term is good for any infinitesimal segment of the bar. Every term in the sum looks just like that one. So we have an expression for every term in the sum. Of course, because the expression is for an infinitesimal length dx of the rod, there will be an infinite number of terms in the sum. So, again we have an infinite sum of infinitesimal terms. That is, again we have an integral. Our expression for the position of the center of mass is:

$$\bar{x} = \int_0 L\frac{\mu dx}{m}x$$

Substituting the given expression $\mu(x) = 0.650\frac{kg}{m^3}x^2$ for μ, which we again write as $\mu = bx^2$ with b being defined by $b = 0.650\frac{kg}{m^3}$, yields

$$\bar{x} = \int_0 L\frac{bx^2dx}{m}x$$

Rearranging and factoring the constants out gives

$$\bar{x} = \frac{b}{m}\int_0 Lx^3dx$$

Next we carry out the integration

$$\bar{x} = \frac{b}{m}\frac{x^4}{4}\Big|_0^L \quad \bar{x} = \frac{b}{m}(\frac{L^4}{4} - \frac{0^4}{4}) \quad \bar{x} = \frac{bL^4}{4m}$$

Now we substitute values with units; the mass m of the rod that we found earlier, the constant b that we defined to simplify the appearance of the linear density function, and the given length L of the rod:

$$\bar{x} = 0.668m$$

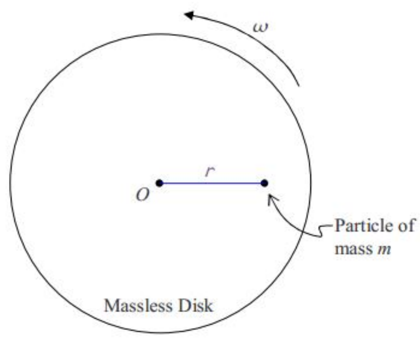
$$\bar{x}=\frac{(0.650\frac{kg}{m^3})(0.890m)^4}{4(0.1527kg)}$$

This is our final answer for the position of the center of mass. Note that it is closer to the denser end of the rod, as we would expect. The reader may also be interested to note that had we substituted the expression $m=\frac{bL^3}{3}$ that we derived for the mass, rather than the value we obtained when we evaluated that expression, our expression for $\bar{x}$ would have simplified to $\frac{3}{4}L$ which evaluates to $\bar{x}=0.668m$, the same result as the one above.

Moment of Inertia—a.k.a. Rotational Inertia

You already know that the moment of inertia of a rigid object, with respect to a specified axis of rotation, depends on the mass of that object, and how that mass is distributed relative to the axis of rotation. In fact, you know that if the mass is packed in close to the axis of rotation, the object will have a smaller moment of inertia than it would if the same mass was more spread out relative to the axis of rotation. Let's quantify these ideas. (Quantify, in this context, means to put into equation form.)

We start by constructing, in our minds, an idealized object for which the mass is all concentrated at a single location which is not on the axis of rotation: Imagine a massless disk rotating with angular velocity w about an axis through the center of the disk and perpendicular to its faces. Let there be a particle of mass m embedded in the disk at a distance r from the axis of rotation. Here's what it looks like from a viewpoint on the axis of rotation, some distance away from the disk:



where the axis of rotation is marked with an O. Because the disk is massless, we call the moment of inertia of the construction, the moment of inertia of a particle, with respect to rotation about an axis from which the particle is a distance r.

Knowing that the velocity of the particle can be expressed as $v=r\omega$ you can show yourself how I must be defined in order for the kinetic energy expression $K=\frac{1}{2}I\omega^2$ for the object, viewed as a spinning rigid body, to be the same as the kinetic energy expression $K=\frac{1}{2}mv^2$ for the particle moving through space in a circle. Either point of view is valid so both viewpoints must yield the same kinetic energy. Please go ahead and derive what I must be and then come back and read the derivation below.

Here is the derivation:

Given that $K=\frac{1}{2}mv^2$, we replace v with $r\omega$.

This gives $K=\frac{1}{2}m(r\omega)^2$

which can be written as

$$K=\frac{1}{2}(mr^2)\omega^2$$

For this to be equivalent to

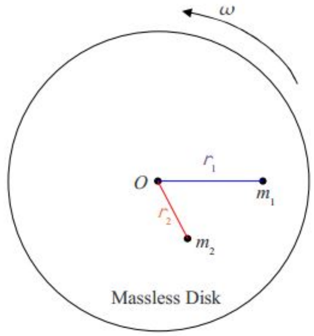
$$K=\frac{1}{2}I\omega^2$$

we must have

$$I=mr^2 \tag{22A.6}$$

This is our result for the moment of inertia of a particle of mass m, with respect to an axis of rotation from which the particle is a distance r.

Now suppose we have two particles embedded in our massless disk, one of mass m_1 at a distance r_1 from the axis of rotation and another of mass m_2 at a distance r_2 from the axis of rotation.



The moment of inertia of the first one by itself would be

$$I_1=m_1r_1^2$$

and the moment of inertia of the second particle by itself would be

$$I_2=m_2r_2^2$$

The total moment of inertia of the two particles embedded in the massless disk is simply the sum of the two individual moments of inertial.

$$I=I_1+I_2$$

$$I=m_1r_1^2+m_2r_2^2$$

This concept can be extended to include any number of particles. For each additional particle, one simply includes another $m_ir_i^2$ term in the sum where m_i is the mass of the additional particle and r_i is the distance that the additional particle is from the axis of rotation. In the case of a rigid object, we subdivide the object up into an infinite set of infinitesimal mass elements dm. Each mass element contributes an amount of moment of inertia

dI=r^2dm \label{22-6}

to the moment of inertia of the object, where r is the distance that the particular mass element is from the axis of rotation.

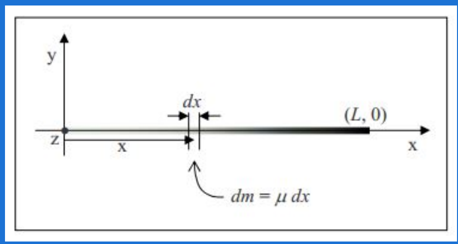
Find the moment of inertia of the rod in Example 22A.5 with respect to rotation about the z axis.

Solution

In Example 22A.5, the linear density of the rod was given as $\mu = 0.650\frac{kg}{m^3}x^2$. To reduce the number of times we have to write the value in that expression, we will write it as $\mu = bx^2$ with b being defined as $b = 0.650\frac{kg}{m^3}$.

The total moment of inertia of the rod is the infinite sum of the infinitesimal contributions

dI=r^2 dm \label{22-6} from each and every mass element dm making up the rod.



In the diagram, we have indicated an infinitesimal element dx of the rod at an arbitrary position on the rod. The z axis, the axis of rotation, looks like a dot in the diagram and the distance r in $dI = r^2 dm$, the distance that the bit of mass under consideration is from the axis of rotation, is simply the abscissa x of the position of the mass element. Hence, equation 22A.4 for the case at hand can be written as

$dI = x^2 dm$ which we copy here $dI = x^2 dm$

By definition of the linear mass density μ, the infinitesimal mass dm can be expressed as $dm = \mu dx$. Substituting this into our expression for dI yields

$$dI = x^2 \mu dx$$

Now μ was given as bx^2 (with b actually being the symbol that I chose to use to represent the given constant $0.650\frac{kg}{m^3}$). Substituting bx^2 in for μ in our expression for dI yields

$$dI = x^2(bx^2)dx \quad dI = bx^4 dx$$

This expression for the contribution of an element dx of the rod to the total moment of inertia of the rod is good for every element dx of the rod. The infinite sum of all such infinitesimal contributions is thus the integral

$$\int dI = \int_0 Lbx^4 dx$$

Again, as with our last integration, on the left, we have not bothered with limits of integration— the infinite sum of all the infinitesimal contributions to the moment of inertia is simply the total moment of inertia.

$$I = \int_0 Lbx^4 dx$$

On the right we use the limits of integration 0 to L to include every element of the rod which extends from $x = 0$ to $x = L$, with L given as $0.890m$. Factoring out the constant b gives us

$I = b \int_0 Lx^4 dx$ Now we carry out the integration: $I = b\frac{x^5}{5}\Big|_0^L$ $I = b(\frac{L^5}{5} - \frac{0^5}{5})$ $I = b\frac{L^5}{5}$

Substituting the given values of b and L yields: $I = 0.650\frac{kg}{m^3}\frac{(0.890m)^5}{5}$ $I = 0.0726kg \cdot m^2$

The Parallel Axis Theorem

We state, without proof , the parallel axis theorem:

$$I = I_{cm} + md^2 \tag{22A.7}$$

in which:

- I is the moment of inertia of an object with respect to an axis from which the center of mass of the object is a distance d.
- I_{cm} is the moment of inertia of the object with respect to an axis that is parallel to the first axis and passes through the center of mass.
- m is the mass of the object
- d is the distance between the two axes.

The parallel axis theorem relates the moment of inertia I_{CM} of an object, with respect to an axis through the center of mass of the object, to the moment of inertia I of the same object, with respect to an axis that is parallel to the axis through the center of mass and is at a distance d from the axis through the center of mass.

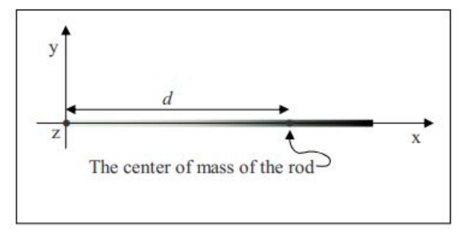
A conceptual statement made by the parallel axis theorem is one that you probably could have arrived at by means of common sense, namely that the moment of inertia of an object with respect to an axis through the center of mass is smaller than the moment of inertia about any axis parallel to that one. As you know, the closer the mass is “packed” to the axis of rotation, the smaller the moment of inertia; and; for a given object, per definition of the center of mass, the mass is packed most closely to the axis of rotation when the axis of rotation passes through the center of mass.

Find the moment of inertia of the rod from examples 22A.5 and 22A.1, with respect to an axis that is perpendicular to the rod and passes through the center of mass of the rod.

Solution

Recall that the rod in question extends along the x axis from $x = 0$ to $x = L$ with $L = 0.890m$ and that the rod has a linear density given by $\mu = bL^2$ with $b = 0.650\frac{kg}{m^3}x^2$.

The axis in question can be chosen to be one that is parallel to the z axis, the axis about which, in solving example 22A.1, we found the moment of inertia to be $I = 0.0726kg \cdot m^2$. In solving example 22A.5 we found the mass of the rod to be $m = 0.1527kg$ and the center of mass of the rod to be at a distance $d = 0.668m$ away from the z axis. Here we present the solution to the problem:



$I = I_{cm} + md^2$ $I_{cm} = I - md^2$

$I_{cm} = 0.0726kg \cdot m^2 - 0.1527kg(0.668m)^2$

$I_{cm} = 0.0047kg \cdot m^2$

23A: Statics

> *It bears repeating: Make sure that any force that enters the torque equilibrium equation is multiplied by a moment arm, and that any pure torque (such as τo in the solution of example 23-2 on page 151) that enters the torque equilibrium equation is NOT multiplied by a moment arm.*

For any rigid body, at any instant in time, Newton's 2nd Law for translational motion

$$\vec{a} = \frac{1}{m}\sum \vec{F}$$

and Newton's 2nd Law for Rotational motion

$$\vec{\propto} = \frac{1}{I}\sum \vec{\tau}$$

both apply. In this chapter we focus on rigid bodies that are in equilibrium. This topic, the study of objects in equilibrium, is referred to as statics. Being in equilibrium means that the acceleration and the angular acceleration of the rigid body in question are both zero. When $\vec{a} = 0$, Newton's 2nd Law for translational motion boils down to

$$\sum \vec{F} = 0 \tag{23A.2}$$

and when $\vec{\propto} = 0$, Newton's 2nd Law for Rotational motion becomes

$$\sum \vec{\tau} = 0 \tag{23A.3}$$

These two vector equations are called the equilibrium equations. They are also known as the equilibrium conditions. In that each of the vectors has three components, the two vector equations actually represent a set of six scalar equations:

$$\sum F_x = 0$$

$$\sum F_y = 0$$

$$\sum F_z = 0$$

$$\sum \tau_{x\circlearrowleft} = 0$$

$$\sum \tau_{y\circlearrowleft} = 0$$

$$\sum \tau_{z\circlearrowleft} = 0$$

In many cases, all the forces lie in one and the same plane, and if there are any torques aside from the torques resulting from the forces, those torques are about an axis perpendicular to that plane. If we define the plane in which the forces lie to be the $x-y$ plane, then for such cases, the set of six scalar equations reduces to a set of 3 scalar equations (in that the other 3 are trivial $0 = 0$ identities):

$$\sum F_x = 0 \tag{23A.4}$$

$$\sum F_y = 0 \tag{23A.5}$$

$$\sum \tau_{z\circlearrowleft} = 0 \tag{23A.6}$$

Statics problems represent a subset of Newton's 2nd Law problems. You already know how to solve Newton's 2nd Law problems so there is not much new for you to learn here, but a couple of details regarding the way in which objects are supported will be useful to you.

Many statics problems involve beams and columns. Beams and columns are referred to collectively as members. The analysis of the equilibrium of a member typically entails some approximations which involve the neglect of some short distances. As long as these distances are small compared to the length of the beam, the approximations are very good. One of these approximations is that, unless otherwise specified, we neglect the dimensions of the cross section of the member (for instance, the width and height of a beam). We do not neglect the length of the member.

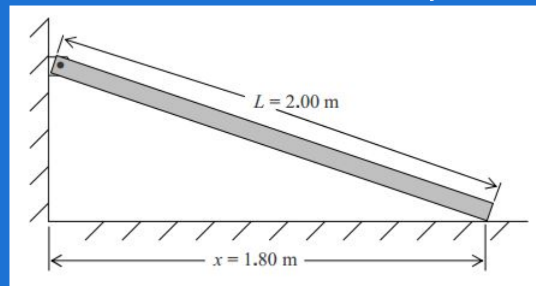
Pin-Connected Members

A pin is a short axle. A member which is pin-connected at one end, is free to rotate about the pin. The pin is perpendicular to the direction in which the member extends. In practice, in the case of a member that is pin-connected at one end, the pin is not really right at the end of the member, but unless the distance from the pin to the end (the end that is very near the pin) of the member is specified, we neglect that distance. Also, the mechanism by which a beam is pin connected to, for instance, a wall, causes the end of the beam to be a short distance from the wall. Unless otherwise specified, we are supposed to neglect this distance as well. A pin exerts a force on the member. The force lies in the plane that contains the member and is perpendicular to the pin. Beyond that, the direction of the force, initially, is unknown. On a free body diagram of the member, one can include the pin force as an unknown force at an unknown angle, or one can include the unknown x and y components of the pin force.

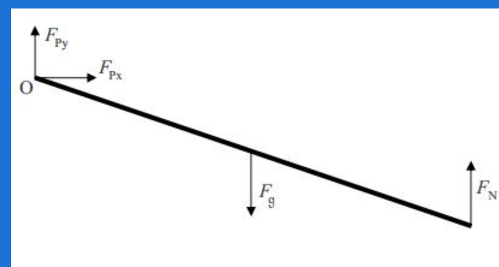
One end of a beam of mass 6.92 kg and of length 2.00 m is pin-connected to a wall. The other end of the beam rests on a frictionless floor at a point that is 1.80 m away from the wall. The beam is in a plane that is perpendicular to both the wall and the floor. The pin is perpendicular to that plane. Find the force exerted by the pin on the beam, and find the normal force exerted on the beam by the floor.

Solution

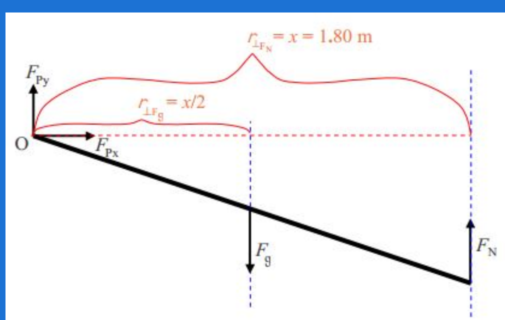
First let's draw a sketch:



Now we draw a free body diagram of the member:



We are going to need to apply the torque equilibrium condition to the beam so I am going to add moment arms to the diagram. My plan is to sum the torques about point O so I will depict moment arms with respect to an axis through point O.



Now let's apply the equilibrium conditions:

$$\sum F_{\rightarrow} = 0$$

$$F_{Px} = 0$$

There are three unknown force values depicted in the free body diagram and we have already found one of them! Let's apply another equilibrium condition:

$$F_{\uparrow} = 0 \quad F_{Py} - F_g + F_N = 0$$ F_{Py}-mg+F_N=0 \label{23-6}

There are two unknowns in this equation. We cannot solve it but it may prove useful later on. Let's apply the torque equilibrium condition.

$$\sum \tau_{\circ\circlearrowleft} = 0 \quad -\frac{x}{2}F_g + xF_N = 0 \quad -\frac{x}{2}mg + xF_N = 0$$

Here, I copy that last line for you before proceeding:

$$-\frac{x}{2}mg + xF_N = 0 \quad F_N = \frac{mg}{2}$$

$$F_N = \frac{6.92kg(9.80N/kg)}{2} \quad F_N = 32.9N$$

We can use this result $(F_N = \frac{mg}{2})$ in Equation 23A.1 to obtain a value for F_{Py}:

$$F_{Py} - mg + F_N = 0$$

$F_{Py} = mg - F_N \quad F_{Py} = mg - \frac{mg}{2} \quad F_{Py} = \frac{mg}{2} \quad F_{Py} = \frac{6.92kg(9.80N/kg)}{2} \quad F_{Py} = 32.9N$

Recalling that we found F_{Px} to be zero, we can write, for our final answer:

$\vec{F_p} = 32.9 \quad \text{newtons, straight upward and} \quad \vec{F_N} \quad \text{newtons, straight upward}$

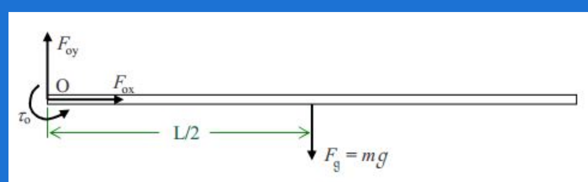
Fix-Connected Members

A fix-connected member is one that is rigidly attached to a structure (such as a wall) that is external to the object whose equilibrium is under study. An example would be a metal rod, one end of which is welded to a metal wall. A fixed connection can apply a force in any direction, and it can apply a torque in any direction. When all the other forces lie in a plane, the force applied by the fixed connection will be in that plane. When all the other torques are along or parallel to a particular line, then the torque exerted by the fixed connection will be along or parallel to that same line.

A horizontal bar of length L and mass m is fix connected to a wall. Find the force and the torque exerted on the bar by the wall.

Solution First a sketch:

then a free body diagram:



followed by the application of the equilibrium conditions to the free body diagram: $\sum F_{\rightarrow} = 0$

$F_{ox} = 0$

That was quick. Let's see what setting the sum of the vertical forces equal to zero yields:

$\sum F_{\uparrow} = 0 \quad F_{oy} - F_g = 0 \quad F_{oy} = F_g \quad F_{oy} = mg$

Now for the torque equilibrium condition:

$\sum \tau_{\circ\circlearrowleft} = 0 \quad \tau_{\circ} - \frac{L}{2} F_g = 0$

$\tau_{\circ} - \frac{L}{2} mg = 0 \quad \tau_{\circ} = \frac{1}{2} mgL$

The wall exerts an upward force of magnitude mg, and a counterclockwise (as viewed from that position for which the free end of the bar is to the right) torque of magnitude $\frac{1}{2} mgL$ on the bar.

24A: Work and Energy

You have done quite a bit of problem solving using energy concepts. Back in chapter 2 we defined energy as a transferable physical quantity that an object can be said to have and we said that if one transfers energy to a material particle that is initially at rest, that particle acquires a speed which is an indicator of how much energy was transferred. We said that an object can have energy because it is moving (kinetic energy), or due to its position relative to some other object (potential energy). We said that energy has units of joules. You have dealt with translational kinetic energy $K = \frac{1}{2}mv^2$, rotational kinetic energy $K = \frac{1}{2}I\omega^2$, spring potential energy $U = \frac{1}{2}kx^2$, near-earth's-surface gravitational potential energy $U = mgy$, and the universal gravitational potential energy $U = -\frac{Gm_1m_2}{r}$ corresponding to the Universal Law of Gravitation. The principle of the conservation of energy is, in the opinion of this author, the central most important concept in physics. Indeed, at least one dictionary defines physics as the study of energy. It is important because it is conserved and the principle of conservation of energy allows us to use simple accounting procedures to make predictions about the outcomes of physical processes that have yet to occur and to understand processes that have already occurred. According to the principle of conservation of energy, any change in the total amount of energy of a system can be accounted for in terms of energy transferred from the immediate surroundings to the system or to the immediate surroundings from the system. Physicists recognize two categories of energy transfer processes. One is called work and the other is called heat flow. In this chapter we focus our attention on work.

Conceptually, positive work is what you are doing on an object when you push or pull on it in the same direction in which the object is moving. You do negative work on an object when you push or pull on it in the direction opposite the direction in which the object is going. The mnemonic for remembering the definition of work that helps you remember how to calculate it is “Work is Force times Distance.” The mnemonic does not tell the whole story. It is good for the case of a constant force acting on an object that moves on a straight line path when the force is in the same exact direction as the direction of motion.

A more general, but still not completely general, “how-to-calculate-it” definition of work applies to the case of a constant force acting on an object that moves along a straight line path (when the force is not necessarily directed along the path). In such a case, the work W done on the object, when it travels a certain distance along the path, is: the along-the-path component of the force $F_{\|}$ times the length of the path segment $\triangle r$.

$$W = F_{\|}\triangle r \tag{24A.1}$$

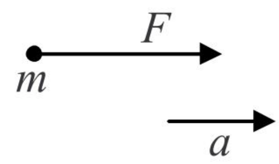
Even this case still needs some additional clarification: If the force component vector along the path is in the same direction as the object’s displacement vector, then $F_{\|}$ is positive, so the work is positive; but if the force component vector along the path is in the opposite direction to that of the object’s displacement vector, then $F_{\|}$ is negative, so the work is negative. Thus, if you are pushing or pulling on an object in a direction that would tend to make it speed up, you are doing positive work on the object. But if you are pushing or pulling on an object in a direction that would tend to slow it down, you are doing negative work on the object.

In the most general case in which the “component of the force along the path” is continually changing because the force is continually changing (such as in the case of an object on the end of a spring) or because the path is not straight, our “how-to-calculate-it” definition of the work becomes: For each infinitesimal path segment making up the path in question, we take the product of the along-the-path force component and the infinitesimal length of the path segment. The work is the sum of all such products. Such a sum would have an infinite number of terms. We refer to such a sum as an integral.

The Relation Between Work and Motion

Let’s go back to the simplest case, the case in which a force $\vec{F}$ is the only force acting on a particle of mass m which moves a distance $\triangle r$ (while the force is acting on it) in a straight line in the exact same direction as the force. The plan here is to investigate the connection between the work on the particle and the motion of the particle. We'll start with Newton's 2nd Law.

Free Body Diagram



$$a_{\rightarrow}=\frac{1}{m}F_{\rightarrow}$$

$$a=\frac{1}{m}F$$

Solving for F, we arrive at:

$$F=ma$$

On the left, we have the magnitude of the force. If we multiply that by the distance $\triangle r$, we get the work done by the force on the particle as it moves the distance $\triangle r$ along the path, in the same direction as the force. If we multiply the left side of the equation by $\triangle r$ then we have to multiply the right by the same thing to maintain the equality.

$$F\triangle r=ma\triangle r$$

On the left we have the work W, so:

$$W=ma\triangle r$$

On the right we have two quantities used to characterize the motion of a particle so we have certainly met our goal of relating work to motion, but we can untangle things on the right a bit if we recognize that, since we have a constant force, we must have a constant acceleration. This means the constant acceleration equations apply, in particular, the one that (in terms of r rather than x) reads:

$$v^2=v_0^2+2a\triangle r$$

Solving this for $a\triangle r$ gives

$$a\triangle r=\frac{1}{2}v^2-\frac{1}{2}v_\circ^2$$

Substituting this into our expression for W above (the one that reads $W=ma\triangle r$) we obtain

$$W=m(\frac{1}{2}v^2-\frac{1}{2}v_\circ^2)$$

which can be written as

$$W=\frac{1}{2}mv^2-\frac{1}{2}mv_\circ^2$$

Of course we recognize the $\frac{1}{2}mv_0^2$ as the kinetic energy of the particle before the work is done on the particle and the $\frac{1}{2}mv^2$ as the kinetic energy of the particle after the work is done on it. To be consistent with the notation we used in our early discussion of the conservation of mechanical energy we change to the notation in which the prime symbol (′) signifies "after" and no super- or subscript at all (rather than the subscript "o") represents "before." Using this notation and the definition of kinetic energy, our expression for W becomes:

$$W=K'-K$$

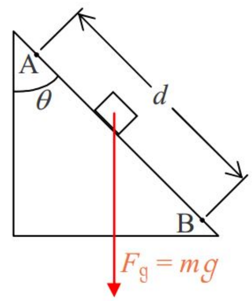
Since the "after" kinetic energy minus the "before" kinetic energy is just the change in kinetic energy δK, we can write the expression for W as:

$$W = \delta K \tag{24A.2}$$

This is indeed a simple relation between work and motion. The cause, work on a particle, on the left, is exactly equal to the effect, a change in the kinetic energy of the particle. This result is so important that we give it a name, it is the Work-Energy Relation. It also goes by the name: The Work-Energy Principle. It works for extended rigid bodies as well. In the case of a rigid body that rotates, it is the displacement of the point of application of the force, along the path of said point of application, that is used (as the δr) in calculating the work done on the object. In the expression $W = \delta K$, the work is the net work (the total work) done by all the forces acting on the particle or rigid body. The net work can be calculated by finding the work done by each force and adding the results, or by finding the net force and using it in the definition of the work.

Calculating the Work as the Force-Along-the-Path Times the Length of the Path

Consider a block on a flat frictionless incline that makes an angle θ with the vertical. The block travels from a point A near the top of the incline to a point B, a distance d in the down-theincline direction from A. Find the work done, by the gravitational force, on the block.



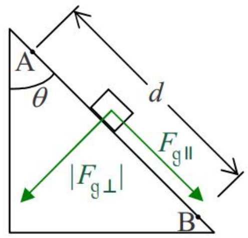
We've drawn a sketch of the situation (not a free body diagram). We note that the force for which we are supposed to calculate the work is not along the path. So, we define a coordinate system with one axis in the down-the-incline direction and the other perpendicular to that axis

and break the gravitational force vector up into its components with respect to that coordinate system.

$$F_{g\|} = F_g \cos\theta = mg\cos\theta$$

$$\left|F_{g\perp}\right| = F_g \sin\theta = mg\sin\theta$$

Now we redraw the sketch with the gravitational force replaced by its components:



$F_{g\perp}$, being perpendicular to the path does no work on the block as the block moves from A to B. The work done by the gravitational force is given by

$$W = F_{\parallel} d$$

$$W = F_{g\parallel} d$$

$$W = mg(\cos\theta)d$$

$$W = mgd\cos\theta$$

While this method for calculating the work done by a force is perfectly valid, there is an easier way. It involves another product operator for vectors (besides the cross product), called the dot product. To use it, we need to recognize that the length of the path, combined with the direction of motion, is none other than the displacement vector (for the point of application of the force). Then we just need to find the dot product of the force vector and the displacement vector.

25A: Potential Energy, Conservation of Energy, Power

The work done on a particle by a force acting on it as that particle moves from point **A** to point **B** under the influence of that force, for some forces, does not depend on the path followed by the particle. For such a force there is an easy way to calculate the work done on the particle as it moves from point **A** to point **B**. One simply has to assign a value of potential energy (of the particle) to point **A** (call that value U_A) and a value of potential energy to point **B** (call that value U_B). One chooses the values such that the work done by the force in question is just the negative of the difference between the two values.

$$W = -(U_B - U_A)$$

$$W = -\triangle U \tag{25A.1}$$

$\triangle U = U_B - U_A$ is the change in the potential energy experienced by the particle as it moves from point **A** to point **B**. The minus sign in equation 25-1 ensures that an increase in potential energy corresponds to negative work done by the corresponding force. For instance for the case of nearearth's-surface gravitational potential energy, the associated force is the gravitational force, a.k.a. the gravitational force. If we lift an object upward near the surface of the earth, the gravitational force does negative work on the object since the (downward) force is in the opposite direction to the (upward) displacement. At the same, time, we are increasing the capacity of the particle to do work so we are increasing the potential energy. Thus, we need the "-" sign in $\omega = -\triangle U$ to ensure that the change in potential energy method of calculating the work gives the same algebraic sign for the value of the work that the force-along-the path times the length of the path gives.

Note that in order for this method of calculating the work to be useful in any case that might arise, one must assign a value of potential energy to every point in space where the force can act on a particle so that the method can be used to calculate the work done on a particle as the particle moves from any point **A** to any point **B**. In general, this means we need a value for each of an infinite set of points in space.

This assignment of a value of potential energy to each of an infinite set of points in space might seem daunting until you realize that it can be done by means of a simple algebraic expression. For instance, we have already written the assignment for a particle of mass m_2 for the case of the universal gravitational force due to a particle of mass m_1. It was equation ???:

$$U = -\frac{Gm_1m_2}{r}$$

in which **G** is the universal gravitational constant $G = 6.67 \times 10^{-11} \frac{N \cdot m^2}{kg^2}$ and r is the distance that particle 2 is from particle 1. Note that considering particle 1 to be at the origin of a coordinate system, this equation assigns a value of potential energy to every point in the universe! The value, for any point, simply depends on the distance that the point is from the origin. Suppose we want to find the work done by the gravitational force due to particle 1, on particle 2 as particle 2 moves from point **A**, a distance r_A from particle 1 to point **B**, a distance r_B, from particle 1. The gravitational force exerted on it (particle 2) by the gravitational field of particle 1 does an amount of work, on particle 2, given by (starting with equation 25A.1):

$$W = \triangle U$$

$$W = -(U_B - U_A)$$

$$W = -\left[(-\frac{Gm_1m_2}{r_B}) - (-\frac{Gm_1m_2}{r_A})\right]$$

$$W = Gm_1m_2(\frac{1}{r_B} - \frac{1}{r_A})$$

The Relation Between a Conservative Force and the Corresponding Potential

While this business of calculating the work done on a particle as the negative of the change in its potential energy does make it a lot easier to calculate the work, we do have to be careful to define the potential such that this method is equivalent to calculating the work as the force-along-the-path times the length of the path. Rather than jump into the problem of finding the potential energy at all points in a three-dimensional region of space for a kind of force known to exist at all points in that three-dimensional region of space, let's look into the simpler problem of finding the potential along a line. We define a coordinate system consisting of a single axis, let's call it the x-axis, with an origin and a positive direction. We put a particle on the line, a particle that can move along the line. We assume that we have a force that acts on the particle wherever the particle is on the line and that the force is directed along the line. While we will also address the case of a force which has the same value at different points along the line, we assume that, in general, the force varies with position. ← Remember this fact so that you can find the flaw discussed below. Because we want to define a potential for it, it is important that the work done on the particle by the force being exerted on the particle, as the particle moves from point A to point B does not depend on how the particle gets from point **A** to point **B**. Our goal is to define a potential energy function for the force such that we get the same value for the work done on the particle by the force whether we use the force-along-the-path method to calculate it or the negative of the change of potential energy method. Suppose the particle undergoes a displacement $\triangle x$ along the line under the influence of the force. See if you can see the flaw in the following, before I point it out: We write $W = F\triangle x$ for the work done by the force, calculated using the force-along-the-path times the length of the path idea, and then $W = -\triangle U$ for the work done by the force calculated using the negative of the change in potential energy concept. Setting the two expressions equal to each other, we have, $F\triangle x = -\triangle U$ which we can write as $F = -\dfrac{\triangle U}{\triangle x}$ for the relation between the potential energy and the x-component of the force.

Do you see where we went wrong? While the method will work for the special case in which the force is a constant, we were supposed to come up with a relation that was good for the general case in which the force varies with position. That means that for each value of x in the range of values extending from the initial value, let's call it x_A, to the value at the end of the displacement $x_A + \triangle x$, there is a different value of force. So the expression $W = F\triangle x$ is inappropriate. Given a numerical problem, there is no one value to plug in for F, because F varies along the $\triangle x$.

To fix things, we can shrink $\triangle x$ to infinitesimal size, so small that, x_A and $x_A + \triangle x$ are, for all practical purposes, one and the same point. That is to say, we take the limit as $\triangle x \to 0$. Then our relation becomes

$$F_x = \lim_{\triangle x \to 0} (-\frac{\triangle U}{\triangle x})$$

which is the same thing as

$$F_x = -\lim_{\triangle x \to 0} (-\frac{\triangle U}{\triangle x})$$

The limit of $\frac{\triangle U}{\triangle x}$ that appears on the right is none other than the derivative $\frac{dU}{dx}$, so:

$$F_x = -\frac{dU}{dx} \tag{25A.2}$$

To emphasize the fact that force is a vector, we write it in unit vector notation as:

$$\vec{F} = -\frac{dU}{dx}\hat{i} \tag{25A.3}$$

Let's make this more concrete by using it to determine the potential energy due to a force with which you are familiar—the force due to a spring.

Consider a block on frictionless horizontal surface. The block is attached to one end of a spring. The other end of the spring is attached to a wall. The spring extends horizontally away from the wall, at right angles to the wall. Define an x-axis with the origin at the equilibrium position of that end of the spring which is attached to the block. Consider the away-from-the-wall direction to be the positive x direction. Experimentally, we find that the force exerted by the spring on the block is given by:

$$\vec{F} = -kx\,\hat{i} \tag{25A.4}$$

where k is the force constant of the spring. (Note: A positive x, corresponding to the block having been pulled away from the wall, thus stretching the spring, results in a force in the negative x direction. A negative x, compressed spring, results in a force in the $+x$ direction, consistent with common sense.) By comparison with equation 25A.3 (the one that reads $\vec{F} = -\dfrac{dU}{dx}\hat{i}$) we note that the potential energy function has to be defined so that

$$\frac{dU}{dx} = kx$$

This is such a simple case that we can pretty much guess what U has to be. U has to be defined such that when we take the derivative of it we get a constant (the k) times x to the power of 1. Now when you take the derivative of x to a power, you reduce the power by one. For that to result in a power of 1, the original power must be 2. Also, the derivative of a constant times something yields that same constant times the derivative, so, there must be a factor of k in the potential energy function. Let's try $U = kx^2$ and see where that gets us. The derivative of kx^2 is $2kx$. Except for that factor of 2 out front, that is exactly what we want. Let's amend our guess by multiplying it by a factor of $\frac{1}{2}$, to eventually cancel out the 2 that comes down when we take the derivative. With $U = \dfrac{1}{2}kx^2$ we get $\frac{dU}{dx} = kx$ which is exactly what we needed. Thus

$$U = \frac{1}{2}kx^2 \tag{25A.5}$$

is indeed the potential energy for the force due to a spring. You used this expression back in chapter 2. Now you know where it comes from.

We have considered two other conservative forces. For each, let's find the potential energy function U that meets the criterion that we have written as, $\vec{F} = -\dfrac{dU}{dx}\hat{i}$.

First, let's consider the near-earth's-surface gravitational force exerted on an object of mass m, by the earth. We choose our single axis to be directed vertically upward with the origin at an arbitrary but clearly specified and fixed elevation for the entire problem that one might solve using the concepts under consideration here. By convention, we call such an axis the y axis rather than the x axis. Now we know that the gravitational force is given simply (again, this is an experimental result) by

$$\vec{F} = -mg\hat{j}$$

where the **mg** is the known magnitude of the gravitational force and the $-\hat{j}$ is the downward direction.

Equation 25A.3, written for the case at hand is:

$$\vec{F} = -\frac{dU}{dy}\hat{j}$$

For the last two equations to be consistent with each other, we need U to be defined such that

$$\frac{dU}{dy} = mg$$

For the derivative of U with respect to y to be the constant "**mg**", U must be given by

$$U = mgy \tag{25A.6}$$

and indeed this is the equation for the earth's near-surface gravitational potential energy. Please verify that when you take the derivative of it with respect to y, you do indeed get the magnitude of the gravitational force, mg.

Now let's turn our attention to the Universal Law of Gravitation. Particle number 1 of mass m_1 creates a gravitational field in the region of space around it. Let's define the position of particle number 1 to be the origin of a three-dimensional Cartesian coordinate system. Now let's assume that particle number 2 is at some position in space, a distance r away from particle 1. Let's define the direction that particle 2 is in, relative to particle 1, as the $+x$ direction. Then, the coordinates of particle 2 are (r,0,0). r is then the x component of the position vector for particle 2, a quantity that we shall now call x. That is, x is defined such that $x = r$. In terms of the coordinate system thus defined, the force exerted by the gravitational field of particle 1, on particle 2, is given by:

$$\vec{F} = -\frac{Gm_1m_2}{x^2}\hat{i}$$

which I rewrite here:

$$\vec{F} = -\frac{Gm_1m_2}{x^2}\hat{i}$$

Compare this with equation 25A.3:

$$\vec{F} = -\frac{dU}{dx}\hat{i}$$

Combining the two equations, we note that our expression for the potential energy U in terms of x must satisfy the equation

$$\frac{dU}{dx} = \frac{Gm_1m_2}{x^2}$$

It's easier to deduce what U must be if we write this as:

$$\frac{dU}{dx} = Gm_1m_2x^{-2}$$

For the derivative of U with respect to x to be a constant (Gm_1m_2) times a power (-2) of x, U itself must be that same constant (Gm_1m_2) times x to the next higher power (-1), divided by the value of the latter power.

$$U = \frac{Gm_1m_2x^{-1}}{-1}$$

which can be written

$$U = -\frac{Gm_1m_2}{r} \tag{25A.7}$$

This is indeed the expression for the gravitational potential that we gave you (without any justification for it) back in Chapter 17, the chapter on the Universal Law of Gravitation.

Conservation of Energy Revisited

Recall the work-energy relation, equation ??? from last chapter,

$$W = \triangle K,$$

the statement that work causes a change in kinetic energy. Now consider a case in which all the work is done by conservative forces, so, the work can be expressed as the negative of the change in potential energy.

$$-\triangle U = \triangle K$$

Further suppose that we are dealing with a situation in which a particle moves from point **A** to point **B** under the influence of the force or forces corresponding to the potential energy U.

Then, the preceding expression can be written as:

$$-(U_B - U_A) = K_B - K_A$$

$$-U_B + U_A = K_B - K_A$$

$$K_A + U_A = K_B + U_B$$

Switching over to notation in which we use primed variables to characterize the particle when it is at point **B** and unprimed variables at **A** we have:

$$K + U = K' + U'$$

Interpreting $E = K + U$ as the energy of the system at the "before" instant, and $E' = K' + U'$ as the energy of the system at the "after" instant, we see that we have derived the conservation of mechanical energy statement for the special case of no net energy transfer to or from the surroundings and no conversion of energy within the system from mechanical energy to other forms or vice versa. In equation form, the statement is

$$E = E' \tag{25A.8}$$

an equation to which you were introduced in chapter 2. Note that you would be well advised to review chapter 2 now, because for the current chapter, you are again responsible for solving any of the "chapter-2-type" problems (remembering to include, and correctly use, before and after diagrams) and answer any of the "chapter-2-type" questions.

Power

In this last section on energy we address a new topic. As a separate and important concept, it would deserve its own chapter except for the fact that it is such a simple, straightforward concept. Power is the rate of energy transfer, energy conversion, and in some cases, the rate at which transfer and conversion of energy are occurring simultaneously. When you do work on an object, you are transferring energy to that object. Suppose for instance that you are pushing a block across a horizontal frictionless surface. You are doing work on the object. The kinetic energy of the object is increasing. The rate at which the kinetic energy is increasing is referred to as power. The rate of change of any quantity (how fast that quantity is changing) can be calculated as the derivative of that quantity with respect to time. In the case at hand, the power **P** can be expressed as

$$P=\frac{dK}{dt} \tag{25A.9}$$

the time derivative of the kinetic energy. Since $K=\frac{1}{2}mV^2$ we have

$$\begin{aligned} P &= \frac{d}{dt}\,\frac{1}{2}m\,V^2 \\ P &= \frac{1}{2}\,m\,\frac{d}{dt}v^2 \\ P &= \frac{1}{2}m\,2v\,\frac{dv}{dt} \\ P &= m\frac{dv}{dt}v \\ P &= m\,a_{\|}v \\ P &= F_{\|}\,v \\ P &= \vec{F}\cdot\vec{v} \end{aligned} \tag{25A.10}$$

where $a_{\|}$ is the acceleration component parallel to the velocity vector. The perpendicular component changes the direction of the velocity but not the magnitude.

Besides the rate at which the kinetic energy is changing, the power is the rate at which work is being done on the object. In an infinitesimal time interval dt, you do an infinitesimal amount of work

$$dW=\vec{F}\cdot\overrightarrow{dx}$$

on the object. Dividing both sides by dt, we have

$$\frac{dW}{dt}=\vec{F}\cdot\frac{\overrightarrow{dx}}{dt}$$

which again is

as it must be since, in accord with the work-energy relation, the rate at which you do work on the object has to be the rate at which the kinetic energy of the object increases.

If you do work at a steady rate for a finite time interval, the power is constant and can simply be calculated as the amount of work done during the time interval divided by the time interval itself. For instance, when you climb stairs, you convert chemical energy stored in your body to gravitational potential energy. The rate at which you do this is power. If you climb at a steady rate for a total increase of gravitational potential energy of $\triangle U$ over a time interval $\triangle t$ then the constant value of your power during that time interval is

$$P=\frac{\triangle U}{\triangle t} \tag{25A.11}$$

If you know that the power is constant, you know the value of the power **P**, and you are asked to find the total amount of work done, the total amount of energy transferred, and/or the total amount of energy converted during a particular time interval $\triangle t$, you just have to multiply the power **P** by the time interval $\triangle t$.

$$\text{Energy}=P\triangle t \tag{25A.12}$$

One could include at least a dozen formulas on your formula sheet for power, but they are all so simple that, if you understand what power is, you can come up with the specific formula you need for the case on which you are working. We include but one formula on the formula sheet,

$$P=\frac{dE}{dt} \tag{25A.13}$$

which should remind you what power is. Since power is the rate of change of energy, the SI units of power must be $\frac{j}{s}$. This combination unit is given a name, the watt, abbreviated **W**.

$$1W=1\frac{J}{s} \tag{25A.14}$$

26A: Impulse and Momentum

27A: Oscillations: Introduction, Mass on a Spring

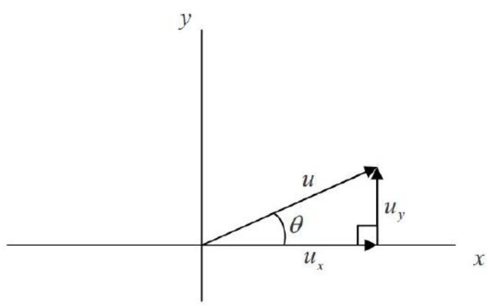
> *If a simple harmonic oscillation problem does not involve the time, you should probably be using conservation of energy to solve it. A common "tactical error" in problems involving oscillations is to manipulate the equations giving the position and velocity as a function of time, $x = x_{max} cos(2\pi ft)$ and $v = -v_{max} sin(2\pi ft)$ rather than applying the principle of conservation of energy. This turns an easy five-minute problem into a difficult fifteen-minute problem.*

When something goes back and forth we say it vibrates or oscillates. In many cases oscillations involve an object whose position as a function of time is well characterized by the sine or cosine function of the product of a constant and elapsed time. Such motion is referred to as sinusoidal oscillation. It is also referred to as simple harmonic motion.

Math Aside: The Cosine Function

By now, you have had a great deal of experience with the cosine function of an angle as the ratio of the adjacent to the hypotenuse of a right triangle. This definition covers angles from 0 radians to $\frac{\pi}{2}$ radians (0° to 90°). In applying the cosine function to simple harmonic motion, we use the extended definition which covers all angles. The extended definition of the cosine of the angle θ is that the cosine of an angle is the x component of a unit vector, the tail of which is on the origin of an x-y coordinate system; a unit vector that originally pointed in the +x direction but has since been rotated counterclockwise from our viewpoint, through the angle θ, about the origin.

Here we show that the extended definition is consistent with the "adjacent over hypotenuse" definition, for angles between 0 radians and $\frac{\pi}{2}$ radians. For such angles, we have:



in which, u, being the magnitude of a unit vector, is of course equal to 1, the pure number 1 with no units. Now, according to the ordinary definition of the cosine of θ as the adjacent over the hypotenuse:

$$\cos\ \theta = \frac{u_x}{u}$$

Solving this for u_x we see that

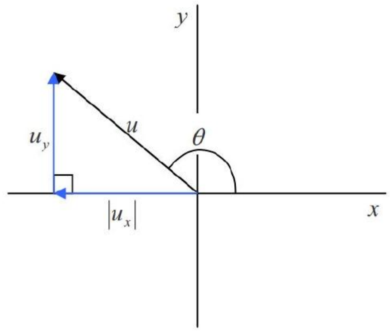
$$u_x = u\ \cos\theta$$

Recalling that $u = 1$, this means that

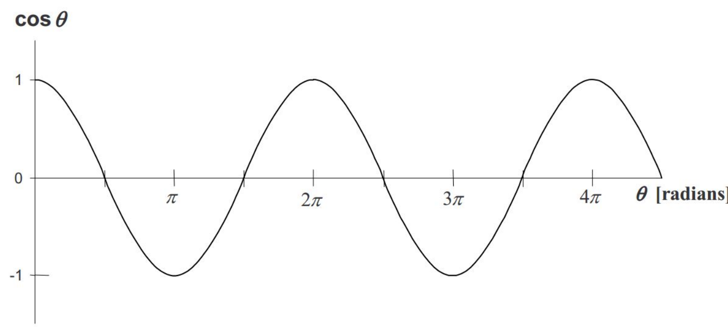
$$u_x = \ \cos\theta$$

Recalling that our extended definition of $\cos\theta$ is, that it is the x component of the unit vector $\hat{u}$ when $\hat{u}$ makes an angle θ with the x-axis, this last equation is just saying that, for the case at hand (θ between 0 and $\frac{\pi}{2}$) our extended definition of $\cos\theta$ is equivalent to our ordinary definition.

At angles between $\frac{\pi}{2}$ and $\frac{3\pi}{2}$ radians (90° and 270°) we see that u_x takes on negative values (when the x component vector is pointing in the negative x direction, the x component value is, by definition, negative). According to our extended definition, $\cos\theta$ takes on negative values at such angles as well.



With our extended definition, valid for any angle θ, a graph of the $\cos\theta$ vs. θ appears as:



Some Calculus Relations Involving the Cosine

The derivative of the cosine of θ, with respect to θ:

$$\frac{d}{d\theta}\cos\theta = -\sin\theta$$

The derivative of the sine of θ, with respect to θ:

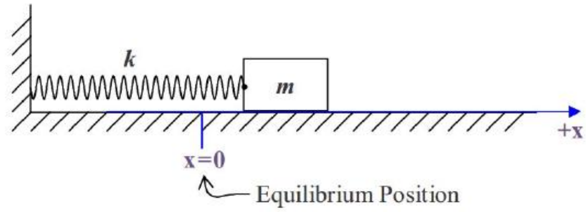
$$\frac{d}{d\theta}\sin\theta = \cos\theta$$

Some Jargon Involving The Sine And Cosine Functions

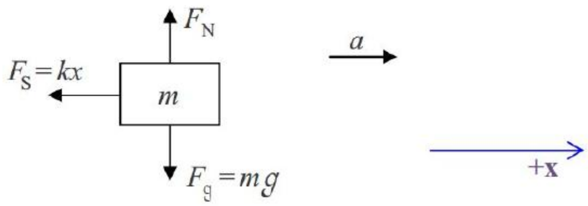
When you express, define, or evaluate the function of something, that something is called the argument of the function. For instance, suppose the function is the square root function and the expression in question is $\sqrt{3x}$. The expression is the square root of $3x$, so, in x that expression, $3x$ is the argument of the square root function. Now when you take the cosine of something, that something is called the argument of the cosine, but in the case of the sine and cosine functions, we give it another name as well, namely, the phase. So, when you write $\cos\theta$, the variable θ is the argument of the cosine function, but it is also referred to as the phase of the cosine function.In order for an expression involving the cosine function to be at all meaningful, the phase of the cosine must have units of angle (for instance, radians or degrees).

A Block Attached to the End of a Spring

Consider a block of mass m on a frictionless horizontal surface. The block is attached, by means of an ideal massless horizontal spring having force constant k, to a wall. A person has pulled the block out, directly away from the wall, and released it from rest. The block oscillates back and forth (toward and away from the wall), on the end of the spring. We would like to find equations that give the block's position, velocity, and acceleration as functions of time. We start by applying Newton's 2^{nd} Law to the block. Before drawing the free body diagram we draw a sketch to help identify our one-dimensional coordinate system. We will call the horizontal position of the point at which the spring is attached, the position x of the block. The origin of our coordinate system will be the position at which the spring is neither stretched nor compressed. When the position x is positive, the spring is stretched and exerts a force, on the block, in the $-x$ direction. When the position of x is negative, the spring is compressed and exerts a force, on the block, in the $+x$ direction.



Now we draw the free body diagram of the block:



and apply Newton's 2nd Law:

$$a_{\rightarrow} = \frac{1}{m}\sum F_{\rightarrow}$$

$$a = \frac{1}{m}(-kx)$$

$$a = -\frac{k}{m}x$$

This equation, relating the acceleration of the block to its position x, can be considered to be an equation relating the position of the block to time if we substitute for a using:

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

and

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

so

$$a = \frac{d}{dt}\frac{dx}{dt}$$

which is usually written

$$a = \frac{d^2x}{dt^2} \tag{27A.1}$$

and read "d squared x by dt squared" or "the second derivative of x with respect to t."

Substituting this expression for a into $a = -\frac{k}{m}x$ (the result we derived from Newton's 2nd Law above) yields

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x \tag{27A.2}$$

We know in advance that the position of the block depends on time. That is to say, x is a function of time. This equation, equation 27A.2, tells us that if you take the second derivative of x with respect to time you get x itself, times a negative constant $(-k/m)$.

We can find the an expression for x in terms of t that solves 27A.2 by the method of "guess and check." Grossly, we're looking for a function whose second derivative is essentially the negative of itself. Two functions meet this criterion, the sine and the cosine. Either will work. We arbitrarily choose to use the cosine function. We include some constants in our trial solution (our guess) to be determined during the "check" part of our procedure. Here's our trial solution:

$$x=x_{max}\cos\left(\frac{2\pi\text{ rad}}{T}t\right)$$

Here's how we have arrived at this trial solution: Having established that x, depends on the cosine of a multiple of the time variable, we let the units be our guide. We need the time t to be part of the argument of the cosine, but we can't take the cosine of something unless that something has units of angle. The constant $\frac{2\pi\,\text{rad}}{T}$, with the constant T having units of time (we'll use seconds), makes it so that the argument of the cosine has units of radians. It is, however, more than just the units that motivates us to choose the ratio $\frac{2\pi\,\text{rad}}{T}$ as the constant. To make the argument of the cosine have units of radians, all we need is a constant with units of radians per second. So why write it as $\frac{2\pi\,\text{rad}}{T}$? Here's the explanation: The block goes back and forth. That is, it repeats its motion over and over again as time goes by. Starting with the block at its maximum distance from the wall, the block moves toward the wall, reaches its closest point of approach to the wall and then comes back out to its maximum distance from the wall. At that point, it's right back where it started from. We define the constant value of time T to be the amount of time that it takes for one iteration of the motion.

Now consider the cosine function. We chose it because its second derivative is the negative of itself, but it is looking better and better as a function that gives the position of the block as a function of time because it too repeats itself as its phase (the argument of the cosine) continually increases. Suppose the phase starts out as 0 at time 0. The cosine of 0 radians is 1, the biggest the cosine ever gets. We can make this correspond to the block being at its maximum distance from the wall. As the phase increases, the cosine gets smaller, then goes negative, eventually reaching the value -1 when the phase is π radians. This could correspond to the block being closest to the wall. Then, as the phase continues to increase, the cosine increases until, when the phase is 2π, the cosine is back up to 1 corresponding to the block being right back where it started from. From here, as the phase of the cosine continues to increase from 2π to 4π, the cosine again takes on all the values that it took on from 0 to 2π. The same thing happens again as the phase increases from 4π to 6π, from 8π to 10π, etc.

Getting back to that constant $\frac{2\pi\,\text{rad}}{T}$ that we "guessed" should be in the phase of the cosine in our trial solution for $x(t)$:

$$x=x_{max}\cos\left(\frac{2\pi\text{ rad}}{T}t\right)$$

With T being defined as the time it takes for the block to go back and forth once, look what happens to the phase of the cosine as the stopwatch reading continually increases. Starting from 0, as t increases from 0 to T, the phase of the cosine, $\frac{2\pi\,\text{rad}}{T}t$, increases from 0 to 2π radians. So, just as the block, from time 0 to time T, goes though one cycle of its motion, the cosine, from time 0 to time T, goes through one cycle of its pattern. As the stopwatch reading increases from T to $2T$, the phase of the cosine increases from 2π rad to 4π rad. The block undergoes the second cycle of its motion and the cosine function used to determine the position of the block goes through the second cycle of its pattern. The idea holds true for any time t —as the stopwatch reading continues to increase, the cosine function keeps repeating its cycle in exact synchronization with the block, as it must if its value is to accurately represent the position of the block as a function of time. Again, it is no coincidence. We chose the constant $\frac{2\pi\,\text{rad}}{T}$ in the phase of the cosine so that things would work out this way.

A few words on jargon are in order before we move on. The time T that it takes for the block to complete one full cycle of its motion is referred to as the period of the oscillations of the block. Now how about that other constant, the "$x_{\max}$" in our educated guess $x=x_{\max}\cos\left(\frac{2\pi\,\text{rad}}{T}t\right)$? Again, the units were our guide. When you take the cosine of an angle, you get a pure number, a value with no units. So, we need the $x_{\max}$ there to give our function units of distance (we'll use meters). We can further relate $x_{\max}$ to the motion of the block. The biggest the cosine of the phase can ever get is 1, thus, the biggest $x_{\max}$ times the cosine of the phase can ever get is $x_{\max}$. So, in the expression $x=x_{\max}\cos(\frac{2\pi\,\text{rad}}{T}t)$, with x being the position of the block at any time t, $x_{\max}$ must be the maximum position of the block, the position of the block, relative to its equilibrium position, when it is as far from the wall as it ever gets.

Okay, we've given a lot of reasons why $x=x_{max}\,cos(\frac{2\pi\,\text{rad}}{T}t)$ should well describe the motion of the block, but unless it is consistent with Newton's 2nd Law, that is, unless it satisfies equation 27A.2:

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

which we derived from Newton's 2nd Law, it is no good. So, let's plug it into equation 27A.2 and see if it works. First, let's take the second derivative $\frac{d^2x}{dt^2}$ of our trial solution with respect to t (so we can plug it and x itself directly into equation 27A.2):

Given

$$x = x_{\max}\cos(\frac{2\pi\, rad}{T}t),$$

the first derivative is

$$\frac{dx}{dt} = x_{\max}\left[-\sin(\frac{2\pi\,\text{rad}}{T}t)\right]\frac{2\pi\,\text{rad}}{T}$$

$$\frac{dx}{dt} = -\frac{2\pi\,\text{rad}}{T}x_{\max}\sin(\frac{2\pi\,\text{rad}}{T}t)$$

The second derivative is then

$$\frac{d^2x}{dt^2} = -\frac{2\pi\,\text{rad}}{T}x_{\max}\cos(\frac{2\pi\,\text{rad}}{T}t)\frac{2\pi\,\text{rad}}{T}t$$

$$\frac{d^2x}{dt^2} = -(\frac{2\pi\,\text{rad}}{T})^2 x_{\max}\cos(\frac{2\pi\,\text{rad}}{T}t)$$

Now we are ready to substitute this and x itself, $x = x_{\max}\cos(\frac{2\pi\text{rad}}{T}t)$, into the differential equation $\frac{d^2x}{dt^2} = -\frac{k}{m}x$ (equation 27A.2) stemming from Newton's 2nd Law of Motion. The substitution yields:

$$-(\frac{2\pi\,\text{rad}}{T})^2 x_{\max}\cos(\frac{2\pi\,\text{rad}}{T}t) = -\frac{k}{m}x_{\max}\cos(\frac{2\pi\,\text{rad}}{T}t)$$

which we copy here for your convenience.

$$-\left(\frac{2\pi\,\text{rad}}{T}\right)^2 x_{max}\cos\left(\frac{2\pi\,\text{rad}}{T}t\right) = -\frac{k}{m}x_{max}\cos\left(\frac{2\pi\,\text{rad}}{T}t\right)$$

The two sides are the same, by inspection, except that where $(\frac{2\pi\text{rad}}{T})^2$ appears on the left, we have $\frac{k}{m}$ on the right. Thus, substituting our guess, $x = x_{\max}\cos\left(\frac{2\pi\text{rad}}{T}t\right)$, into the differential equation that we are trying to solve, $\frac{d^2x}{dt^2} = -\frac{k}{m}x$ (equation 27A.2) leads to an identity if and only if $\left(\frac{2\pi\text{rad}}{T}\right)^2 = \frac{k}{m}$. This means that the period T is determined by the characteristics of the spring and the block, more specifically by the force constant (the "stiffness factor") k of the spring, and the mass (the inertia) of the block. Let's solve for T in terms of these quantities. From $(\frac{2\pi\text{rad}}{T})^2 = \frac{k}{m}$ we find:

$$\frac{2\pi\,\text{rad}}{T} = \sqrt{\frac{k}{m}}$$

$$T = 2\pi\,\text{rad}\sqrt{\frac{m}{k}}$$

$$T = 2\pi\sqrt{\frac{m}{k}} \tag{27A.3}$$

where we have taken advantage of the fact that a radian is, by definition, 1 m/m by simply deleting the "rad" from our result.

The presence of the m in the numerator means that the greater the mass, the longer the period. That makes sense: we would expect the block to be more "sluggish" when it has more mass. On the other hand, the presence of the k in the denominator means that the stiffer the spring, the shorter the period. This makes sense too in that we would expect a stiff spring to result in quicker oscillations. Note the absence of $x_{\max}$ in the result for the period T. Many folks would expect that the bigger the oscillations, the longer it would take the block to complete each oscillation, but the absence of $x_{\max}$ in our result for T shows that it just isn't so. The period T does not depend on the size of the oscillations.

So, our end result is that a block of mass m, on a frictionless horizontal surface, a block that is attached to a wall by an ideal massless horizontal spring, and released, at time $t = 0$, from rest, from a position $x = x_{\max}$, a distance $x_{\max}$ from its equilibrium position; will oscillate about the equilibrium position with a period $T=2\pi\sqrt{\frac{m}{k}}$. Furthermore, the block's position as a function of time will be given by

$$x=x_{\max}\cos\left(\frac{2\pi\text{ rad}}{T}t\right) \tag{27A.4}$$

From this expression for $x(t)$ we can derive an expression for the velocity $v(t)$ as follows:

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

$$v=\frac{d}{dt}\left[x_{\max}\cos(\frac{2\pi\,\text{rad}}{T}t)\right]$$

$$x=x_{\max}\left[-\sin\left(\frac{2\pi\,\text{rad}}{T}t\right)\right]\frac{2\pi\,\text{rad}}{T}$$

$$x=-x_{\max}\frac{2\pi\,\text{rad}}{T}\sin\left(\frac{2\pi\,\text{rad}}{T}t\right) \tag{27A.5}$$

And from this expression for $v\,(t)$ we can get the acceleration $a\,(t)$ as follows:

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

$$a=\frac{d}{dt}\left[-x_{\max}\frac{2\pi\,\text{rad}}{T}\sin\left(\frac{2\pi\,\text{rad}}{T}t\right)\right]$$

$$a=-x_{\max}\frac{2\pi\,\text{rad}}{T}\left[\cos\left(\frac{2\pi\,\text{rad}d}{T}t\right)\right]\frac{2\pi\,\text{rad}}{T}$$

$$a=-x_{\max}=\left(\frac{2\pi\text{ rad}}{T}\right)^2\cos\left(\frac{2\pi\text{ rad}}{T}t\right) \tag{27A.6}$$

Note that this latter result is consistent with the relation $a=-\frac{k}{m}x$ between a and x that we derived from Newton's 2nd Law near the start of this chapter. Recognizing that the $x_{\max}\cos\left(\frac{2\pi\text{rad}}{T}t\right)$ is x and that the $\left(\frac{2\pi\text{rad}}{T}\right)^2$ is $\frac{k}{m}$, it is clear that equation 27A.6 is the same thing as

$$a=-\frac{k}{m}x \tag{27A.7}$$

Frequency

The period T has been defined to be the time that it takes for one complete oscillation. In SI units we can think of it as the number of seconds per oscillation. The reciprocal of T is thus the number of oscillations per second. This is the rate at which oscillations occur. We give it a name, frequency, and a symbol, f.

$$f=\frac{1}{T} \tag{27A.8}$$

The units work out to be $\frac{1}{s}$ which we can think of as $\frac{\text{oscillations}}{s}$ as the oscillation, much like the radian is a marker rather than a true unit. A special name has been assigned to the SI unit of frequency, $1\frac{\text{oscillations}}{s}$ is defined to be 1 hertz, abbreviated 1 Hz. You can think of 1 Hz as either $1\frac{\text{oscillations}}{s}$ or simply $1\frac{1}{s}$.

In terms of frequency, rather than period, we can use $f=\frac{1}{T}$ to express all our previous results in terms of f rather than t.

$$f=\frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

$$x=x_{\max}\cos\left(2\pi\text{ rad }f\,t\right)$$

$$v=-2\pi f x_{\max}\sin\left(2\pi\text{ rad }f\,t\right)$$

$$a = -(2\pi f)^2 \cos\left(2\pi \text{ rad } f\, t\right)$$

By inspection of the expressions for the velocity and acceleration above we see that the greatest possible value for the velocity is $2\pi f x_{\text{max}}$ and the greatest possible value for the acceleration is $(2\pi f)^2 x_{\text{max}}$. Defining

$$v_{\text{max}} = x_{\text{max}}(2\pi f) \tag{27A.9}$$

and

$$a_{\text{max}} = x_{\text{max}}(2\pi f)^2 \tag{27A.10}$$

and, omitting the units "rad" from the phase (thus burdening the user with remembering that the units of the phase are radians while making the expressions a bit more concise) we have:

$$x = x_{\text{max}} \cos(2\pi f\, t) \tag{27A.11}$$

$$v = -v_{\text{max}} \sin(2\pi f\, t) \tag{27A.12}$$

$$a = -a_{\text{max}} \cos(2\pi f\, t) \tag{27A.13}$$

The Simple Harmonic Equation

When the motion of an object is sinusoidal as in $x = x_{max} \cos(2\pi f\, t)$, we refer to the motion as simple harmonic motion. In the case of a block on a spring we found that

$$a = -\left|\text{constant}\right| x \tag{27A.14}$$

where the $\left|\text{constant}\right|$ was $\frac{k}{m}$ and was shown to be equal to $(2\pi\ f)^2$. Written as

$$\frac{d^2x}{dt^2} = -(2\pi\ f)^2\, x \tag{27A.15}$$

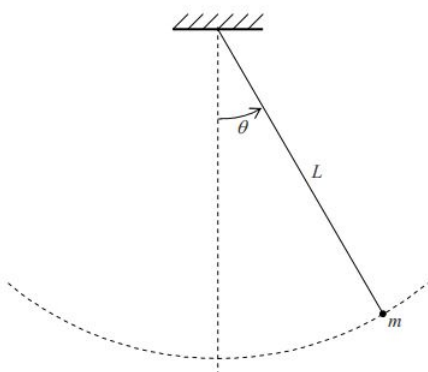
the equation is a completely general equation, not specific to a block on a spring. Indeed, any time you find that, for any system, the second derivative of the position variable, with respect to time, is equal to a negative constant times the position variable itself, you are dealing with a case of simple harmonic motion, and you can equate the absolute value of the constant to $(2\pi\ f)^2$.

28A: Oscillations: The Simple Pendulum, Energy in Simple Harmonic Motion

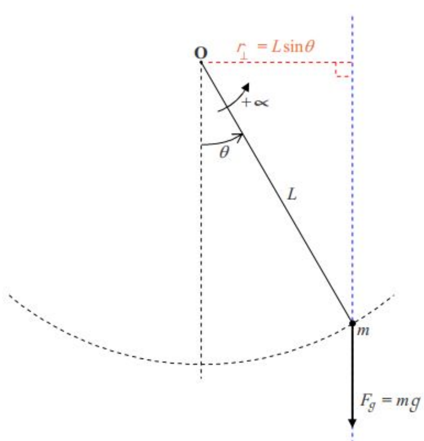
> *Starting with the pendulum bob at its highest position on one side, the period of oscillations is the time it takes for the bob to swing all the way to its highest position on the other side and back again. Don't forget that part about "**and** back again."*

By definition, a simple pendulum consists of a particle of mass m suspended by a massless unstretchable string of length L in a region of space in which there is a uniform constant gravitational field, e.g. near the surface of the earth. The suspended particle is called the pendulum bob. Here we discuss the motion of the bob. While the results to be revealed here are most precise for the case of a point particle, they are good as long as the length of the pendulum (from the fixed top end of the string to the center of mass of the bob) is large compared to a characteristic dimension (such as the diameter if the bob is a sphere or the edge length if it is a cube) of the bob. (Using a pendulum bob whose diameter is 10% of the length of the pendulum (as opposed to a point particle) introduces a 0.05% error. You have to make the diameter of the bob 45% of the pendulum length to get the error up to 1%.)

If you pull the pendulum bob to one side and release it, you find that it swings back and forth. It oscillates. At this point, you don't know whether or not the bob undergoes simple harmonic motion, but you certainly know that it oscillates. To find out if it undergoes simple harmonic motion, all we have to do is to determine whether its acceleration is a negative constant times its position. Because the bob moves on an arc rather than a line, it is easier to analyze the motion using angular variables.



The bob moves on the lower part of a vertical circle that is centered at the fixed upper end of the string. We'll position ourselves such that we are viewing the circle, face on, and adopt a coordinate system, based on our point of view, which has the reference direction straight downward, and for which positive angles are measured counterclockwise from the reference direction. Referring to the diagram above, we now draw a pseudo free-body diagram (the kind we use when dealing with torque) for the string-plus-bob system.



We consider the counterclockwise direction to be the positive direction for all the rotational motion variables. Applying Newton's 2nd Law for Rotational Motion, yields:

$$\propto_{\circ\circlearrowleft} = \frac{\sum \tau_{\circ\circlearrowleft}}{I}$$

$$\propto = \frac{-mgL\,\theta}{I}$$

Next we implement the small angle approximation. Doing so means our result is approximate and that the smaller the maximum angle achieved during the oscillations, the better the approximation. According to the small angle approximation, with it understood that θ must be in radians, $\sin\theta \approx \theta$. Substituting this into our expression for $\propto$, we obtain:

$$\propto = -\frac{mgL\theta}{I}$$

Here comes the part where we treat the bob as a point particle. The moment of inertia of a point particle, with respect to an axis that is a distance L away, is given by $I = mL^2$. Substituting this into our expression for $\propto$ we arrive at:

$$\propto = -\frac{-mgL}{mL^2}\theta$$

Something profound occurs in our simplification of this equation. The masses cancel out. The mass that determines the driving force behind the motion of the pendulum (the gravitational force $F_g = mg$) in the numerator, is exactly canceled by the inertial mass of the bob in the denominator. The motion of the bob does not depend on the mass of the bob! Simplifying the expression for $\propto$ yields:

$$\propto = -\frac{g}{L}\theta$$

Recalling that $\propto \equiv \frac{d^2\theta}{dt^2}$, we have:

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

Hey, this is the simple harmonic motion equation, which, in generic form, appears as $\frac{d^2x}{dt^2} = -|constant|x$ (equation ???) in which the $|constant|$ can be equated to $(2\pi f)^2$ where f is the frequency of oscillations. The position variable in our equation may not be x, but we still have the second derivative of the position variable being equal to the negative of a constant times the position variable itself. That being the case, number 1: we do have simple harmonic motion, and number 2: the constant $\frac{g}{L}$ must be equal to $(2\pi f)^2$.

$$\frac{g}{L} = (2\pi f)^2$$

Solving this for f, we find that the frequency of oscillations of a simple pendulum is given by

$$f = \frac{1}{2\pi}\sqrt{\frac{g}{L}} \tag{28A.1}$$

Again we call your attention to the fact that the frequency does not depend on the mass of the bob!

$T = \frac{1}{f}$ as in the case of the block on a spring. This relation between T and f is a definition that applies to any oscillatory motion (even if the motion is not simple harmonic motion).

All the other formulas for the simple pendulum can be transcribed from the results for the block on a spring by writing

θ in place of x,

ω in place of v, and

$\propto$ in place of a.

Thus,

$$\theta = \theta_{max}\, cos(2\pi f\, t) \tag{28A.2}$$

$$\omega = -\omega_{max}\, sin(2\pi f\, t) \tag{28A.3}$$

$$\propto = -\propto_{max}\, cos(2\pi f\, t) \tag{28A.4}$$

$$\omega_{max} = (2\pi f)\theta_{max} \tag{28A.5}$$

$$\propto_{max} = (2\pi f)^2\theta_{max} \tag{28A.6}$$

Energy Considerations in Simple Harmonic Motion

Let's return our attention to the block on a spring. A person pulls the block out away from the wall a distance x_{max} from the equilibrium position, and releases the block from rest. At that instant, before the block picks up any speed at all, (but when the person is no longer affecting the motion of the block) the block has a certain amount of energy E. And since we are dealing with an ideal system (no friction, no air resistance) the system has that same amount of energy from then on. In general, while the block is oscillating, the energy

$$E = K + U$$

is partly kinetic energy $K = \frac{1}{2}mv^2$ and partly spring potential energy $U = \frac{1}{2}kx^2$. The amount of each varies, but the total remains the same. At time 0, the K in $E = K + U$ is zero since the velocity of the block is zero. So, at time 0:

$$E = U$$

$$E = \frac{1}{2}k\, x^2_{\text{max}}$$

An endpoint in the motion of the block is a particularly easy position at which to calculate the total energy since all of it is potential energy.

As the spring contracts, pulling the block toward the wall, the speed of the block increases so, the kinetic energy increases while the potential energy $U = \frac{1}{2}kx^2$ decreases because the spring becomes less and less stretched. On its way toward the equilibrium position, the system has both kinetic and potential energy

$$E = K + U$$

with the kinetic energy K increasing and the potential energy U decreasing. Eventually the block reaches the equilibrium position. For an instant, the spring is neither stretched nor compressed and hence it has no potential energy stored in it. All the energy (the same total that we started with) is in the form of kinetic energy, $K = \frac{1}{2}mV^2$.

$$E = K + 0$$

$$E = K$$

The block keeps on moving. It overshoots the equilibrium position and starts compressing the spring. As it compresses the spring, it slows down. Kinetic energy is being converted into spring potential energy. As the block continues to move toward the wall, the ever-the-same value of total energy represents a combination of kinetic energy and potential energy with the kinetic energy decreasing and the potential energy increasing. Eventually, at its closest point of approach to the wall, its maximum displacement in the $-x$ direction from its equilibrium position, at its turning point, the block, just for an instant has a velocity of zero. At that instant, the kinetic energy is zero and the potential energy is at its maximum value:

$$E = 0 + U$$

$$E = U$$

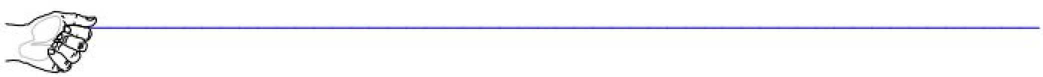
Then the block starts moving out away from the wall. Its kinetic energy increases as its potential energy decreases until it again arrives at the equilibrium position. At that point, by definition, the spring is neither stretched nor compressed so the potential energy is zero. All the energy is in the form of kinetic energy. Because of its inertia, the block continues past the equilibrium position, stretching the spring and slowing down as the kinetic energy decreases while, at the same rate, the potential energy increases. Eventually, the block is at its starting point, again just for an instant, at rest, with no kinetic energy. The total energy is the same total as it has been throughout the oscillatory motion. At that instant, the total energy is all in the form of potential energy. The conversion of energy, back and forth between the kinetic energy of the block and the potential energy stored in the spring, repeats itself over and over again as long as the block continues to oscillate (with—and this is indeed an idealization—no loss of mechanical energy).

A similar description, in terms of energy, can be given for the motion of an ideal (no air resistance, completely unstretchable string) simple pendulum. The potential energy, in the case of the simple pendulum, is in the form of gravitational potential energy $U = mgy$ rather than spring potential energy. The one value of total energy that the pendulum has throughout its oscillations is all potential energy at the endpoints of the oscillations, all kinetic energy at the midpoint, and a mix of potential and kinetic energy at locations in between.

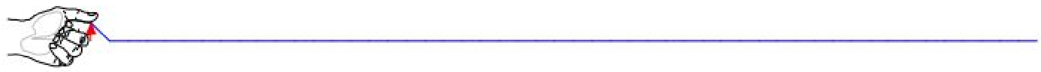
29A: Waves: Characteristics, Types, Energy

Consider a long taut horizontal string of great length. Suppose one end is in the hand of a person and the other is fixed to an immobile object. Now suppose that the person moves her hand up and down. The person causes her hand, and her end of the string, to oscillate up and down. To discuss what happens, we, in our mind, consider the string to consist of a large number of very short string segments. It is important to keep in mind that the force of tension of a string segment exerted on any object, including another segment of the string, is directed away from the object along the string segment that is exerting the force. (The following discussion and diagrams are intentionally oversimplified. The discussion does correctly give the gross idea of how oscillations at one end of a taut string can cause a pattern to move along the length of the string despite the fact that the individual bits of string are essentially doing nothing more than moving up and down.

The person is holding one end of the first segment. She first moves her hand upward.



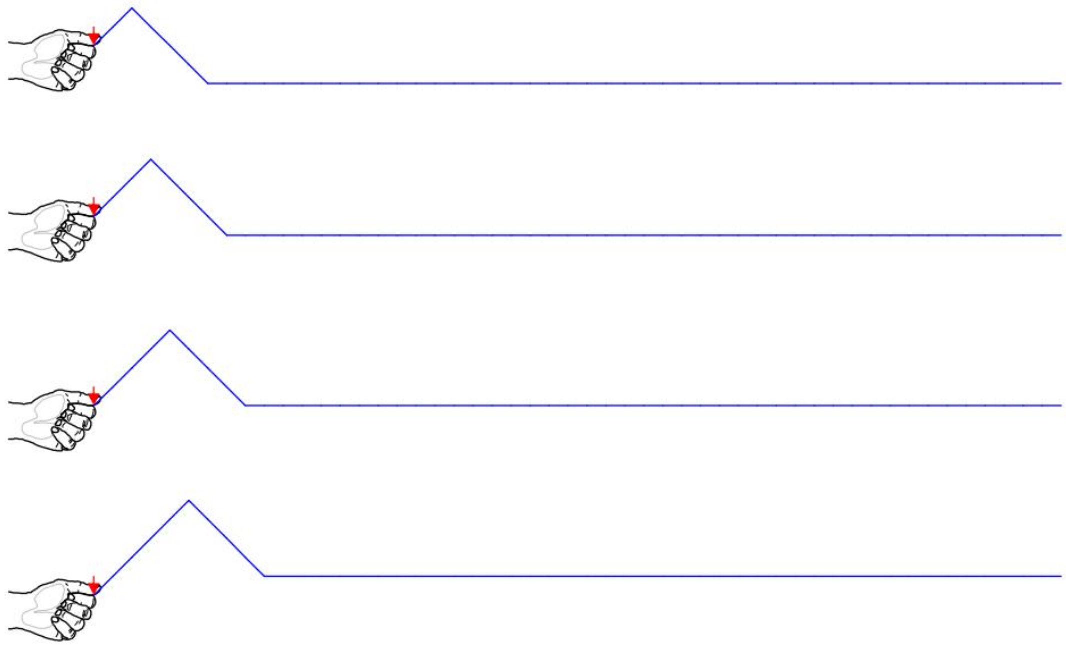
This tilts the first segment so that the force of tension that it is exerting on the second segment has an upward component.

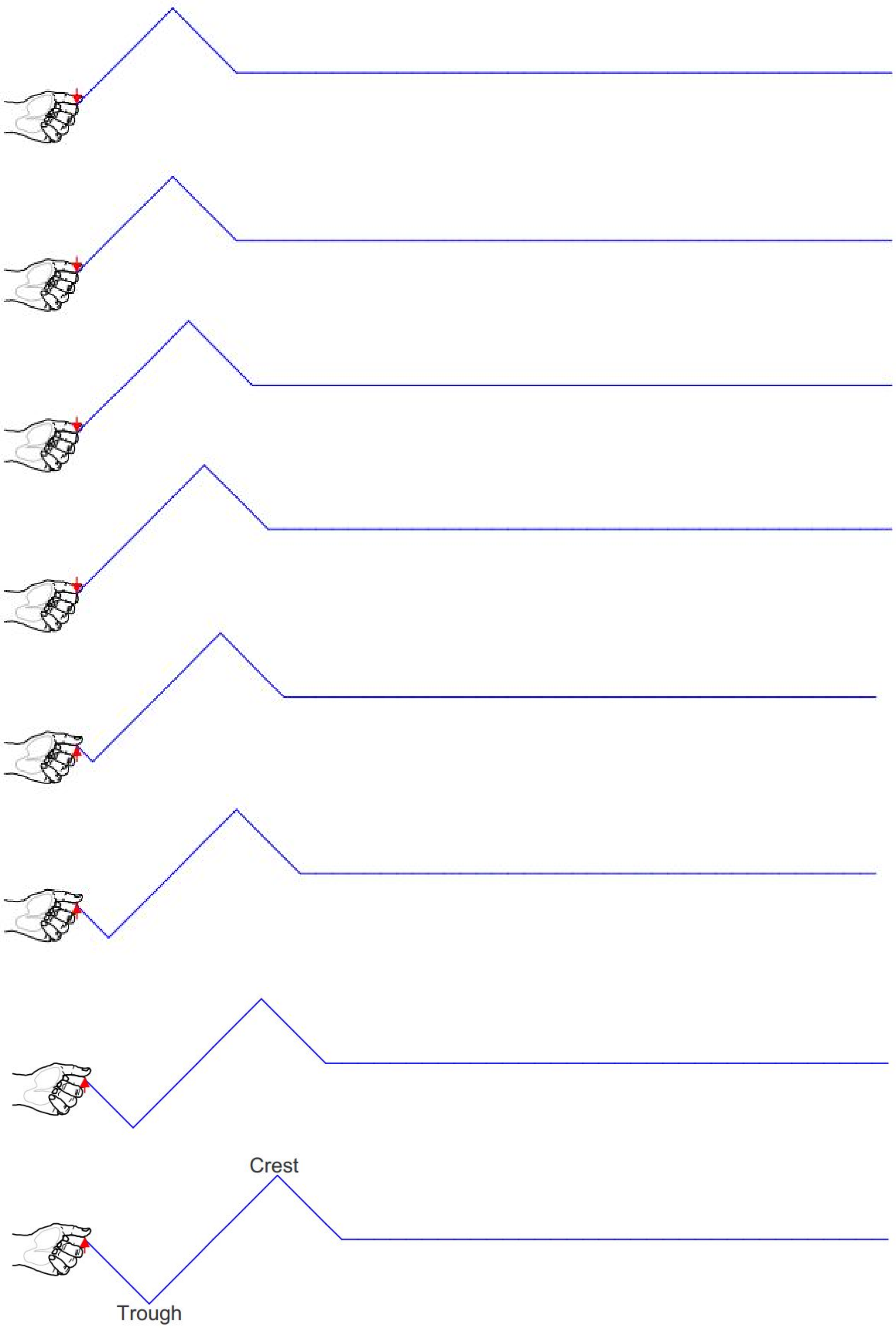


This, in turn, tilts the second segment so that its force of tension on the third segment now has an upward component. The process continues with the 3rd segment, the 4th segment, etc.

After reaching the top of the oscillation, the person starts moving her hand downward. She moves the left end of the first segment downward, but by this time, the first four segments have an upward velocity. Due to their inertia, they continue to move upward. The downward pull of the first segment on the left end of the second segment causes it to slow down, come to rest,

and eventually start moving downward. Inertia plays a huge role in wave propagation. “To propagate” means “to go” or “to travel.” Waves propagate through a medium.





Each very short segment of the string undergoes oscillatory motion like that of the hand, but for any given section, the motion is delayed relative to the motion of the neighboring segment that is closer to the hand. The net effect of all these string segments oscillating up and down, each with the same frequency but slightly out of synchronization with its nearest neighbor, is to create a disturbance in the string. Without the disturbance, the string would just remain on the original horizontal line. The disturbance moves along the length of the string, away from the hand. The disturbance is called a wave. An observer, looking at the string from the side sees crests and troughs of the disturbance, moving along the length of the string, away from the hand. Despite appearances, no material is moving along the length of the string, just a disturbance. The illusion that actual material is moving along the string can be explained by the timing with which the individual segments move up and down, each about its own equilibrium position, the position it was in before the person started making waves.

30A: Wave Function, Interference, Standing Waves

In that two of our five senses (sight and sound) depend on our ability to sense and interpret waves, and in that waves are ubiquitous, waves are of immense importance to human beings. Waves in physical media conform to a wave equation that can be derived from Newton's Second Law of motion. The wave equation reads:

$$\frac{\partial^2 y}{\partial x^2}=\frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} \tag{30A.1}$$

where:

y is the displacement of a point on the medium from its equilibrium position,
x is the position along the length of the medium,
t is time, and
v is the wave velocity.

Take a good look at this important equation. Because it involves derivatives, the wave equation is a differential equation. The wave equation says that the second derivative of the displacement with respect to position (treating the time t as a constant) is directly proportional to the second derivative of the displacement with respect to time (treating the position x as a constant). When you see an equation for which this is the case, you should recognize it as the wave equation.

In general, when the analysis of a continuous medium, e.g. the application of Newton's second law to the elements making up that medium, leads to an equation of the form

$$\frac{\partial^2 y}{\partial x^2}=\left|\text{constant}\right|\frac{\partial^2 y}{\partial x^2}$$

,

the constant will be an algebraic combination of physical quantities representing properties of the medium. That combination can be related to the wave velocity by

$$\left|\text{constant}\right|=\frac{1}{v^2}$$

For instance, application of Newton's Second Law to the case of a string results in a wave equation in which the constant of proportionality depends on the linear mass density μ and the string tension F_T:

$$\frac{\partial^2 y}{\partial x^2}=\frac{\mu}{F_T}\frac{\partial^2 y}{\partial t^2}$$

Recognizing that the constant of proportionality $\frac{\mu}{F_T}$ has to be equal to the reciprocal of the square of the wave velocity, we have

$$\frac{\mu}{F_T}=\frac{1}{v^2}$$

$$v=\sqrt{\frac{F_T}{\mu}} \tag{30A.2}$$

relating the wave velocity to the properties of the string. The solution of the wave equation $\frac{\partial^2 y}{\partial x^2}=\frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}$ can be expressed as

$$y\;=\;y_{\text{max}}\cos(\frac{2\pi}{\lambda}x\pm\frac{2\pi}{\lambda}t+\phi) \tag{30A.3}$$

where:

y is the displacement of a point in the medium from its equilibrium position,
y_{max} is the amplitude of the wave,
x is the position along the length of the medium,
λ is the wavelength,
t is time,

T is the period, and
ϕ is a constant having units of radians. ϕ is called the phase constant.

A "$-$" in the location of the "$\pm$" is used in the case of a wave traveling in the $+x$ direction and a "$+$" for one traveling in the $-x$ direction. Equation 30A.3, the solution to the wave equation $\frac{\partial^2 y}{\partial x^2}=\frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}$, is known as the wave function. Substitute the wave function into the wave equation and verify that you arrive at

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

a necessary condition for the wave function to actually solve the wave equation. $v=\frac{\lambda}{T}$ is the statement that the wave speed is equal to the ratio of the wavelength to the period, a relation that we derived in a conceptual fashion in the last chapter.

At position $x=0$ in the medium, at time $t=0$, the wave function, equation 30A.3, evaluates to

$$y=y_{\max}\cos(\phi).$$

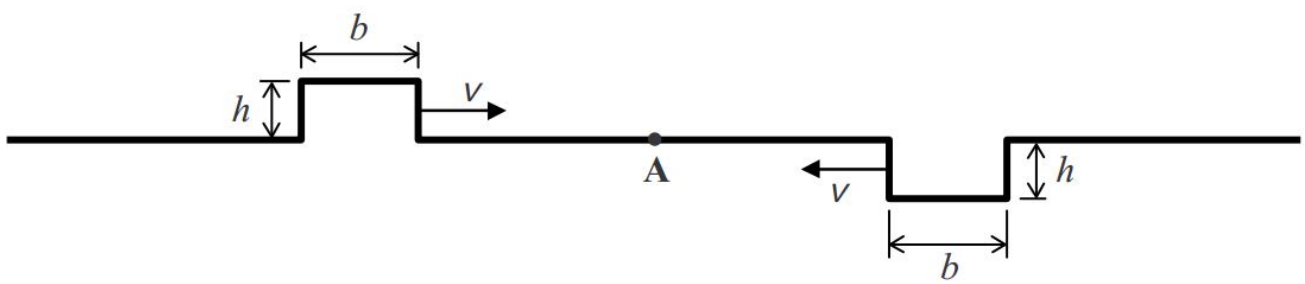
The phase of the cosine boils down to the phase constant ϕ whose value thus determines the value of y at $x=0, t=0$. (Note that the "phase" of the cosine is the argument of the cosine—that which you are taking the cosine of.) The value of the phase constant ϕ is of no relevance to our present discussion so we arbitrarily set $\phi=0$. Also, on your formula sheet, we write the wave function only for the case of a wave traveling in the $+x$ direction, that is, we replace the "$\pm$" with a "$-$". The wave function becomes

$$y=y_{\max}\cos\left(\frac{2\pi}{\lambda}x-\frac{2\pi}{T}t\right) \tag{30A.4}$$

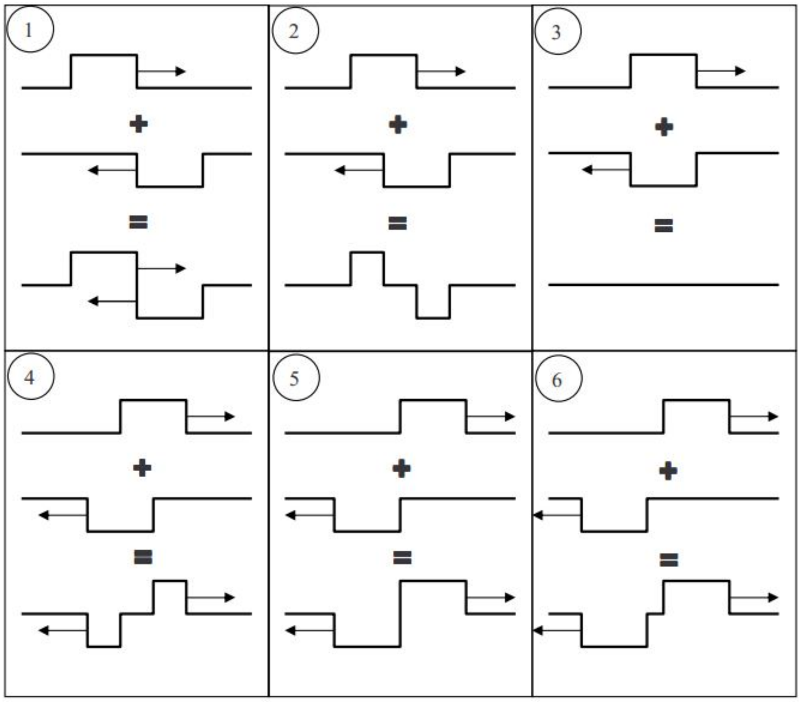
This is the way it appears on your formula sheet. You are supposed to know that this corresponds to a wave traveling in the +x direction and that the expression for a wave traveling in the $-x$ direction can be arrived at by replacing the "$-$ " with a "$+$".

Interference

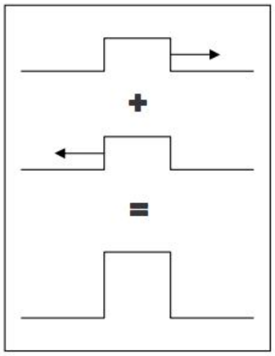
Consider a case in which two waveforms arrive at the same point in a medium at the same time. We'll use idealized waveforms in a string to make our points here. In the case of a string, the only way two waveforms can arrive at the same point in the medium at the same time is for the waveforms to be traveling in opposite directions:



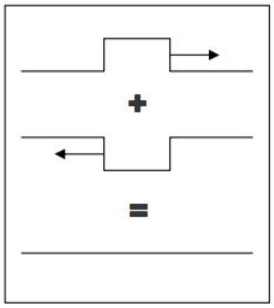
The two waveforms depicted in the diagram above are "scheduled" to arrive at point A at the same time. At that time, based on the waveform on the left alone, point A would have a displacement $+h$, and based on the waveform on the right alone, point A would have the displacement $-h$. So, what is the actual displacement of point A when both waveforms are at point A at the same time? To answer that, you simply add the would-be single-waveform displacements together algebraically (taking the sign into account). One does this point for point over the entire length of the string for any given instant in time. In the following series of diagrams we show the point-for-point addition of displacements for several instants in time.



The phenomenon in which waves traveling in different directions simultaneously arrive at one and the same point in the wave medium is referred to as interference. When the waveforms add together to yield a bigger waveform,



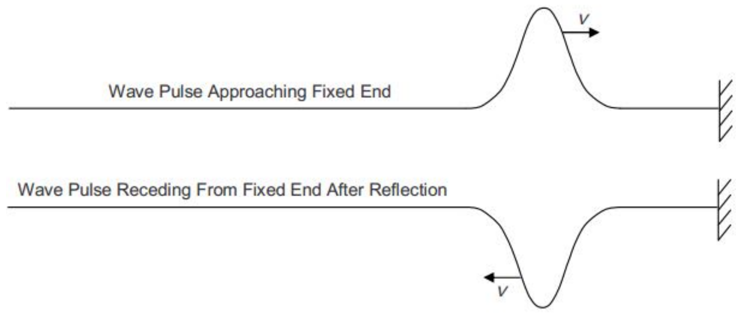
the interference is referred to as constructive interference. When the two waveforms tend to cancel each other out,



the interference is referred to as destructive interference.

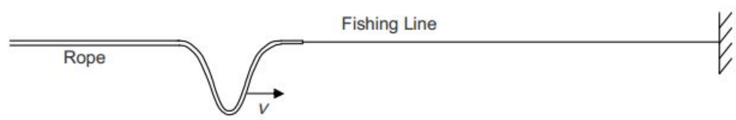
Reflection of a Wave from the End of a Medium

Upon reflection from the fixed end of a string, the displacement of the points on a traveling waveform is reversed.



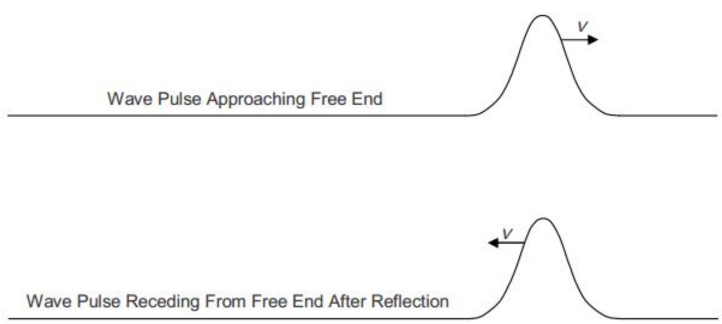
The fixed end, by definition, never undergoes any displacement.

Now we consider a free end. A fixed end is a natural feature of a taut string. A free end, on the other hand, an idealization, is at best an approximation in the case of a taut string. We approximate a free end in a physical string by means of a drastic and abrupt change in linear density. Consider a rope, one of which is attached to the wave source, and the other end of which is attached to one end of a piece of thin, but strong, fishing line. Assume that the fishing line extends through some large distance to a fixed point so that the whole system of rope plus fishing line is taut. A wave traveling along the rope, upon encountering the end of the rope attached to the thin fishing line, behaves approximately as if it has encountered a free end of a taut rope.



In the case of sound waves in a pipe, a free end can be approximated by an open end of the pipe.

Enough said about how one might set up a physical free end of a wave medium, what happens when a wave pulse encounters a free end? The answer is, as in the case of the fixed end, the waveform is reflected, but this time, there is no reversal of displacements.



Standing Waves

Consider a piece of string, fixed at both ends, into which waves have been introduced. The configuration is rich with interference. A wave traveling toward one end of the string reflects off the fixed end and interferes with the waves that were trailing it. Then it reflects off the other end and interferes with them again. This is true for every wave and it repeats itself continuously. For any given length, linear density, and tension of the string, there are certain frequencies for which, at, at least one point on the string, the interference is always constructive. When this is the case, the oscillations at that point (or those points) on the string are maximal and the string is said to have standing waves in it. Again, standing waves result from the interference of the reflected waves with the transmitted waves and with each other. A point on the string at which the interference is always constructive is called an antinode. Any string in which standing waves exist also has at least one point at which the interference is always destructive. Such a point on the string does not move from its equilibrium position. Such a point on the string is called a node.

It might seem that it would be a daunting task to determine the frequencies that result in standing waves. Suppose you want to investigate whether a point on a string could be an antinode. Consider an instant in time when a wave crest is at that position. You need to find the conditions that would make it so that in the time it takes for the crest to travel to one fixed end of the string, reflect back as a trough and arrive back at the location in question; a trough, e.g. one that was trailing the original crest, propagates just the right distance so that it arrives at the location in question at the same time. As illustrated in the next chapter, the analysis that yields the frequencies of standing waves is easier than these timing considerations would suggest.

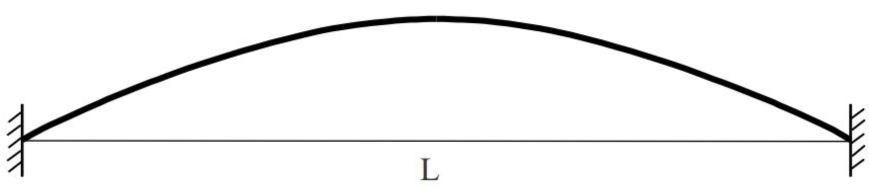
31A: Strings, Air Columns

> *Be careful not to jump to any conclusions about the wavelength of a standing wave. Folks will do a nice job drawing a graph of Displacement vs. Position Along the Medium and then interpret it incorrectly. For instance, look at the diagram on this page. Folks see that a half wavelength fits in the string segment and quickly write the wavelength as $\lambda=\frac{1}{2}L$. But this equation says that a whole wavelength fits in half the length of the string. This is not at all the case. Rather than recognizing that the fraction $\frac{1}{2}$ is relevant and quickly using that fraction in an expression for the wavelength, one needs to be more systematic. First write what you see, in the form of an equation, and then solve that equation for the wavelength. For instance, in the diagram below we see that one half a wavelength λ fits in the length L of the string. Writing this in equation form yields $\frac{1}{2}\lambda=L$. Solving this for λ yields $\lambda=2L$.*

One can determine the wavelengths of standing waves in a straightforward manner and obtain the frequencies from

$$v=\lambda f$$

where the wave speed v is determined by the tension and linear mass density of the string. The method depends on the boundary conditions—the conditions at the ends of the wave medium. (The wave medium is the substance [string, air, water, etc.] through which the wave is traveling. The wave medium is what is "waving.") Consider the case of waves in a string. A fixed end forces there to be a node at that end because the end of the string cannot move. (A node is a point on the string at which the interference is always destructive, resulting in no oscillations. An antinode is a point at which the interference is always constructive, resulting in maximal oscillations.) A free end forces there to be an antinode at that end because at a free end the wave reflects back on itself without phase reversal (a crest reflects as a crest and a trough reflects as a trough) so at a free end you have one and the same part of the wave traveling in both directions along the string. The wavelength condition for standing waves is that the wave must "fit" in the string segment in a manner consistent with the boundary conditions. For a string of length L fixed at both ends, we can meet the boundary conditions if half a wavelength is equal to the length of the string.



Such a wave "fits" the string in the sense that whenever a zero-displacement part of the wave is aligned with one fixed end of the string another zero-displacement part of the wave is aligned with the other fixed end of the string.

Since half a wavelength fits in the string segment we have:

$$\frac{1}{2}\lambda=L$$

$$\lambda=2L$$

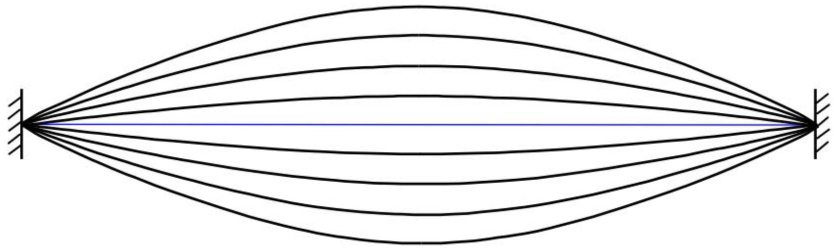
Given the wave speed v, the frequency can be solved for as follows:

$$v=\lambda f$$

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

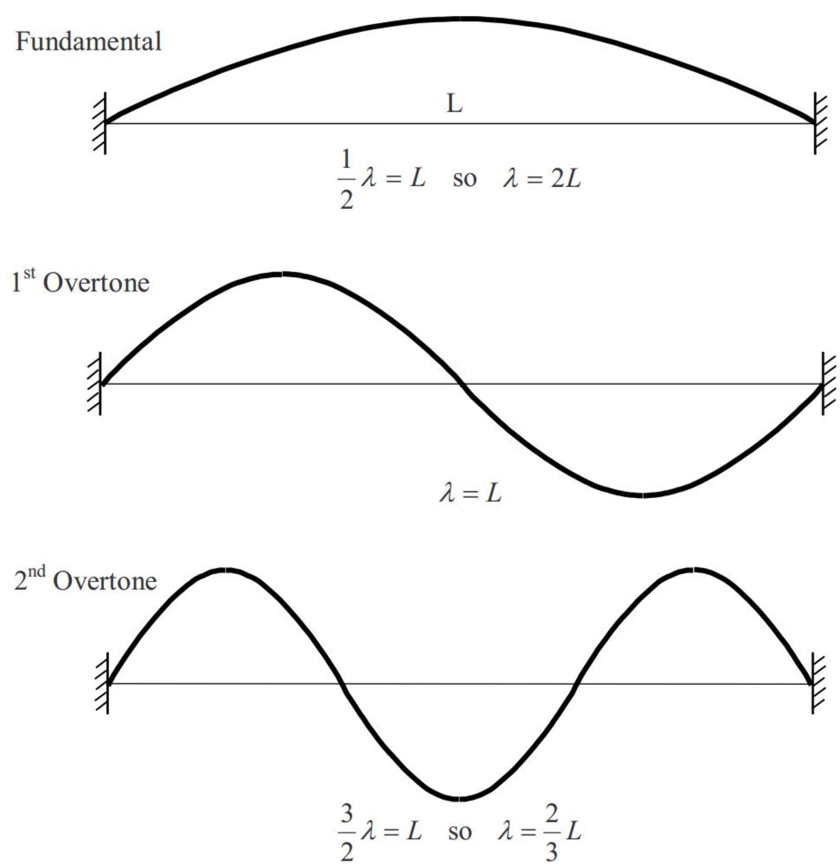
$$f=\frac{v}{2L}$$

It should be noted that despite the fact that the wave is called a standing wave and the fact that it is typically depicted at an instant in time when an antinode on the string is at its maximum displacement from its equilibrium position, all parts of the string (except the nodes) do oscillate about their equilibrium position.



Note that, while the interference at the antinode, the point in the middle of the string in the case at hand, is always as constructive as possible, that does not mean that the string at that point is always at maximum displacement. At times, at that location, there is indeed a crest interfering with a crest, but at other times, there is a zero displacement part of the wave interfering with a zero-displacement part of the wave, at times a trough interfering with a trough, and at times, an intermediate-displacement part of the wave interfering with the same intermediate-displacement part of the wave traveling in the opposite direction. All of this corresponds to the antinode oscillating about its equilibrium position.

The $\lambda = 2L$ wave is not the only wave that will fit in the string. It is, however, the longest wavelength standing wave possible and hence is referred to as the fundamental. There is an entire sequence of standing waves. They are named: the fundamental, the first overtone, the second overtone, the third overtone, etc, in order of decreasing wavelength, and hence, increasing frequency.



Each successive waveform can be obtained from the preceding one by including one more node.

A wave in the series is said to be a harmonic if its frequency can be expressed as an integer times the fundamental frequency. The value of the integer determines which harmonic (1^{st}, 2^{nd}, 3^{rd}, etc.) the wave is. The frequency of the fundamental wave is, of course, 1 times itself. The number 1 is an integer so the fundamental is a harmonic. It is the 1st harmonic.

Starting with the wavelengths in the series of diagrams above, we have, for the frequencies, using $v=\lambda f$ which can be rearranged to read

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

The Fundamental

$$\lambda_{\text{FUND}}=2L$$

$$f_{\text{FUND}}=\frac{v}{\lambda_{\text{FUND}}}$$

$$f_{\text{FUND}}=\frac{v}{2L}$$

The 1st Overtone

$$\lambda_{\text{1st O.T.}}=L$$

$$f_{\text{1st O.T.}}=\frac{v}{\lambda_{\text{1st O.T.}}}$$

$$f_{\text{1st O.T.}}=\frac{v}{L}$$

The 2nd Overtone

$$\lambda_{\text{2nd O.T.}}=\frac{2}{3}L$$

$$f_{\text{2nd O.T.}}=\frac{v}{\lambda_{\text{2nd O.T.}}}$$

$$f_{\text{2nd O.T.}}=\frac{v}{\frac{2}{3}L}$$

$$f_{\text{2nd O.T.}}=\frac{3}{2}\frac{v}{L}$$

Expressing the frequencies in terms of the fundamental frequency $f_{\text{FUND}}=\frac{v}{2L}$ we have

$$f_{\text{FUND}}=\frac{v}{2L}=1\left(\frac{v}{2L}\right)=1f_{\text{FUND}}$$

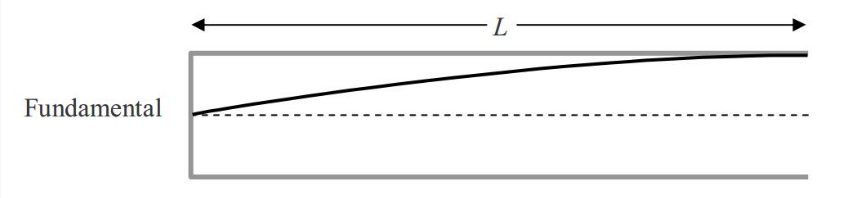
$$f_{\text{1st O.T.}}=\frac{v}{L}=2\left(\frac{v}{2L}\right)=2f_{\text{FUND}}$$

$$f_{\text{2nd O.T.}}=\frac{3}{2}\frac{v}{L}=3\left(\frac{v}{2L}\right)=3f_{\text{FUND}}$$

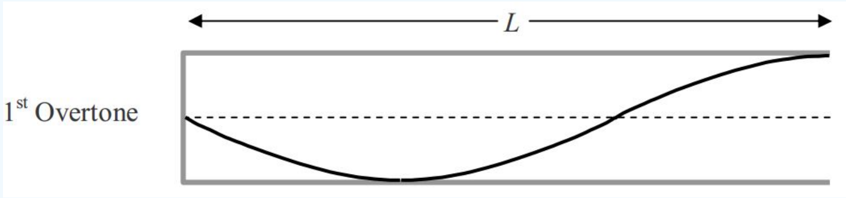
Note that the fundamental is (as always) the 1st harmonic; the 1st overtone is the 2nd harmonic; and the 2nd overtone is the 3rd harmonic. While it is true for the case of a string that is fixed at both ends (the system we have been analyzing), it is not always true that the set of all overtones plus fundamental includes all the harmonics. For instance, consider the following example:

An organ pipe of length L is closed at one end and open at the other. Given that the speed of sound in air is vs , find the frequencies of the fundamental and the first three overtones.

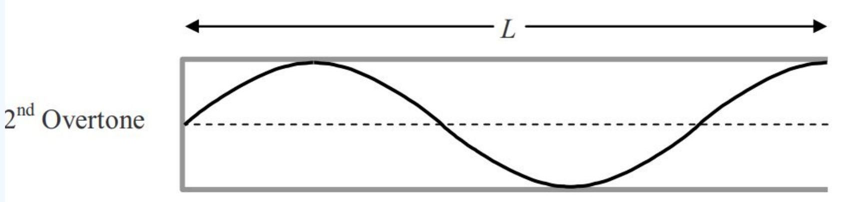
Solution



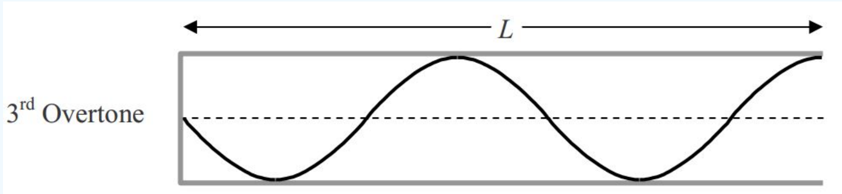
$$\frac{1}{4}\lambda = L \quad \text{so} \quad \lambda = 4L$$



$$\frac{3}{4}\lambda = L \quad \text{so} \quad \lambda = \frac{4}{3}L$$



$$\frac{5}{4}\lambda = L \quad \text{so} \quad \lambda = \frac{4}{5}L$$



$$\frac{7}{4}\lambda = L \quad \text{so} \quad \lambda = \frac{4}{7}L$$

In the preceding sequence of diagrams, a graph of displacement vs. position along the pipe, for an instant in time when the air molecules at an antinode are at their maximum displacement from equilibrium, is a more abstract representation then the corresponding graph for a string. The sound wave in air is a longitudinal wave, so, as the sound waves travel back and forth along the length of the pipe, the air molecules oscillate back and forth (rather than up and down as in the case of the string) about their equilibrium positions. Thus, how high up on the graph a point on the graph is, corresponds to how far to the right (using the viewpoint from which the pipe is depicted in the diagrams) of its equilibrium position the thin layer of air molecules, at the corresponding position in the pipe, is. It is conventional to draw the waveform right inside the outline of the pipe. The boundary conditions are that a closed end is a node and an open end is an antinode.

Starting with the wavelengths in the series of diagrams above, we have, for the frequencies, using $v_s = \lambda f$ which can be rearranged to read

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

The Fundamental

$$\lambda_{\text{FUND}} = 4L$$

$$f_{\text{FUND}}=\frac{v_s}{\lambda_{\text{FUND}}}$$

$$f_{\text{FUND}}=\frac{v_s}{4L}$$

The 1st Overtone

$$\lambda_{\text{1st O.T.}}=\frac{4}{3}L$$

$$f_{\text{1st O.T.}}=\frac{v_s}{\lambda_{\text{1st O.T.}}}$$

$$f_{\text{1st O.T.}}=\frac{v_s}{\frac{4}{3}L}$$

$$f_{\text{1st O.T.}}=\frac{3}{4}\frac{v_s}{L}$$

The 2nd Overtone

$$\lambda_{\text{2nd O.T.}}=\frac{4}{5}L$$

$$f_{\text{2nd O.T.}}=\frac{v_s}{\lambda_{\text{2nd O.T.}}}$$

$$f_{\text{2nd O.T.}}=\frac{v_s}{\frac{4}{5}L}$$

$$f_{\text{2nd O.T.}}=\frac{5}{4}\frac{v_s}{L}$$

The 3rd Overtone

$$\lambda_{\text{3rd O.T.}}=\frac{4}{7}L$$

$$f_{\text{3rd O.T.}}=\frac{v_s}{\lambda_{\text{3rd O.T.}}}$$

$$f_{\text{3rd O.T.}}=\frac{v_s}{\frac{4}{7}L}$$

$$f_{\text{3rd O.T.}}=\frac{7}{4}\frac{v_s}{L}$$

Expressing the frequencies in terms of the fundamental frequency $f_{\text{FUND}}=\frac{v_s}{4L}$ we have

$$f_{\text{FUND}}=\frac{v_s}{4L}=1\left(\frac{v_s}{4L}\right)=1f_{\text{FUND}}$$

$$f_{\text{1st O.T.}}=\frac{3}{4}\frac{v_s}{4L}=3\left(\frac{v_s}{4L}\right)=3f_{\text{FUND}}$$

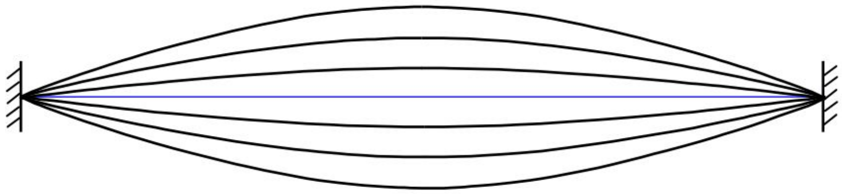
$$f_{\text{2nd O.T.}}=\frac{5}{4}\frac{v_s}{4L}=5\left(\frac{v_s}{4L}\right)=5f_{\text{FUND}}$$

$$f_{\text{3rd O.T.}}=\frac{7}{4}\frac{v_s}{4L}=7\left(\frac{v_s}{4L}\right)=7f_{\text{FUND}}$$

Note that the frequencies of the standing waves are odd integer multiples of the fundamental frequency. That is to say that only odd harmonics, the 1st, 3rd, 5th, etc. occur in the case of a pipe closed at one end and open at the other.

Regarding, Waves, in a Medium that is in Contact with a 2nd Medium

Consider a violin string oscillating at its fundamental frequency, in air. For convenience of discussion, assume the violin to be oriented so that the oscillations are up and down.



Each time the string goes up it pushes air molecules up. This results in sound waves in air. The violin with the standing wave in it can be considered to be the "something oscillating" that is the cause of the waves in air. Recall that the frequency of the waves is identical to the frequency of the source. Thus, the frequency of the sound waves in air will be identical to the frequency of the waves in the string. In general, the speed of the waves in air is different from the speed of waves in the string. From $v = \lambda f$, this means that the wavelengths will be different as well.

32A: Beats and the Doppler Effect

Beats

Consider two sound sources, in the vicinity of each other, each producing sound waves at its own single frequency. Any point in the air-filled region of space around the sources will receive sound waves from both the sources. The amplitude of the sound at any position in space will be the amplitude of the sum of the displacements of the two waves at that point. This amplitude will vary because the interference will alternate between constructive interference and destructive interference. Suppose the two frequencies do not differ by much. Consider the displacements at a particular point in space. Let's start at an instant when two sound wave crests are arriving at that point, one from each source. At that instant the waves are interfering constructively, resulting in a large total amplitude. If your ear were at that location, you would find the sound relatively loud. Let's mark the passage of time by means of the shorter period, the period of the higher-frequency waves. One period after the instant just discussed, the next crest (call it the second crest) from the higher-frequency source is at the point in question, but the peak of the next crest from the lower-frequency source is not there yet. Rather than a crest interfering with a crest, we have a crest interfering with an intermediate-displacement part of the wave. The interference is still constructive but not to the degree that it was. When the third crest from the higher-frequency source arrives, the corresponding crest from the lower-frequency source is even farther behind. Eventually, a crest from the higher-frequency source is arriving at the point in question at the same time as a trough from the lower-frequency source. At that instant in time, the interference is as destructive as it gets. If your ear were at the point in question, you would find the sound to be inaudible or of very low volume. Then the trough from the lower-frequency source starts "falling behind" until, eventually a crest from the higher-frequency source is interfering with the crest preceding the corresponding crest from the lower-frequency source and the interference is again as constructive as possible.

To a person whose ear is at a location at which waves from both sources exist, the sound gets loud, soft, loud, soft, etc. The frequency with which the loudness pattern repeats itself is called the beat frequency. Experimentally, we can determine the beat frequency by timing how long it takes for the sound to get loud N times and then dividing that time by N (where N is an arbitrary integer chosen by the experimenter—the bigger the N the more precise the result). This gives the beat period. Taking the reciprocal of the beat period yields the beat frequency.

The beat frequency is to be contrasted with the ordinary frequency of the waves. In sound, we hear the beat frequency as the rate at which the loudness of the sound varies whereas we hear the ordinary frequency of the waves as the pitch of the sound.

Derivation of the Beat Frequency Formula

Consider sound from two different sources impinging on one point, call it point P, in airoccupied space. Assume that one source has a shorter period T_{SHORT} and hence a higher frequency f_{HIGH} than the other (which has period and frequency T_{LONG} and f_{LOW} respectively). The plan here is to express the beat frequency in terms of the frequencies of the sources—we get there by relating the periods to each other. As in our conceptual discussion, let's start at an instant when a crest from each source is at point P. When, after an amount of time T_{SHORT} passes, the next crest from the shorter-period source arrives, the corresponding crest from the longer-period source won't arrive for an amount of time $\Delta T = T_{\text{LONG}} - T_{\text{SHORT}}$. In fact, with the arrival of each successive short-period crest, the corresponding long-period crest is another ΔT behind. Eventually, after some number n of short periods, the long-period crest will arrive a full long period T_{LONG} after the corresponding short-period crest arrives.

$$n\Delta T = T_{\text{LONG}} \tag{32A.1}$$

This means that as the short-period crest arrives, the long-period crest that precedes the corresponding long-period crest is arriving. This results in constructive interference (loud sound). The time it takes, starting when the interference is maximally constructive, for the interference to again become maximally constructive is the beat period

$$T_{\text{BEAT}} = nT_{\text{SHORT}} \tag{32A.2}$$

Let's use Equation 32A.1 to eliminate the n in this expression. Solving Equation 32A.1 for n we find that

$$n = \frac{T_{\text{LONG}}}{\Delta T}$$

Substituting this into Equation 32A.2 yields

$$T_{\text{BEAT}} = \frac{T_{\text{LONG}}}{\Delta T} T_{\text{SHORT}}$$

ΔT is just $T_{\text{LONG}} - T_{\text{SHORT}}$ so

$$T_{\text{BEAT}} = \frac{T_{\text{LONG}}}{T_{\text{LONG}} - T_{\text{SHORT}}} T_{\text{SHORT}}$$

$$T_{\text{BEAT}} = \frac{T_{\text{LONG}} T_{\text{SHORT}}}{T_{\text{LONG}} - T_{\text{SHORT}}}$$

Dividing top and bottom by the product $T_{\text{LONG}} T_{\text{SHORT}}$ yields

$$T_{\text{BEAT}} = \frac{1}{\frac{1}{T_{\text{SHORT}}} - \frac{1}{T_{\text{LONG}}}}$$

Taking the reciprocal of both sides results in

$$\frac{1}{T_{\text{BEAT}}} = \frac{1}{T_{\text{SHORT}}} - \frac{1}{T_{\text{LONG}}}$$

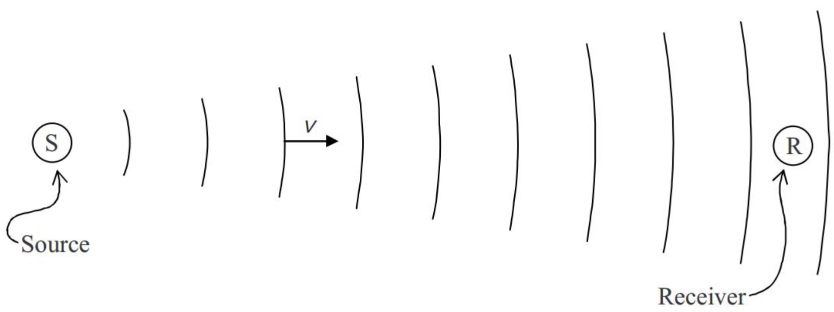
Now we use the frequency-period relation $F = \frac{1}{T}$ to replace each reciprocal period with its corresponding frequency. This yields:

$$f_{\text{BEAT}} = f_{\text{HIGH}} - f_{\text{LOW}} \tag{32A.3}$$

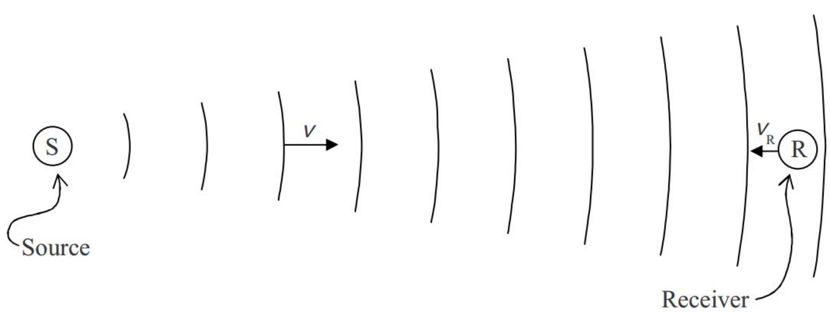
for the beat frequency in terms of the frequencies of the two sources.

The Doppler Effect

Consider a single-frequency sound source and a receiver. The source is something oscillating. It produces sound waves. They travel through air, at speedv, the speed of sound in air, to the receiver and cause some part of the receiver to oscillate. (For instance, if the receiver is your ear, the sound waves cause your eardrum to oscillate.) If the receiver and the source are at rest relative to the air, then the received frequency is the same as the source frequency.



But if the source is moving toward or away from the receiver, and/or the receiver is moving toward or away from the source, the received frequency will be different from the source frequency. Suppose for instance, the receiver is moving toward the source with speed V_R.



The receiver meets wave crests more frequently than it would if it were still. Starting at an instant when a wavefront is at the receiver, the receiver and the next wavefront are coming together at the rate $V + V_R$ (where V is the speed of sound in air). The distance between the wavefronts is just the wavelength λ which is related to the source frequency f by $v = \lambda f$ meaning that $\lambda = \frac{V}{f}$. From the fact that, in the case of constant velocity, distance is just speed times time, we have:

$$\lambda = (V + V_R)T'$$

$$T' = \frac{\lambda}{V + V_R} \tag{32A.4}$$

for the period of the received oscillations. Using $T' = \frac{1}{f'}$ and $\lambda = \frac{V}{f}$ Equation 32A.4 can be written as:

$$\frac{1}{f'} = \frac{V/f}{V + V_R}$$

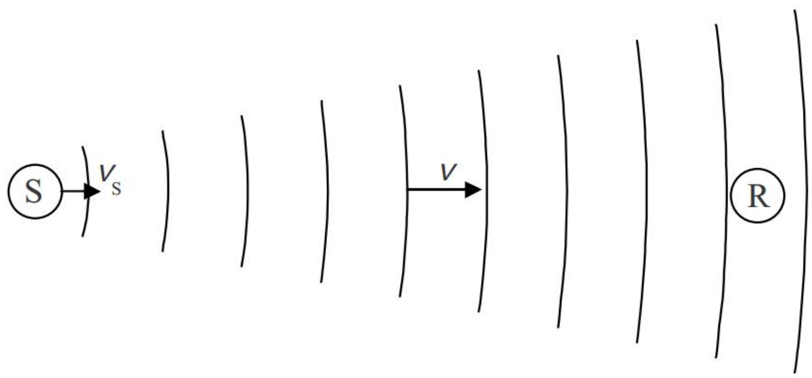
$$\frac{1}{f'} = \frac{V}{V + V_R}\frac{1}{f}$$

$$f' = \frac{V + V_R}{V} f \quad \text{(Receiver Approaching Source)}$$

This equation states that the received frequency f' is a factor times the source frequency. The expression $V + V_R$ is the speed at which the sound wave in air and the receiver are approaching each other. If the receiver is moving away from the source at speed V_R, the speed at which the sound waves are "catching up with" the receiver is $V - V_R$ and our expression for the received frequency becomes

$$f' = \frac{V - V_R}{V} f \quad \text{(Receiver Receding from Source)}$$

Now consider the case in which the source is moving toward the receiver.



The source produces a crest which moves toward the receiver at the speed of sound. But the source moves along behind that crest so the next crest it produces is closer to the first crest than it would be if the source was at rest. This is true of all the subsequent crests as well. They are all closer together than they would be if the source was at rest. This means that the wavelength of the sound waves traveling in the direction of the source is reduced relative to what the wavelength would be if the source was at rest.

The distance d that the source travels toward the receiver in the time from the emission of one crest to the emission of the next crest, that is in period T of the source oscillations, is

$$d = v_s T$$

where v_s is the speed of the source. The wavelength is what the wavelength would be (λ) if the source was at rest, minus the distance $d = v_s T$ that the source travels in one period

$$\lambda' = \lambda - d$$

$$\lambda' = \lambda - v_s T \tag{32A.5}$$

Now we'll use $v = \lambda f$ solved for $\lambda = \frac{v}{f}$ to eliminate the wavelengths and $f = \frac{1}{T}$ solved for the period $T = \frac{1}{f}$ to eliminate the period. With these substitutions, Equation 32A.5 becomes

$$\frac{v}{f'} = \frac{v}{f} - v_s \frac{1}{f}$$

$$\frac{v}{f'} = \frac{1}{f}(v - v_s)$$

$$\frac{f'}{v} = f\frac{1}{v - v_s}$$

$$f' = \frac{v}{v - v_s} f \quad \text{Source Approaching Receiver}$$

If the source is moving away from the receiver, the sign in front of the speed of the source is reversed meaning that

$$f' = \frac{v}{v + v_s} f \quad \text{Source Receding from Receiver}$$

The four expressions for the received frequency as a function of the source frequency are combined on your formula sheet where they are written as:

$$f' = \frac{v \pm v_R}{v \mp v_s} f \tag{32A.6}$$

In solving a Doppler Effect problem, rather than copying this expression directly from your formula sheet, you need to be able to pick out the actual formula that you need. For instance, if the receiver is not moving relative to the air you should omit the $\pm v_R$. If the source is not moving relative to the air, you need to omit the $\mp v_s$. To get the formula just right you need to recognize that when either the source is moving toward the receiver or the receiver is moving toward the source, the Doppler-shifted received frequency is higher (and you need to recognize that when either is moving away from the other, the Doppler-shifted received frequency is lower). You also need enough mathematical savvy to know which sign to choose to make the received frequency f' come out right.

> *Some people get mixed up about the Doppler Effect. They think it's about position rather than about velocity. (It is really about velocity.) If a single frequency sound source is coming at you at constant speed, the pitch (frequency) you hear is higher than the frequency of the source. How much higher depends on how fast the source is coming at you. Folks make the mistake of thinking that the pitch gets higher as the source approaches the receiver. No. That would be the case if the frequency depended on how close the source was to the receiver. It doesn't. The frequency stays the same. The Doppler Effect is about velocity, not position. The whole time the source is moving straight at you, it will sound like it has one single unchanging pitch that is higher than the frequency of the source.* ***Now duck!*** *Once the object passes your position and it is heading away from you it will have one single unchanging pitch that is lower than the frequency of the source*

33A: Fluids: Pressure, Density, Archimedes' Principle

> *One mistake you see in solutions to submerged-object static fluid problems, is the inclusion, in the free body diagram for the problem, in addition to the buoyant force, of a pressure-times-area force typically expressed as $F_P = PA$. This is double counting. Folks that include such a force, in addition to the buoyant force, don't realize that the buoyant force is the net sum of all the pressure-times-area forces exerted, on the submerged object by the fluid in which it is submerged.*

Gases and liquids are fluids. Unlike solids, they flow. A fluid is a liquid or a gas.

Pressure

A fluid exerts pressure on the surface of any substance with which the fluid is in contact. Pressure is force-per-area. In the case of a fluid in contact with a flat surface over which the pressure of the fluid is constant, the magnitude of the force on that surface is the pressure times the area of the surface. Pressure has units of N/m^2.

Never say that pressure is the amount of force exerted on a certain amount of area. Pressure is not an amount of force. Even in the special case in which the pressure over the "certain amount of area" is constant, the pressure is not the amount of force. In such a case, the pressure is what you have to multiply the area by to determine the amount of force.

The fact that the pressure in a fluid is $5N/m^2$ in no way implies that there is a force of 5N acting on a square meter of surface (any more than the fact that the speedometer in your car reads 35 mph implies that you are traveling 35 miles or that you have been traveling for an hour). In fact, if you say that the pressure at a particular point underwater in a swimming pool is $15,000N/m^2$ (fifteen thousand newtons per square meter), you are not specifying any area whatsoever. What you are saying is that any infinitesimal surface element that may be exposed to the fluid at that point will experience an infinitesimal force of magnitude dF that is equal to 15,000 N/m^2 times the area dA of the surface. When we specify a pressure, we're talking about a would-be effect on a would-be surface element.

We talk about an infinitesimal area element because it is entirely possible that the pressure varies with position. If the pressure at one point in a liquid is 15,000 N/m^2 it could very well be 16,000 N/m^2 at a point that's less than a millimeter away in one direction and 14,000 N/m^2 at a point that's less than a millimeter away in another direction.

Let's talk about direction. Pressure itself has no direction. But the force that a fluid exerts on a surface element, because of the pressure of the fluid, does have direction. The force is perpendicular to, and toward, the surface. Isn't that interesting? The direction of the force resulting from some pressure (let's call that the pressure-times-area force) on a surface element is determined by the victim (the surface element) rather than the agent (the fluid).

Pressure Dependence on Depth

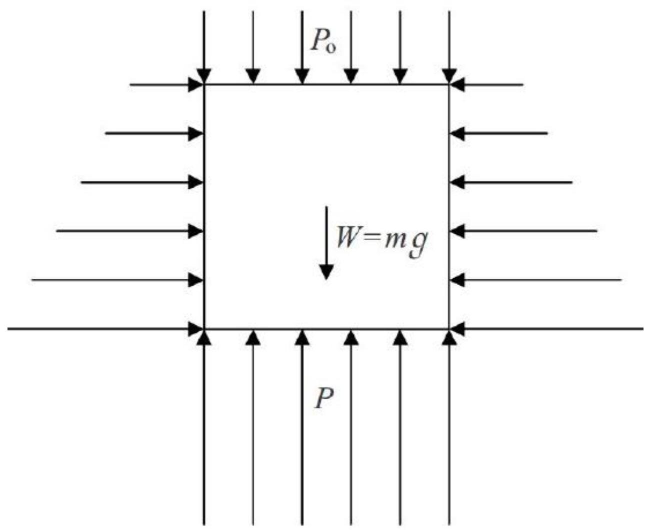
For a fluid near the surface of the earth, the pressure in the fluid increases with depth. You may have noticed this, if you have ever gone deep under water, because you can feel the effect of the pressure on your ear drums. Before we investigate this phenomenon in depth, I need to point out that in the case of a gas, this pressure dependence on depth is, for many practical purposes, negligible. In discussing a container of a gas for instance, we typically state a single value for the pressure of the gas in the container, neglecting the fact that the pressure is greater at the bottom of the container. We neglect this fact because the difference in the pressure at the bottom and the pressure at the top is so very small compared to the pressure itself at the top. We do this when the pressure difference is too small to be relevant, but it should be noted that even a very small pressure difference can be significant. For instance, a helium-filled balloon, released from rest near the surface of the earth would fall to the ground if it weren't for the fact that the air pressure in the vicinity of the lower part of the balloon is greater (albeit only slightly greater) than the air pressure in the vicinity of the upper part of the balloon.

Let's do a thought experiment. (Einstein was fond of thought experiments. They are also called Gedanken experiments. Gedanken is the German word for thought.) Imagine that we construct a pressure gauge as follows: We cap one end of a piece of thin pipe and put a spring completely inside the pipe with one end in contact with the end cap. Now we put a disk whose diameter is equal to the inside diameter of the pipe, in the pipe and bring it into contact with the other end of the spring. We grease the inside walls of the pipe so that the disk can slide freely along the length of the pipe, but we make the fit exact so that no fluid can get past the disk.

Now we drill a hole in the end cap, remove all the air from the region of the pipe between the disk and the end cap, and seal up the hole. The position of the disk in the pipe, relative to its position when the spring is neither stretched nor compressed, is directly proportional to the pressure on the outer surface, the side facing away from the spring, of the disk. We calibrate (mark a scale on) the pressure gauge that we have just manufactured, and use it to investigate the pressure in the water of a swimming pool. First we note that, as soon as we removed the air, the gauge started to indicate a significant pressure (around 1.013× 105N/m^2), namely the air pressure in the atmosphere. Now we move the gauge around and watch the gauge reading. Wherever we put the gauge (we define the location of the gauge to be the position of the center point on the outer surface of the disk) on the surface of the water, we get one and the same reading, (the air pressure reading). Next we verify that the pressure reading does indeed increase as we lower the gauge deeper and deeper into the water. Then we find, the point I wrote this paragraph to make, that if we move the gauge around horizontally at one particular depth, the pressure reading does not change. That's the experimental result I want to use in the following development, the experimental fact that the pressure has one and the same value at all points that are at one and the same depth in a fluid.

Here we derive a formula that gives the pressure in an incompressible static fluid as a function of the depth in the fluid. Let's get back into the swimming pool. Now imagine a closed surface enclosing a volume, a region in space, that is full of water. I'm going to call the water in such a volume, "a volume of water," and I'm going to give it another name as well. If it were ice, I would call it a chunk of ice, but since it is liquid water, I'll call it a "slug" of water. We're going to derive the pressure vs. depth relation by investigating the equilibrium of an "object" which is a slug of water.

Consider a cylindrical slug of water whose top is part of the surface of the swimming pool and whose bottom is at some arbitrary depth h below the surface. I'm going to draw the slug here, isolated from its surroundings. The slug itself is, of course, surrounded by the rest of the water in the pool.



In the diagram, we use arrows to convey the fact that there is pressure-times-area force on every element of the surface of the slug. Now the downward pressure-times-area force on the top of the slug is easy to express in terms of the pressure because the pressure on every infinitesimal area element making up the top of the slug has one and the same value. In terms of the determination of the pressure-times-area, this is the easy case. The magnitude of the force, F_o, is just the pressure P_o times the area A of the top of the cylinder.

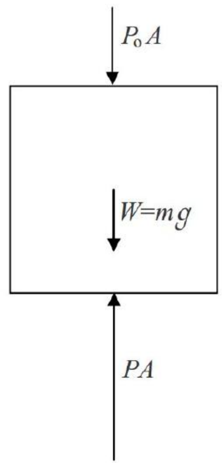
$$F_o = P_o A$$

A similar argument can be made for the bottom of the cylinder. All points on the bottom of the cylinder are at the same depth in the water so all points are at one and the same pressure P. The bottom of the cylinder has the same area A as the top so the magnitude of the upward force F on the bottom of the cylinder is given by

$$F = PA$$

As to the sides, if we divide the sidewalls of the cylinder up into an infinite set of equal-sized infinitesimal area elements, for every sidewall area element, there is a corresponding area element on the opposite side of the cylinder. The pressure is the same on both elements because they are at the same depth. The two forces then have the same magnitude, but because the elements face in opposite directions, the forces have opposite directions. Two opposite but equal forces add up to zero. In such a manner, all the forces on the sidewall area elements cancel each other out.

Now we are in a position to draw a free body diagram of the cylindrical slug of water.



Applying the equilibrium condition

$$\sum F_{\uparrow} = 0$$

yields

$$PA - mg - P_o A = 0 \tag{33A.1}$$

At this point in our derivation of the relation between pressure and depth, the depth does not explicitly appear in the equation. The mass of the slug of water, however, does depend on the length of the slug which is indeed the depth h. First we note that

$$m = \varrho V \tag{33A.2}$$

where ϱ is the density, the mass-per-volume, of the water making up the slug and V is the volume of the slug. The volume of a cylinder is its height times its face area so we can write

$$m = \varrho h A$$

Substituting this expression for the mass of the slug into equation 33A.1 yields

$$PA - \varrho h A g - P_o A = 0$$

$$P - \varrho h g - P_o = 0$$

$$P = P_o + \varrho g h \tag{33A.3}$$

While we have been writing specifically about water, the only thing in the analysis that depends on the identity of the incompressible fluid is the density ϱ. Hence, as long as we use the density of the fluid in question, equation 33A.3 ($P = P_o + \varrho g h$) applies to any incompressible fluid. It says that the pressure at any depth h is the pressure at the surface plus $\varrho g h$.

A few words on the units of pressure are in order. We have stated that the units of pressure are N/m^2. This combination of units is given a name. It is called the *pascal*, abbreviated Pa.

$$1 Pa = 1 \frac{N}{m^2}$$

Pressures are often quoted in terms of the non-SI unit of pressure, the atmosphere, abbreviated atm and defined such that, on the average, the pressure of the earth's atmosphere at sea level is 1 atm. In terms of the pascal,

$$1 \text{ atm} = 1.013 \times 10^5 \text{Pa}$$

The big mistake that folks make in applying equation $33A.3$ ($P = P_o + \varrho gh$) is to ignore the units. They'll use 1 atm for P_o and without converting that to pascals, they'll add the product ϱgh to it. Of course, if one uses SI units for ϱ, g, and h, the product ϱgh comes out in N/m^2 which is a pascal which is definitely not an atmosphere (but rather, about a hundred-thousandth of an atmosphere). Of course one can't add a value in pascals to a value in atmospheres. The way to go is to convert the value of P_o that was given to you in units of atmospheres, to pascals, and then add the product rgh (in SI units) to your result so that your final answer cmes out in pascals.

Gauge Pressure

Remember the gauge we constructed for our thought experiment? That part about evacuating the inside of the pipe presents quite the manufacturing challenge. The gauge would become inaccurate as air leaked in by the disk. As regards function, the description is fairly realistic in terms of actual pressure gauges in use, except for the pumping of the air out the pipe. To make it more like an actual gauge that one might purchase, we would have to leave the interior open to the atmosphere. In use then, the gauge reads zero when the pressure on the sensor end is 1 atmosphere, and in general, indicates the amount by which the pressure being measured exceeds atmospheric pressure. This quantity, the amount by which a pressure exceeds atmospheric pressure, is called gauge pressure (since it is the value registered by a typical pressure gauge.) When it needs to be contrasted with gauge pressure, the actual pressure that we have been discussing up to this point is called absolute pressure. The absolute pressure and the gauge pressure are related by:

$$P = P_G + P_o \tag{33A.4}$$

where:

P is the absolute pressure,

P_G is the gauge pressure, and

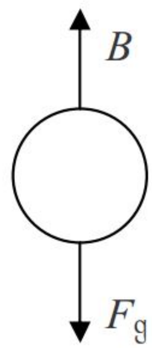
P_o is atmospheric pressure.

When you hear a value of pressure (other than the so-called barometric pressure of the earth's atmosphere) in your everyday life, it is typically a gauge pressure (even though one does not use the adjective "gauge" in discussing it.) For instance, if you hear that the recommended tire pressure for your tires is 32 psi (pounds per square inch) what is being quoted is a gauge pressure. Folks that work on ventilation systems often speak of negative air pressure. Again, they are actually talking about gauge pressure, and a negative value of gauge pressure in a ventilation line just means that the absolute pressure is less than atmospheric pressure.

Archimedes' Principle

The net pressure-times-area force on an object submerged in a fluid, the vector sum of the forces on all the infinite number of infinitesimal surface area elements making up the surface of an object, is upward because of the fact that pressure increases with depth. The upward pressuretimes-area force on the bottom of an object is greater than the downward pressure-times-area force on the top of the object. The result is a net upward force on any object that is either partly or totally submerged in a fluid. The force is called the buoyant force on the object. The agent of the buoyant force is the fluid.

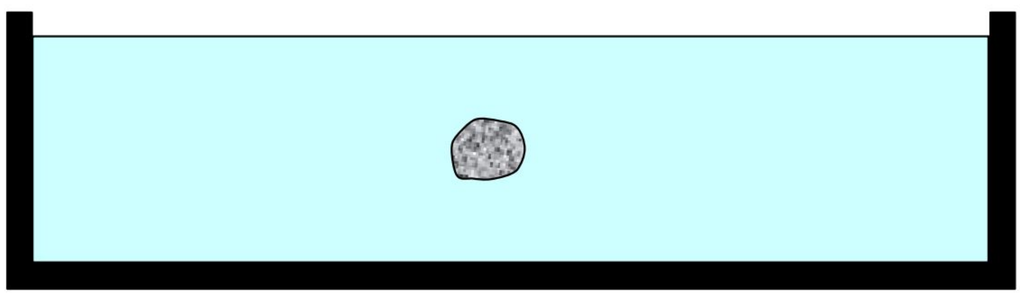
If you take an object in your hand, submerge the object in still water, and release the object from rest, one of three things will happen: The object will experience an upward acceleration and bob to the surface, the object will remain at rest, or the object will experience a downward acceleration and sink. We have emphasized that the buoyant force is always upward. So why on earth would the object ever sink? The reason is, of course, that after you release the object, the buoyant force is not the only force acting on the object. The gravitational force still acts on the object when the object is submerged. Recall that the earth's gravitational field permeates everything. For an object that is touching nothing of substance but the fluid it is in, the free body diagram (without the acceleration vector being included) is always the same (except for the relative lengths of the arrows):



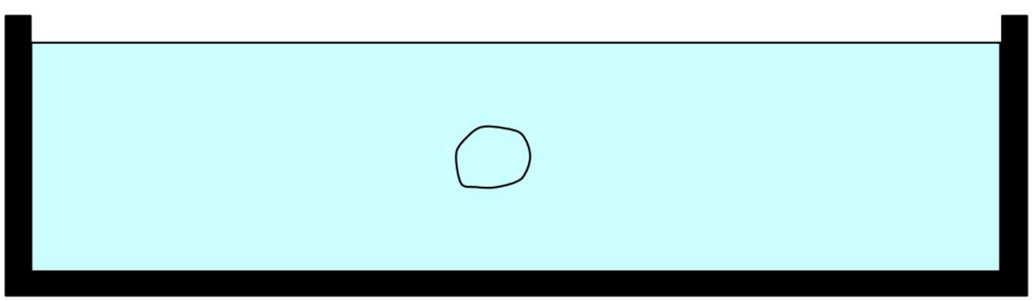
and the whole question as to whether the object (released from rest in the fluid) sinks, stays put, or bobs to the surface, is determined by how the magnitude of the buoyant force compares with that of the gravitational force. If the buoyant force is greater, the net force is upward and the object bobs toward the surface. If the buoyant force and the gravitational force are equal in magnitude, the object stays put. And if the gravitational force is greater, the object sinks.

So how does one determine how big the buoyant force on an object is? First, the trivial case: If the only forces on the object are the buoyant force and the gravitational force, and the object remains at rest, then the buoyant force must be equal in magnitude to the gravitational force. This is the case for an object such as a boat or a log which is floating on the surface of the fluid it is in.

But suppose the object is not freely floating at rest. Consider an object that is submerged in a fluid. We have no information on the acceleration of the object, but we cannot assume it to be zero. Assume that a person has, while maintaining a firm grasp on the object, submerged the object in fluid, and then, released it from rest. We don't know which way it is going from there, but we can not assume that it is going to stay put.



To derive our expression for the buoyant force, we do a little thought experiment. Imagine replacing the object with a slug of fluid (the same kind of fluid as that in which the object is submerged), where the slug of fluid has the exact same size and shape as the object.



From our experience with still water we know that the slug of fluid would indeed stay put, meaning that it is in equilibrium.

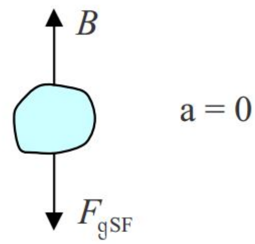


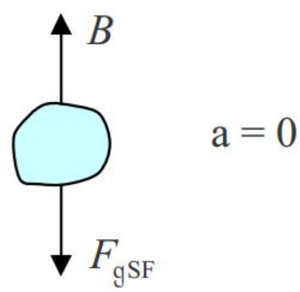
Table of Forces			
Symbol=?	Name	Agent	Victim
B	Buoyant Force	The Surrounding Fluid	The Slug of Fluid
$F_{g\,SF} = m_{SF}\,g$	Gravitational Force on the Slug of Fluid	The Earth Gravitational Force	The Slug of Fluid

Applying the equilibrium equation $\sum F_{\uparrow} = 0$ to the slug of fluid yields:

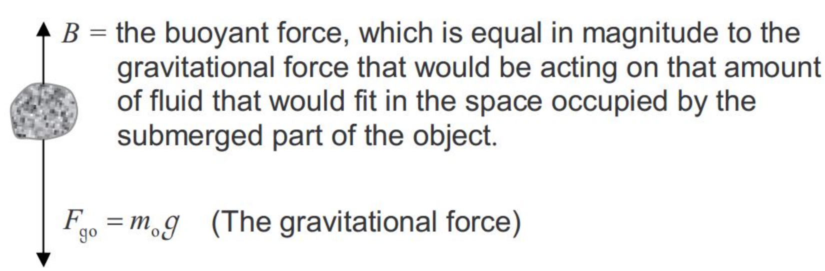
$$\sum F_{\uparrow} = 0$$

$$B - F_{g\,SF} = 0$$

$$B = F_{g\,SF}$$



The last equation states that the buoyant force on the slug of fluid is equal to the gravitational force on the slug of fluid. Now get this; this is the crux of the derivation: Because the slug of fluid has the exact same size and shape as the original object, it presents the exact same surface to the surrounding fluid, and hence, the surrounding fluid exerts the same buoyant force on the slug of fluid as it does on the original object. Since the buoyant force on the slug of fluid is equal in magnitude to the gravitational force acting on the slug of fluid, the buoyant force on the original object is equal in magnitude to the gravitational force acting on the slug of fluid. This is Archimedes' principle.



Archimedes' Principle states that: The buoyant force on an object that is either partly or totally submerged in a fluid is upward, and is equal in magnitude to the gravitational force that would be acting on that amount of fluid that would be where the object is if the object wasn't there. For an object that is totally submerged, the volume of that amount of fluid that would be where the object is if the object wasn't there is equal to the volume of the object itself. But for an object hat is only partly submerged, the volume of that amount of fluid that would be where the object is if the object wasn't there is equal to the (typically unknown) volume of the submerged part of the object. However, if the object is freely floating at rest, the equilibrium equation (instead of Archimedes' Principle) can be used to quickly establish that the buoyant force (of a freely floating object such as a boat) is equal in magnitude to the gravitational force acting on the object itself.

34A: Pascal’s Principle, the Continuity Equation, and Bernoulli’s Principle

There are a couple of mistakes that tend to crop up with some regularity in the application of the Bernoulli equation $P+\frac{1}{2}\varrho v^2+\varrho gh=\text{constant}$. Firstly, folks tend to forget to create a diagram in order to identify point 1 and point 2 in the diagram so that they can write the Bernoulli equation in its useful form:

$$P_1+\frac{1}{2}\varrho v_1^2+\varrho gh_1=P_2+\frac{1}{2}\varrho v_2^2+\varrho gh_2.$$

Secondly, when both the velocities in Bernoulli’s equation are unknown, they forget that there is another equation that relates the velocities, namely, the continuity equation in the form $A_1v_1=A_2v_2$ which states that the flow rate at position 1 is equal to the flow rate at position 2.

Pascal’s Principle

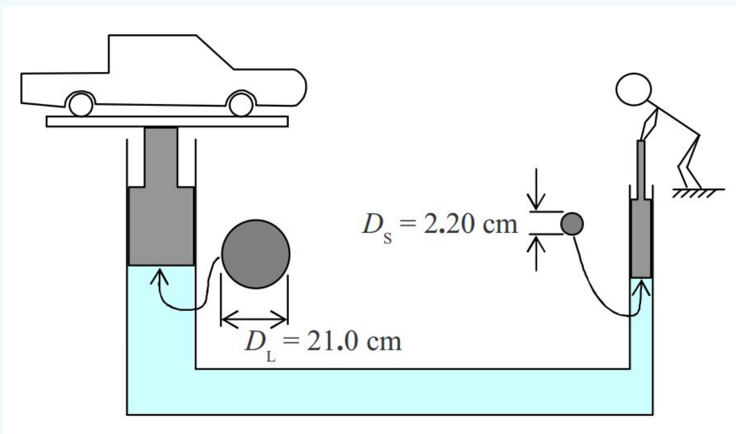
Experimentally, we find that if you increase the pressure by some given amount at one location in a fluid, the pressure increases by that same amount everywhere in the fluid. This experimental result is known as Pascal’s Principle.

We take advantage of Pascal’s principle every time we step on the brakes of our cars and trucks. The brake system is a hydraulic system. The fluid is oil that is called hydraulic fluid. When you depress the brake pedal you increase the pressure everywhere in the fluid in the hydraulic line. At the wheels, the increased pressure acting on pistons attached to the brake pads pushes them against disks or drums connected to the wheels.

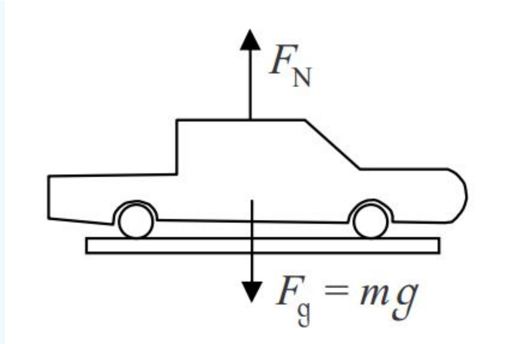
A simple hydraulic lift consists of two pistons, one larger than the other, in cylinders connected by a pipe. The cylinders and pipe are filled with water. In use, a person pushes down upon the smaller piston and the water pushes upward on the larger piston. The diameter of the smaller piston is 2.20 centimeters. The diameter of the larger piston is 21.0 centimeters. On top of the larger piston is a metal support and on top of that is a car. The combined mass of the support-plus-car is 998 kg. Find the force that the person must exert on the smaller piston to raise the car at a constant velocity. Neglect the masses of the pistons.

Solution

We start our solution with a sketch.



Now, let’s find the force R_N exerted on the larger piston by the car support. By Newton’s 3rd Law, it is the same as the normal force F_N exerted by the larger piston on the car support. We’ll draw and analyze the free body diagram of the car-plus-support to get that.



$$\sum F_{\uparrow} = 0$$

$$F_N - mg = 0$$

$$F_N = mg$$

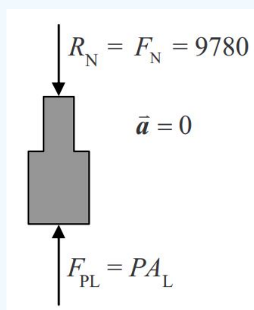
$$F_N = 998 kg\,(9.80 \frac{\text{newtons}}{kg})$$

$$F_N = 9780\,\text{newtons}$$

Table of Forces

Symbol	Name	Agent	Victim
$F_g = mg$	Gravitational Force on Support-Plus Car	The Earth's Gravitational Field	The Support-Plus Car
F_N	Normal Force	The Large Piston	The Support-Plus Car

Now we analyze the equilibrium of the larger piston to determine what the pressure in the fluid must be in order for the fluid to exert enough force on the piston (with the car-plus-support on it) to keep it moving at constant velocity.



$$\sum F_{\uparrow} = 0$$

$$F_{PL} - R_N = 0$$

$$PA_L - R_N = 0$$

$$P = \frac{R_N}{A_L} \tag{34A.1}$$

Table of Forces

Symbol	Name	Agent	Victim
$R_N = F_N = 9780\,\text{newtons}$	Interaction Partner to Normal Force (see above)	Support (That part, of the hydraulic lift, that the car is on).	Large Piston

Symbol	Name	Agent	Victim
F_{PL}	Pressure-Related Force on Large Piston	The Water	Large Piston

A_L is the area of the face of the larger piston. We can use the given larger piston diameter $D_L = 0.210\ m$ to determine the area of the face of the larger piston as follows:

$$A_L = \pi r_L^2$$

where $r_L = \frac{D_L}{2}$ is the radius of the larger piston.

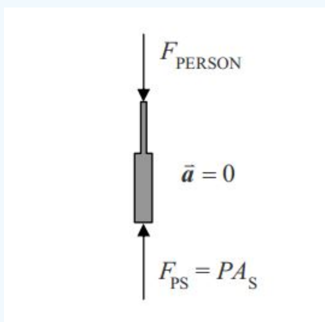
$$\begin{aligned} A_L &= \pi\left(\frac{D_L}{2}\right)^2 \\ &= \pi\left(\frac{0.210\mathrm{m}}{2}\right)^2 \\ &= 0.03464\, m^2 \end{aligned}$$

Substituting this and the value $R_N = F_N = 9780\ \text{newtons}$ into equation 34A.1 above yields

$$\begin{aligned} P &= \frac{9780\ \text{newtons}}{0.03464\ \mathrm{m}^2} \\ &= 282333\,\frac{N}{m^2} \end{aligned}$$

We intentionally keep 3 too many significant figures in this intermediate result.

Now we just have to analyze the equilibrium of the smaller piston to determine the force that the person must exert on the smaller piston.



$$\sum F_{\uparrow} = 0$$

$$F_{PS} - F_{PERSON} = 0$$

$$PA_S - F_{PERSON} = 0$$

$$F_{PERSON} = PA_S$$

The area A_S of the face of the smaller piston is just π times the square of the radius of the smaller piston where the radius is $\frac{D_S}{2}$, half the diameter of the smaller piston. So:

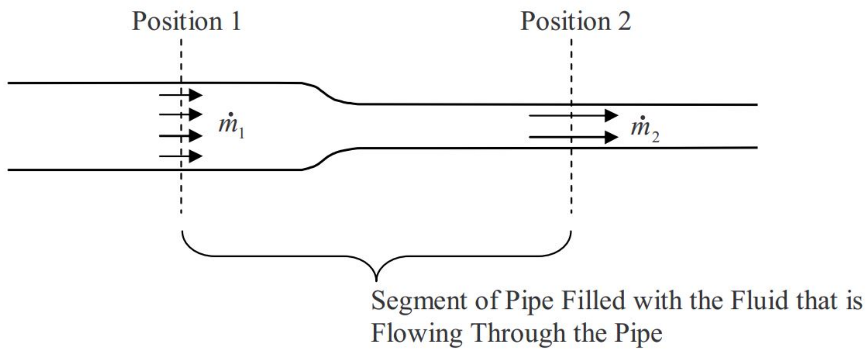
$$F_{PERSON} = P\pi\left(\frac{D_S}{2}\right)^2$$

$$F_{PERSON} = 282333\frac{N}{m^2}\pi\left(\frac{0.0220m}{2}\right)^2$$

$$F_{PERSON} = 107N$$

Fluid in Motion—the Continuity Principle

The Continuity Principle is a fancy name for something that common sense will tell you has to be the case. It is simply a statement of the fact that for any section of a single pipe, filled with an incompressible fluid (an idealization approached by liquids), through which the fluid with which the pipe is filled is flowing, the amount of fluid that goes in one end in any specified amount of time is equal to the amount that comes out the other end in the same amount of time. If we quantify the amount of fluid in terms of the mass, this is a statement of the conservation of mass. Having stipulated that the segment is filled with fluid, the incoming fluid has no room to expand in the segment. Having stipulated that the fluid is incompressible, the molecules making up the fluid cannot be packed closer together; that is, the density of the fluid cannot change. With these stipulations, the total mass of the fluid in the segment of pipe cannot change, so, any time a certain mass of the fluid flows in one end of the segment, the same mass of the fluid must flow out the other end.

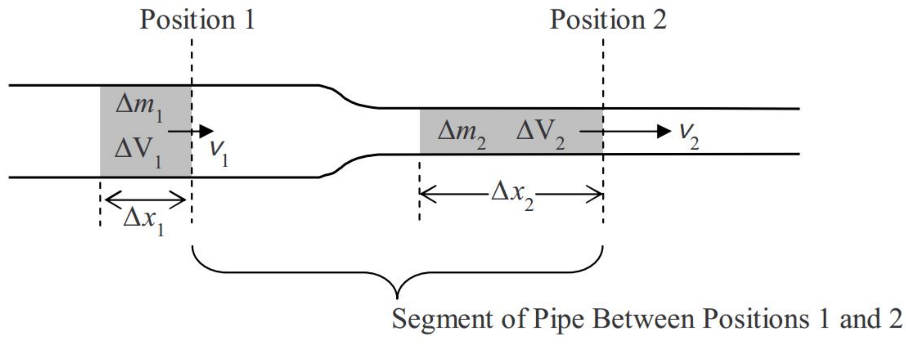


This can only be the case if the mass flow rate, the number of kilograms-per-second passing a given position in the pipe, is the same at both ends of the pipe segment.

$$\dot{m_1} = \dot{m_2} \tag{34A.2}$$

An interesting consequence of the continuity principle is the fact that, in order for the mass flow rate (the number of kilograms per second passing a given position in the pipe) to be the same in a fat part of the pipe as it is in a skinny part of the pipe, the velocity of the fluid (i.e. the velocity of the molecules of the fluid) must be greater in the skinny part of the pipe. Let's see why this is the case.

Here, we again depict a pipe in which an incompressible fluid is flowing.



Keeping in mind that the entire pipe is filled with the fluid, the shaded region on the left represents the fluid that will flow past position 1 in time Δt and the shaded region on the right represents the fluid that will flow past position 2 in the same time Δt. In both cases, in order for the entire slug of fluid to cross the relevant position line, the slug must travel a distance equal to its length. Now the slug labeled Δm_2 has to be longer than the slug labeled Δm_1 since the pipe is skinnier at position 2 and by the continuity equation $\Delta m_1 = \Delta m_2$ (the amount of fluid that flows into the segment of the pipe between position 1 and position 2 is equal to the amount of fluid that flows out of it). So, if the slug at position 2 is longer and it has to travel past the position line in the same amount of time as it takes for the slug at position 1 to travel past its position line, the fluid velocity at position 2 must be greater. The fluid velocity is greater at a skinnier position in the pipe.

Let's get a quantitative relation between the velocity at position 1 and the velocity at position 2.

Starting with

$$\Delta m_1 = \Delta m_2$$

we use the definition of density to replace each mass with the density of the fluid times the relevant volume:

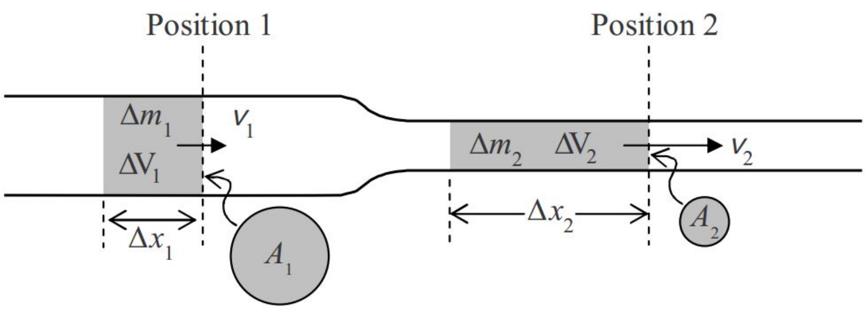
$$\varrho \Delta V_1 = \varrho \Delta V_2$$

Dividing both sides by the density tells us something you already know:

$$\Delta V_1 = \Delta V_2 \tag{34A.3}$$

> If you divide both sides of Equation 34A.3 by Δt and take the limit as Δt goes to zero, we have $\dot{V}_1 = \dot{V}_2$ which is an expression of the continuity principle in terms of volume flow rate. The volume flow rate is typically referred to simply as the flow rate. While we use the SI units $\frac{\text{m}^3}{\text{s}}$ for flow rate, the reader may be more familiar with flow rate expressed in units of gallons per minute.

Now back to our goal of finding a mathematical relation between the velocities of the fluid at the two positions in the pipe. Here we copy the diagram of the pipe and add, to the copy, a depiction of the face of slug 1 of area A_1 and the face of slug 2 of area A_2.



We left off with the fact that $\Delta V_1 = \Delta V_2$. Each volume can be replaced with the area of the face of the corresponding slug times the length of that slug. So,

$$A_1 \Delta x_1 = A_2 \Delta x_2$$

Recall that Δx_1 is not only the length of slug 1, it is also how far slug 1 must travel in order for the entire slug of fluid to get past the position 1 line. The same is true for slug 2 and position 2. Dividing both sides by the one time interval Δt yields:

$$A_1 \frac{\Delta x_1}{\Delta t} = A_2 \frac{\Delta x_2}{\Delta t}$$

Taking the limit as Δt goes to zero results in:

$$A_1 v_1 = A_2 v_2 \tag{34A.4}$$

This is the relation, between the velocities, that we have been looking for. It applies to any pair of positions in a pipe completely filled with an incompressible fluid. It can be written as

$$Av = \text{constant} \tag{34A.5}$$

which means that the product of the cross-sectional area of the pipe and the velocity of the fluid at that cross section is the same for every position along the fluid-filled pipe. To take advantage of this fact, one typically writes, in equation form, that the product Av at one location is equal tothe same product at another location. In other words, one writes equation 34A.4.

Note that the expression Av, the product of the cross-sectional area of the pipe, at a particular position, and the velocity of the fluid at that same position, having been derived by dividing an expression for the volume of fluid ΔV that would flow past a given position of the pipe in time Δt, by Δt, and taking the limit as Δt goes to zero, is none other than the flow rate (the volume flow rate) discussed in the aside above.

$$\text{Flow Rate} = Av$$

Further note that if we multiply the flow rate by the density of the fluid, we get the mass flow rate.

$$\dot{m} = \varrho A v \tag{34A.6}$$

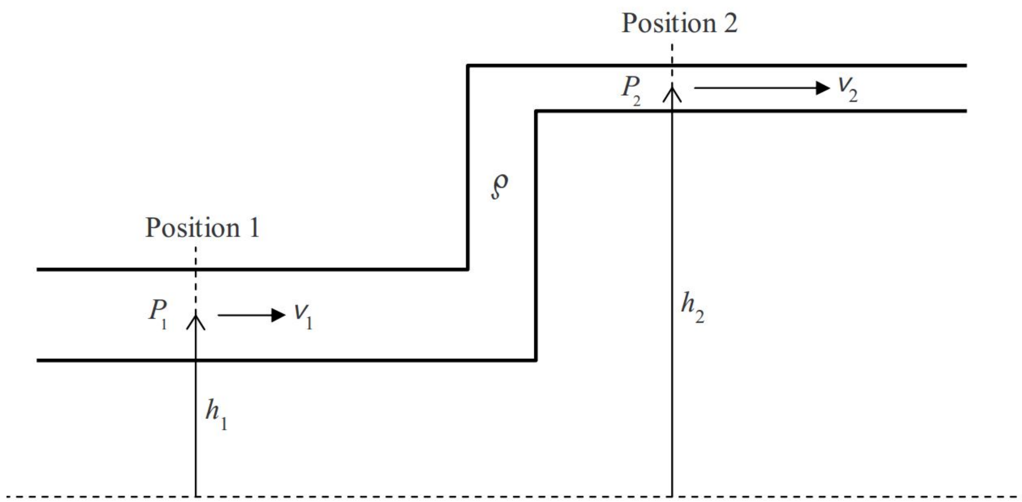
Fluid in Motion—Bernoulli's Principle

The derivation of Bernoulli's Equation represents an elegant application of the Work-Energy Theorem. Here we discuss the conditions under which Bernoulli's Equation applies and then simply state and discuss the result.

Bernoulli's Equation applies to a fluid flowing through a full pipe. The degree to which Bernoulli's Equation is accurate depends on the degree to which the following conditions are met:

1. The fluid must be experiencing steady state flow. This means that the flow rate at all positions in the pipe is not changing with time.
2. The fluid must be experiencing streamline flow. Pick any point in the fluid. The infinitesimal fluid element at that point, at an instant in time, traveled along a certain path to arrive at that point in the fluid. In the case of streamline flow, every infinitesimal element of fluid that ever finds itself at that same point traveled the same path. (Streamline flow is the opposite of turbulent flow.)
3. The fluid must be non-viscous. This means that the fluid has no tendency to "stick to" either the sides of the pipe or to itself. (Molasses has high viscosity. Alcohol has low viscosity.)

Consider a pipe full of a fluid that is flowing through the pipe. In the most general case, the cross-sectional area of the pipe is not the same at all positions along the pipe and different parts of the pipe are at different elevations relative to an arbitrary, but fixed, reference level.



Pick any two positions along the pipe, e.g. positions 1 and 2 in the diagram above. (You already know that, in accord with the continuity principle, $A_1 v_1 = A_2 v_2$.) Consider the following unnamed sum of terms:

$$P + \frac{1}{2}\varrho v^2 + \varrho g h$$

where, at the position under consideration:

- P is the pressure of the fluid,
- ϱ (the Greek letter rho) is the density of the fluid,
- v is the magnitude of the velocity of the fluid,
- $g = 9.80 \frac{N}{kg}$ is the near-surface magnitude of the earth's gravitational field, and
- h is the elevation, relative to a fixed reference level, of the position in the pipe.

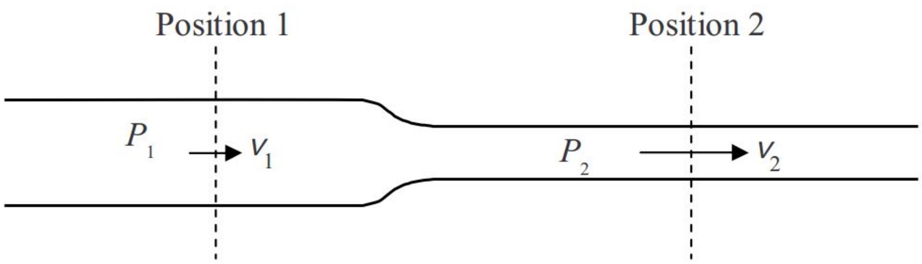
The Bernoulli Principle states that this unnamed sum of terms has the same value at each and every position along the pipe. Bernoulli's equation is typically written:

$$P+\frac{1}{2}\varrho v^2+\varrho gh=\text{constant} \tag{34A.7}$$

but to use it, you have to pick two positions along the pipe and write an equation stating that the value of the unnamed sum of terms is the same at one of the positions as it is at the other.

$$P_1+\frac{1}{2}\varrho v_1^2+\varrho gh_1=P_2+\frac{1}{2}\varrho v_2^2+\varrho gh_2 \tag{34A.8}$$

A particularly interesting characteristic of fluids is incorporated in this equation. Suppose positions 1 and 2 are at one and the same elevation as depicted in the following diagram:



Then $h_1=h_2$ in equation 34A.8 and equation 34A.8 becomes:

$$P_1+\frac{1}{2}\varrho v_1^2=P_2+\frac{1}{2}\varrho v_2^2$$

Check it out. If $v_2>v_1$ then P_2 must be less than P_1 in order for the equality to hold. This equation is saying that, where the velocity of the fluid is high, the pressure is low.

35A: Temperature, Internal Energy, Heat and Specific Heat Capacity

As you know, temperature is a measure of how hot something is. Rub two sticks together and you will notice that the temperature of each increases. You did work on the sticks and their temperature increased. Doing work is transferring energy. So you transferred energy to the sticks and their temperature increased. This means that an increase in the temperature of a system is an indication of an increase in the internal energy (a.k.a. thermal energy) of the system. (In this context the word system is thermodynamics jargon for the generalization of the word object. Indeed an object, say an iron ball, could be a system. A system is just the subject of our investigations or considerations. A system can be as simple as a sample of one kind of gas or a chunk of one kind of metal, or it can be more complicated as in the case of a can plus some water in the can plus a thermometer in the water plus a lid on the can. For the case at hand, the system is the two sticks.) The internal energy of a system is energy associated with the motion of molecules, atoms, and the particles making up atoms relative to the center of mass of the system, and the potential energy corresponding to the positions and velocities of the aforementioned submicroscopic constituents of the system relative to each other. As usual with energy accounting, the absolute zero of energy in the case of internal energy doesn't matter—only changes in internal energy have any relevance. As such, you or the publisher of a table of internal energy values (for a given substance, publishers actually list the internal energy per mass or the internal energy per mole of the substance under specified conditions rather that the internal energy of a sample of such a substance), are free to choose the zero of internal energy for a given system. In making any predictions regarding a physical process involving that system, as long as you stick with the same zero of internal energy throughout your analysis, the measurable results of your prediction or explanation will not depend on your choice of the zero of internal energy.

Another way of increasing the temperature of a pair of sticks is to bring them into contact with something hotter than the sticks are. When you do that, the temperature of the sticks automatically increases—you don't have to do any work on them. Again, the increase in the temperature of either stick indicates an increase in the internal energy of that stick. Where did that energy come from? It must have come from the hotter object. You may also notice that the hotter object's temperature decreased when you brought it into contact with the sticks. The decrease in temperature of the hotter object is an indication that the amount of internal energy in the hotter object decreased. You brought the hotter object in contact with the sticks and energy was automatically transferred from the hotter object to the sticks. The energy transfer in this case is referred to as the flow of heat. Heat is energy that is automatically transferred from a hotter object to a cooler object when you bring the two objects in contact with each other. Heat is not something that a system has but rather energy that is transferred or is being transferred. Once it gets to the system to which it is transferred we call it internal energy. The idea is to distinguish between what is being done to a system, "Work is done on the system and/or heat is caused to flow into it", with how the system changes as a result of what was done to it, "The internal energy of the system increases."

The fact that an increase in the temperature of an object is an indication of energy transferred to that object might suggest that anytime you transfer energy to an object its temperature increases. But this is not the case. Try putting a hot spoon in a glass of ice water. (Here we consider a case for which there is enough ice so that not all of the ice melts.) The spoon gets as cold as the ice water and some of the ice melts, but the temperature of the ice water remains the same (0 °C). The cooling of the spoon indicates that energy was transferred from it, and since the spoon was in contact with the ice water the energy must have been transferred to the ice water. Indeed the ice does undergo an observable change; some of it melts. The presence of more liquid water and less ice is an indication that there is more energy in the ice water. Again there has been a transfer of energy from the spoon to the ice water. This transfer is an automatic flow of heat that takes place when the two systems are brought into contact with one another. Evidently, heat flow does not always result in a temperature increase.

Experiment shows that when a higher temperature object is in contact with a lower temperature object, heat is flowing from the higher temperature object to the lower temperature object. The flow of heat persists until the two objects are at one and the same temperature. We define the average translational kinetic energy of a molecule of a system as the sum of the translational kinetic energies of all the molecules making up the system divided by the total number of molecules. When two simple ideal gas systems, each involving a multitude of single atom molecules interacting via elastic collisions, are brought together, we find that heat flows from the system in which the average translational kinetic energy per molecule is greater to the system in which the average translational kinetic energy per molecule is lesser. This means the former system is at a higher temperature. That is to say that the higher the translational kinetic energy, on the average, of the particles making up the system, the higher the temperature. This is true for many systems.

Solids consist of atoms that are bound to neighboring atoms such that molecules tend to be held in their position, relative to the bulk of the solid, by electrostatic forces. A pair of molecules that are bound to each other has a lower amount of internal potential energy relative to the same pair of molecules when they are not bound together because we have to add energy to the bound pair at rest to yield the free pair at rest. In the case of ice water, the transfer of energy into the ice water results in the breaking of bonds between water molecules, which we see as the melting of the ice. As such, the transfer of energy into the ice water results in an increase in the internal potential energy of the system.

The two different kinds of internal energy that we have discussed are internal potential energy and internal kinetic energy. When there is a net transfer of energy into a system, and the macroscopic mechanical energy of the system doesn't change (e.g. for the case of an object near the surface of the earth, the speed of object as a whole does not increase, and the elevation of the object does not increase), the internal energy (the internal kinetic energy, the internal potential energy, or both) of the system increases. In some, but not all, cases, the increase in the internal energy is accompanied by an increase in the temperature of the system. If the temperature doesn't increase, then we are probably dealing with a case in which it is the internal potential energy of the system that increases.

Heat Capacity and Specific Heat Capacity

Let's focus our attention on cases in which heat flow into a sample of matter is accompanied by an increase in the temperature of the sample. For many substances, over certain temperature ranges, the temperature change is (at least approximately) proportional to the amount of heat that flows into the substance.

$$\Delta T \propto Q$$

Traditionally, the constant of proportionality is written as $\frac{1}{C}$ so that

$$\Delta T = \frac{1}{C} Q$$

where the upper case C is the heat capacity. This equation is more commonly written as

$$Q = C\Delta T \tag{35A.1}$$

which states that the amount of heat that must flow into a system to change the temperature of that system by ΔT is the heat capacity C times the desired temperature change ΔT. Thus the heat capacity C is the "heat-per-temperature-change." It's reciprocal is a measure of a system's temperature sensitivity to heat flow.

Let's focus our attention on the simplest kind of system, a sample of one kind of matter, such as a certain amount of water. The amount of heat that is required to change the temperature of the sample by a certain amount is directly proportional to the mass of the single substance; e.g., if you double the mass of the sample it will take twice as much heat to raise its temperature by, for instance, 1 C°. Mathematically, we can write this fact as

$$C \propto m$$

It is traditional to use a lower case c for the constant of proportionality. Then

$$C = cm$$

where the constant of proportionality C is the heat-capacity-per-mass of the substance in question. The heat-capacity-per-mass C is referred to as the mass specific heat capacity or simply the specific heat capacity of the substance in question. (In this context, the adjective specific means "per amount." Because the amount can be specified in more than one way we have the expression "mass specific" meaning "per amount of mass" and the expression "molar specific" meaning "per number of moles." Here, since we are only dealing with mass specific heat, we can omit the word "mass" without generating confusion.) The specific heat capacity c has a different value for each different kind of substance in the universe. (Okay, there might be some coincidental duplication but you get the idea.) In terms of the mass specific heat capacity, equation 35A.1 ($Q = C\Delta T$), for the case of a system consisting only of a sample of a single substance, can be written as

$$Q = mc\Delta T \tag{35A.2}$$

The specific heat capacity C is a property of the kind of matter of which a substance consists. As such, the values of specific heat for various substances can be tabulated.

Substance	Specific Heat Capacity* $\left[\frac{J}{kg \cdot C^\circ}\right]$

Substance	Specific Heat Capacity* $\left[\frac{J}{kg \cdot C^{\circ}}\right]$
Ice (solid water)	2090
Liquid Water	4186
Water Vapor (gas)	2000
Solid Copper	387
Solid Aluminum	902
Solid Iron	448

∗The specific heat capacity of a substance varies with temperature and pressure. The values given correspond to atmospheric pressure. Use of these representative constant values for cases involving atmospheric pressure and temperature ranges between -100°C and +600°C , as applicable for the phase of the material, can be expected to yield reasonable results but if precision is required, or information on how reasonable your results are is needed, you should consult a thermodynamics textbook and thermodynamics tables and carry out a more sophisticated analysis.

Note how many more Joules of energy are needed to raise the temperature of 1 kg of liquid water 1 C° than are required to raise the temperature of 1 kg of a metal 1 C°.

Temperature

Despite the fact that you are quite familiar with it, some more discussion of temperature is in order. Whenever you measure something, you are really just comparing that something with an arbitrarily-established standard. For instance, when you measure the length of a table with a meter stick, you are comparing the length of the table with the modern day equivalent of what was historically established as one ten-thousandth of the distance from the earth's north pole to the equator. In the case of temperature, a standard, now called the "degree Celsius" was established as follows: At 1 atmosphere of pressure, the temperature at which water freezes was defined to be 0 °C and the temperature at which water boils was defined to be 100 °C. Then a substance with a temperature-dependent measurable characteristic, such as the length of a column of liquid mercury, was used to interpolate and extrapolate the temperature range. (Mark the position of the end of the column of mercury on the tube containing that mercury when it is at the temperature of freezing water and again when it is at the temperature of boiling water. Divide the interval between the two marks into a hundred parts. Use the same length of each of those parts to extend the scale in both directions and call it a temperature scale.)

Note the arbitrary manner in which the zero of the Celsius scale has been established. The choice of zero is irrelevant for our purposes since equations $35A.1$ ($Q = C\Delta T$) and $35A.2$ ($Q = mc\Delta T$) relate temperature change, rather than temperature itself, to the amount of heat flow. An absolute temperature scale has been established for the SI system of units. The zero of temperature on this scale is set at the greatest possible temperature such that it is theoretically impossible for the temperature of any system in equilibrium to be as low as the zero of the Kelvin scale. The unit of temperature on the Kelvin scale is the Kelvin, abbreviated K. Note the absence of the degree symbol in the unit. The Kelvin scale is similar to the Celsius scale in that a change in temperature of, say, 1 K, is equivalent to a change in temperature of 1 C°. (Note regarding units notation: The units °C are used for a temperature on the Celsius scale, but the units C° are used for a temperature change on the Celsius scale.)

On the Kelvin scale, at a pressure of one atmosphere, water freezes at $273.15K$. So, a temperature in kelvin is related to a temperature in °C by

$$\text{Temperature in K} = (\text{Temperature in }^{\circ}\text{C}) \cdot \left(\frac{1K}{C}\right) + 273.15K$$

36A: Heat: Phase Changes

> *There is a tendency to believe that any time heat is flowing into ice, the ice is melting. NOT SO. When heat is flowing into ice, the ice will be melting only if the ice is already at the melting temperature. When heat is flowing into the ice that is below the melting temperature, the temperature of the ice is increasing.*

As mentioned in the preceding chapter, there are times when you bring a hot object into contact with a cooler sample, that heat flows from the hot object to the cooler sample, but the temperature of the cooler sample does not increase, even though no heat flows out of the cooler sample (e.g. into an even colder object). This occurs when the cooler sample undergoes a phase change. For instance, if the cooler sample happens to be H_2O ice or H_2O ice plus liquid water, at 0°C and atmospheric pressure, when heat is flowing into the sample, the ice is melting with no increase in temperature. This will continue until all the ice is melted (assuming enough heat flows into the sample to melt all the ice). Then, after the last bit of ice melts at 0°C, if heat continues to flow into the sample, the temperature of the sample will be increasing.

Lets review the question about how it can be that heat flows into the cooler sample without causing the cooler sample to warm up. Energy flows from the hotter object to the cooler sample, but the internal kinetic energy of the cooler sample does not increase. Again, how can that be? What happens is that the energy flow into the cooler sample is accompanied by an increase in the internal potential energy of the sample. It is associated with the breaking of electrostatic bonds between molecules where the negative part of one molecule is bonded to the positive part of another. The separating of the molecules corresponds to an increase in the potential energy of the system. This is similar to a book resting on a table. It is gravitationally bound to the earth. If you lift the book and put it on a shelf that is higher than the tabletop, you have added some energy to the earth/book system, but you have increased the potential energy with no net increase in the kinetic energy. In the case of melting ice, heat flow into the sample manifests itself as an increase in the potential energy of the molecules without an increase in the kinetic energy of the molecules (which would be accompanied by a temperature increase).

The amount of heat that must flow into a single-substance solid sample that is already at its melting temperature in order to melt the whole sample depends on a property of the substance of which the sample consists, and on the mass of the substance. The relevant substance property is called the latent heat of melting. The latent heat of melting is the heat-per-mass needed to melt the substance at the melting temperature. Note that, despite the name, the latent heat is not an amount of heat but rather a ratio of heat to mass. The symbol used to represent latent heat in general is L, and we use the subscript m for melting. In terms of the latent heat of melting, the amount of heat, Q, that must flow into a sample of a single-substance solid that is at the melting temperature, in order to melt the entire sample is given by:

$$Q = mL_m$$

Note the absence of a ΔT in the expression $Q = mL_m$. There is no ΔT in the expression because there is no temperature change in the process. The whole phase change takes place at one temperature.

So far, we have talked about the case of a solid sample, at the melting temperature, which is in contact with a hotter object. Heat flows into the sample, melting it. Now consider a sample of the same substance in liquid form at the same temperature but in contact with a colder object. In this case, heat will flow from the sample to the colder object. This heat loss from the sample does not result in a decrease in the temperature. Rather, it results in a phase change of the substance of which the sample consists, from liquid to solid. This phase change is called freezing. It also goes by the name of solidification. The temperature at which freezing takes place is called the freezing temperature, but it is important to remember that the freezing temperature has the same value as the melting temperature. The heat-per-mass that must flow out of the substance to freeze it (assuming the substance to be at the freezing temperature already) is called the latent heat of fusion, or L_f. The latent heat of fusion for a given substance has the same value as the latent heat of melting for that substance:

$$L_f = L_m$$

The amount of heat that must flow out of a sample of mass m in order to convert the entire sample from liquid to solid is given by:

$$Q = mL_f$$

Again, there is no temperature change.

The other two phase changes we need to consider are vaporization and condensation. Vaporization is also known as boiling. It is the phase change in which liquid turns into gas. It too (as in the case of freezing and melting), occurs at a single temperature, but for a given substance, the boiling temperature is higher than the freezing temperature. The heat-per-mass that must flow into a liquid to convert it to gas is called the latent heat of vaporization L_v. The heat that must flow into mass m of a liquid that is already at its boiling temperature (a.k.a. its vaporization temperature) to convert it entirely into gas is given by:

$$Q = mL_v$$

Condensation is the phase change in which gas turns into liquid. In order for condensation to occur, the gas must be at the condensation temperature, the same temperature as the boiling temperature (a.k.a. the vaporization temperature). Furthermore, heat must flow out of the gas, as it does when the gas is in contact with a colder object. Condensation takes place at a fixed temperature known as the condensation temperature. (The melting temperature, the freezing temperature, the boiling temperature, and the condensation temperature are also referred to as the melting point, the freezing point, the boiling point, and the condensation point, respectively.) The heat-per-mass that must be extracted from a particular kind of gas that is already at the condensation temperature, to convert that gas to liquid at the same temperature, is called the latent heat of condensation L_c. For a given substance, the latent heat of condensation has the same value as the latent heat of vaporization. For a sample of mass m of a gas at its condensation temperature, the amount of heat that must flow out of the sample to convert the entire sample to liquid is given by:

$$Q = mL_c$$

It is important to note that the actual values of the freezing temperature, the boiling temperature, the latent heat of melting, and the latent heat of vaporization are different for different substances. For water we have:

Phase Change	Temperature	Latent Heat
Melting / Freezing	0°C	$0.334 \frac{MJ}{kg}$
Boiling or Vaporization / Condensation	100°C	$2.26 \frac{MJ}{kg}$

How much heat does it take to convert 444 grams of H_2O ice at –9.0°C to steam (H_2O gas) at 128.0 °C ?

Discussion of Solution

Rather than solve this one for you, we simply explain how to solve it.

To convert the ice at –9.0°C to steam at 128.0°C, we first have to cause enough heat to flow into the ice to warm it up to the melting temperature, 0°C. This step is a specific heat capacity problem. We use

$$Q_1 = mc_{ice}\Delta T$$

where ΔT is [0°C – (–9.0°C)]=9.0C°.

Now that we have the ice at the melting temperature, we have to add enough heat to melt it. This step is a latent heat problem.

$$Q_2 = mL_m$$

After $Q_1 + Q_2$ flows into the H_2O, we have liquid water at 0°C. Next, we have to find how much heat must flow into the liquid water to warm it up to the boiling point, 100°C.

$$Q_3 = mc_{liquid\ water}\Delta T'$$

where $\Delta T' = (100^\circ C - 0^\circ C) = 100^\circ C$.

After $Q_1 + Q_2 + Q_3$ flows into the H_2O, we have liquid water at 100°C. Next, we have to find how much heat must flow into the liquid water at 100°C to convert it to steam at 100°C.

$$Q_4 = mL_v$$

After $Q_1 + Q_2 + Q_3 + Q_4$ flows into the H_20, we have water vapor (gas) at 100°C. Now, all we need to do is to find out how much heat must flow into the water gas at 100°C to warm it up to 128°C.

$$Q_5 = mc_{stream}\Delta T''$$

where $\Delta T'' = 128^\circ C - 100^\circ C = 28^\circ C$.

So the amount of heat that must flow into the sample of solid ice at –9.0°C in order for sample to become steam at 128°C (the answer to the question) is:

$$Q_{total} = Q_1 + Q_2 + Q_3 + Q_4 + Q_5$$

37A: The First Law of Thermodynamics

> *We use the symbol U to represent internal energy. That is the same symbol that we used to represent the mechanical potential energy of an object. Do not confuse the two different quantities with each other. In problems, questions, and discussion, the context will tell you whether the U represents internal energy or it represents mechanical potential energy.*

We end this physics textbook as we began the physics part of it (Chapter 1 was a mathematics review), with a discussion of conservation of energy. Back in Chapter 2, the focus was on the conservation of mechanical energy; here we focus our attention on thermal energy.

In the case of a deformable system, it is possible to do some net work on the system without causing its mechanical kinetic energy $\frac{1}{2}mv^2+\frac{1}{2}I\omega^2$ to change (where m is the mass of the system, v is the speed of the center of mass of the system, I is the moment of inertia of the system, and ω is the magnitude of the angular velocity of the system). Examples of such work would be: the bending of a coat hanger, the stretching of a rubber band, the squeezing of a lump of clay, the compression of a gas, and the stirring of a fluid.

When you do work on something, you transfer energy to that something. For instance, consider a case in which you push on a cart that is initially at rest. Within your body, you convert chemical potential energy into mechanical energy, which, by pushing the cart, you give to the cart. After you have been pushing on it for a while, the cart is moving, meaning that it has some kinetic energy. So, in the end, the cart has some kinetic energy that was originally chemical potential energy stored in you. Energy has been transferred from you to the cart.

In the case of the cart, what happens to the energy that you transfer to the cart is clear. But how about the case of a deformable system whose center of mass stays put? When you do work on such a system, you transfer energy to that system. So what happens to the energy?

Experimentally, we find that the energy becomes part of the internal energy of the system. The internal energy of the system increases by an amount that is equal to the work done on the system.

This increase in the internal energy can be an increase in the internal potential energy, an increase in the internal kinetic energy, or both. An increase in the internal kinetic energy would manifest itself as an increase in temperature.

Doing work on a system represents the second way, which we have considered, of causing an increase in the internal energy of the system. The other way was for heat to flow into the system. The fact that doing work on a system and/or having heat flow into that system will increase the internal energy of that system, is represented, in equation form, by:

$$\Delta U=Q+W_{IN}$$

which we copy here for your convenience:

$$\Delta U=Q+W_{IN} \tag{37A.1}$$

In this equation, ΔU is the change in the internal energy of the system, Q is the amount of heat that flows into the system, and W_{IN} is the amount of work that is done on the system. This equation is referred to as the First Law of Thermodynamics. Chemists typically write it without the subscript IN on the symbol W representing the work done on the system. (The subscript IN is there to remind us that the W_{IN} represents a transfer of energy into the system. In the chemistry convention, it is understood that W represents the work done on the system—no subscript is necessary.)

Historically, physicists and engineers have studied and developed thermodynamics with the goal of building a better heat engine, a device, such as a steam engine, designed to produce work from heat. That is, a device for which heat goes in and work comes out. It is probably for this reason that physicists and engineers almost always write the first law as:

$$\Delta U=Q-W \tag{37A.2}$$

where the symbol W represents the amount of work done by the system on the external world. (This is just the opposite of the chemistry convention.) Because this is a physics course, this $(\Delta U=Q-W)$ is the form in which the first law appears on your formula sheet. I suggest making the first law as explicit as possible by writing it as $\Delta U=Q_{IN}-W_{OUT}$ or, better yet:

$$\Delta U=Q_{IN}+W_{IN} \tag{37A.3}$$

In this form, the equation is saying that you can increase the internal energy of a system by causing heat to flow into that system and/or by doing work on that system. Note that any one of the quantities in the equation can be negative. A negative value of Q_{IN} means that heat actually flows out of the system. A negative value of W_{IN} means that work is actually done by the system on the surroundings. Finally, a negative value of ΔU means that the internal energy of the system decreases.

Again, the real tip here is to use subscripts and common sense. Write the First Law of Thermodynamics in a manner consistent with the facts that heat or work into a system will increase the internal energy of the system, and heat or work out of the system will decrease the internal energy of the system.

CHAPTER OVERVIEW

Volume B: Electricity, Magnetism, and Optics

B1: Charge & Coulomb's Law
B2: The Electric Field - Description and Effect
B3: The Electric Field Due to one or more Point Charges
B4: Conductors and the Electric Field
B5: Work Done by the Electric Field and the Electric Potential
B6: The Electric Potential Due to One or More Point Charges
B7: Equipotential Surfaces, Conductors, and Voltage
B8: Capacitors, Dielectrics, and Energy in Capacitors
9B: Electric Current, EMF, and Ohm's Law
B10: Resistors in Series and Parallel; Measuring I & V
B11: Resistivity and Power
B12: Kirchhoff’s Rules, Terminal Voltage
B13: RC Circuit
B14: Capacitors in Series & Parallel
B15: Magnetic Field Introduction - Effects
B16: Magnetic Field - More Effects
B17: Magnetic Field: Causes
B18: Faraday's Law and Lenz's Law
B19: Induction, Transformers, and Generators
B20: Faraday’s Law and Maxwell’s Extension to Ampere’s Law
B21: The Nature of Electromagnetic Waves
B22: Huygens’s Principle and 2-Slit Interference
B23: Single-Slit Diffraction
B24: Thin Film Interference
B25: Polarization
B26: Geometric Optics, Reflection
B27: Refraction, Dispersion, Internal Reflection
B28: Thin Lenses - Ray Tracing
B29: Thin Lenses - Lens Equation, Optical Power
B30: The Electric Field Due to a Continuous Distribution of Charge on a Line
B31: The Electric Potential due to a Continuous Charge Distribution
B32: Calculating the Electric Field from the Electric Potential
B33: Gauss’s Law
B34: Gauss’s Law Example
B35: Gauss’s Law for the Magnetic Field and Ampere’s Law Revisited
B36: The Biot-Savart Law
B37: Maxwell’s Equations

Thumbnail: Lightning over the outskirts of Oradea, Romania, during the August 17, 2005 thunderstorm which went on to cause major flash floods over southern Romania. (Public Domani; Nelumadau).

B1: Charge & Coulomb's Law

Charge is a property of matter. There are two kinds of charge, positive "+" and negative "-". An object can have positive charge, negative charge, or no charge at all. A particle which has charge causes a force-per-charge-of-would-be-victim vector to exist at each point in the region of space around itself. The infinite set of force-per-charge-of-would be-victim vectors is called a vector field. Any charged particle that finds itself in the region of space where the force-per-charge-of-would-be-victim vector field exists will have a force exerted upon it by the force-per-charge-of-would-be-victim field. The force-per-charge-of-would-be-victim field is called the electric field. The charged particle causing the electric field to exist is called the source charge. (Regarding jargon: A charged particle is a particle that has charge. A charged particle is often referred to simply as "a charge.")

The source charge causes an electric field which exerts a force on the victim charge. The net effect is that the source charge causes a force to be exerted on the victim. While we have much to discuss about the electric field, for now, we focus on the net effect, which we state simply (neglecting the "middle man", the electric field) as, "A charged particle exerts a force on another charged particle." This statement is *Coulomb's Law* in its conceptual form. The force is called the *Coulomb force*, a.k.a. the *electrostatic force*.

Note that either charge can be viewed as the source charge and either can be viewed as the victim charge. Identifying one charge as the victim charge is equivalent to establishing a point of view, similar to identifying an object whose motion or equilibrium is under study for purposes of applying Newton's 2nd Law of motion. $\vec{a} = \frac{\sum \vec{F}}{m}$. In Coulomb's Law, the force exerted on one charged particle by another is directed along the line connecting the two particles, and, away from the other particle if both particles have the same kind of charge (both positive, or, both negative) but, toward the other particle if the kind of charge differs (one positive and the other negative). This fact is probably familiar to you as, "like charges repel and unlike attract."

The SI unit of charge is the coulomb, abbreviated C. One coulomb of charge is a lot of charge, so much that, two particles, each having a charge of +1 C and separated by a distance of 1 meter exert a force of $9 \times 10^9 N$, that is 9 billion newtons on each other.

This brings us to the equation form of Coulomb's Law which can be written to give the magnitude of the force exerted by one charged particle on another as:

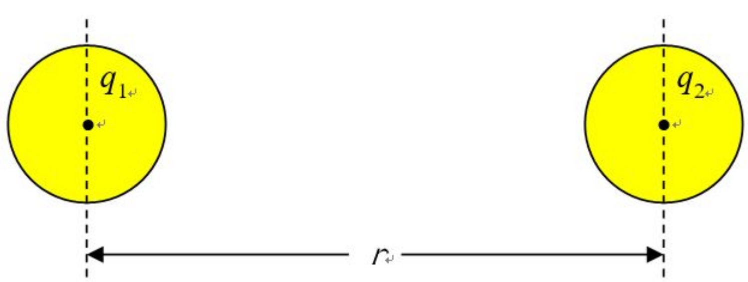
$$F = k\frac{|q_1||q_2|}{r^2} \tag{B1.1}$$

where:

- $k = 8.99 \times 10^9 \frac{N \cdot m^2}{C^2}$, a universal constant called the Coulomb constant,
- q_1 is the charge of particle 1,
- q_2 is the charge of particle 2, and
- r is the distance between the two particles.

The user of the equation (we are still talking about equation B1.1, $F = k\frac{|q_1||q_2|}{r^2}$) is expected to establish the direction of the force by means of "common sense" (the user's understanding of what it means for like charges to repel and unlike charges to attract each other).

While Coulomb's Law in equation form is designed to be exact for point particles, it is also exact for spherically symmetric charge distributions (such as uniform balls of charge) as long as one uses the center-to-center distance for \(r\).



Coulomb’s Law is also a good approximation in the case of objects on which the charge is not spherically symmetric as long as the objects’ dimensions are small compared to the separation of the objects (the truer this is, the better the approximation). Again, one uses the separation of the centers of the charge distributions in the Coulomb’s Law equation.

Coulomb’s Law can be written in vector form as:

$$\overrightarrow{F_{12}} = k\frac{q_1 q_2}{r^2}\hat{r}_{12} \tag{B1.2}$$

where:

- $\overrightarrow{F_{12}}$ is the force “of 1 on 2”, that is, the force exerted by particle 1 on particle 2,
- $\hat{r}_{12}$ is a unit vector in the direction “from 1 to 2”, and
- k, q_1 and q_2 are defined as before (the Coulomb constant, the charge on particle 1, and the charge on particle 2 respectively).

Note the absence of the absolute value signs around q_1 and q_2. A particle which has a certain amount, say, 5 coulombs of the negative kind of charge is said to have a charge of -5 coulombs and one with 5 coulombs of the positive kind of charge is said to have a charge of +5 coulombs) and indeed the plus and minus signs designating the kind of charge have the usual arithmetic meaning when the charges enter into equations. For instance, if you create a composite object by combining an object that has a charge of $q_1 = +3C$ with an object that has a charge of $q_2 = -5C$, then the composite object has a charge of

$$q = q_1 + q_2$$

$$q = +3C + (-5C)$$

$$q = -2C$$

Note that the arithmetic interpretation of the kind of charge in the vector form of Coulomb’s Law causes that equation to give the correct direction of the force for any combination of kinds of charge. For instance, if one of the particles has positive charge and the other negative, then the value of the product \(q_1q_2\) in equation B1.2

$$\overrightarrow{F_{12}} = k\frac{q_1 q_2}{r^2}\hat{r}_{12}$$

has a negative sign which we can associate with the unit vector. Now $-\hat{r}_{12}$ is in the direction opposite “from 1 to 2” meaning it is in the direction “from 2 to 1.” This means that $\overrightarrow{F_{12}}$, the force of 1 on 2, is directed toward particle 1. This is consistent with our understanding that opposites attract. Similarly, if q_1 and q_2 2 are both positive, or both negative in $\overrightarrow{F_{12}} = k\frac{q_1 q_2}{r^2}\hat{r}_{12}$ then the value of the product \(q_1q_2\) is positive meaning that the direction of the force of 1 on 2 is $\hat{r}_{12}$ (from 1 to 2), that is, away from 1, consistent with the fact that like charges repel. We’ve been talking about the force of 1 on 2. Particle 2 exerts a force on particle 1 as well. It is given by $\overrightarrow{F_{21}} = k\frac{q_1 q_2}{r^2}\hat{r}_{21}$. The unit vector $\hat{r}_{21}$, pointing from 2 to 1, is just the negative of the unit vector pointing from 1 to 2:

$$\hat{r_{21}} = -\hat{r}_{12}$$

If we make this substitution into our expression for the force exerted by particle 2 on particle 1, we obtain:

$$\overrightarrow{F_{21}} = k\frac{q_1 q_2}{r^2}(-\hat{r}_{12})$$

$$\overrightarrow{F_{21}} = -k\frac{q_1 q_2}{r^2}\hat{r}_{12}$$

Comparing the right side with our expression for the force of 1 on 2 (namely, $\overrightarrow{F_{12}} = k\frac{q_1 q_2}{r^2}\hat{r}_{12}$), we see that

$$\overrightarrow{F_{21}} = -\overrightarrow{F_{12}}$$

So, according to Coulomb’s Law, if particle 1 is exerting a force $\overrightarrow{F_{12}}$ on particle 2, then particle 2 is, at the same time, exerting an equal but opposite force $-\overrightarrow{F_{12}}$ back on particle 2, which, as we know, by Newton’s 3^{rd} Law, it must.

In our macroscopic world we find that charge is not an inherent fixed property of an object but, rather, something that we can change. Rub a neutral rubber rod with animal fur, for instance, and you'll find that afterwards, the rod has some charge and the fur has the opposite kind of charge. Ben Franklin defined the kind of charge that appears on the rubber rod to be negative charge and the other kind to be positive charge. To provide some understanding of how the rod comes to have negative charge, we delve briefly into the atomic world and even the subatomic world.

The stable matter with which we are familiar consists of protons, neutrons, and electrons. Neutrons are neutral, protons have a fixed amount of positive charge, and electrons have the same fixed amount of negative charge. Unlike the rubber rod of our macroscopic world, you cannot give charge to the neutron and you can neither add charge to, nor remove charge from, either the proton or the electron. Every proton has the same fixed amount of charge, namely $1.60 \times 10^{-19} C$. Scientists have never been able to isolate any smaller amount of charge. That amount of charge is given a name. It is called the e, abbreviated e and pronounced "ee". The e is a non-SI unit of charge. As stated $1e = 1.60 \times 10^{-19} C$. In units of e, the charge of a proton is $1e$ (exactly) and the charge of an electron is $-1e$. For some reason, there is a tendency among humans to interpret the fact that the unit the e is equivalent to $1.60 \times 10^{-19} C$ to mean that $1e$ equals $-1.60 \times 10^{-19} C$. This is wrong! Rather,

$$1e = 1.60 \times 10^{-19} C$$

A typical neutral atom consists of a nucleus made up of neutrons and protons surrounded by orbiting electrons such that the number of electrons in orbit about the nucleus is equal to the number of protons in the nucleus. Let's see what this means in terms of an everyday object such as a polystyrene cup. A typical polystyrene cup has a mass of about 2 grams. It consists of roughly: 6×10^{23} neutrons, 6×10^{23} protons, and, when neutral, 6×10^{23} electrons. Thus, when neutral it has about $1 \times 10^{5} C$ of positive charge and $1 \times 10^{5} C$ of negative charge, for a total of 0 charge. Now if you rub a polystyrene cup with animal fur you can give it a noticeable charge. If you rub it all over with the fur on a dry day and then experimentally determine the charge on the cup, you will find it to be about $-5 \times 10^{-8} C$. This represents an increase of about 0.00000000005%in the number of electrons on the cup. They were transferred from the fur to the cup. We are talking about 3×10^{10} electrons, which sure would be a lot of marbles but represents a minuscule fraction of the total number of electrons in the material of the cup.

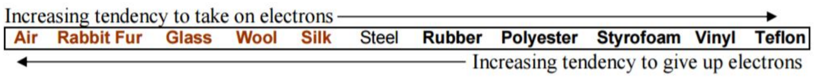
The main points of the preceding discussion are:

- A typical neutral macroscopic object consists of incredibly huge amounts of both kinds of charge (about 50 million coulombs of each for every kilogram of matter), the same amount of each kind.
- When we charge an object, we transfer a relatively minuscule amount of charge to or from that object.
- A typical everyday amount of charge (such as the amount of charge on a clingy sock just out of the dryer) is 10^{-7} coulombs.
- When we transfer charge from one object to another, we are actually moving charged particles, typically electrons, from one object to the other.

One point that we did not make in the discussion above is that charge is conserved. For instance, if, by rubbing a rubber rod with fur, we transfer a certain amount of negative charge to the rubber rod, then, the originally-neutral fur is left with the exact same amount of positive charge. Recalling the exact balance between the incredibly huge amount of negative charge and the incredibly huge amount of positive charge in any macroscopic object, we recognize that, in charging the rubber rod, the fur becomes positively charged not because it somehow gains positive charge, but, because it loses negative charge, meaning that the original incredibly huge amount of positive charge now (slightly) exceeds the (still incredibly huge) amount of negative charge remaining on and in the fur.

Charging by Rubbing

One might well wonder why rubbing a rubber rod with animal fur would cause electrons to be transferred from the fur to the rod. If one could imagine some way that even one electron might, by chance, find its way from the fur to the rod, it would seem that, then, the rod would be negatively charged and the fur positively charged so that any electron that got free from the fur would be attracted back to the fur by the positive charge on it and repelled by the negative charge on the rod. So why would any more charge ever be transferred from the fur to the rod? The answer comes under the heading of "distance matters." In rubbing the rod with the fur you bring lots of fur molecules very close to rubber molecules. In some cases, the outer electrons in the atoms of the fur come so close to nuclei of the atoms on the surface of the rubber that the force of attraction of these positive nuclei is greater than the force of attraction of the nucleus of the atom of which they are a part. The net force is then toward the rod, the electrons in question experience acceleration toward the rod that changes the velocity such that the electrons move to the rod. Charging by rubbing depends strongly on the molecular structure of the materials in question. One interesting aspect of the process is that the rubbing only causes lots of molecules in the fur to come very close to molecules in the rubber. It is not as if the energy associated with the rubbing motion is somehow given to the electrons causing them to jump from the fur to the rubber. It should be noted that fur is not the only material that has a tendency to give up electrons and rubber is not the only material with a tendency to acquire them. The phenomenon of charging by rubbing is called triboelectrification. The following ordered list of the tendency of (a limited number of) materials to give up or accept electrons is called a triboelectric sequence:



The presence and position of air on the list suggests that it is easier to maintain a negative charge on objects in air than it is to maintain a positive charge on them.

Conductors and Insulators

Suppose you charge a rubber rod and then touch it to a neutral object. Some charge, repelled by the negative charge on the rod, will be transferred to the originally-neutral object. What happens to that charge then depends on the material of which the originally-neutral object consists. In the case of some materials, the charge will stay on the spot where the originally neutral object is touched by the charged rod. Such materials are referred to as insulators, materials through which charge cannot move, or, through which the movement of charge is very limited. Examples of good insulators are quartz, glass, and air. In the case of other materials, the charge, almost instantly spreads out all over the material in question, in response to the force of repulsion (recalling that force causes acceleration which leads to the movement) that each elementary particle of the charge exerts on every other elementary particle of charge. Materials in which the charge is free to move about are referred to as conductors. Examples of good conductors are metals and saltwater.

When you put some charge on a conductor, it immediately spreads out all over the conductor. The larger the conductor, the more it spreads out. In the case of a very large object, the charge can spread out so much that any chunk of the object has a negligible amount of charge and hence, behaves as if were neutral. Near the surface of the earth, the earth itself is large enough to play such a role. If we bury a good conductor such as a long copper rod or pipe, in the earth, and connect to it another good conductor such as a copper wire, which we might connect to another metal object, such as a cover plate for an electrical socket, above but near the surface of the earth, we can take advantage of the earth's nature as a huge object made largely of conducting material. If we touch a charged rubber rod to the metal cover plate just mentioned, and then withdraw the rod, the charge that is transferred to the metal plate spreads out over the earth to the extent that the cover plate is neutral. We use the expression "the charge that was transferred to the cover plate has flowed into the earth." A conductor that is connected to the earth in the manner that the cover plate just discussed is connected is called "ground." The act of touching a charged object to ground is referred to as grounding the object. If the object itself is a conductor, grounding it (in the absence of other charged objects) causes it to become neutral.

Charging by Induction

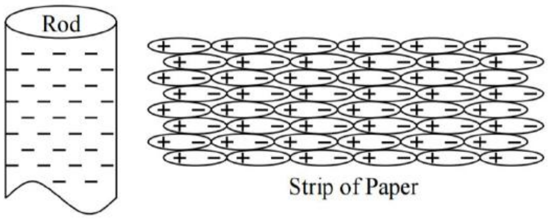
If you hold one side of a conductor in contact with ground and bring a charged object very near the other side of the conductor, and then, keeping the charged object close to the conductor without touching it, break the contact of the conductor with ground, you will find that the conductor is charged with the opposite kind of the charge that was originally on the charged object. Here's why. When you bring the charged object near the conductor, it repels charge in the conductor right out of the conductor and into the earth. Then, with those charges gone, if you break the path to ground, the conductor is stuck with the absence of those charged particles that were repelled into the ground. Since the original charged object repels the same kind of charge that it has, the conductor is left with the opposite kind of charge.

Polarization

Let's rub that rubber rod with fur again and bring the rubber rod near one end of a small strip of neutral aluminum foil. We find that the foil is attracted to the rubber rod, even though the foil remains neutral. Here's why:

The negatively charged rubber rod repels the free-to-move negative charge in the strip to the other end of the strip. As a result, the near end of the aluminum strip is positively charged and the far end is negatively charged. So, the rubber rod attracts the near end of the rod and repels the far end. But, because the near end is nearer, the force of attraction is greater than the force of repulsion and the net force is toward the rod. The separation of charge that occurs in the neutral strip of aluminum is called polarization, and, when the neutral aluminum strip is positive on one end and negative on the other, we say that it is polarized.

Polarization takes place in the case of insulators as well, despite the fact that charge is not free to move about within an insulator. Let's bring a negatively-charged rod near one end of a piece of paper. Every molecule in the paper has a positive part and a negative part. The positive part is attracted to the rod and the negative part is repelled. The effect is that each molecule in the paper is polarized and stretched. Now, if every bit of positive charge gets pulled just a little bit closer to the rod and every bit of negative charge gets pushed a little farther away, the net effect in the bulk of the paper is to leave it neutral, but, at the ends there is a net charge. On the near end, the repelled negative charge leaves the attracted positive charge all by itself, and, on the far end, the attracted positive charge leaves the repelled negative charge all by itself.



As in the case of the aluminum strip, the negative rubber rod attracts the near, positive, end and repels the far, negative, end, but, the near end is closer so the attractive force is greater, meaning that the net force on the strip of paper is attractive. Again, the separation of the charge in the paper is called polarization and the fact that one end of the neutral strip of paper is negative and the other is positive means that the strip of paper is polarized.

B2: The Electric Field - Description and Effect

An electric field is an invisible entity which exists in the region around a charged particle. It is caused to exist by the charged particle. The effect of an electric field is to exert a force on any charged particle (other than the charged particle causing the electric field to exist) that finds itself at a point in space at which the electric field exists. The electric field at an empty point in space is the force-per-charge-of-would-be-victim at that empty point in space. The charged particle that is causing the electric field to exist is called a source charge. The electric field exists in the region around the source charge whether or not there is a victim charged particle for the electric field to exert a force upon. At every point in space where the electric field exists, it has both magnitude and direction. Hence, the electric field is a vector at each point in space at which it exists. We call the force-per-charge-of-would-be-victim vector at a particular point in space the "electric field" at that point. We also call the infinite set of all such vectors, in the region around the source charge, the electric field of the source charge. We use the symbol $\vec{E}$ to represent the electric field. I am using the word "victim" for any particle upon which an electric field is exerting a force. The electric field will only exert a force on a particle if that particle has charge. So all "victims" of an electric field have charge. If there does happen to be a charged particle in an electric field, then that charged particle (the victim) will experience a force

$$\vec{F} = q\vec{E} \tag{B2.1}$$

where q is the charge of the victim and $\vec{E}$ is the electric field vector at the location of the victim. We can think of the electric field as a characteristic of space. The force experienced by the victim charged particle is the product of a characteristic of the victim (its charge) and a characteristic of the point in space (the electric field) at which the victim happens to be.

The electric field is not matter. It is not "stuff." It is not charge. It has no charge. It neither attracts nor repels charged particles. It cannot do that because its "victims", the charged particles upon which the electric field exerts force, are within it. To say that the electric field attracts or repels a charged particle would be analogous to saying that the water in the ocean attracts or repels a submarine that is submerged in the ocean. Yes, the ocean water exerts an upward buoyant force on the submarine. But, it neither attracts nor repels the submarine. In like manner, the electric field never attracts nor repels any charged particles. It is nonsense to say that it does.

If you have two source charge particles, e.g. one at point A and another at point B, each creating its own electric field vector at one and the same point P, the actual electric field vector at point P is the vector sum of the two electric field vectors. If you have a multitude of charged particles contributing to the electric field at point P, the electric field at point P is the vector sum of all the electric field vectors at P. Thus, by means of a variety of source charge distributions, one can create a wide variety of electric field vector sets in some chosen region of space. In the next chapter, we discuss the relation between the source charges that cause an electric field to exist, and the electric field itself. In this chapter, we focus our attention on the relation between an existing electric field (with no concern for how it came to exist) and the effect of that electric field on any charged particle in the electric field. To do so, it is important for you to be able to accept a given electric field as specified, without worrying about how the electric field is caused to exist in a region of space. (The latter is an important topic which we deal with at length in the next chapter.)

Suppose for instance that at a particular point in an empty region in space, let's call it point P, there is an eastward-directed electric field of magnitude $0.32N/C$. Remember, initially, we are talking about the electric field at an empty point in space. Now, let's imagine that we put a particle that has $+2.0$ coulombs of charge at point P. The electric field at point P will exert a force on our $2.0C$ victim:

$$\vec{F} = q\vec{E}$$

$$\vec{F} = 2.0C(0.32\frac{N}{C}\text{ eastward})$$

Note that we are dealing with vectors so we did include both magnitude and direction when we substituted for $\vec{E}$. Calculating the product on the right side of the equation, and including the direction in our final answer yields:

$$\vec{F} = 0.64N\text{ eastward}$$

We see that the force is in the same direction as the electric field. Indeed, the point I want to make here is about the direction of the electric field: The electric field at any location is defined to be in the direction of the force that the electric field would exert on a positively charged victim if there was a positively charged victim at that location.

Told that there is an electric field in a given empty region in space and asked to determine its direction at the various points in space at which the electric exists, what you should do is to put a single positively-charged particle at each of the various points in the region in turn, and find out which way the force that the particle experiences at each location is directed. Such a positivelycharged particle is called a positive test charge. At each location you place it, the direction of the force experienced by the positive test charge is the direction of the electric field at that location.

Having defined the electric field to be in the direction of the force that it would exert on a positive test charge, what does this mean for the case of a negative test charge? Suppose that, in the example of the empty point in space at which there was a $0.32 N/C$ eastward electric field, we place a particle with charge -2.0 coulombs (instead of $+2.0$ coulombs as we did before). This particle would experience a force:

$$\vec{F} = q\vec{E} \tag{B2.2}$$

$$= -2.0C\,(0.32\frac{N}{C}\text{ eastward}) \tag{B2.3}$$

$$= -0.64N\text{ eastward} \tag{B2.4}$$

A negative eastward force is a positive westward force of the same magnitude:

$$\vec{F} = 0.64N\text{ westward}$$

In fact, any time the victim particle has negative charge, the effect of the minus sign in the value of the charge q in Equation B2.1 is to make the force vector have the direction opposite that of the electric field vector. So the force exerted by an electric field on a negatively charged particle that is at any location in that field, is always in the exact opposite direction to the direction of the electric field itself at that location.

Let's investigate this direction business for cases in which the direction is specified in terms of unit vectors. Suppose that a Cartesian reference frame has been established in an empty region of space in which there is an electric field. Further assume that the electric field at a particular point, call it point P, is:

$$\vec{E} = 5.0\frac{kN}{C}\hat{k}$$

Now suppose that a proton ($q = 1.60 \times 10^{-19}C$) is placed at point P. What force would the electric field exert on the proton?

$$\vec{F} = q\vec{E} \tag{B2.5}$$

$$= (1.60 \times 10^{-19}C)5.0 \times 10^3\frac{N}{C}\hat{k} \tag{B2.6}$$

$$= 8.0 \times 10^{-16}N\hat{k} \tag{B2.7}$$

The force on the proton is in the same direction as that of the electric field at the location at which the proton was placed (the electric field is in the +z direction and so is the force on the proton), as it must be for the case of a positive victim.

If, in the preceding example, instead of a proton, an electron ($q = -1.60 \times 10^{-19}C$) is placed at point P, recalling that in the example $\vec{E} = 5.0\frac{kN}{C}\hat{k}$, we have

$$\vec{F} = q\vec{E} \tag{B2.8}$$

$$= (-1.60 \times 10^{-19}C)5.0 \times 10^3\frac{N}{C}\hat{k} \tag{B2.9}$$

$$= -8.0 \times 10^{-16}N\hat{k} \tag{B2.10}$$

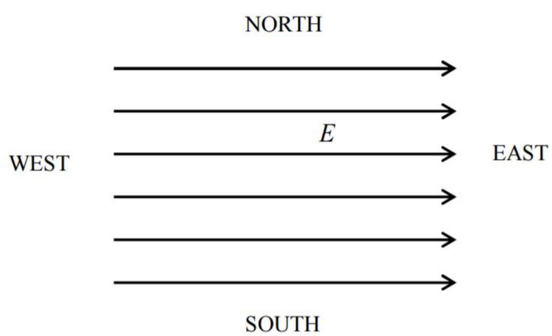
The negative sign is to be associated with the unit vector. This means that the force has a magnitude of $8.0 \times 10^{-16}N$ and a direction of $-\hat{k}$. The latter means that the force is in the –z direction which is the opposite direction to that of the electric field. Again, this is as expected.

The force exerted on a negatively charged particle by the electric field is always in the direction opposite that of the electric field itself.

In the context of the electric field as the set of all electric field vectors in a region of space, the simplest kind of an electric field is a uniform electric field. A uniform electric field is one in which every electric field vector has one and the same magnitude and one and the same direction. So, we have an infinite set of electric field vectors, one at every point in the region of space where the uniform electric field is said to exist, and every one of them has the same magnitude and direction as every other one. A charged particle victim that is either released from rest within such an electric field, or launched with some initial velocity within such a field, will have one and the same force exerted upon it, no matter where it is in the electric field. By Newton's 2^{nd} Law, this means that the particle will experience a constant acceleration. If the particle is released from rest, or, if the initial velocity of the particle is in the same direction as, or the exact opposite direction to, the electric field, the particle will experience constant acceleration motion in one dimension. If the initial velocity of the particle is in a direction that is not collinear with the electric field, then the particle will experience constant acceleration motion in two dimensions. The reader should review these topics from Calculus-Based Physics I.

Electric Field Diagrams

Consider a region in space in which there is a uniform, eastward-directed field. Suppose we want to depict this situation, as viewed from above, in a diagram. At every point in the region of space where the electric field exists, there is an electric field vector. Because the electric field is uniform, all the vectors are of the same magnitude and hence, we would draw all the arrows representing the electric field vectors, the same length. Since the field is uniform and eastward, we would draw all the arrows so that they would be pointing eastward. The problem is that it is not humanly possible to draw an arrow at every point on the region of a page used to depict a region of space in which there is an electric field. Another difficulty is that in using the convention that the length of a vector is representative of its magnitude, the arrows tend to run into each other and overlap.



Physicists have adopted a set of conventions for depicting electric fields. The result of the application of the conventions is known as an *electric field diagram*. According to the convention, the drawer creates a set of curves or lines, with arrowheads, such that, at every point on each curve, the electric field is, at every point on the curve, directed tangent to the curve, in the direction agreeing with that depicted by the arrowhead on that curve. Furthermore, the spacing of the lines in one region of the diagram as compared to other regions in the diagram is representative of the magnitude of the electric field relative to the magnitude at other locations in the same diagram. The closer the lines are, the stronger the electric field they represent. In the case of the uniform electric field in question, because the magnitude of the electric field is the same everywhere (which is what we mean by "uniform"), the line spacing must be the same everywhere. Furthermore, because the electric field in this example has a single direction, namely eastward, the electric field lines will be straight lines, with arrowheads.

B3: The Electric Field Due to one or more Point Charges

A charged particle (a.k.a. a point charge, a.k.a. a source charge) causes an electric field to exist in the region of space around itself. This is Coulomb's Law for the Electric Field in conceptual form. The region of space around a charged particle is actually the rest of the universe. In practice, the electric field at points in space that are far from the source charge is negligible because the electric field due to a point charge "dies off like one over r-squared." In other words, the electric field due to a point charge obeys an inverse square law, which means, that the electric field due to a point charge is proportional to the reciprocal of the square of the distance that the point in space, at which we wish to know the electric field, is from the point charge that is causing the electric field to exist. In equation form, Coulomb's Law for the magnitude of the electric field due to a point charge reads

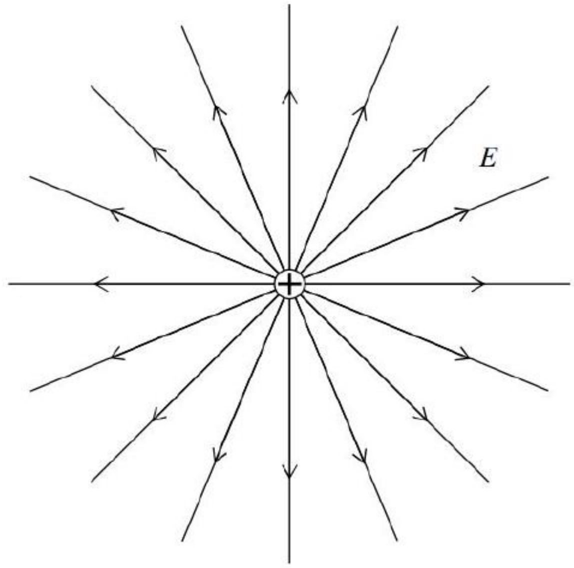
$$E=\frac{k|q|}{r^2} \tag{B3.1}$$

where

> E is the magnitude of the electric field at a point in space,
>
> k is the universal Coulomb constant $k=8.99\times 10^9\,\frac{N\cdot m^2}{C^2}$,
>
> q is the charge of the particle that we have been calling the point charge, and
>
> r is the distance that the point in space, at which we want to know E, is from the point charge that is causing E.

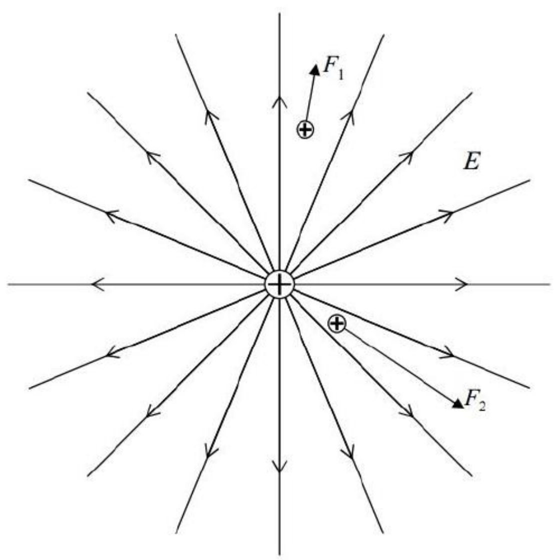
Again, Coulomb's Law is referred to as an inverse square law because of the way the magnitude of the electric field depends on the distance that the point of interest is from the source charge.

Now let's talk about direction. Remember, the electric field at any point in space is a force-per-charge-of-would-be-victim vector and as a vector, it always has direction. We have already discussed the defining statement for the direction of the electric field: The electric field at a point in space is in the direction of the force that the electric field would exert on a positive victim if there were a positive victim at that point in space. This defining statement for the direction of the electric field is about the effect of the electric field. We need to relate this to the cause of the electric field. Let's use some grade-school knowledge and common sense to find the direction of the electric field due to a positive source charge. First, we just have to obtain an imaginary positive test charge. I recommend that you keep one in your pocket at all times (when not in use) for just this kind of situation. Place your positive test charge in the vicinity of the source charge, at the location at which you wish to know the direction of the electric field. We know that like charges repel, so, the positive source charge repels our test charge. This means that the source charge, the point charge that is causing the electric field under investigation to exist, exerts a force on the test charge that is directly away from the source charge. Again, the electric field at any point is in the direction of the force that would be exerted on a positive test charge if that charge was at that point, so, the direction of the electric field is "directly away from the positive source charge." You get the same result no matter where, in the region of space around the source charge, you put the positive test charge. So, put your imaginary positive test charge back in your pocket. It has done its job. We know what we needed to know. The electric field due to a positive source charge, at any point in the region of space around that positive source charge, is directed directly away from the positive source charge. At every point in space, around the positive source charge, we have an electric field vector (a force-per-charge-of-would be-victim vector) pointing directly away from the positive source charge. So, how do we draw the electric field diagram for that? We are supposed to draw a set of lines or curves with arrowheads (NEVER OMIT THE ARROWHEADS!), such that, at every point on each line or curve, the electric field vector at that point is directed along the line or curve in the direction specified by the arrowhead or arrowheads on that line or curve. Let's give it a try.

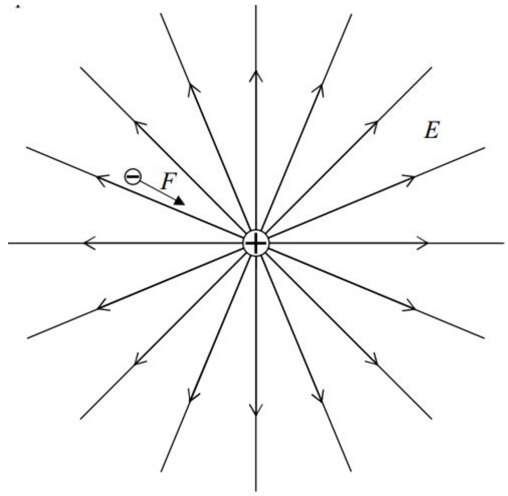


The number of lines drawn extending out of the positive source charge is chosen arbitrarily, but, if there was another positively charged particle, with twice the charge of the first one, in the same diagram, I would need to have twice as many lines extending out of it. That is to say that the line spacing has no absolute meaning overall, but it does have some relative meaning within a single electric field diagram. Recall the convention that the closer together the electric field lines are, the stronger the electric field. Note that in the case of a field diagram for a single source charge, the lines turn out to be closer together near the charged particle than they are farther away. It turned out this way when we created the diagram to be consistent with the fact that the electric field is always directed directly away from the source charge. The bunching of the lines close to the source charge (signifying that the electric field is strong there) is consistent with the inverse square dependence of the electric field magnitude on the distance of the point of interest from the source charge.

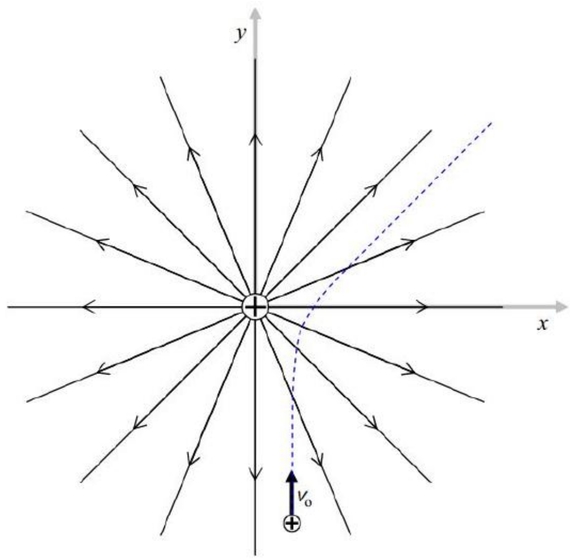
There are a few of important points to be made here. The first one is probably pretty obvious to you, but, just to make sure: The electric field exists between the electric field lines—its existence there is implied by the lines that are drawn—we simply can't draw lines everywhere that the electric field does exist without completely blackening every square inch of the diagram. Thus, a charged victim that finds itself at a position in between the lines will experience a force as depicted below for each of two different positively-charged victims.



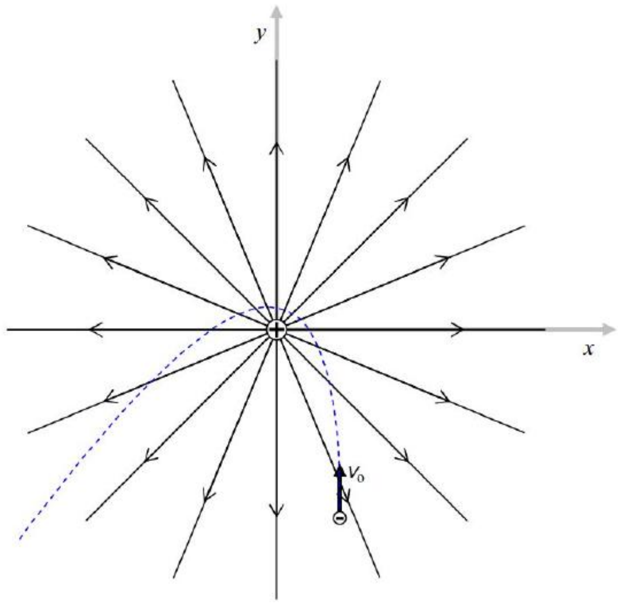
The next point is a reminder that a negatively-charged particle that finds itself at a position at which an electric field exists, experiences a force in the direction exactly opposite that of the electric field at that position.



The third and final point that should be made here is a reminder that the direction of the force experienced by a particle, is not, in general, the direction in which the particle moves. To be sure, the expression “in general” implies that there are special circumstances in which the particle would move in the same direction as that of the electric field but these are indeed special. For a particle on which the force of the electric field is the only force acting, there is no way it will stay on one and the same electric field line (drawn or implied) unless that electric field line is straight (as in the case of the electric field due to a single particle). Even in the case of straight field lines, the only way a particle will stay on one and the same electric field line is if the particle’s initial velocity is zero, or if the particle’s initial velocity is in the exact same direction as that of the straight electric field line. The following diagram depicts a positively-charged particle, with an initial velocity directed in the $+y$ direction. The dashed line depicts the trajectory for the particle (for one set of initial velocity, charge, and mass values). The source charge at the origin is fixed in position by forces not specified.

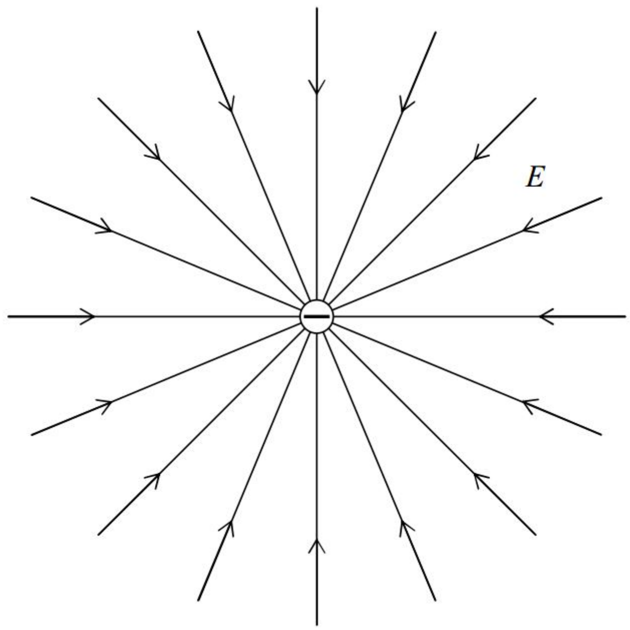


Here is an example of a trajectory of a negatively-charged particle, again for one set of values of source charge, victim charge, victim mass, and victim initial velocity:



Again, the point here is that, in general, charged particles do not move along the electric field lines, rather, they experience a force along (or, in the case of negative particles, in the exact opposite direction to) the electric field lines.

At this point, you should know enough about electric field diagrams to construct the electric field diagram due to a single negatively-charged particle. Please do so and then compare your work with the following diagram:



Some General Statements that can be made about Electric Field Lines

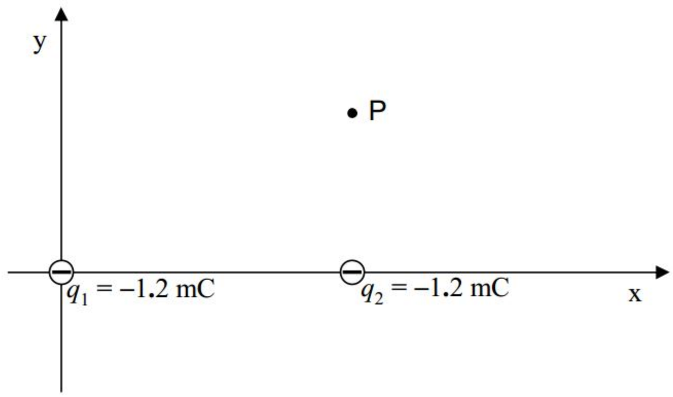
The following useful facts about electric field lines can be deduced from the definitions you have already been provided:

1. Every electric field line begins either at infinity or at a positive source charge.
2. Every electric field line ends either at infinity or at a negative source charge.
3. Electric field lines never cross each other or themselves.

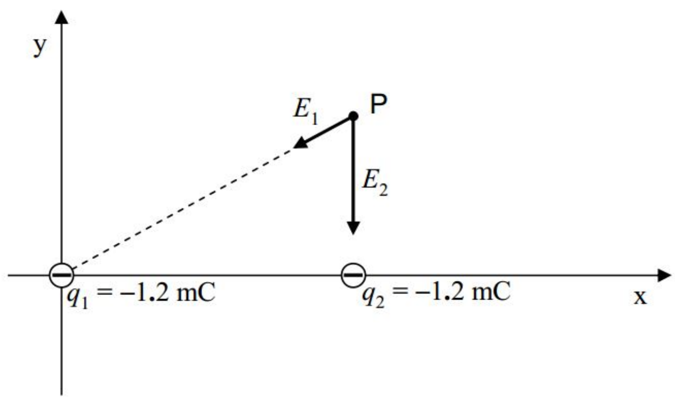
Superposition

If there is more than one source charge, each source charge contributes to the electric field at every point in the vicinity of the source charges. The electric field at a point in space in the vicinity of the source charges is the vector sum of the electric field at that point due to each source charge. For instance, suppose the set of source charges consists of two charged particles. The electric field at some point P will be the electric field vector at point P due to the first charged particle plus the electric field vector at point P due to the second particle. The determination of the total electric field at point P is a vector addition problem because the two electric field vectors contributing to it are, as the name implies, vectors.

Suppose, for instance, that you were asked to find the magnitude and direction of the electric field vector at point P due to the two charges depicted in the diagram below:

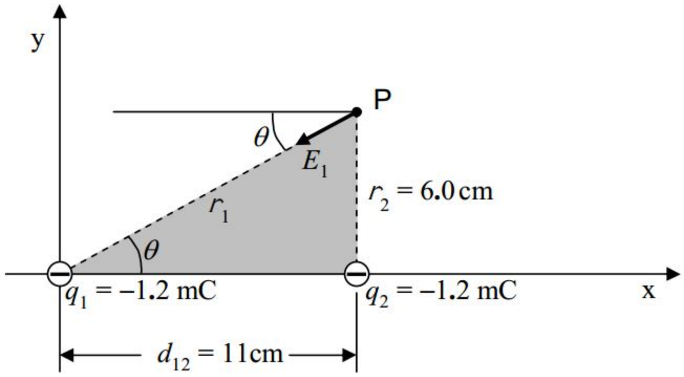


given that charge q_1 is at $(0, 0)$, q_2 is at $(11\text{cm}, 0)$ and point P is at $(11\text{cm}, 6.0\text{cm})$. The first thing that you would have to do is to find the direction and magnitude of $\vec{E}_1$ (the electric field vector due to q_1) and the direction and magnitude of $\vec{E}_2$ (the electric field vector due to q_2).



Referring to the diagram above, the direction of $\vec{E}_2$ is "the $-y$ direction" by inspection.

The angle θ specifying the direction of $\vec{E}_1$ can be determined by analyzing the shaded triangle in the following diagram.



Analysis of the shaded triangle will also give the distance r_1 that point P is from charge q_1. The value of r_1 can then be substituted into

$$E_1 = \frac{k|q_1|}{r_2^1}$$

to get the magnitude of $\vec{E}_1$. Based on the given coordinates, the value of r_2 is apparent by inspection and we can use it in

$$E_2 = \frac{k|q_2|}{r_2^2}$$

to get the magnitude of $\vec{E}_2$. With the magnitude and direction for both $\vec{E}_1$ and $\vec{E}_2$, you follow the vector addition recipe to arrive at your answer:

For each vector:
a. Draw a vector component diagram.
b. Analyze the vector component diagram to get the components of the vector.

- Add the x components to get the x component of the resultant.
- Add the y components to get the y component of the resultant.
- For the resultant:

a. Draw a vector component diagram.
b. Analyze the vector component diagram to get the magnitude and direction of the resultant.

B4: Conductors and the Electric Field

An ideal conductor is chock full of charged particles that are perfectly free to move around within the conductor. Like all macroscopic samples of material, an ideal conductor consists of a huge amount of positive charge, and, when neutral, the same amount of negative charge. When not neutral, there is a tiny fractional imbalance one way or the other. In an ideal conductor, some appreciable fraction of the charge is completely free to move around within the conducting material. The ideal (perfect) conductor is well approximated by some materials familiar to you, in particular, metals. In some materials, it is positive charge that is free to move about, in some, it is negative, and in others, it is both. For our purposes, the observable effects of positive charge moving in one direction are so close to being indistinguishable from negative charge moving in the opposite direction that, we will typically treat the charge carriers as being positive without concern for what the actual charge carriers are.

Here, we make one point about conductors by means of an analogy. The analogy involves a lake full of fish. Let the lake represent the conductor and the fish the charge carriers. The fish are free to move around anywhere within the lake, but, and this is the point, they can't, under ordinary circumstances, escape the lake. They can go to every boundary of the body of water, you might even see some on the surface, but, they cannot leave the water. This is similar to the charge carriers in a conductor surrounded by vacuum or an insulating medium such as air. The charges can go everywhere in and on the conductor, but, they cannot leave the conductor.

The facts we have presented on the nature of charge, electric fields, and conductors allow one to draw some definite conclusions about the electric field and unbalanced charge within the material of, and at or on the surface of, an ideal conductor. Please try to reason out the answers to the following questions:

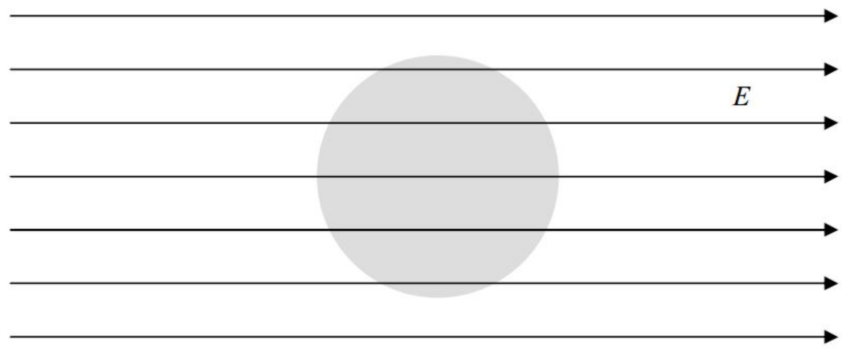
1. Suppose you put a neutral ideal conducting solid sphere in a region of space in which there is, initially, a uniform electric field. Describe (as specifically as possible) the electric field inside the conductor and the electric field at the surface of the conductor. Describe the distribution of charge in and on the conductor.
2. Repeat question 1 for the case of a non-uniform field.
3. Suppose you put some charge on an initially-neutral, solid, perfectly-conducting sphere (where the sphere is not in a pre-existing electric field). Describe the electric field inside the conductor, at the surface of the conductor, and outside the conductor as a result of the unbalanced charge. Describe the distribution of the charge in and on the conductor.
4. Repeat questions 1-3 for the case of a hollow perfectly-conducting spherical shell (with the interior being vacuum).
5. How would your answers to questions 1-4 change if the conductor had some shape other than spherical?

Here we provide the answers (preceded in each case, with the corresponding question).

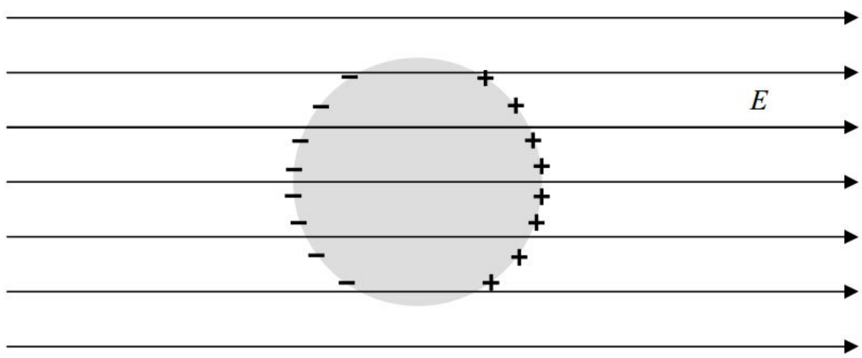
> 1) Suppose you put a neutral ideal conducting solid sphere in a region of space in which there is, initially, a uniform electric field. Describe (as specifically as possible) the electric field inside the conductor and the electric field at the surface of the conductor. Describe the distribution of charge in and on the conductor.

Answer: We start with a uniform electric field.

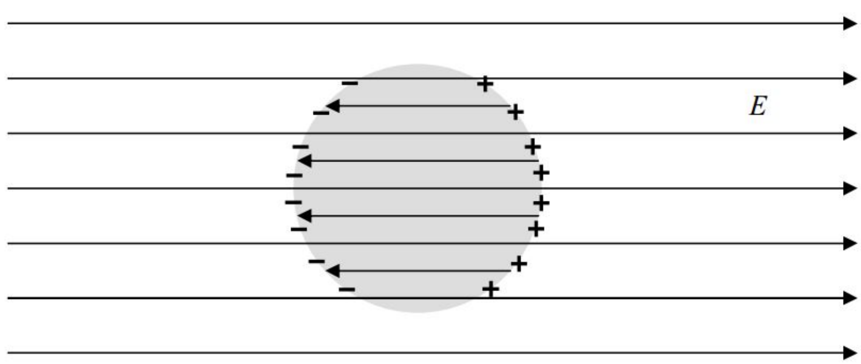
We put a solid, ideal conductor in it. The electric field permeates everything, including the conductor.



The charged particles in the conductor respond to the force exerted on them by the electric field. (The force causes acceleration, the acceleration of particles that are initially at rest causes them to acquire some velocity. In short, they move.) All this occurs in less than a microsecond. The net effect is a redistribution of the charged particles.

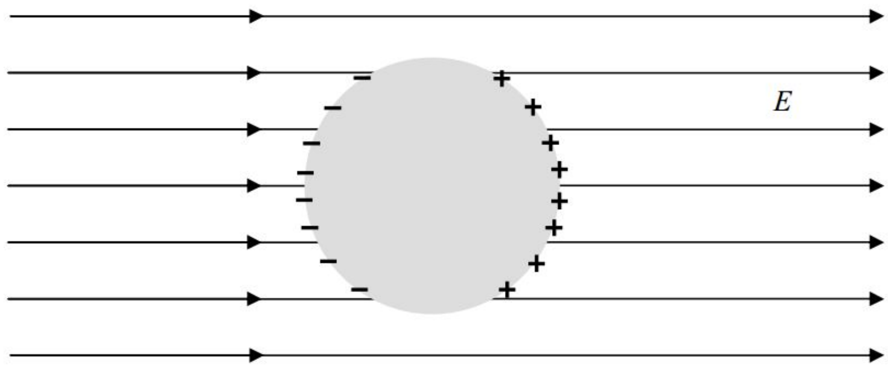


Now, get this! The charged particles create their own electric field.

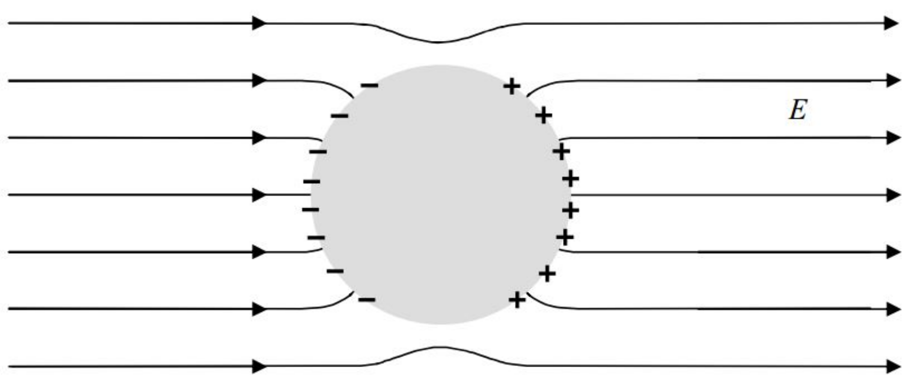


The total electric field at any point in the conductor is the vector sum of the original electric field and the electric field due to the redistributed charged particles. Since they are oppositely directed, the two contributions to the electric field inside the conductor tend to cancel each other. Now comes the profound part of the argument: the two contributions to the electric field at any point in the conductor exactly cancel. We know they have to completely cancel because, if they didn't, the free-to-move-charge in the conductor would move as a result of the force exerted on it by the electric field. And the force on the charge is always in a direction that causes the charge to be redistributed to positions in which it will create its own electric field that tends to cancel the electric field that caused the charge to move. The point is that the charge will not stop responding to the electric field until the net electric field at every point in the conductor is zero.

So far, in answer to the question, we have: The electric field is zero at all points inside the conductor, and, while the total charge is still zero, the charge has been redistributed as in the following diagram:

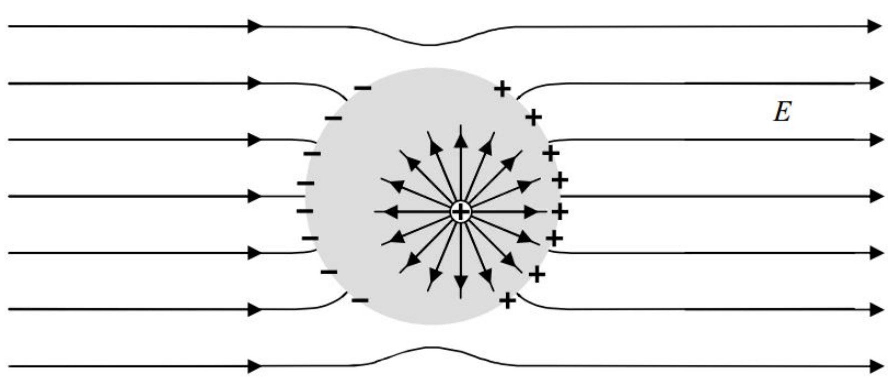


Recall that we were also called upon to describe the electric field at the surface of the conductor. Note that the charge on the surface of the sphere will not only contribute to the electric field inside the conductor, it will also contribute to the electric field outside. The net effect of all the contributions to the electric field in the near vicinity of the sphere is to cause the electric field to be normal to (perpendicular to) the surface of the sphere at all points where it meets the sphere.



How is it that we are able to assert this without doing any calculations? Here's the argument: If the electric field at the surface had a component parallel to the surface, then the charged particles on the surface of the conductor would experience a force directed along the surface. Since those particles are free to move anywhere in the conductor, they would be redistributed. In their new positions, they would make their own contribution to the electric field in the surface and their contribution would cancel the electric field that caused the charge redistribution.

About the charge distribution: The object started out neutral and no charge has left or entered the conductor from the outside world so it is still neutral. But we do see a separation of the two different kinds of charge. Something that we have depicted but not discussed is the assertion that all the charge resides on the surface. (In the picture above, there is positive charge on the right surface of the sphere and an equal amount of negative charge on the left side.) How do we know that all charge must be on the surface? Assume that there was a positive point charge at some location within the conductor:

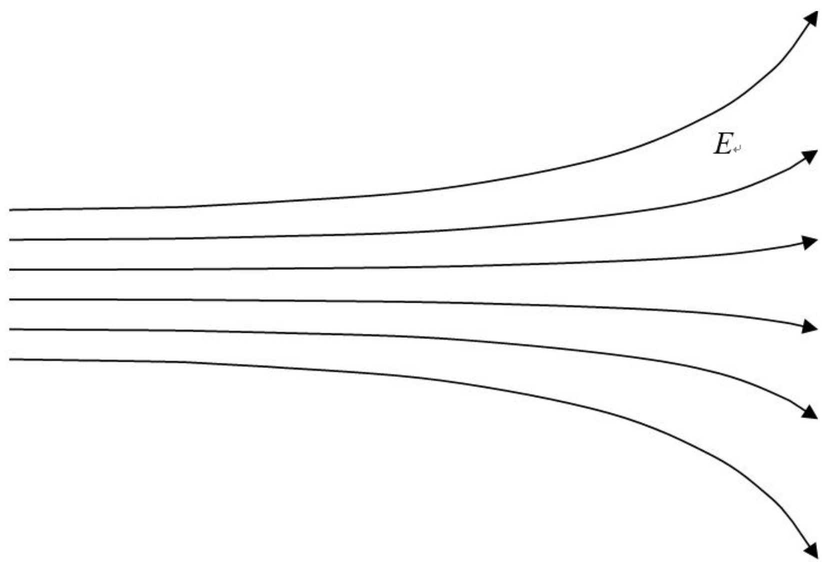


The electric field of that point charge would cause the free-to-move charge in the conductor to move, and it would keep moving as long as there was an electric field. So where would the charge move in order to cancel out the electric field of the positive point charge. You can try any arrangement of charge that you want to, around that positive point charge, but, if it is stipulated that there be a net positive charge at that location, there is no way to cancel out the electric field of that positive charge. So the situation doesn't even occur. If it did happen, the particle would repel the conductor's free-to-move-positive charge away from the stipulated positive charge, so that (excluding the stipulated positive charge under consideration) the conductor would have a net negative charge at that location, an amount of negative charge exactly equal to the originally stipulated positive charge. Taking the positive charge into account as well, the point, after the redistribution of charge, would be neutral. The point of our argument is that, under static conditions, there can be no net charge inside the material of a perfect conductor. Even if you assume there to be some, it would soon be neutralized by the nearly instantaneous charge redistribution that it would cause.

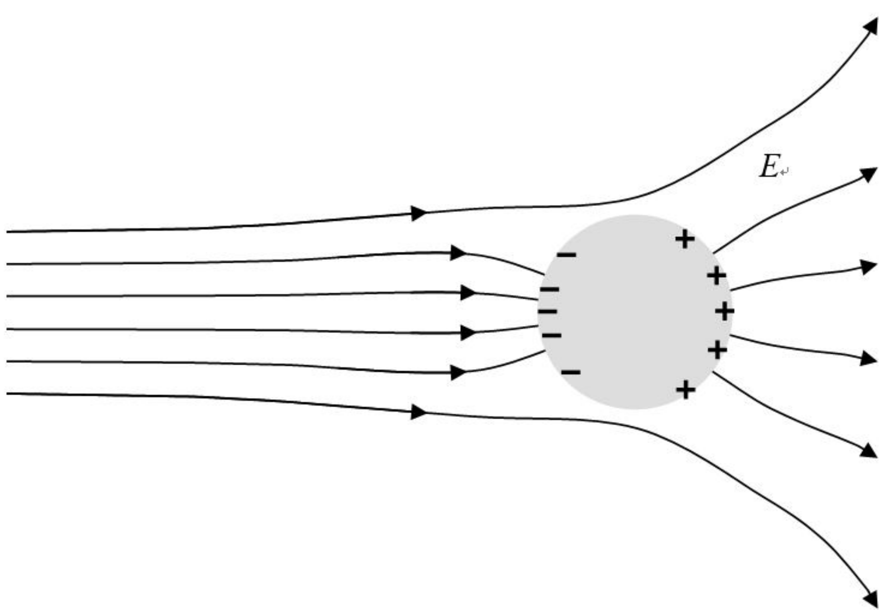
Next question:

> 2) Repeat question 1 for the case of a non-uniform field. (Question 1 asked for a description of the charge distribution that develops on a solid neutral conducting sphere when you place it in a uniform electric field.)

Answer: Here is a depiction of an example of a non-uniform field:



If we put a solid, perfectly-conducting sphere in it we get:

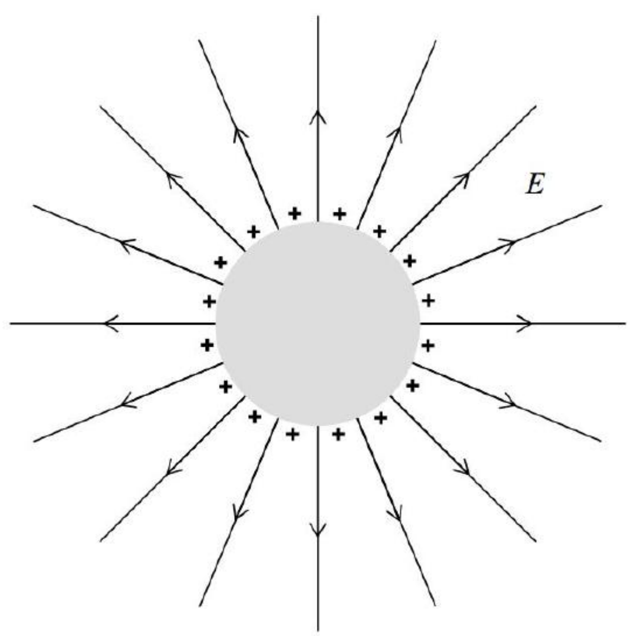


The same arguments lead to the same conclusions. When, after less than a microsecond, the new static conditions are achieved: There can be no electric field inside the conductor or else the free-to-move-charges in it would still be moving around within the volume of the conductor. There can be no unbalanced charge within the volume of the conductor or else there would be an electric field inside the conductor. Hence, any locally unbalanced charge (overall, the initially neutral sphere remains neutral) must be on the surface. The electric field has to be normal to the surface of the sphere or else the free-to-move-charge at the surface would still be moving around on the surface. The only thing that is different in this case, as compared to the initially-uniform electric field case, is the way the charge is distributed on the surface. We see that the negative charge is more bunched up than the positive charge in the case at hand. In the initially-uniform electric field case, the positive charge distribution was the mirror image of the negative charge distribution.

Next Question:

> 3) Suppose you put some charge on an initially-neutral, solid, perfectly-conducting sphere (where the sphere is not in a pre-existing electric field). Describe the electric field inside the conductor, at the surface of the conductor, and outside the conductor as a result of the unbalanced charge. Describe the distribution of the charge in and on the conductor.

Again, we assume that we have waited long enough (less than a microsecond) for static conditions to have been achieved. There can be no charge within the bulk of the conductor or else there would be an electric field in the conductor and there can't be an electric field in the conductor or else the conductor's free-to-move charge would move and static conditions would not be prevailing. So, all the unbalanced charge must be on the surface. It can't be bunched up more at any location on the surface than it is at any other location on the surface or else the charge on the edge of the bunch would be repelled by the bunch and it would move, again in violation of our stipulation that we have waited until charge stopped moving. So, the charge must be distributed uniformly over the surface of the sphere. Inside the sphere there is no electric field. Where the outside electric field meets the surface of the sphere, the electric field must be normal to the surface of the sphere. Otherwise, the electric field at the surface would have a vector component parallel to the surface which would cause charge to move along the surface, again in violation of our static conditions stipulations. Now, electric field lines that are perpendicular to the surface of a sphere lie on lines that pass through the center of the sphere. Hence, outside the sphere, the electric field lines form the same pattern as the pattern that would be formed by a point charge at the location of the center of the sphere (with the sphere gone). Furthermore, if you go so far away from the sphere that the sphere "looks like" a point, the electric field will be the same as that due to a point charge at the location of the center of the sphere. Given that outside the sphere, it has the same pattern as the field due to a point charge at the center of the sphere, the only way it can match up with the point charge field at a great distance from the sphere, is if it is identical to the point charge field everywhere that it exists. So, outside the sphere, the electric field is indistinguishable from the electric field due to the same amount of charge that you put on the sphere, all concentrated at the location of the center of the sphere (with the sphere gone).



Next question:

> 4) Repeat questions 1-3 for the case of a hollow perfectly-conducting spherical shell (with the interior being vacuum).

In all three cases we have considered so far, the interior of the sphere has played no role. It is initially neutral and it is neutral after the sphere is placed in a pre-existing electric field or some charge is placed on it. Nothing would change if we removed all that neutral material making up the bulk of the conductor, leaving nothing but a hollow shell of a sphere. Hence all the results that we found for the solid sphere apply to the hollow sphere. In particular, the electric field at all points inside an empty hollow perfectly-conducting spherical shell is, under all conditions, zero.

Last question:

> 5) How would your answers to questions 1-4 change if the conductor had some shape other than spherical?

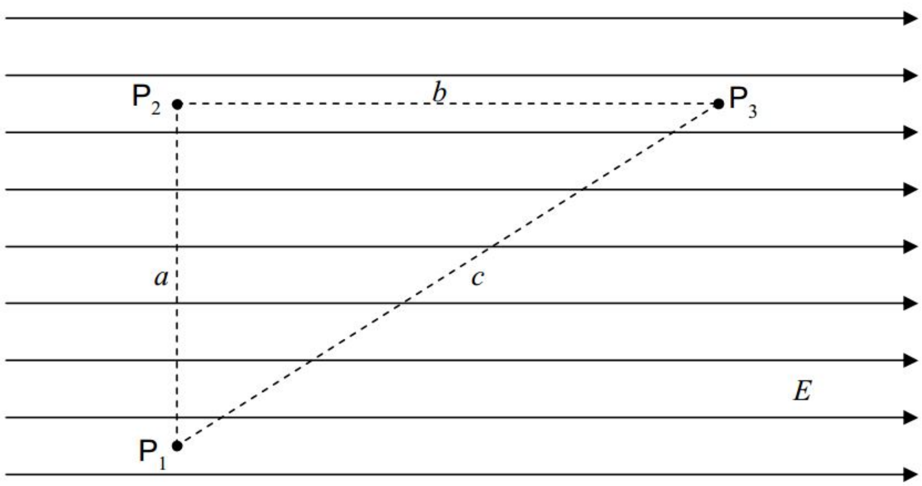
For a solid perfect conductor, the electric field and the charge everywhere inside would have to be zero for the same reasons discussed above. Furthermore, the electric field would have to be normal to the surface for the same reasons as before. Again, it would not make any difference if we hollow out the conductor by removing a bunch of neutral material. The only things that would be different for a non-spherical conductor are the way the charge would be distributed on the surface, and, the outside electric field. In particular, if you put some charge on a perfectlyconducting object that is not a sphere, the electric field in the vicinity of the object will not be the same as the electric field due to a point charge at the center of the object (although the difference would be negligible at great enough distances from the object).

B5: Work Done by the Electric Field and the Electric Potential

When a charged particle moves from one position in an electric field to another position in that same electric field, the electric field does work on the particle. The work done is conservative; hence, we can define a potential energy for the case of the force exerted by an electric field. This allows us to use the concepts of work, energy, and the conservation of energy, in the analysis of physical processes involving charged particles and electric fields.

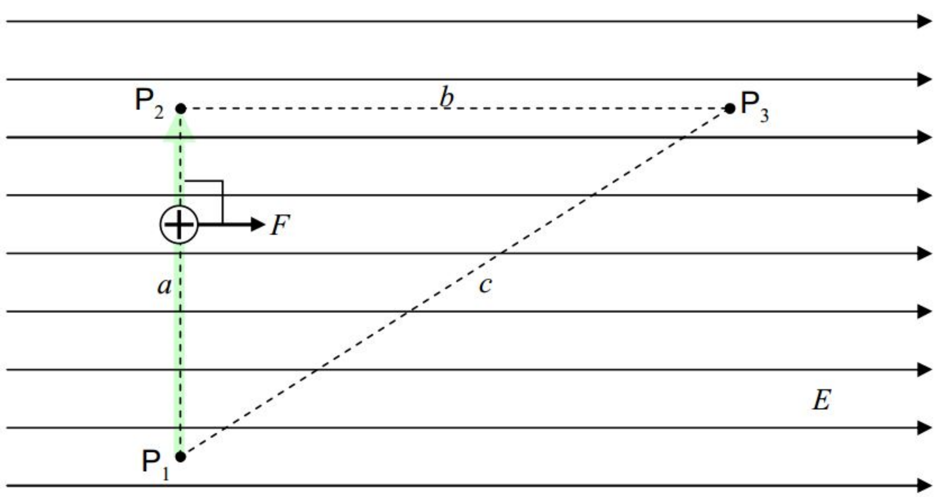
We have defined the work done on a particle by a force, to be the force-along-the-path times the length of the path, with the stipulation that when the component of the force along the path is different on different segments of the path, one has to divide up the path into segments on each of which the force-along-the-path has one value for the whole segment, calculate the work done on each segment, and add up the results.

Let's investigate the work done by the electric field on a charged particle as it moves in the electric field in the rather simple case of a uniform electric field. For instance, let's calculate the work done on a positively-charged particle of charge q as it moves from point P_1 to point P_3



along the path: "From P_1 straight to point P_2 and from there, straight to P_3." Note that we are not told what it is that makes the particle move. We don't care about that in this problem. Perhaps the charged particle is on the end of a quartz rod (quartz is a good insulator) and a person who is holding the rod by the other end moves the rod so the charged particle moves as specified.

Along the first part of the path, from P_1 to P_2, the force on the charged particle is perpendicular to the path.

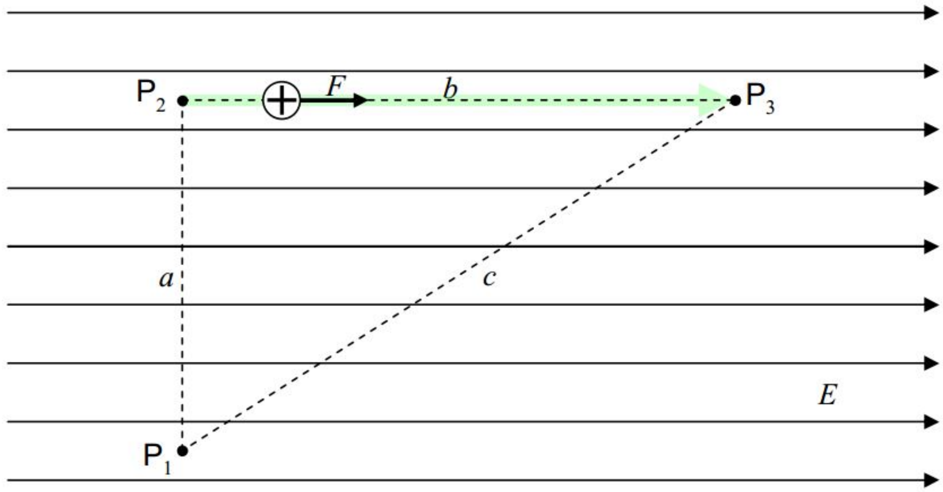


The force has no component along the path so it does no work on the charged particle at all as the charged particle moves from point P_1 to point P_2.

$$W_{12} = 0$$

From P_2, the particle goes straight to P_3.

On that segment of the path (from P_2 to P_3) the force is in exactly the same direction as the direction in which the particle is going.



As such, the work is just the magnitude of the force times the length of the path segment:

$$W_{23} = Fb$$

The magnitude of the force is the charge of the particle times the magnitude of the electric field $F = qE$, so,

$$W_{23} = qEb$$

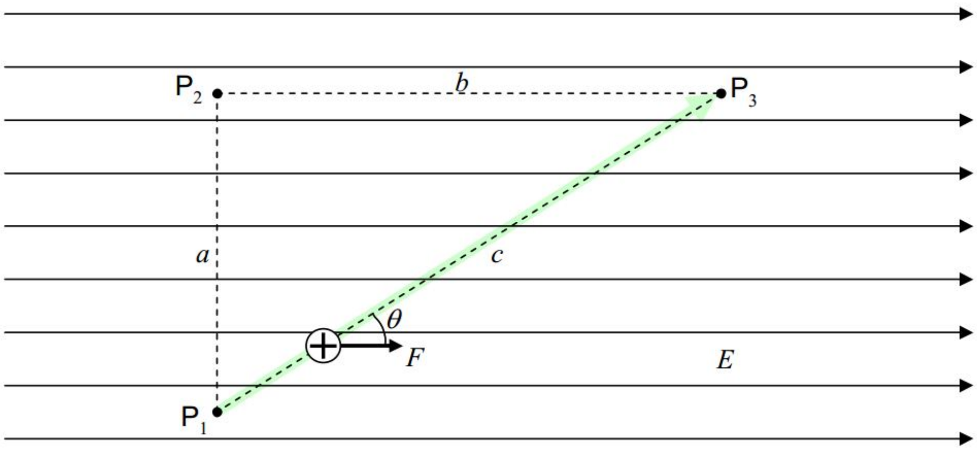
Thus, the work done on the charged particle by the electric field, as the particle moves from point P_1 to P_3 along the specified path is

$$W_{123} = W_{12} + W_{23}$$

$$W_{123} = 0 + qEb$$

$$W_{123} = qEb$$

Now let's calculate the work done on the charged particle if it undergoes the same displacement (from P_1 to P_3) but does so by moving along the direct path, straight from P_1 to P_3.



The force on a positively-charged particle being in the same direction as the electric field, the force vector makes an angle θ with the path direction and the expression

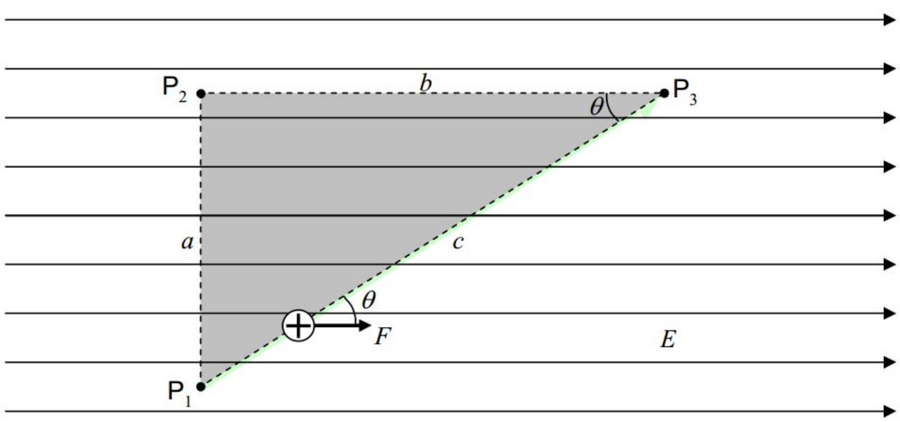
$$W=\vec{F}\cdot\overrightarrow{\Delta r}$$

for the work becomes

$$W_{13}=Fc\,cos\theta$$

$$W_{13}=qEc\,cos\theta$$

Analyzing the shaded triangle in the following diagram:

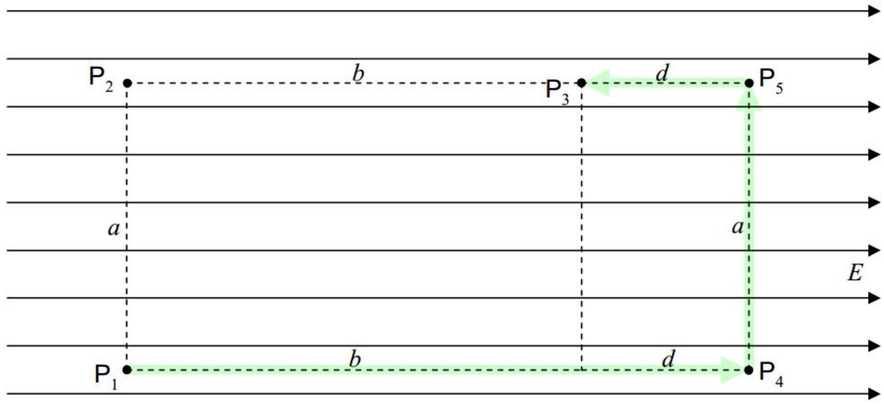


we find that $cos\theta=\frac{b}{c}$. Substituting this into our expression for the work ($W_{13}=qEc\,cos\theta$) yields

$$W_{13}=qEc\frac{b}{c}$$

$$W_{13}=qEb$$

This is the same result we got for the work done on the charged particle by the electric field as the particle moved between the same two points (from P_1 to P_3) along the other path (P_1 to P_2 to P_3). As it turns out, the work done is the same no matter what path the particle takes on its way from P_1 to P_3. I don't want to take the time to prove that here but I would like to investigate one more path (not so much to get the result, but rather, to review an important point about how to calculate work). Referring to the diagram:



Let's calculate the work done on a particle with charge q, by the electric field, as the particle moves from P_1 to P_3 along the path "from P_1 straight to P_4, from P_4 straight to P_5, and from P_5 straight to P_3." On P_1 to P_4, the force is in the exact same direction as the direction in which the particle moves along the path, so,

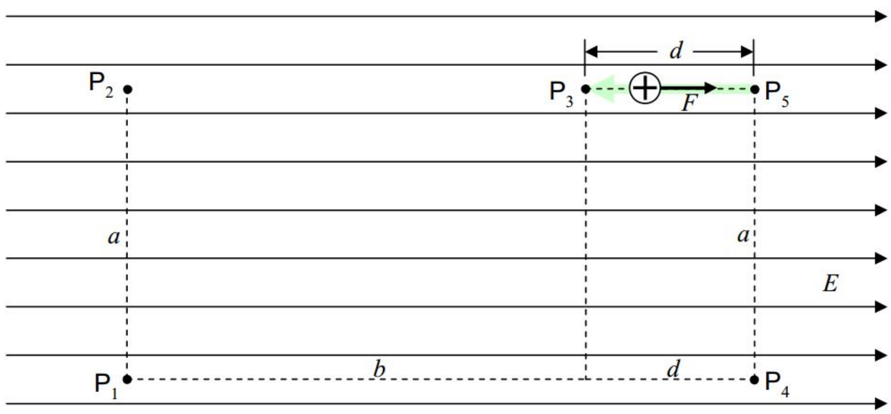
$$W_{14} = F(b+d)$$

$$W_{14} = qE(b+d)$$

From point P_4 to P_5, the force exerted on the charged particle by the electric field is at right angles to the path, so, the force does no work on the charged particle on segment P_4 to P_5.

$$W_{45} = 0$$

On the segment from P_5 to P_3,



the force is in the exact opposite direction to the direction in which the particle moves. This means that the work done by the force of the electric field on the charged particle as the particle moves form P_5 to P_3 is the negative of the magnitude of the force times the length of the path segment. Thus

$$W_{53} = -Fd$$

$$W_{53} = -qEd$$

and

$$W_{1453} = W_{14} + W_{45} + W_{53}$$

$$W_{1453} = qE(b+d) + 0 + (-qEd)$$

$$W_{1453} = qEb$$

As advertised, we obtain the same result for the work done on the particle as it moves from P_1 to P_3 along "P_1 to P_4 to P_5 to P_3" as we did on the other two paths.

Whenever the work done on a particle by a force acting on that particle, when that particle moves from point P_1 to point P_3, is the same no matter what path the particle takes on the way from P_1 to P_3, we can define a potential energy function for the force. The potential energy function is an assignment of a value of potential energy to every point in space. Such an assignment allows us to calculate the work done on the particle by the force when the particle moves from point P_1 to point P_3 simply by subtracting the value of the potential energy of the particle at P_1 from the value of the potential energy of the particle at P_3 and taking the negative of the result. In other words, the work done on the particle by the force of the electric field when the particle goes from one point to another is just the negative of the change in the potential energy of the particle.

In determining the potential energy function for the case of a particle of charge q in a uniform electric field $\vec{E}$, (an infinite set of vectors, each pointing in one and the same direction and each having one and the same magnitude E) we rely heavily on your understanding of the nearearth's-surface gravitational potential energy. Near the surface of the earth, we said back in volume 1 of this book, there is a uniform gravitational field, (a force-per-mass vector field) in the downward direction. A particle of mass m in that field has a force "mg downward" exerted upon it at any location in the vicinity of the surface of the earth. For that case, the potential energy of a particle of mass m is given by mgy where mg is the magnitude of the downward force and y is the height that the particle is above an arbitrarily-chosen reference level. For ease of comparison with the case of the electric field, we now describe the reference level for gravitational potential energy as a plane, perpendicular to the gravitational field g, the force-per mass vector field; and; we call the variable y the "upfield" distance (the distance in the direction opposite that of the gravitational field) that the particle is from the reference plane. (So, we're calling the direction in which the gravitational field points, the direction you know to be downward, the "downfield" direction.)

Now let's switch over to the case of the uniform electric field. As in the case of the near-earth's surface gravitational field, the force exerted on its victim by a uniform electric field has one and the same magnitude and direction at any point in space. Of course, in the electric field case, the force is qE rather than mg and the characteristic of the victim that matters is the charge q rather than the mass m. We call the direction in which the electric field points, the "downfield" direction, and the opposite direction, the "upfield" direction. Now we arbitrarily define a plane that is perpendicular to the electric field to be the reference plane for the electric potential energy of a particle of charge q in the electric field. If we call d the distance that the charged particle is away from the plane in the upfield direction, then the potential energy of the particle with charge q is given by

$$U = qEd$$

where

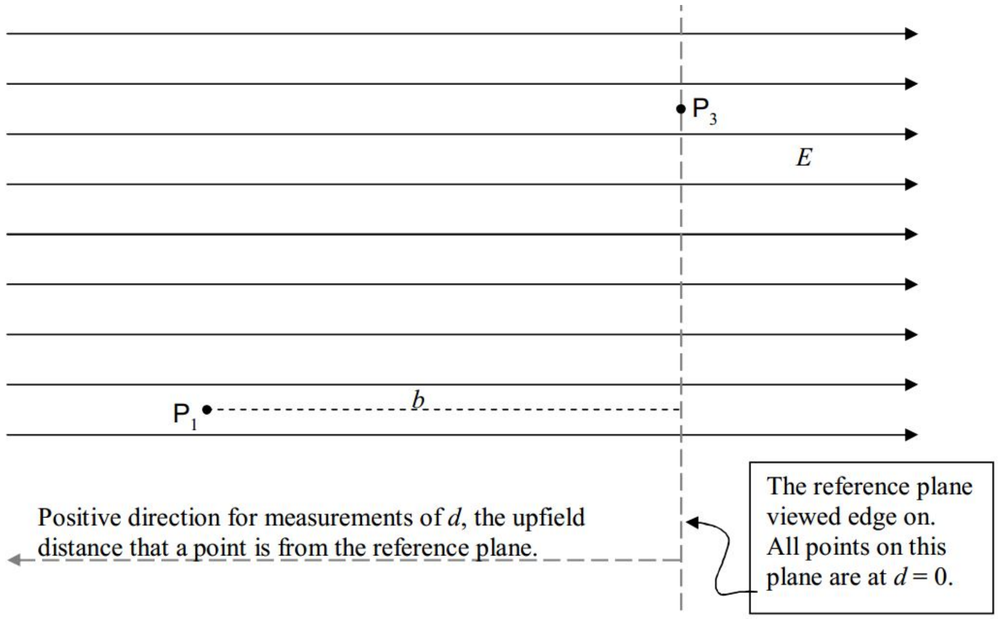
U is the electric potential energy of the charged particle,

q is the charge of the particle,

E is the magnitude of every electric field vector making up the uniform electric field, and

d is the "upfield" distance that the particle is from the $U = 0$ reference plane.

Let's make sure this expression for the potential energy function gives the result we obtained previously for the work done on a particle with charge q, by the uniform electric field depicted in the following diagram, when the particle moves from P_1 to P_3



As you can see, I have chosen (for my own convenience) to define the reference plane to be at the most downfield position relevant to the problem. With that choice, the particle of charge q, when it is at P_1 has potential energy qEb (since point P_1 is a distance b "upfield" from the reference plane) and, when it is at P_3, the particle of charge q has potential energy 0 since P_3 is on the reference plane.

$$W_{13} = -\Delta U$$

$$W_{13} = -(U_3 - U_1)$$

$$W_{13} = -(0 - qEb)$$

$$W_{13} = qEb$$

This is indeed the result we got (for the work done by the electric field on the particle with charge q as that particle was moved from P_1 to P_3) the other three ways that we calculated this work.

B6: The Electric Potential Due to One or More Point Charges

The electric potential due to a point charge is given by

$$\varphi = \frac{kq}{r} \tag{B6.1}$$

where

- φ is the electric potential due to the point charge,
- $k = 8.99 \times 10^9 \frac{Nm^2}{C^2}$ is the Coulomb constant,
- q is the charge of the particle (the source charge, a.k.a. the point charge) causing the electric field for which the electric potential applies, and,
- r is the distance that the point of interest is from the point charge.

In the case of a non-uniform electric field (such as the electric field due to a point charge), the electric potential method for calculating the work done on a charged particle is much easier than direct application of the force-along-the-path times the length of the path. Suppose, for instance, a particle of charge q' is fixed at the origin and we need to find the work done by the electric field of that particle on a victim of charge q as the victim moves along the x axis from x_1 to x_2. We can't simply calculate the work as

$$F \cdot (x_2 - x_1)$$

even though the force is in the same direction as the displacement, because the force F takes on a different value at every different point on the x axis from $x = x_1$ to $x = x_2$. So, we need to do an integral:

$$dW = F dx$$

$$dW = qE dx$$

$$dW = q\frac{kq'}{x^2} dx$$

$$\int dW = \int_{x_1}^{x_2} q\frac{kq'}{x^2} dx$$

$$W = kq'q \int_{x_1}^{x_2} x^{-2} dx$$

$$W = kq'q \frac{x^{-1}}{-1}\Big|_{x_1}^{x_2}$$

$$W = -kq'q(\frac{1}{x_2} - \frac{1}{x_1})$$

$$W = -(\frac{kq'q}{x_2} - \frac{kq'q}{x_1})$$

Compare this with the following solution to the same problem (a particle of charge q' is fixed at the origin and we need to find the work done by the electric field of that particle on a victim of charge q as the victim moves along the x axis from x_1 to x_2):

$$W = -\Delta U$$

$$W = -q\Delta\varphi$$

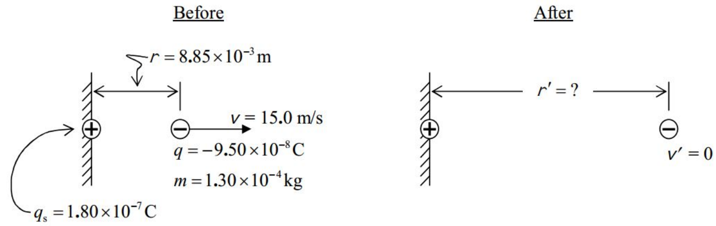
$$W = -q(\varphi_2 - \varphi_1)$$

$$W = -(\frac{kq'q}{x_2} - \frac{kq'q}{x_1})$$

The electric potential energy of a particle, used in conjunction with the principle of the conservation of mechanical energy, is a powerful problem-solving tool. The following example makes this evident:

A particle of charge 0.180 μC is fixed in space by unspecified means. A particle of charge -0.0950 μC and mass 0.130 grams is 0.885 cm away from the first particle and moving directly away from the first particle with a speed of 15.0 m/s. How far away from the first particle does the second particle get?

This is a conservation of energy problem. As required for all conservation of energy problems, we start with a before and after diagram:



$\text{Energy Before} = \text{Energy After}$

$$K + U = K' + U'$$

$$K + q\varphi = 0 + q\varphi'$$

$$\frac{1}{2}mv^2 + q\frac{kq_s}{r} = q\frac{kq_s}{r'}$$

$$\frac{1}{r'} = \frac{1}{r} + \frac{mv^2}{2kq_sq}$$

$$r' = \frac{1}{\frac{1}{r} + \frac{mv^2}{2kq_sq}}$$

$$r' = \frac{1}{\frac{1}{8.85\times10^{-3}m} + \frac{1.30\times10^{-4}kg(15.0m/s)^2}{2(8.99\times10^9\frac{N\cdot m^2}{C^2})1.80\times10^{-7}C(-9.50\times10^{-8}C)}}$$

$$r' = 0.05599m$$

$$r' = 0.0560m$$

$$r' = 5.6cm$$

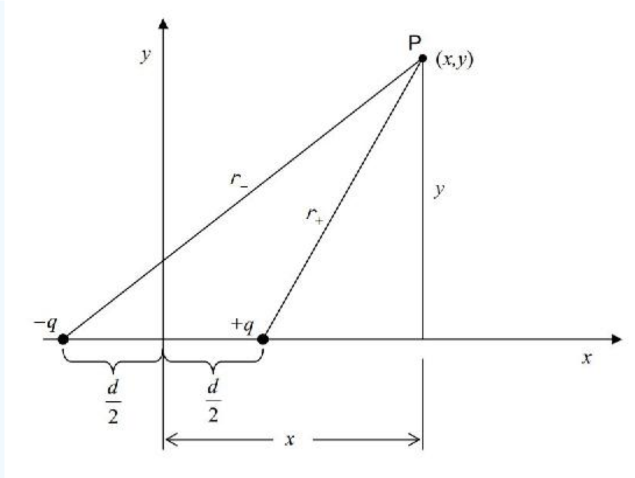
Superposition in the Case of the Electric Potential

When there is more than one charged particle contributing to the electric potential at a point in space, the electric potential at that point is the sum of the contributions due to the individual charged particles. The electric potential at a point in space, due to a set of several charged particles, is easier to calculate than the electric field due to the same set of charged particles is. This is true because the sum of electric potential contributions is an ordinary arithmetic sum, whereas, the sum of electric field contributions is a vector sum.

Find a formula that gives the electric potential at any point (x, y) on the x-y plane, due to a pair of particles: one of charge $-q$ at $(-\frac{d}{2}, 0)$ and the other of charge $+q$ at $(\frac{d}{2}, 0)$.

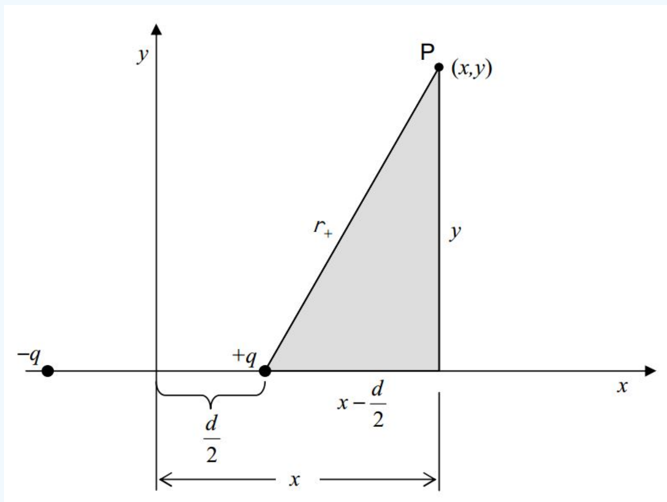
Solution

We establish a point P at an arbitrary position (x, y) on the x-y plain and determine the distance that point P is from each of the charged particles. In the following diagram, I use the symbol r_+ to represent the distance that point P is from the positively-charged particle, and r_- to represent the distance that point P is from the negatively-charged particle.



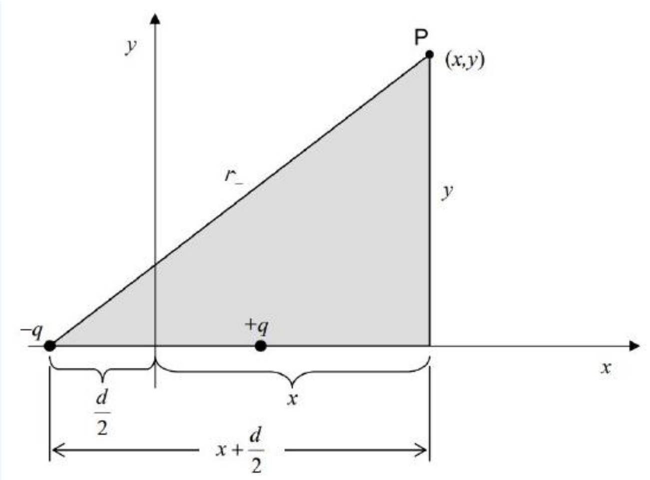
Analysis of the shaded triangle in the diagram at right gives us r_{+}.

$$r_{+} = \sqrt{(x - \frac{d}{2})^2 + y^2}$$



Analysis of the shaded triangle in the diagram at right gives us r_{-}.

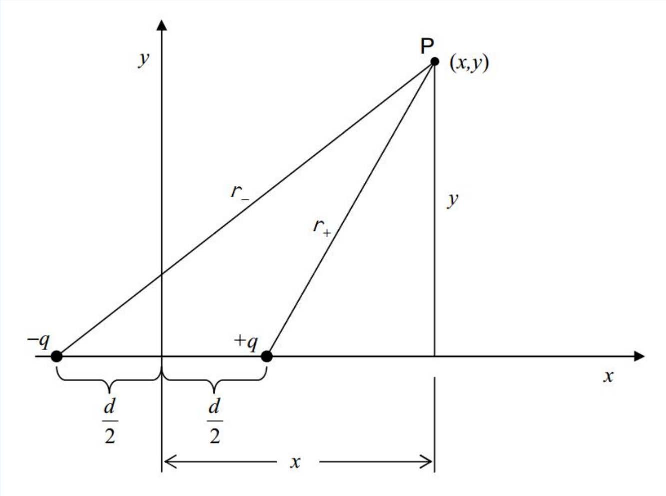
$$r_{-} = \sqrt{(x + \frac{d}{2})^2 + y^2}$$



With the distances that point P is from each of the charged particles in hand, we are ready to determine the potential:

$$r_+ = \sqrt{(x - \frac{d}{2})^2 + y^2}$$

$$r_- = \sqrt{(x + \frac{d}{2})^2 + y^2}$$



$$\varphi(x, y) = \varphi_+ + \varphi_-$$

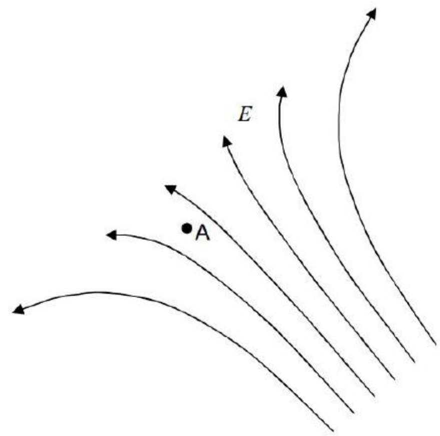
$$\varphi(x, y) = \frac{kq}{r_+} + \frac{k(-q)}{r_-}$$

$$\varphi(x, y) = \frac{kq}{r_+} - \frac{kq}{r_-}$$

$$\varphi(x, y) = \frac{kq}{\sqrt{(x - \frac{d}{2})^2 + y^2}} - \frac{kq}{\sqrt{(x + \frac{d}{2})^2 + y^2}}$$

B7: Equipotential Surfaces, Conductors, and Voltage

Consider a region of space in which there exists an electric field. Focus your attention on a specific point in that electric field, call it point A.



Imagine placing a positive test charge at point A. (Assume that, by means not specified, you can move the test charge anywhere you want to.) Please think about the answer to the following question before reading on: Is it possible for you to move the test charge around in the electric field in such a manner that the electric field does no work on the test charge?

If we move the positive test charge in the “downfield” direction (toward the upper left corner of the diagram), there will be a positive amount of work (force-along-the-path times the length of the path) done on the test charge. And, if we move the positive test charge in the “upfield” direction there will be a negative amount of work done on it. But, if we move the positive test charge at right angles to the electric field, no work is done on it. That is, if we choose a path for the positive test charge such that every infinitesimal displacement of the particle is normal to the electric field at the location of the particle when it (the particle) undergoes said infinitesimal displacement, then the work done on the test charge, by the electric field, is zero. The set of all points that can be reached by such paths makes up an infinitesimally thin shell, a surface, which is everywhere perpendicular to the electric field. In moving a test charge along the surface from one point (call it point A) to another point (call it point B) on the surface, the work done is zero because the electric field is perpendicular to the path at all points along the path. Let’s (momentarily) call the kind of surface we have been discussing a “zero-work surface.” We have constructed the surface by means of force-along-the-path times the length-of-the-path work considerations. But the work done by the electric field when a test charge is moved from point A on the surface to point B on the surface must also turn out to be zero if we calculate it as the negative of the change in the potential energy of the test charge. Let’s do that and see where it leads us. We know that the work $W=0$.

Also

$$W=-\Delta U \tag{B7.1}$$

$$=-(U_B-U_A) \tag{B7.2}$$

In terms of the electric potential φ, $U=q\varphi$ so the work can be expressed as

$$W=-(q\varphi_B-q\varphi_A) \tag{B7.3}$$

$$=-q(\varphi_B-\varphi_A) \tag{B7.4}$$

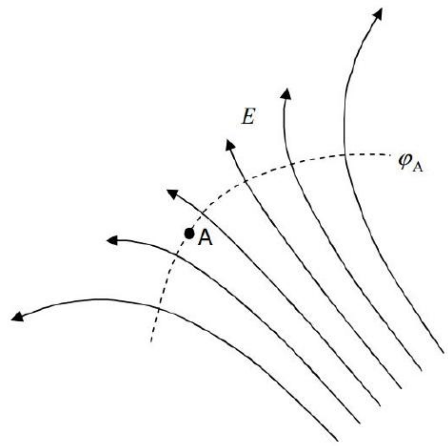
Given that $W=0$, this means that

$$0=-q(\varphi_B-\varphi_A)$$

$$\varphi_B-\varphi_A=0$$

$$\varphi_B=\varphi_A$$

This is true for any point B on the entire “zero-work” surface. This means that every point on the entire surface is at the same value of electric potential. Thus a “zero-work” surface is also an equipotential surface. Indeed, this is the name (equipotential surface) that physicists use for such a surface. An equipotential surface is typically labeled with the corresponding potential value (φ_A in the case at hand). In the following diagram, the dashed curve represents the equipotential surface viewed edge on.



Summarizing:

- An equipotential surface is an imaginary surface on which every point has one and the same value of electric potential.
- An equipotential surface is everywhere perpendicular to the electric field that it characterizes.
- The work done by the electric field on a particle when it is moved from one point on an equipotential surface to another point on the same equipotential surface is always zero.

Perfect Conductors and the Electric Potential

Please recall what you know about perfect conductors and the electric field. Namely, that everywhere inside and on a perfect conductor, the electric field is zero. This goes for solid conductors as well as hollow, empty shells of perfectly conducting material. This means that the work done by the electric field on a test charge that is moved from one point in or on a perfect conductor (consider this to be a thought experiment), to another point in or on the same conductor, is zero. This means that the difference in the electric potential between any two points in or on a perfect conductor must be zero. This means that the electric potential at every point in and on a perfect conductor must have one and the same value. Note that the value is not, in general, zero.

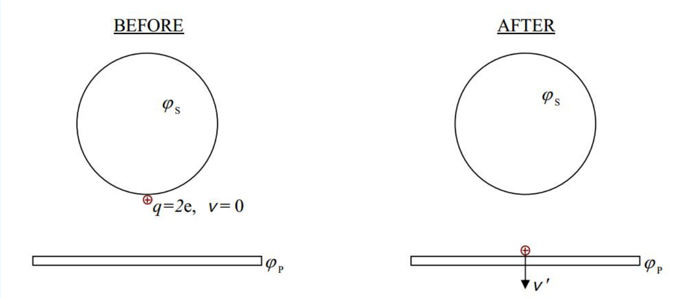
Some Electric Potential Jargon

When we talk about the electric potential in the context of a perfect conductor (or an object that approximates a perfect conductor), because every point in and on the conductor has the same value of electric potential, we typically call that value the electric potential of the conductor. We also use expressions such as, "the conductor is at a potential of 25 volts," meaning that the value of electric potential at every point in and on the conductor is 25 volts with respect to infinity (meaning that the zero of electric potential is at an infinite distance from the conductor) and/or with respect to "ground" (meaning that the potential of the earth is the zero of electric potential).

Electric Potential Difference, a.k.a. Voltage

In general, what is at issue when one talks about conductors and electric potential is not the value of the electric potential of a conductor, but rather, the electric potential difference between one conductor and another.

A hollow metal sphere is at a potential that is 472 volts higher than that of a nearby metal plate. A particle of charge $2e$ is released from rest at the surface of the sphere. It subsequently strikes the plate. With what kinetic energy does the charged particle strike the plate? (Assume that the only force acting on the particle is that due to the electric field corresponding to the given information.)



(Given $\varphi_S - \varphi_P = \Delta\varphi = 472$ volts)

$$\text{Energy Before} = \text{Energy After}$$

$$0 + U = K' + U'$$

$$q\varphi_S = K' + q\varphi_P$$

$$K' = q\varphi_S - q\varphi_P$$

$$K' = q(\varphi_S - \varphi_P)$$

$$K' = q\Delta\varphi$$

$$K' = 2e(472\text{volts})$$

$$K' = 944eV$$

Note that in the solution to the example problem, we never needed to know the value of the electric potential of either the sphere or the plate, only the difference between the two potentials. There is a device which can be used to measure the potential difference between two points in space. The device is called a voltmeter. A typical voltmeter consists of a box with two wires extending from it. On the end of each wire is a short metal wand called a probe. Each wire and each probe, except for the tip of the probe, is covered with insulating material. The box displays, either by means of a digital readout or the position of a needle, the potential difference between the two wires. In typical use, one presses the metal tip of one probe against a conductor of interest and holds the tip there. That causes that probe and wire to be at the same potential as the conductor. One presses the tip of the other probe against another conductor. This causes that probe and wire to be at the potential of the second conductor. With each probe in contact with a conductor, the voltmeter continually displays the potential difference between the two conductors.

Based on the SI units of measurement, the electric potential difference between two points in space goes by another name, namely, voltage. Voltage means electric potential difference which means, the difference between the electric-potential-energy-per-charge-of-would-be-victim at one point in space and the electric-potential-energy-per-charge-of-would-be-victim at another point in space. While voltage literally means potential difference, the word is also, quite often used to mean electric potential itself, where, one particular conductor or point in space is defined to be the zero of potential. If no conductor or point in space has been defined to be the zero, then it is understood that “infinity” is considered to be at the zero of electric potential. So, if you read that a metal object is at a potential of 230 volts (when no conductor or point in space has been identified as the zero of electric potential), you can interpret the statement to mean the same thing as a statement that the electric potential of the metal object is 230 volts higher than the electric potential at any point that is an infinite distance away from the object.

As you move on in your study of physics, onward to your study and work with electric circuits, it is important to keep in mind that voltage, in a circuit, is the difference in the value of a characteristic (the electric potential) of one conductor, and the value of the same characteristic (electric potential) of another conductor.

Analogy Between Voltage and Altitude

One can draw a pretty good analogy between voltage (electric potential) and altitude. Consider a particular altitude above the surface of the earth (measured, for instance, from sea level). The value of the altitude characterizes a point in space or a set of points in space. In fact, the set of all points in space that are at the same altitude above the surface of the earth forms an “equi-altitude” surface. On a local scale, we can think of that “equi-altitude” surface as a plane. On a global scale, looking at the big picture, we recognize it to be a spheroidal shell. Flocks of birds can be at that altitude and when they are, we attribute the altitude to the flock of birds. We say that the flock of birds has such and such an altitude. But, whether or not the flock of birds is there, the altitude exists. Regarding a particular altitude, we can have birds and air and clouds moving or flowing through space at that altitude, but the altitude itself just exists—it doesn’t flow or go anywhere. This is like the voltage in a circuit. The voltage in a circuit exists. The voltage characterizes a conductor in a circuit. Charged particles can move and flow in and through a conductor that is at that voltage, but, the voltage doesn’t flow or go anywhere, any more than altitude flows or goes anywhere.

B8: Capacitors, Dielectrics, and Energy in Capacitors

Capacitance is a characteristic of a conducting object. Capacitance is also a characteristic of a pair of conducting objects.

Let's start with the capacitance of a single conducting object, isolated from its surroundings. Assume the object to be neutral. Now put some positive charge on the object. The electric potential of the object is no longer zero. Put some more charge on the object and the object will have a higher value of electric potential. What's interesting is, no matter how much, or how little charge you put on the object, the ratio of the amount of charge q on the object to the resulting electric potential φ of the object has one and the same value.

$$\frac{q}{\varphi} \text{ have the same value for any value of } q$$

You double the charge, and, the electric potential doubles. You reduce the amount of charge to one tenth of what it was, and, the electric potential becomes one tenth of what it was. The actual value of the unchanging ratio is called the capacitance C_{sc} of the object (where the subscript "sc" stands for "single conductor").

$$C_{sc} = \frac{q}{\varphi} \tag{B8.1}$$

where:

- C_{sc} is the capacitance of a single conductor, isolated (distant from) its surroundings,
- q is the charge on the conductor, and,
- φ is the electric potential of the conductor relative to the electric potential at infinity (the position defined for us to be our zero level of electric potential).

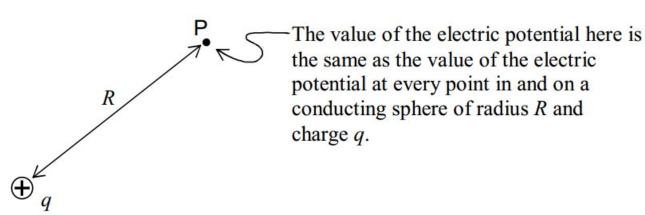
The capacitance of a conducting object is a property that an object has even if it has no charge at all. It depends on the size and shape of the object.

The more positive charge you need to add to an object to raise the potential of that object 1 volt, the greater the capacitance of the object. In fact, if you define q_1 to be the amount of charge you must add to a particular conducting object to increase the electric potential of that object by one volt, then the capacitance of the object is $\frac{q_1}{1 \text{ volt}}$.

The Capacitance of a Spherical Conductor

Consider a sphere (either an empty spherical shell or a solid sphere) of radius R made out of a perfectly-conducting material. Suppose that the sphere has a positive charge q and that it is isolated from its surroundings. We have already covered the fact that the electric field of the charged sphere, from an infinite distance away, all the way to the surface of the sphere, is indistinguishable from the electric field due to a point charge q at the position of the center of the sphere; and; everywhere inside the surface of the sphere, the electric field is zero. Thus, outside the sphere, the electric potential must be identical to the electric potential due to a point charge at the center of the sphere (instead of the sphere). Working your way in from infinity, however, as you pass the surface of the sphere, the electric potential no longer changes. Whatever the value of electric potential at the surface of the sphere is, that is the value of electric potential at every point inside the sphere.

This means that the electric potential of the sphere is equal to the electric potential that would be caused by a point charge (all by itself) at a point in space a distance R from the point charge (where R is the radius of the sphere).



Thus, $\varphi = \frac{kq}{R}$ is the electric potential of a conducting sphere of radius R and charge q.

Solving this expression for $\frac{q}{\varphi}$ yields:

$$\frac{q}{\varphi}=\frac{R}{k}$$

Since, by definition, the capacitance $C_{sc}=\frac{q}{\varphi}$, we have:

$$C_{sc}=\frac{R}{k} \tag{B8.2}$$

The capacitance of a conducting sphere is directly proportional to the radius of the sphere. The bigger the sphere, the more charge you have to put on it to raise its potential one volt (in other words, the bigger the capacitance of the sphere). This is true of conducting objects in general. Since all the unbalanced charge on a conductor resides on the surface of the conductor, it really has to do with the amount of surface area of the object. The more surface area, the more room the charge has to spread out and, therefore, the more charge you have to put on the object to raise its potential one volt (in other words, the bigger the capacitance of the object).

Consider, for instance, a typical paper clip. It only takes an amount of charge on the order of a pC (picocoulomb, 1×10^{-12} coulombs) to raise the potential of a paper clip 10 volts.

Units: the Farad

The unit of capacitance is the coulomb-per-volt, $\frac{C}{V}$. That combination unit is given a name, the farad, abbreviated F.

$$1F=1\frac{C}{V}$$

The Capacitance of a Pair of Conducting Objects

So far, we've been talking about the capacitance of a conducting object that is isolated from its surroundings. You put some charge on such an object, and, as a result, the object takes on a certain value of electric potential. The charge-to-potential ratio is called the capacitance of the object. But get this, if the conductor is near another conductor when you put the charge on it, the conductor takes on a different value of electric potential (compared to the value it takes on when it is far from all other conductors) for the exact same amount of charge. This means that just being in the vicinity of another conductor changes the effective capacitance of the conductor in question. In fact, if you put some charge on an isolated conductor, and then bring another conductor into the vicinity of the first conductor, the electric potential of the first conductor will change, meaning, its effective capacitance changes. Let's investigate a particular case to see how this comes about.

Consider a conducting sphere with a certain amount of charge, q, on it. Suppose that, initially, the sphere is far from its surroundings and, as a result of the charge on it, it is at a potential φ.

Let's take a moment to review what we mean when we say that the sphere is at a potential φ. Imagine that you take a test charge q_T from a great distance away from the sphere and take it to the surface of the sphere. Then you will have changed the potential energy of the test charge from zero to $q_T\varphi$. To do that, you have to do an amount of work $q_T\varphi$ on the test charge. We're assuming that the test charge was initially at rest and is finally at rest. You have to push the charge onto the sphere. You apply a force over a distance to give that particle the potential energy $q_T\varphi$. You do positive work on it. The electric field of the sphere exerts a force on the test charge in the opposite direction to the direction in which you are moving the test charge. The electric field does a negative amount of work on the test charge such that the total work, the work done by you plus the work done by the electric field, is zero (as it must be since the kinetic energy of the test charge does not change). But I want you to focus your attention on the amount of work that you must do, pushing the test charge in the same direction in which it is going, to bring the test charge from infinity to the surface of the sphere. That amount of work is $q_T\varphi$ because $q_T\varphi$ is the amount by which you increase the potential energy of the charged particle. If you were to repeat the experiment under different circumstances and you found that you did not have to do as much work to bring the test charge from infinity to the surface of the sphere, then you would know that the sphere is at a lower potential than it was the first time.

Now, we are ready to explore the case that will illustrate that the charge-to-voltage ratio of the conducting object depends on whether or not there is another conductor in the vicinity. Let's bring an identical conducting sphere near one side of the first sphere. The first sphere still has the same amount of charge q on it that it always had, and, the second sphere is neutral. The question is, "Is the potential of the original sphere still the same as what it was when it was all alone?" Let's test it by bringing a charge in from an infinite distance on the opposite side of the first sphere (as opposed to the side to which the second sphere now resides). Experimentally we find that it takes less work to bring the test charge to the original sphere than it did before, meaning that the original sphere now has a lower value of electric potential. How can that be? Well, when we brought the second sphere in close to the original sphere, the second sphere became polarized. (Despite the fact that it is neutral, it is a conductor so the balanced charge in it is free to move around.) The original sphere, having positive charge q, attracts the negative charge in the second sphere and repels the positive charge. The near side of the second sphere winds up with a negative charge and the far side, with the same amount of positive charge. (The second sphere remains neutral overall.) Now the negative charge on the near side of the second sphere attracts the (unbalanced) positive charge on the original sphere to it. So the charge on the original sphere, instead of being spread out uniformly over the surface as it was before the second sphere was introduced, is bunched up on the side of the original sphere that is closer to the second sphere. This leaves the other side of the original sphere, if not neutral, at least less charged than it was before. As a result, it takes less work to bring the positive test charge in from infinity to that side of the original sphere. As mentioned, this means that the electric potential of the original sphere must be lower than it was before the second sphere was brought into the picture. Since it still has the same charge that it always had, the new, lower potential, means that the original sphere has a greater charge-to-potential ratio, and hence a greater effective capacitance.

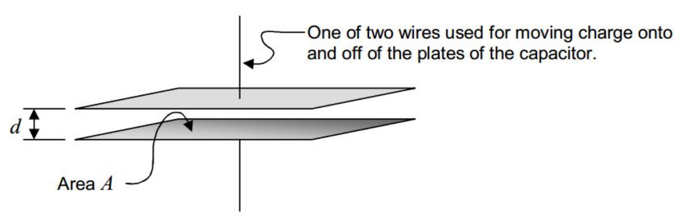
In practice, rather than call the charge-to-potential ratio of a conductor that is near another conductor, the "effective capacitance" of the first conductor, we define a capacitance for the pair of conductors. Consider a pair of conductors, separated by vacuum or insulating material, with a given position relative to each other. We call such a configuration a capacitor. Start with both conductors being neutral. Take some charge from one conductor and put it on the other. The amount of charge moved from one conductor to the other is called the charge of the capacitor. (Contrast this with the actual total charge of the device which is still zero.) As a result of the repositioning of the charge, there is a potential difference between the two conductors. This potential difference $\Delta\varphi$ is called the voltage of the capacitor or, more often, the voltage across the capacitor. We use the symbol V to represent the voltage across the capacitor. In other words, $V \equiv \Delta\varphi$. The ratio of the amount of charge moved from one conductor to the other, to, the resulting potential difference of the capacitor, is the capacitance of the capacitor (the pair of conductors separated by vacuum or insulator).

$$C=\frac{q}{V} \tag{B8.3}$$

where:

- C is the capacitance of a capacitor, a pair of conductors separated by vacuum or an insulating material,
- q is the "charge on the capacitor," the amount of charge that has been moved from one initially neutral conductor to the other. One conductor of the capacitor actually has an amount of charge q on it and the other actually has an amount of charge $-q$ on it.
- V is the electric potential difference $\Delta\varphi$ between the conductors. It is known as the voltage of the capacitor. It is also known as the voltage across the capacitor.

A two-conductor capacitor plays an important role as a component in electric circuits. The simplest kind of capacitor is the parallel-plate capacitor. It consists of two identical sheets of conducting material (called plates), arranged such that the two sheets are parallel to each other. In the simplest version of the parallel-plate capacitor, the two plates are separated by vacuum.



The capacitance of such a capacitor is given by

$$C=\epsilon_o\frac{A}{d}$$

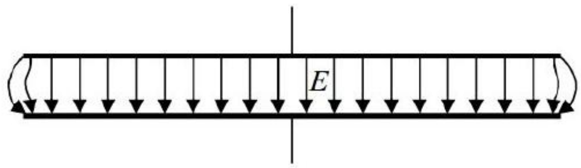
where:

- C is the capacitance of the parallel-plate capacitor whose plates are separated by vacuum,
- d is the distance between the plates,
- A is the area of one face of one of the plates,

- ϵ_o is a universal constant called the permittivity of free space. ϵ_o is closely related to the Coulomb constant k. In fact, $k = \frac{1}{4\pi\epsilon_o}$.

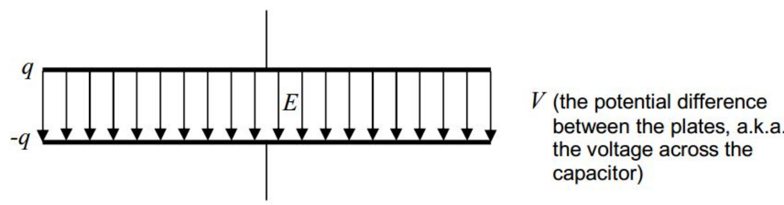
Thus, $\epsilon = 8.85 \times 10^{-12} \frac{\mathrm{C}^2}{\mathrm{N} \cdot \mathrm{m}^2}$. Our equation for the capacitance can be expressed in terms of the Coulomb constant k as $C = \frac{1}{4\pi k}\frac{A}{d}$, but, it is more conventional to express the capacitance in terms of ϵ_o.

This equation for the capacitance is an approximate formula. It is a good approximation as long as the plate separation d is small compared to a representative plate dimension (the diameter in the case of circular plates, the smaller edge length in the case of rectangular plates). The derivation of the formula is based on the assumption that the electric field, in the region between the plates is uniform, and the electric field outside that region is zero. In fact, the electric field is not uniform in the vicinity of the edges of the plates. As long as the region in which the electric field is not well-approximated by a uniform electric field is small compared to the region in which it is, our formula for the capacitance is good.



The Effect of Insulating Material Between the Plates of a Capacitor

To get at the effect of insulating material, rather than vacuum, between the plates of a capacitor, I need to at least outline the derivation of the formula $C = \epsilon_o \frac{A}{d}$. Keep in mind that the capacitance is the charge-per-voltage of the capacitor. Suppose that we move charge q from one initially-neutral plate to the other. We assume that the electric field is uniform between the plates of the capacitor and zero elsewhere.



By means that you will learn about later in this book we establish that the value of the electric field (valid everywhere between the plates) is given by:

$$E = \frac{q}{A\epsilon_o} \tag{B8.4}$$

Also, we know that the work done on a test charge q_T by the electric field when the test charge is moved from the higher-potential plate to the lower-potential plate is the same whether we calculate it as force-along the path times the length of the path, or, as the negative of the change in the potential energy. This results in a relation between the electric field and the electric potential as follows:

$$\text{W calcualted as force times distance} = \text{W calculated as minus change in potential energy}$$

$$F\Delta x = -\Delta U$$

$$q_T E d = -q_T \Delta\varphi$$

$$Ed = -(-V)$$

$$V = Ed$$

Using Equation B8.4 ($E = \frac{q}{A\epsilon_o}$) to replace the E in $V = Ed$ with $\frac{q}{A\epsilon_o}$ gives us:

$$V = \frac{q}{A\epsilon_o}d$$

Solving this for q/V yields

$$\frac{q}{V} = \epsilon_o \frac{A}{d}$$

for the charge-to-voltage ratio. Since the capacitance is the charge-to-voltage ratio, this means

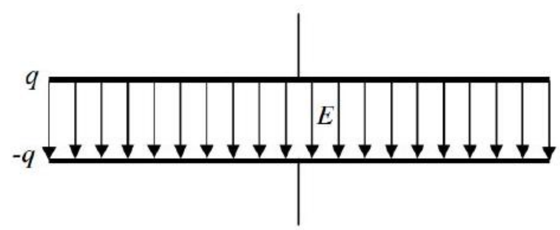
$$C = \epsilon_o \frac{A}{d}$$

which is what we set out to derive.

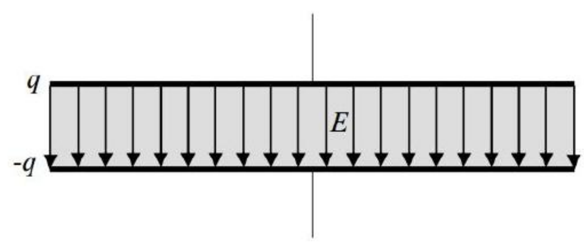
Okay now, here's the deal on having an insulator between the plates: Consider a capacitor that is identical in all respects to the one we just dealt with, except that there is an insulating material between the plates, rather than vacuum. Further suppose that the capacitor has the same amount of charge q on it as the vacuum-between-the-plates capacitor had on it. The presence of the insulator between the plates results in a weaker electric field between the plates. This means that a test charge moved from one plate to another would have less work done on it by the electric field, meaning that it would experience a smaller change in potential energy, meaning the electric potential difference between the plates is smaller. So, with the same charge, but a smaller potential difference, the charge-to-voltage ratio (that is, the capacitance of the capacitor) must be bigger.

The presence of the insulating material makes the capacitance bigger. The part of the preceding argument that still needs explaining is that part about the insulating material weakening the electric field. Why does the insulating material make the field weaker? Here's the answer:

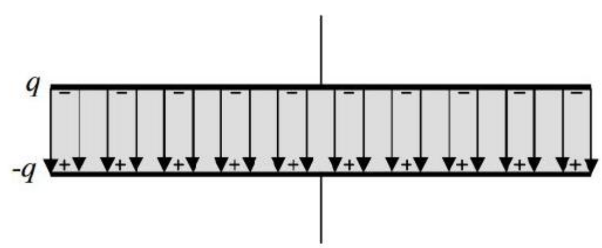
Starting with vacuum between the plates,



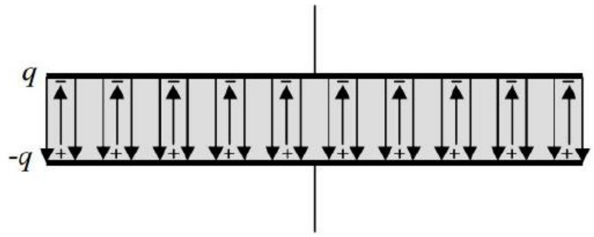
we insert some insulating material:



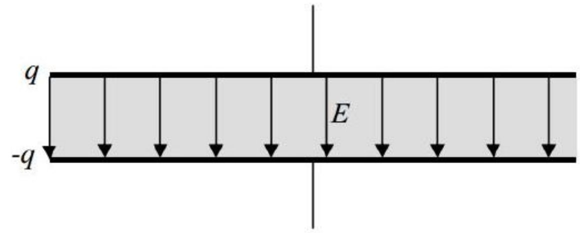
The original electric field polarizes the insulating material:



The displaced charge creates an electric field of its own, in the direction opposite that of the original electric field:



The net electric field, being at each point in space, the vector sum of the two contributions to it, is in the same direction as the original electric field, but weaker than the original electric field:



This is what we wanted to show. The presence of the insulating material makes for a weaker electric field (for the same charge on the capacitor), meaning a smaller potential difference, meaning a bigger charge-to-voltage ratio, meaning a bigger capacitance. How much bigger depends on how much the insulator is polarized which depends on what kind of material the insulator consists of. An insulating material, when placed between the plates of a capacitor is called a dielectric. The net effect of using a dielectric instead of vacuum between the plates is to multiply the capacitance by a factor known as the dielectric constant. Each dielectric is characterized by a unitless dielectric constant specific to the material of which the dielectric is made. The capacitance of a parallel-plate capacitor which has a dielectric in between the plates, rather than vacuum, is just the dielectric constant κ times the capacitance of the same capacitor with vacuum in between the plates.

$$C = \kappa \epsilon_o \frac{A}{d} \tag{B8.5}$$

where:

- C is the capacitance of the parallel-plate capacitor whose plates are separated by an insulating material,
- κ is the dielectric constant characterizing the insulating material between the plates,
- d is the distance between the plates,
- A is the area of one face of one of the plates, and
- ϵ_o is a universal constant called the permittivity of free space.

Calling the dielectric constant for vacuum 1 (exactly one), we can consider this equation to apply to all parallel-plate capacitors. Some dielectric constants of materials used in manufactured capacitors are provided in the following table:

Substance	Dielectric Constant
Air	1.00
Aluminium Oxide (a corrosion product found in many electrolytic capacitors)	7
Mica	3-8
Titanium Dioxide	114
Vacuum	1 (exactly)
Waxed Paper	2.5-3.5

Energy Stored in a Capacitor

Moving charge from one initially-neutral capacitor plate to the other is called **charging the capacitor**. When you charge a capacitor, you are storing energy in that capacitor. Providing a conducting path for the charge to go back to the plate it came from is called discharging the capacitor. If you discharge the capacitor through an electric motor, you can definitely have that charge do some work on the surroundings. So, how much energy is stored in a charged capacitor? Imagine the charging process. You use some force to move some charge over a distance from one plate to another. At first, it doesn't take much force because both plates are neutral. But the more charge that you have already relocated, the harder it is to move more charge. Think about it. If you are moving positive charge, you are pulling positive charge from a negatively charged plate and pushing it onto a positively charged plate. The total amount of work you do in moving the charge is the amount of energy you store in the capacitor. Let's calculate that amount of work.

In this derivation, a lower case q represents the variable amount of charge on the capacitor plate (it increases as we charge the capacitor), and an upper case Q represents the final amount of charge. Similarly, a lower case v represents the variable amount of voltage across the capacitor (it too increases as we charge the capacitor), and the upper case V represents the final voltage across the capacitor. Let U represent the energy stored in the capacitor:

$$dU = vdp$$

but the voltage across the capacitor is related to the charge of the capacitor by $C = q/v$ (Equation B8.3), which, solved for v is $v = q/C$, so:

$$dU = \frac{q}{C}dq \tag{B8.6}$$

$$\int dU = \frac{1}{C}\int_0^Q qdq \tag{B8.7}$$

$$U = \frac{1}{C}\frac{q^2}{2}\Big|_0^Q \tag{B8.8}$$

$$U = \frac{1}{C}\left(\frac{Q^2}{2} - \frac{0^2}{2}\right) \tag{B8.9}$$

$$U = \frac{1}{2}\frac{1}{C}Q^2 \tag{B8.10}$$

Using $C = Q/V$, we can also express the energy stored in the capacitor as $U = \frac{1}{2}QV$, or

$$U = \frac{1}{2}CV^2 \tag{B8.11}$$

9B: Electric Current, EMF, and Ohm's Law

We now begin our study of electric circuits. A circuit is a closed conducting path through which charge flows. In circuits, charge goes around in loops. The charge flow rate is called electric current. A circuit consists of circuit elements connected together by wires. A capacitor is an example of a circuit element with which you are already familiar. We introduce some more circuit elements in this chapter. In analyzing circuits, we treat the wires as perfect conductors and the circuit elements as ideal circuit elements. There is a great deal of variety in the complexity of circuits. A computer is a complicated circuit. A flashlight is a simple circuit.

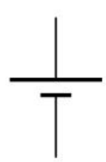
The kind of circuit elements that you will be dealing with in this course are two-terminal circuit elements. There are several different kinds of two-terminal circuit elements but they all have some things in common. A two-terminal circuit element is a device with two ends, each of which is a conductor. The two conductors are called terminals. The terminals can have many different forms. Some are wires, some are metal plates, some are metal buttons, and some are metal posts. One connects wires to the terminals to make a circuit element part of a circuit.

An important two-terminal circuit element is a seat of EMF. You can think of a seat of EMF as an ideal battery or as an ideal power supply. What it does is to maintain a constant potential difference (a.k.a. a constant voltage) between its terminals. One uses either the constant name ε (script E) or the constant name V to represent that potential difference.

To achieve a potential difference E between its terminals, a seat of EMF, when it first comes into existence, has to move some charge (we treat the movement of charge as the movement of positive charge) from one terminal to the other. The "one terminal" is left with a net negative charge and "the other" acquires a net positive charge. The seat of EMF moves charge until the positive terminal is at a potential E higher than the negative terminal. Note that the seat of EMF does not produce charge; it just pushes existing charge around. If you connect an isolated wire to the positive terminal, then it is going to be at the same potential as the positive terminal, and, because the charge on the positive terminal will spread out over the wire, the seat of EMF is going to have to move some more charge from the lower-potential terminal to maintain the potential difference. One rarely talks about the charge on either terminal of a seat of EMF or on a wire connected to either terminal. A typical seat of EMF maintains a potential difference between its terminals on the order of 10 volts and the amount of charge that has to be moved, from one wire whose dimensions are similar to that of a paper clip, to another of the same sort, is on the order of a pC ($1 \times 10^{-12} C$). Also, the charge pileup is almost instantaneous, so, by the time you finish connecting a wire to a terminal, that wire already has the charge we are talking about. In general, we don't know how much charge is on the positive terminal and whatever wire might be connected to it, and we don't care. It is minuscule. But, it is enough for the potential difference between the terminals to be the rated voltage of the seat of EMF.

You'll recall that electric potential is something that is used to characterize an electric field. In causing there to be a potential difference between its terminals and between any pair of wires that might be connected to its terminals, the seat of EMF creates an electric field. The electric field depends on the arrangement of the wires that are connected to the terminals of the seat of EMF. The electric field is another quantity that we rarely discuss in analyzing circuits. We can typically find out what we need to find out from the value of the potential differenceE that the seat of EMF maintains between its terminals. But, the electric field does exist, and, in circuits, the electric field of the charge on the wires connected to the seat of EMF is what causes charge to flow in a circuit, and charge flow in a circuit is a huge part of what a circuit is all about.

We use the symbol

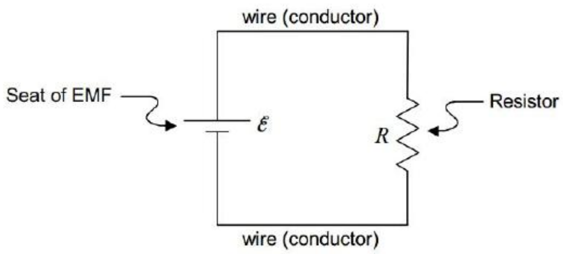


to represent a seat of EMF in a circuit diagram (a.k.a. a schematic diagram of a circuit) where the two collinear line segments represent the terminals of the seat of EMF, the one connected to the shorter of the parallel line segments being the negative, lower-potential, terminal; and; the one connected to the longer of the parallel line segments being the positive, higher-potential, terminal.

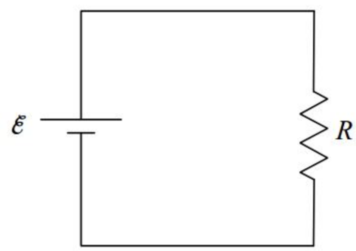
The other circuit element that I want to introduce in this chapter is the resistor. A resistor is a poor conductor. The resistance of a resistor is a measure of how poor a conductor the resistor is. The bigger the value of resistance, the more poorly the circuit element allows charge to flow through itself. Resistors come in many forms. The filament of a light bulb is a resistor. A toaster element (the part that glows red when the toaster is on) is a resistor. Humans manufacture small ceramic cylinders (with a coating of carbon and a wire sticking out each end) to have certain values of resistance. Each one has its value of resistance indicated on the resistor itself. The symbol

is used to represent a resistor in a circuit diagram. The symbol R is typically used to represent the value of the resistance of a resistor.

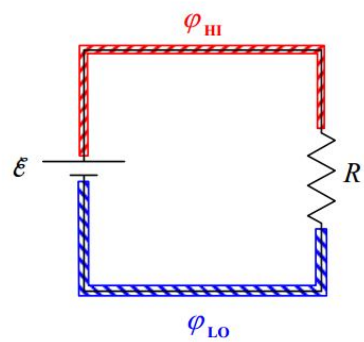
We are now ready to consider the following simple circuit:



Here it is again without so many labels:

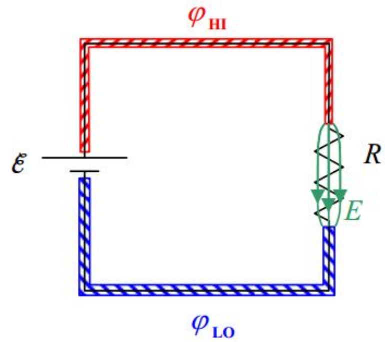


The upper wire (conductor) has one value of electric potential (call it φ_{HI}) and the lower wire has another value of electric potential (call it φ_{LOW}) such that the difference $\varphi_{HI} - \varphi_{LOW}$ is ε.



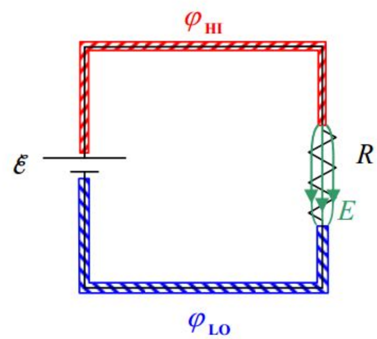
$$\varphi_{HI} - \varphi_{LOW} = \varepsilon$$

In order to maintain the potential difference ε between the two conductors, the seat of EMF causes there to be a minuscule amount of positive charge on the upper wire and the same amount of negative charge on the lower wire. This charge separation causes an electric field in the resistor.



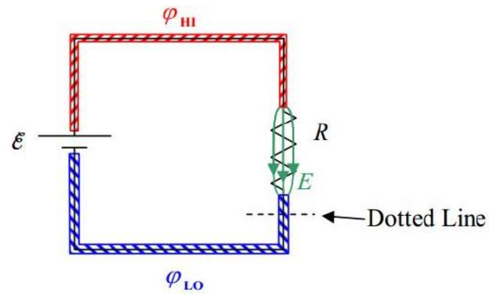
(We carry out this argument in the positive charge carrier model. While it makes no difference for the circuit, as a point of fact, it is actually negatively charged particles moving in the opposite direction. The effect is the same.)

It is important to realize that every part of the circuit is chock full of both kinds of charge. The wire, the resistor, everything is incredibly crowded with both positive and negative charge. One kind of charge can move against the background of the other. Now the electric field in the resistor pushes the positive charge in the resistor in the direction from the higher-potential terminal toward the lower-potential terminal.



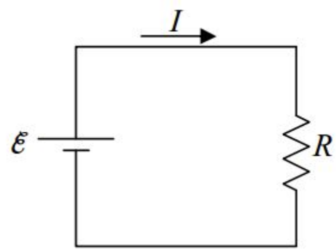
Pushing positive charge onto the lower-potential wire would tend to raise the potential of the lower-potential wire and leave the upper end of the resistor with a negative charge. I say "would" because any tendency for a change in the relative potential of the two wires is immediately compensated for by the seat of EMF. Remember, that's what the seat of EMF does, it maintains a constant potential difference between the wires. To do so in the case at hand, the seat of EMF must pull some positive charges from the lower-potential wire and push them onto the higher-potential wire. Also, any tendency of the upper end of the resistor to become negative immediately results in an attractive force on the positive charge in the higher-potential wire. This causes that positive charge to move down into the resistor in the place of the charge that just moved along the resistor toward the lower-potential wire. The net effect is a continual movement of charge, clockwise around the loop, as we view it in the diagram, with the net amount of charge in any short section of the circuit never changing. Pick a spot anywhere in the circuit. Just as fast as positive charge moves out of that spot, more positive charge from a neighboring spot moves in. What we have is this whole crowded mass of positive charge carriers moving clockwise around the loop, all because of the electric field in the resistor, and the EMF's "insistence" on maintaining a constant potential difference between the wires.

Now draw a dotted line across the path of the circuit, at any point in the circuit, as indicated below.



The rate at which charge crosses that line is the charge flow rate at that point (the point at which you drew the dotted line) in the circuit. The charge flow rate, how many coulombs-of charge per-second are crossing that line is called the electric current at that point. In the case at hand, because the whole circuit consists of a single loop, the current is the same at every point in the circuit—it doesn't matter where you "draw the line." The symbol that one typically uses to represent the value of the current is I.

In analyzing a circuit, if the current variable is not already defined for you, you should define it by drawing an arrow on the circuit and labeling it I or I with a subscript.



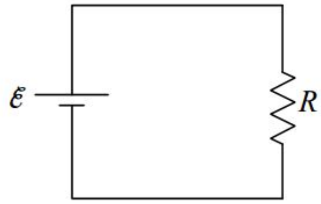
The units for current are coulombs per second (C/s). That combination of units is given a name: the ampere, abbreviated A.

$$1A = 1\frac{C}{s}$$

Now about that resistor: In our positive charge carrier model, the charged particles that are free to move in the resistor experience a force exerted on them by the electric field, in the direction of the electric field. As a result, they experience acceleration. But, the background material making up the substance of which the charge carriers are a part, exerts a velocity dependent retarding force on the charge carriers. The faster they go, the bigger the retarding force. Upon completion of the circuit (making that final wire-to-terminal connection), the charge carriers in the resistor, almost instantaneously, reach a terminal velocity at which the retarding force on a given charge carrier is just as great as the force exerted by the electric field on that charge carrier. The value of the terminal velocity, along with the number-of-charge-carriers-per-volume in the resistor, and the cross-sectional area of the poorly-conducting material making up the resistor, determine the charge flow rate, the current, in the resistor. In the simple circuit under consideration, the charge flow rate in the resistor is the charge flow rate everywhere in the circuit.

The value of the terminal velocity itself depends on how strong the electric field is, and, on the nature of the retarding force. The nature of the retarding force depends on what kind of material the resistor is made of. One kind of material will result in a bigger terminal velocity for the same electric field as another kind of material. Even with one kind of material, there's the question of how the retarding force depends on the velocity. Is it proportional to the square of the velocity, the log of the velocity, or what? Experiment shows that in an important subset of materials, over certain ranges of the terminal velocity, the retarding force is proportional to the velocity itself. Such materials are said to obey Ohm's law and are referred to as ohmic materials.

Consider the resistor in the simple circuit we have been dealing with.



If you double the voltage across the resistor (by using a seat of EMF that maintains twice the potential difference between its terminals as the original seat of EMF) then you double the electric field in the resistor. This doubles the force exerted on each charge carrier. This means that, at the terminal velocity of any charge carrier, the retarding force has to be twice as great. (Since, upon making that final circuit connection, the velocity of the charge carriers increases until the retarding force on each charge carrier is equal in magnitude to the applied force.) In an ohmic material, if the retarding force is twice as great, then the velocity is twice as great. If the velocity is twice as great, then the charge flow rate, the electric current, is twice as great. So, doubling the voltage across the resistor doubles the current. Indeed, for a resistor that obeys Ohm's Law, the current in a resistor is directly proportional to the voltage across the resistor.

Summarizing: When you put a voltage across a resistor, there is a current in that resistor. The ratio of the voltage to the current is called the resistance of the resistor.

$$R = \frac{V}{I}$$

This definition of resistance is consistent with our understanding that the resistance of a resistor is a measure of how lousy a conductor it is. Check it out. If, for a given voltage across the resistor, you get a tiny little current (meaning the resistor is a very poor conductor), the value of resistance $R = \frac{V}{I}$ with that small value of current in the denominator, is very big. If, on the other hand, for the same voltage, you get a big current (meaning the resistor is a good conductor), then the value of resistance $R = \frac{V}{I}$ is small.

If the material of which the resistor is made obeys Ohm's Law, then the resistance R is a constant, meaning that its value is the same for different voltages. The relation $R = \frac{V}{I}$ is typically written in the form $V = IR$.

Ohm's Law:

The resistance R, in the expression $V = IR$, is a constant.

Ohm's Law is good for resistors made of certain materials (called ohmic materials) over a limited range of voltages.

Units of Resistance

Given that the resistance of a resistor is defined as the ratio of the voltage across that resistor to the resulting current in that resistor,

$$R = \frac{V}{I}$$

it is evident that the unit of resistance is the volt per ampere, $\frac{V}{A}$. This combination unit is given a name. We call it the ohm, abbreviated Ω, the Greek letter upper-case omega.

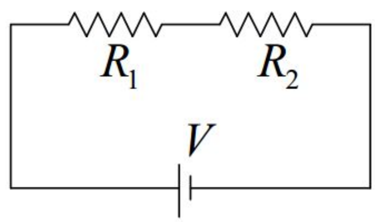
$$1\Omega = 1\frac{\text{volt}}{\text{ampere}}$$

B10: Resistors in Series and Parallel; Measuring I & V

The analysis of a circuit involves the determination of the voltage across, and the current through, circuit elements in that circuit. A method that I call "the method of ever simpler circuits" can be used to simplify the analysis of many circuits that have more than one resistor. The method involves the replacement of a combination of resistors with a single resistor, carefully chosen so that the replacement does not change the voltage across, nor the current through, the other circuit elements in the circuit. The resulting circuit is easier to analyze, and, the results of its analysis apply to the original circuit. Because the single carefully-chosen resistor has the same effect on the rest of the circuit as the original combination of resistors, we call the single resistor the equivalent resistance of the combination, or, simply, the equivalent resistor.

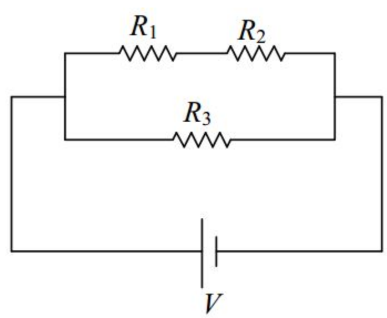
Resistors in Series

One combination of resistors that can be replaced with a single effective resistor is a series combination of resistors. Two two-terminal circuit elements in a circuit are in series with each other when one end of one is connected with one end of the other with nothing else connected to the connection. For instance, R_1 and R_2 in the following circuit are in series with each other.

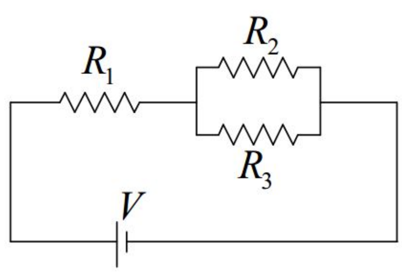


From our viewpoint, the right end of R_1 is connected to the left end of R_2 and nothing else is connected to the point in the circuit where they are connected.

R_1 and R_2 in the following circuit are also in series with each other:



But, R_1 and R_2 in the following circuit are not in series with each other:



While it is true that the right end of R_1 is connected to the left end of R_2, it is not true that "nothing else is connected to the connection." Indeed, the left end of R_3 is connected to the point in the circuit at which R_1 and R_2 are connected to each other.

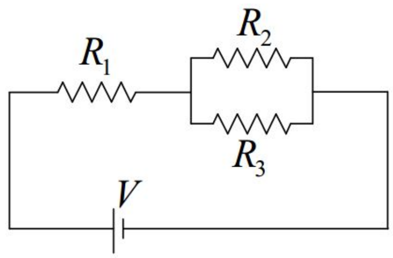
In implementing the method of ever simpler circuits, the plan is to replace a combination of resistors that are in series with each other with a single, well-chosen equivalent resistor. The question is, what value must the resistance of the single resistor be in order for it to be equivalent to the set of series resistors it replaces? For now, we simply give you the result. The derivation will be provided in the next chapter.

The equivalent resistance of resistors in series is simply the sum of the resistances.

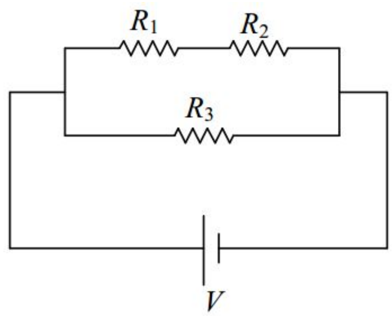
$$R_S = R_1 + R_2 + R_3 + \ldots$$

Resistors in Parallel

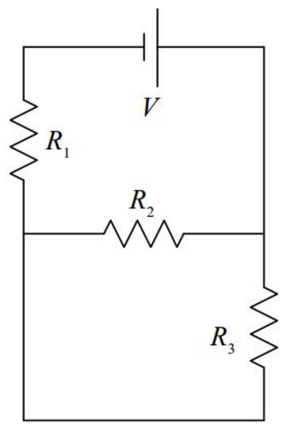
Circuit elements are in parallel with each other if they are connected together (by nothing but "perfect" conductor) at both ends. So, for instance, R_2 and R_3 in the following circuit are in parallel with each other.



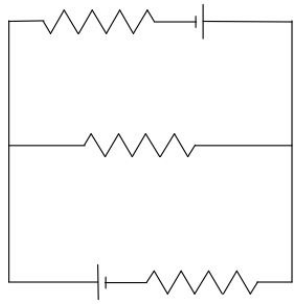
On the other hand, R_1 and R_3 in the following circuit are not in parallel with each other.



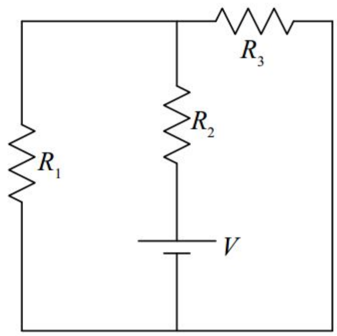
Resistors R_2 and R_3 in the following circuit are in parallel with each other:



But, none of the resistors in the following circuit are in parallel with each other:



whereas R_1 and R_3 in the following circuit are in parallel with each other:



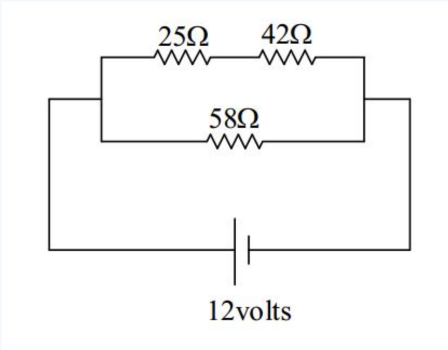
So what is the equivalent resistor for resistors in parallel? Here we provide the result. We save the derivation for the next chapter.

The equivalent resistance of resistors in parallel is the reciprocal of the sum of the reciprocals of the resistances of the resistors making up the parallel combination:

$$R_P = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots..}$$

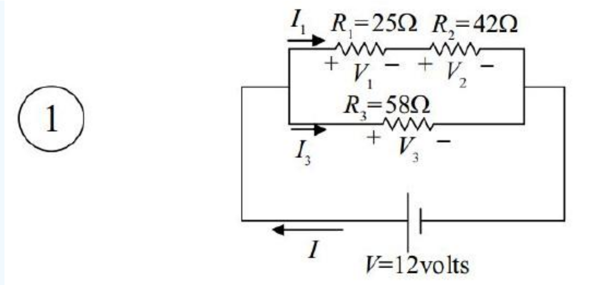
✔ Example

Find the voltage across, and the current through, each of the circuit elements in the diagram below.



Solution

First we add some notation to the diagram to define our variables (do not omit this step):

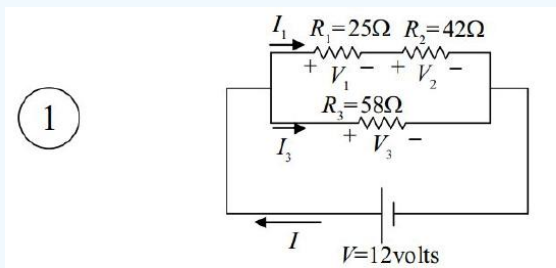


The $+$ and $-$ signs on the resistors (indicating the high potential side and the low potential side of each resistor), are an important part of the definition of the voltages. If you are given values, and the value you calculate for V1 turns out to be positive, e.g. $+5.0$ volts, then the reader of your solution knows that the potential of the left end of R_1 is 5.0 volts higher than that of the right end. But, if the value that you calculate for V_1 is negative, e.g. -5.0 volts, then the reader knows that the potential of the left end of R_1 is 5.0 volts lower than that of the right end.

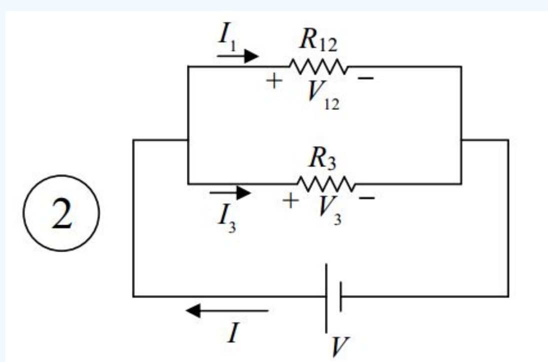
The "$+$" and "–" labels on the resistors must be consistent with the current direction. In fact, one first draws and labels the current arrows, and then puts the "$+$" on the end of the resister that the current enters (and the "–" on the other end).

Next we draw a sequence of circuits. Each new diagram includes an equivalent resistor in place of one series combination or one parallel combination. (Do not omit any diagrams, and, do not replace anything more than a single series combination or a single parallel combination of resistors in any one step.) As you draw each circuit, calculate the value of the equivalent resistance.

First, we copy the diagram from the preceding page.



Next, we replace the series combination of R_1 and R_2 with the equivalent resistor R_{12}.



$$R_{12} = R_1 + R_2$$

$$R_{12} = 25\Omega + 42\Omega$$

$$R_{12} = 67\Omega$$

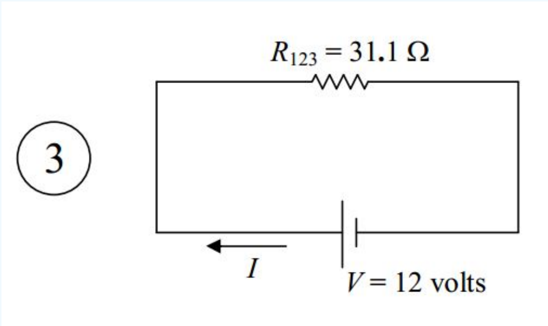
Finally, we replace the parallel combination of R_{12} and R_3 with the equivalent resistor R_{123}.

$$R_{123} = \frac{1}{\frac{1}{R_{12}} + \frac{1}{R_3}}$$

$$R_{123}=\frac{1}{\frac{1}{67\Omega}+\frac{1}{58\Omega}}$$

$$R_{123}=31.1\Omega$$

Now we analyze the simplest circuit, the one I have labeled "3" above.



One of the most common mistakes that folks make in analyzing circuits is using any old voltage in $V = IR$. You have to use the voltage across the resistor. In analyzing circuit 3, however, we can use the one voltage in the diagram because the voltage across the seat of EMF is the voltage across the resistor. The terminals of the resistor are connected to the same two conductors that the terminals of the seat of EMF are connected to. Thus,

$$V = IR_{123}$$

$$I=\frac{V}{R_{123}}$$

$$I=\frac{12\text{volts}}{31.1\Omega}$$

$$I = 0.386A$$

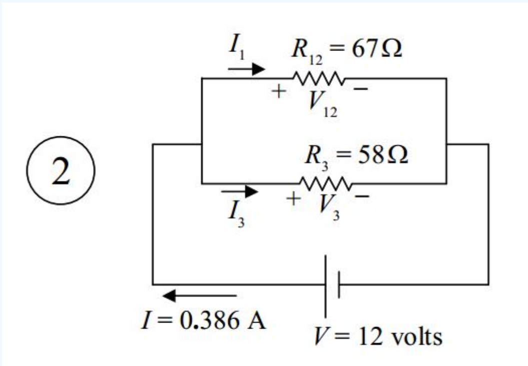
At this point, we've got two of the answers. The voltage across the seat of EMF was asked for, but it is also given, so we don't have to show any work for it. And now we have the current through the seat of EMF.

$$V = 12 \text{ volts}$$

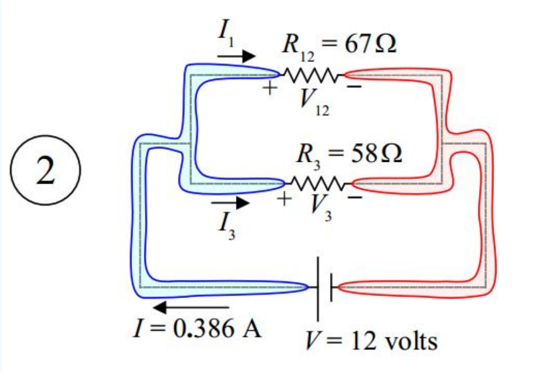
$$I = 0.39 \text{ amperes}$$

Note that the arrow labeled I in our diagram is part of our answer. It tells the reader what I means, including the direction of charge flow for a positive value of I.

Our next step is to take the information that we have learned here, to our next more complicated circuit. That would be the one I labeled "2" above.



There are only two wires (conductors) in this circuit. I am going to highlight them in order to make my next point:



Highlighting the conductors makes it obvious that the voltage across R_{12} is the same as the voltage across the seat of EMF because, in both cases, the voltage is the potential difference between one and the same pair of conductors. Likewise, the voltage across R_3 is the same as the voltage across the seat of EMF. Hence, we have,

$$V_{12} = V$$

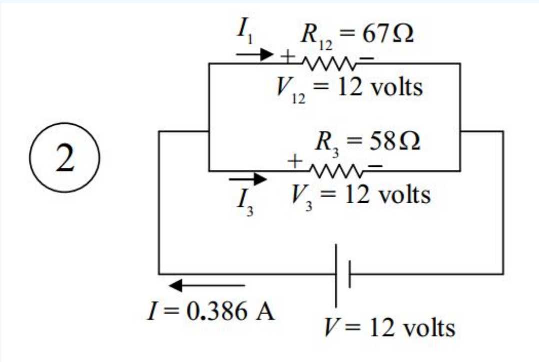
$$V_{12} = 12 \text{ volts}$$

and,

$$V_3 = V$$

$$V_3 = 12 \text{ volts}$$

The last value is one of our answers. We were supposed to find V_3. Now that we know the voltage across R_3, we can use it in $V = IR$ to get I_3.



For resistor R_3, we have:

$$V_3 = I_3 R_3$$

$$I_3 = \frac{V_3}{R_3}$$

$$I_3 = \frac{12 \text{ volts}}{58\Omega}$$

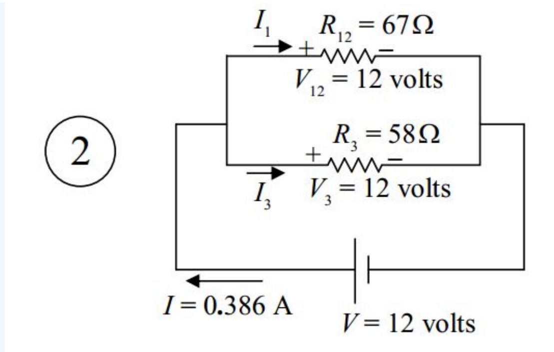
$$I_3 = 0.207\ A$$

The voltage and current through resistor R_3 are answers to the problem:

$$V_3 = 12 \text{volts}$$

$$I_3 = 0.21 \text{ amperes}$$

Now let’s get the current through R_{12}. I’ve labeled that current I_1 in diagram 2.



For resistor R_{12}, we have:

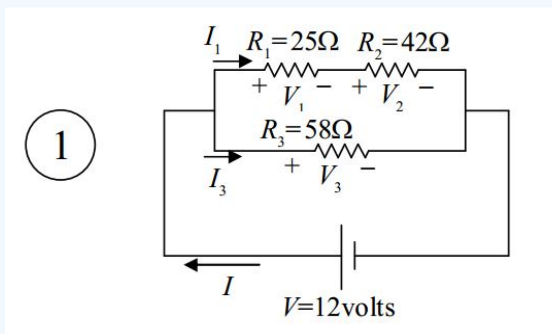
$$V_{12} = I_1 R_{12}$$

$$I_1 = \frac{V_{12}}{R_{12}}$$

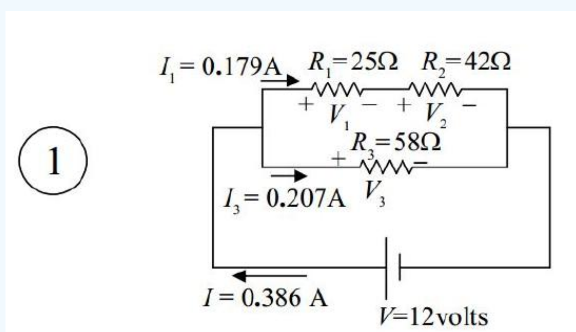
$$I_1 = \frac{12 \text{ volts}}{67\ \Omega}$$

$$I_1 = 0.179\ A$$

Now it is time to take what we have learned here up to the next more complicated circuit (which is the original circuit).



I copy that here with the values of the current included:



It is clear from this diagram that the current I_1 that we just found (the current through R_{12}) is the current through R_1, and, it is the current through R_2.

$$I_2 = I_1$$

$$I_2 = 0.179\ A$$

These are answers to the problem.

With the current through R_1 known, we can now solve for V_1:

$$V_1 = I_1 R_1$$

$$V_1 = 0.179\ (25\Omega)$$

$$V_1 = 4.5 \text{ volts}$$

Thus, our answers for resistor R_1 are:

$$V_1 = 4.5 \text{ volts}$$

$$I_1 = 0.18 \text{ amperes}$$

And, with the current through R_2 known, we can solve for V_2:

$$V_2 = I_2 R_2$$

$$V_2 = 0.179\ A(42\Omega)$$

$$V_2 = 7.5 \text{ volts}$$

Thus, our answers for resistor R_2 are:

$$V_2 = 7.5 \text{ volts}$$

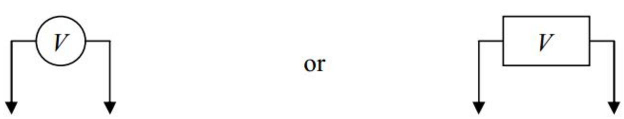
$$I_2 = 0.18 \text{ amperes}$$

How to Connect a Voltmeter in a Circuit

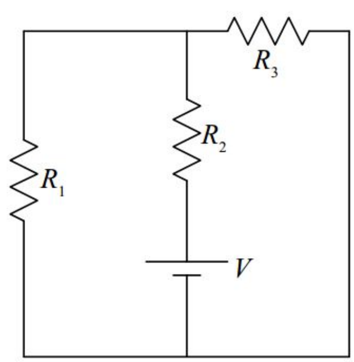
As discussed earlier in this book, a voltmeter is a device used for measuring the potential difference between two different points in space. In a circuit, we use it to measure the potential difference between two conductors (wires) in the circuit. When you do that, the voltmeter becomes a two-terminal circuit element of the circuit. The ideal voltmeter, as a circuit element, can be considered to be a resistor with infinite resistance. As such, it has no effect on the circuit. This is good. We don't want the measuring device to change the value of that which you are trying to measure.

A voltmeter consists of a box with two wires coming out of it. Typically, each wire ends in a metal-tipped wand (called a probe) or some kind of metal clip. The box has a gauge on it which displays the potential difference between the two wires. Touch the tip of one wire to one point in the circuit and the tip of the other wire to another point in the circuit (being sure to establish good metal-to-metal contact at both points) and the voltmeter will display the potential difference (the voltage) between those two points in the circuit.

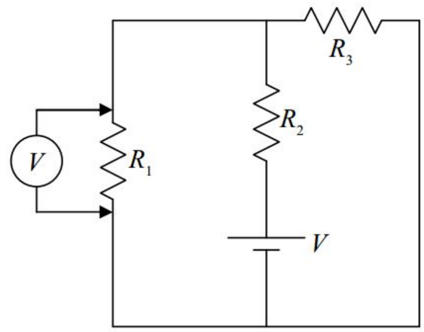
A typical manner of depicting a voltmeter in a circuit is to draw it as



To connect a voltmeter to measure the voltage across R_1 in the following circuit:



hook it up as indicated in the following diagram.

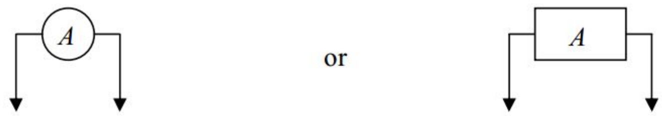


As far as its role as a circuit element (a side effect), the ideal voltmeter has as much effect on the circuit it is used on, as the air around the circuit has.

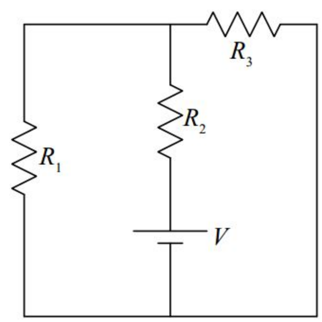
How to Connect an Ammeter in a Circuit

The ammeter, a device used to measure current, is a totally different beast. The ideal ammeter acts like a perfectly-conducting piece of wire that monitors the charge flow through itself. Connecting it in a circuit as you would a voltmeter (don't do it!) will drastically change the circuit (and could cause damage to the meter).

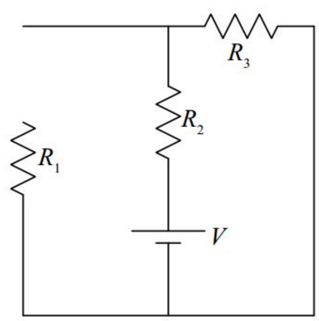
A typical manner of depicting an ammeter in a circuit is to draw it as



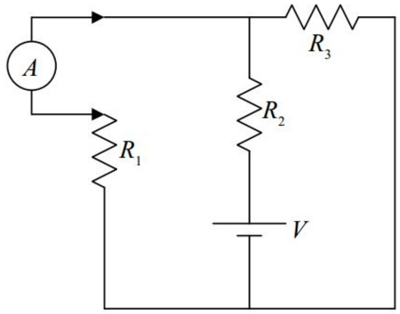
To connect an ammeter to measure the current in R_1 in the following circuit:



You have to first break the circuit,



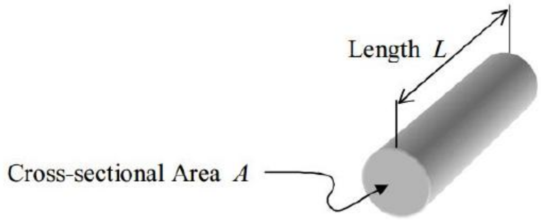
and then connect the ammeter in series with the circuit element whose current you wish to measure.



Remember, to measure current with an ammeter, some disassembly is required!

B11: Resistivity and Power

In chapter 9 we discussed resistors that conform to Ohm's Law. From the discussion, one could deduce that the resistance of such a resistor depends on the nature of the material of which the resistor is made and on the size and shape of the resistor. In fact, for resistors made out of a single kind of material, in the shape of a wire with a terminal at each end,



the resistance is given by:

$$R = \varrho \frac{L}{A} \tag{B11.1}$$

where:

- R is the resistance of the resistor as measured between the ends,
- ϱ is the resistivity of the substance of which the resistor is made,
- A is the cross-sectional area of the wire-shaped resistor, and
- L is the length of the resistor.

The values of resistivity for several common materials are provided in the following table:

Material	**Resisitivity ϱ**
Silver	$1.6 \times 10^{-8}\ \Omega \cdot m$
Copper	$1.7 \times 10^{-8}\ \Omega \cdot m$
Gold	$2.4 \times 10^{-8}\ \Omega \cdot m$
Aluminum	$3 \times 10^{-8}\ \Omega \cdot m$
Tungsten	$5.6 \times 10^{-8}\ \Omega \cdot m$
Nichrome	$1.0 \times 10^{-6}\ \Omega \cdot m$
Seawater	$0.25\ \Omega \cdot m$
Rubber	$1 \times 10^{13}\ \Omega \cdot m$
Glass	1×10^{10} to $1 \times 10^{14}\ \Omega \cdot m$
Quartz	5×10^{15} to $7.5 \times 10^{17}\ \Omega \cdot m$

In the expression $R = \varrho \frac{L}{A}$, the resistivity ϱ depends on the charge carrier density, that is, the number-of-charge-carriers-per-volume. The more charge carriers per volume, the smaller the resistance since, for a given velocity of the charge carriers, more of them will be passing any point along the length of the resistor every second for a given voltage across the resistor. The resistivity also depends on the retarding force factor. We said that the retarding force on each charge carrier is proportional to the velocity of that charge carrier.

$$\text{Retarding Force} = \text{-(factor) times (charge carrier velocity)}$$

(The minus sign is there because the retarding force is in the direction opposite that of the charge-carrier velocity.) The bigger the retarding force factor, the greater the resistivity of the material for which the factor applies.

The charge carrier density and the retarding force factor determine the value of ϱ. The effect of ϱ on the resistance is evident in the expression $R=\varrho \frac{L}{A}$. The bigger ϱ is, the greater the resistance is.

Why the factor of L in $R=\varrho \frac{L}{A}$? It's saying that the greater the length of the single-substance resistor in the shape of a wire, the greater the resistance of the resistor, all other things being equal (same substance, same cross-sectional area). It means, for instance, that if you have two resistors, identical in all respects except that one is twice as long as the other, and you put the same voltage across each of the resistors, you'll get half the current in the longer resistor. Why is that?

To get at the answer, we need to consider the electric field inside the wire-shaped resistor when we have a voltage V across the resistor. The thing is, the electric field inside the resistor is directed along the length of the resistor, and, it has the same magnitude everywhere along the length of the resistor. Evidence for this can be obtained by means of simple voltage measurements. Use a voltmeter to measure the potential difference $\Delta \varphi$ between two points on the resistor that are separated by a certain distance Δx, say 2 mm (measured along the length of the resistor) for instance. It turns out that no matter where along the length you pick the pair of points (separated from each other by the Δx), you always get the same voltage reading. Imagine (this part is a thought experiment) moving a positive test charge q_T that distanceΔx along the resistor from high potential toward low potential. No matter where along the length of the resistor you do that, the work done (by the electric field characterized by the potential) $q_T \Delta \varphi$ (calculated as the negative of the change of the potential energy of the test charge) is the same. The work, calculated as force times distance, is $q_T E \Delta x$. For that to be the same at every point along the length of the resistor, the electric field E has to have the same value everywhere along the length of the resistor. Furthermore, setting the two expressions for the work equal to each other yields:

$$q_T E \Delta x=q_T \Delta \varphi$$

$$E=\frac{\Delta \varphi}{\Delta x}$$

E being constant thus means that $\frac{\Delta \varphi}{\Delta x}$ is constant which means that a graph of φ vs. x is a straight line with slope $\frac{\Delta V}{\Delta x}$. But, in calculating that slope, since it is a straight line, we don't have to use a tiny little Δx. We can use the entire length of the resistor and the corresponding potential difference, which is the voltage V across the resistor. Thus,

$$E=\frac{V}{L}$$

where:

- E is the magnitude of the electric field everywhere in the single-substance wire-shaped resistor,
- V is the voltage across the resistor, and
- L is the length of the resistor.

This result ($E=\frac{V}{L}$) is profound in and of itself, but, if you recall, we were working on answering the question about why the resistance R, of a single-substance wire-shaped resistor, is proportional to the length of the resistor. We are almost there. The resistance is the ratio of the voltage across the resistor to the current in it. According to $E=\frac{V}{L}$, the longer the resistor, the weaker the electric field in the resistor is for a given voltage across it. A weaker E results in a smaller terminal velocity for the charge carriers in the resistor, which results in a smaller current. Thus, the longer the resistor, the smaller the current is; and; the smaller the current, the greater the voltage-to-current ratio is; meaning, the greater the resistance.

The next resistance-affecting characteristic in $R=\varrho \frac{L}{A}$ that I want to discuss is the area A. Why should that affect the resistance the way it does? Its presence in the denominator means that the bigger the cross-sectional area of the wire-shaped resistor, the smaller the resistance. Why is that?

If we compare two different resistors made of the same material and having the same length (but different cross-sectional areas) both having the same voltage across them, they will have the same electric field $E=\frac{V}{L}$ in them. As a result, the charge carriers will have the same velocity v. In an amount of time Δt,

$$L=v \Delta t$$

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

all the free-to-move charge carriers in either resistor will flow out the lower potential end of the resistor (while the same amount of charge flows in the higher potential end). This time Δt is the same for the two different resistors because both resistors have the same length, and the charge carriers in them have the same v. The number of charge carriers in either resistor is proportional to the volume of the resistor. Since the volume is given by volume = LA, the number of charge carriers in either resistor is proportional to the cross-sectional area A of the resistor. Since the number of charge carriers in either resistor, divided by the time Δt is the current in that resistor, this means that the current is proportional to the area.

If the current is proportional to the area, then the resistance, being the ratio of the voltage to the current, must be inversely proportional to the area. And so ends our explanation regarding the presence of the A in the denominator in the expression

$$R = \varrho \frac{L}{A}$$

Power

You were introduced to power in Volume I of this book. It is the rate at which work is done. It is the rate at which energy is transferred. And, it is the rate at which energy is transformed from one form of energy into another form of energy. The unit of power is the watt, W.

$$1W = 1\frac{\mathrm{J}}{\mathrm{s}}$$

In a case in which the power is the rate that energy is transformed from one form to another, the amount of energy that is transformed from time 0 to time t:

- if the power is constant, is simply the power times the duration of the time interval:

$$\mathrm{Energy} = \mathrm{Pt}$$

- if the power is a function of time, letting t′ be the time variable that changes from 0 to t, is:

$$\mathrm{Energy} = \int_0^t P(t')dt'$$

The Power of a Resistor

In a resistor across which there is a voltage V, energy is transformed from electric potential energy into thermal energy. A particle of charge q, passing through the resistor, loses an amount of potential energy qV but it does not gain any kinetic energy. As it passes through the resistor, the electric field in the resistor does an amount of work qV on the charged particle, but, at a same time, the retarding force exerted on the charged particle by the background material of the resistor, does the negative of that same amount of work. The retarding force, like friction, is a non-conservative force. It is exerted on the charge carrier when the charge carrier collides with impurities and ions (especially at sites of defects and imperfections in the structure of the material). During those collisions, the charge carriers impart energy to the ions with which they collide. This gives the ions vibrational energy which manifests itself, on a macroscopic scale, (early in the process) as an increase in temperature. Some of the thermal energy is continually transferred to the surroundings. Under steady state conditions, arrived at after the resistor has warmed up, thermal energy is transferred to the surroundings at the same rate that it is being transformed from electrical potential energy in the resistor.

The rate at which electric potential energy is converted to thermal energy in the resistor is the power of the resistor (a.k.a. the power dissipated by the resistor). It is the rate at which the energy is being delivered to the resistor. The energy conversion that occurs in the resistor is sometimes referred to as the dissipation of energy. One says that the resistor power is the rate at which energy is dissipated in the resistor. It's pretty easy to arrive at an expression for the power of a resistor in terms of circuit quantities. Each time a coulomb of charge passes through a resistor that has a voltage V across it, an amount of energy equal to one coulomb times V is converted to thermal energy. The current I is the number of coulombs-per-second passing through the resistor. Hence V times I is the number of joules-per-second converted to thermal energy. That's the power of the resistor. In short,

$$P = IV \tag{B11.2}$$

where:

- P is the power of the resistor. It is the rate at which the resistor is converting electrical potential energy into thermal energy. The unit of power is the watt. $1W = 1\frac{J}{s}$.
- I is the current in the resistor. It is the rate at which charge is flowing through the resistor. The unit of current is the ampere. $1A = 1\frac{C}{s}$.

- V is the voltage across the resistor. It is the amount by which the value of electric potential (the electric potential energy per charge) at one terminal of the resistor exceeds that at the other terminal. The unit of voltage is the volt. $1 \text{ volt} = 1\frac{J}{C}$.

The Power of a Seat of EMF

In a typical circuit, a seat of EMF causes positive charge carriers (in our positive-charge-carrier model) to go from a lower-potential conductor, through itself, to a higher-potential conductor. The electric field of the conductors exerts a force on the charge carriers inside the seat of EMF in the direction opposite to the direction in which the charge carriers are going. The charged particles gain electric potential energy in moving from the lower-potential terminal of the seat of EMF to the higher-potential terminal. Where does that energy come from?

In the case of a battery, the energy comes from chemical potential energy stored in the battery and released in chemical reactions that occur as the battery moves charge from one terminal to the other. In the case of a power supply, the power supply, when plugged into a wall outlet and turned on, becomes part of a huge circuit including transmission wires extending all the way back to a power plant. At the power plant, depending on the kind of power plant, kinetic energy of moving water, or thermal energy used to make steam to turn turbines, or chemical potential energy stored in wood, coal, or oil; is converted to electric potential energy. Whether it is part of a battery, or a part of a power supply, the seat of EMF converts energy into electric potential energy. It keeps one of its terminals at a potential ε higher than the other terminal. Each time it moves a coulomb of charge from the lower potential terminal to the higher potential terminal, it increases the potential energy of that charge by one coulomb times ε. Since the current I is the number of coulombs per second that the seat of EMF moves from one terminal to the other, the power, the rate at which the seat of EMF delivers energy to the circuit, is given by:

$$P = I\varepsilon$$

Recall that it is common to use the symbol V (as well as ε) to represent the voltage across a seat of EMF. If you use V, then the power of the seat of EMF is given by:

$$P = IV$$

where:

- P is the rate at which a seat of EMF delivers energy to a circuit,
- I is the current in the seat of EMF (the rate at which charge flows through the seat of EMF), and
- V is the voltage across the seat of EMF.

This is the same expression as the expression for the power of a resistor (Equation B11.2).

B12: Kirchhoff's Rules, Terminal Voltage

There are two circuit-analysis laws that are so simple that you may consider them "statements of the obvious" and yet so powerful as to facilitate the analysis of circuits of great complexity. The laws are known as Kirchhoff's Laws. The first one, known both as "Kirchhoff's Voltage Law" and "The Loop Rule" states that, starting on a conductor, if you drag the tip of your finger around any loop in the circuit back to the original conductor, the sum of the voltage changes experienced by your fingertip will be zero. (To avoid electrocution, please think of the finger dragging in an actual circuit as a thought experiment.)

Kirchhoff's Voltage Law (a.k.a. the Loop Rule)

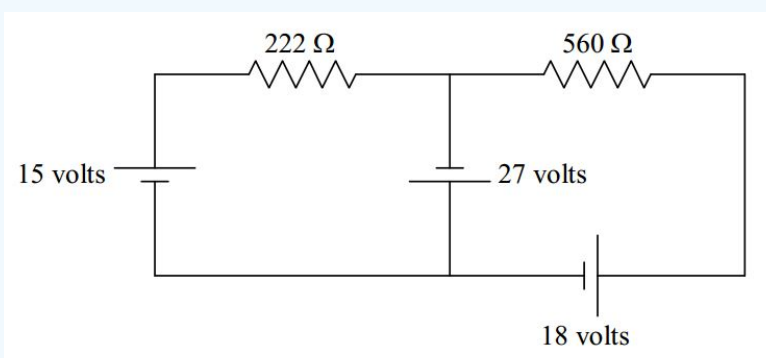
To convey the idea behind Kirchhoff's Voltage Law, I provide an analogy. Imagine that you are exploring a six-story mansion that has 20 staircases. Suppose that you start out on the first floor. As you wander around the mansion, you sometimes go up stairs and sometimes go down stairs. Each time you go up stairs, you experience a positive change in your elevation. Each time you go down stairs, you experience a negative change in your elevation. No matter how convoluted the path of your explorations might be, if you again find yourself on the first floor of the mansion, you can rest assured that the algebraic sum of all your elevation changes is zero.

To relate the analogy to a circuit, it is best to view the circuit as a bunch of conductors connected by circuit elements (rather than the other way around as we usually view a circuit). Each conductor in the circuit is at a different value of electric potential (just as each floor in the mansion is at a different value of elevation). You start with your fingertip on a particular conductor in the circuit, analogous to starting on a particular floor of the mansion. The conductor is at a particular potential. You probably don't know the value of that potential any more than you know the elevation that the first floor of the mansion is above sea level. You don't need that information. Now, as you drag your finger around the loop, as long as you stay on the same conductor, your fingertip will stay at the same potential. But, as you drag your fingertip from that conductor, through a circuit element, to the next conductor on your path, the potential of your fingertip will change by an amount equal to the voltage across the circuit element (the potential difference between the two conductors). This is analogous to climbing or descending a flight of stairs and experiencing a change in elevation equal to the elevation difference between the two floors.

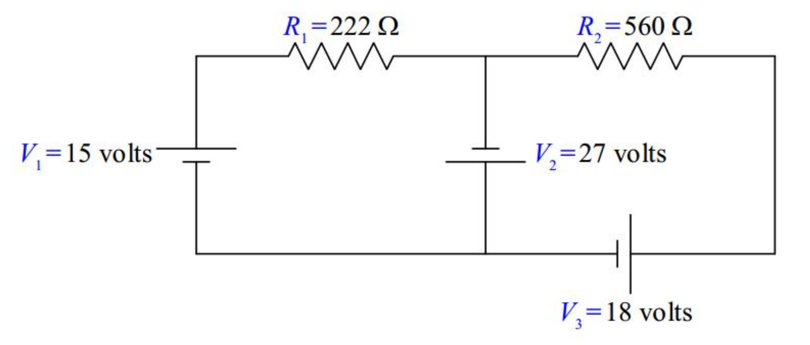
If you drag your fingertip around the circuit in a loop, back to the original conductor, your finger is again at the potential of that conductor. As such, the sum of the changes in electric potential experienced by your finger on its traversal of the loop must be zero. This is analogous to stating that if you start on one floor of the mansion, and, after wandering through the mansion, up and down staircases, you end up on the same floor of the mansion, your total elevation change is zero. In dragging your finger around a closed loop of a circuit (in any direction you want, regardless of the current direction) and adding each of the voltage changes to a running total, the critical issue is the algebraic sign of each voltage change. In the following example we show the steps that you need to take to get those signs right, and to prove to the reader of your solution that they are correct.

Example

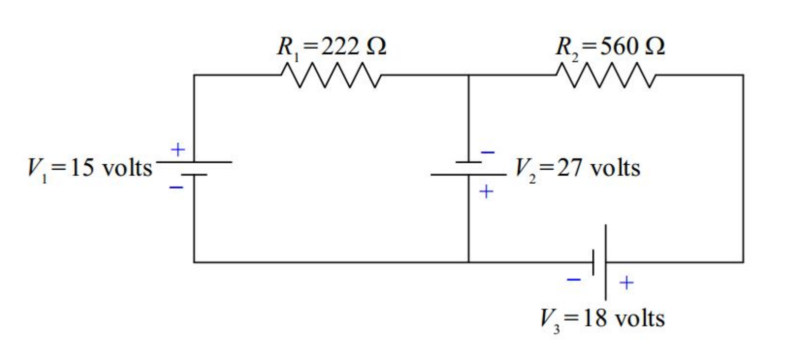
Find the current through each of the resistors in the following circuit.



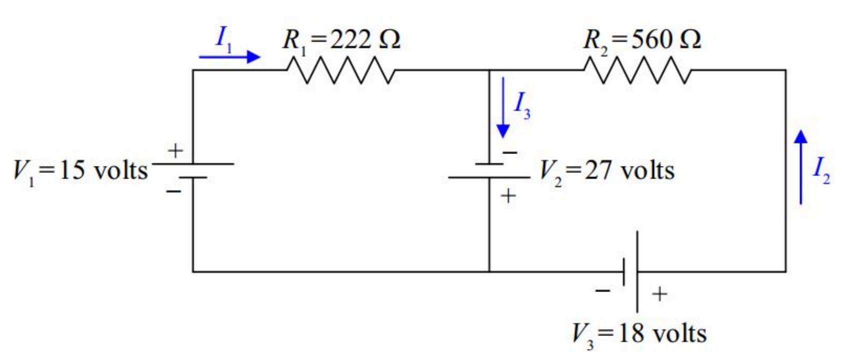
Before we get started, let's define some names for the given quantities:



Each two-terminal circuit element has one terminal that is at a higher potential than the other terminal. The next thing we want to do is to label each higher potential terminal with a "+" and each lower-potential terminal with a "−". We start with the seats of EMF. They are trivial. By definition, the longer parallel line segment, in the symbol used to depict a seat of EMF, is at the higher potential.



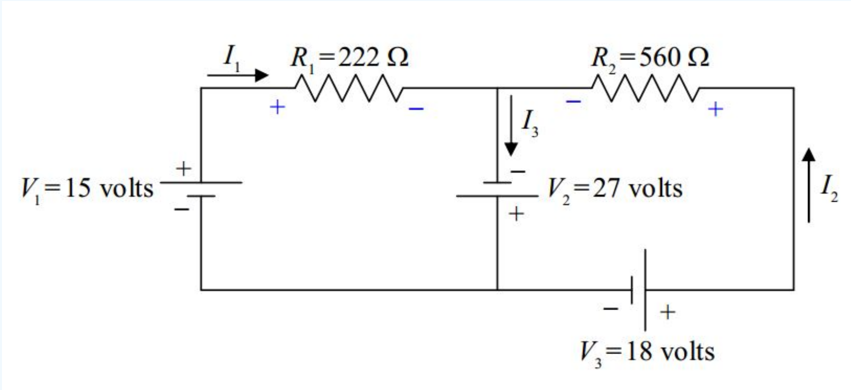
Next we define a current variable for each "leg" of the circuit. A "leg" of the circuit extends from a point in the circuit where three or more wires are joined (called a junction) to the next junction. All the circuit elements in any one leg of the circuit are in series with each other, so, they all have the same current through them.



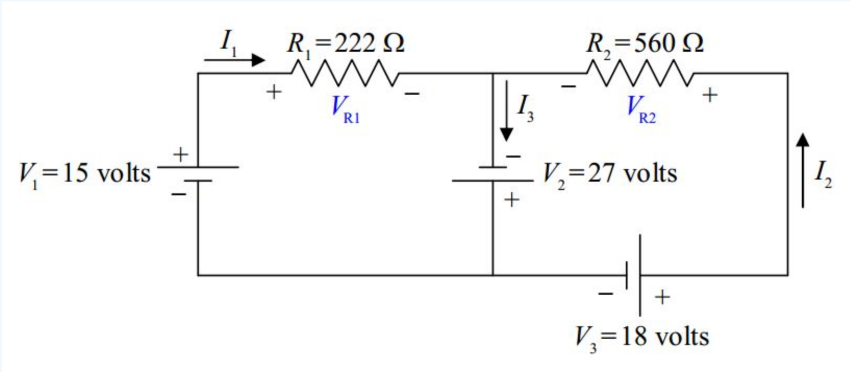
Note

In defining your current variables, the direction in which you draw the arrow in a particular leg of the circuit, is just a guess. Don't spend a lot of time on your guess. It doesn't matter. If the current is actually in the direction opposite that in which your arrow points, you will simply get a negative value for the current variable. The reader of your solution is responsible for looking at your diagram to see how you have defined the current direction and for interpreting the algebraic sign of the current value accordingly.

Now, by definition, the current is the direction in which positive charge carriers are flowing. The charge carriers lose electric potential energy when they go through a resistor, so, they go from a higher-potential conductor, to a lower-potential conductor when they go through a resistor. That means that the end of the resistor at which the current enters the resistor is the higher potential terminal ($+$), and, the end at which the current exits the resistor is the lowerpotential terminal ($-$) of the resistor.

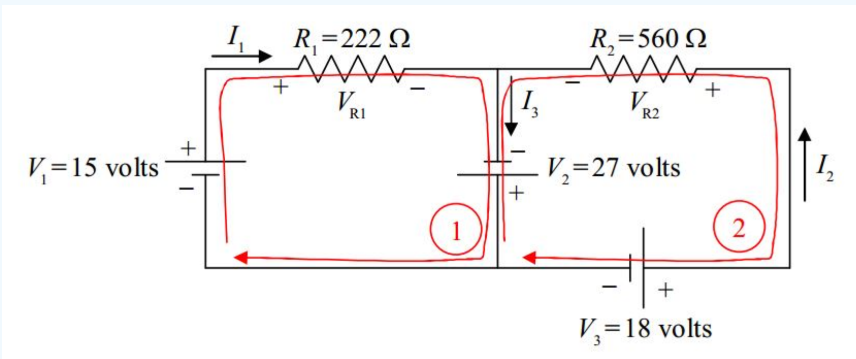


Now let's define some variable names for the resistor voltages:



Note that the $+$ and – signs on the resistors are important parts of our definitions of V_{R1} and V_{R2}. If, for instance, we calculate V_{R1} to have a positive value, then, that means that the left (as we view it) end of V_{R1} is at a higher potential than the right end (as indicated in our diagram). If V_{R1} turns out to be negative, that means that the left end of R_1 is actually at a lower potential than the right end. We do not have to do any more work if V_{R1} turns out to be negative. It is incumbent upon the reader of our solution to look at our circuit diagram to see what the algebraic sign of our value for V_{R1} means.

With all the circuit-element terminals labeled "+" for "higher potential" or "–" for "lower potential," we are now ready to apply the Loop Rule. I'm going to draw two loops with arrowheads. The loop that one draws is not supposed to be a vague indicator of direction but a specific statement that says, "Start at this point in the circuit. Go around this loop in this direction, and, end at this point in the circuit." Also, the starting point and the ending point should be the same. In particular, they must be on the same conductor. (Never start the loop on a circuit element.) In the following diagram are the two loops, one labeled 1 and the other labeled 2.



Now we write KVL 1 to tell the reader that we are applying the Loop Rule (Kirchhoff's Voltage Law) using loop 1, and transcribe the loop equation from the circuit diagram:

$$KVL1 + V_1 - V_{R1} + V_2 = 0$$

The equation is obtained by dragging your fingertip around the exact loop indicated and recording the voltage changes experienced by your fingertip, and then, remembering to write "$=0$." Starting at the point on the circuit closest to the tail of the loop 1 arrow, as we drag our finger around the loop, we first traverse the seat of EMF, V_1. In traversing V_1 we go from lower potential $(-)$ to higher potential $(+)$. That means that the finger experiences a positive change in potential, hence, V_1 enters the equation with a positive sign. Next we come to resistor R_1. In traversing R_1 we go from higher potential $(+)$ to lower potential $(-)$. That's a negative change in potential. Hence, V_{R1} enters our loop equation with a negative sign. As we continue our way about the loop we come to the seat of EMF V_2 and go from lower potential $(-)$ to higher potential $(+)$ as we traverse it. Thus, V_2 enters the loop equation with a positive sign. Finally, we arrive back at the starting point. That means that it is time to write "$=0$."

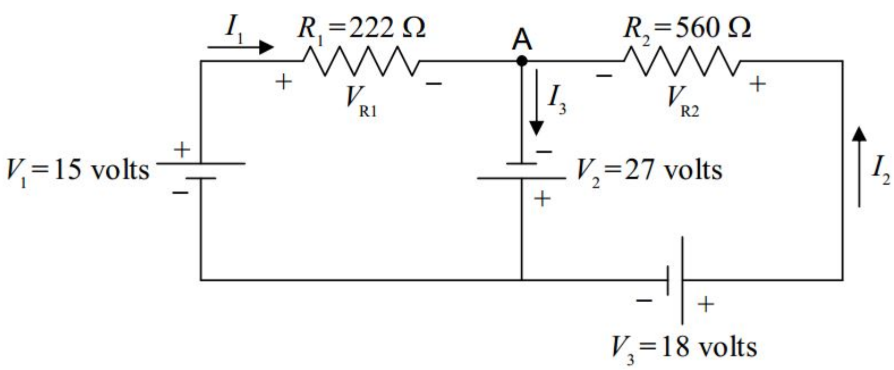
We transcribe the second loop equation in the same fashion:

$$KVL2 - V_2 + V_{R2} - V_3 = 0$$

With these two equations in hand, and knowing that $V_{R1} = I_1 R_1$ and $V_{R2} = I_2 R_2$, the solution to the example problem is straightforward. (We leave it as an exercise for the reader.) It is now time to move on to Kirchhoff's other law.

Kirchhoff's Current Law (a.k.a. the Junction Rule)

Kirchhoff's junction rule is a simple statement of the fact that charge does not pile up at a junction. (Recall that a junction is a point in a circuit where three or more wires are joined together.) I'm going to state it two ways and ask you to pick the one you prefer and use that one. One way of stating it is to say that the net current into a junction is zero. Check out the circuit from the example problem:



In this copy of the diagram of that circuit, I put a dot at the junction at which I wish to apply Kirchhoff's Current Law, and, I labeled that junction "A."

Note that there are three legs of the circuit attached to junction A. In one of them, current I_1 flows toward the junction. In another, current I_2 flows toward the junction. In the third leg, current I_3 flows away from the junction. A current away from the junction counts as the negative of that value of current, toward the junction. So, applying Kirchhoff's Current Law in the form, "The net current into any junction is zero," to junction A yields:

$$\frac{\text{KCL A}}{I_1} + I_2 - I_3 = 0$$

Note the negative sign in front of I_3. A current of $-I_3$ into junction A is the same thing as a current of I_3 out of that junction, which is exactly what we have.

The other way of stating Kirchhoff's Current Law is, "The current into a junction is equal to the current out of that junction." In this form, in applying Kirchhoff's Current Law to junction A in the circuit above, one would write:

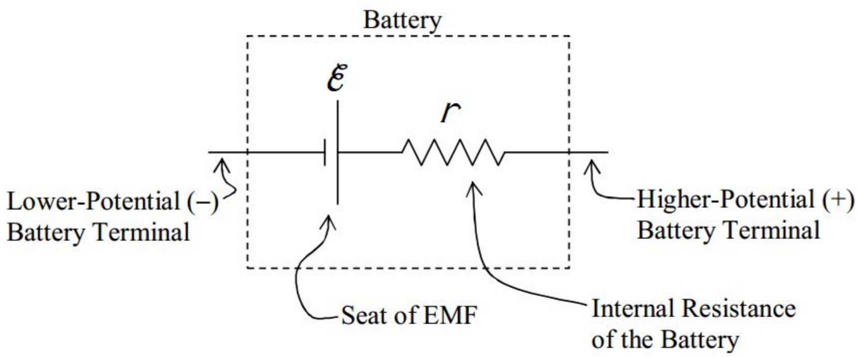
$$\frac{\text{KCL A}}{I_1} + I_2 = I_3$$

Obviously, the two results are the same.

Terminal Voltage – A More Realistic Model for a Battery or DC Electrical Power Source

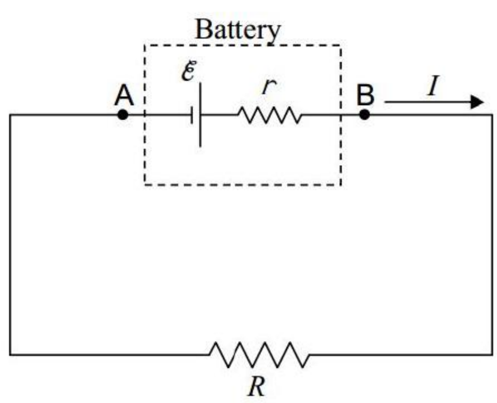
Our model for a battery up to this point has been a seat of EMF. I said that a seat of EMF can be considered to be an ideal battery. This model for a battery is good as long as the battery is fairly new and unused and the current through it is small. Small compared to what? How small? Well, small enough so that the voltage across the battery when it is in the circuit is about the same as it is when it is not in any circuit. How close to being the same? That depends on how accurate you want your results to be. The voltage across a battery decreases when you connect the battery in a circuit. If it decreases by five percent and you calculate values based on the voltage across the battery when it is in no circuit, your results will probably be about 5 off.

A better model for a battery is an ideal seat of EMF in series with a resistor. A battery behaves very much as if it consisted of a seat of EMF in series with a resistor, but, you can never separate the seat of EMF from the resistor, and if you open up a battery you will never find a resistor in there. Think of a battery as a black box containing a seat of EMF and a resistor. The resistor in this model is called the internal resistance of the battery.



The point at which the seat of EMF is connected to the internal resistance of the battery is inaccessible. The potential difference between the terminals of the battery is called the terminal voltage of the battery. When the battery is not part of a circuit, the terminal voltage is equal to the EMF. You can deduce this from the fact that when the battery is not part of a circuit, there can be no current through the resistor. If there is no current through the resistor than the two terminals of the resistor must be at one and the same value of electric potential. Thus, in the diagram above, the right end of the resistor is at the same potential as the high-potential terminal of the seat of EMF.

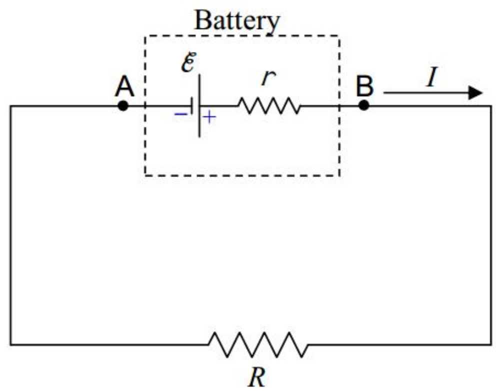
Now, let's put the battery in a circuit:



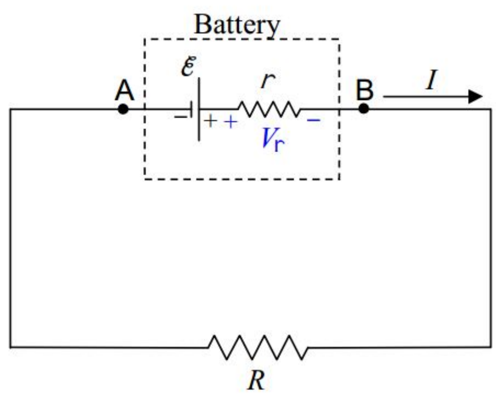
I've indicated the two points A and B on the circuit for communication purposes. The terminal voltage is the voltage from A to B (V_{AB}). If you trace the circuit, with your fingertip, from A to B, the terminal voltage (how much higher the potential is at B than it is at A) is just the sum of the voltage changes your finger experiences along the path. (Note that this time, we are not going all the way around a loop. We do not wind up on the same conductor upon which we started. So, the sum of the voltage changes from A to

B is not zero.) To sum the voltage changes from A to B, I will mark the terminals of the components between A and B with "+" for higher potential and "−" for lower potential.

First the seat of EMF: That's trivial. The shorter side of the EMF symbol is the lower potential (−) side and the longer side is the higher potential (+) side.



Now, for the internal resistance of the battery: The end of the internal resistance r that the current enters is the higher-potential (+) end, and, the end that it exits is the lower potential (−) end.



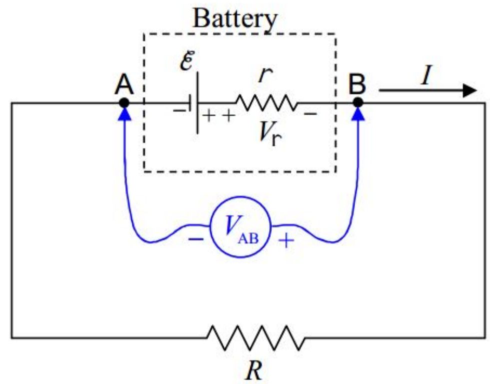
Note that I have also defined, in the preceding diagram, the variable V_r for the voltage across the internal resistance of the battery. Remember, to get the terminal voltage V_{AB} of the battery, all we have to do is to sum the potential changes that our fingertip would experience if we were to drag it from A to B in the circuit. (This is definitely a thought experiment because we can't get our fingertip inside the battery.)

$$V_{AB} = \varepsilon - V_r$$

$$V_{AB} = \varepsilon - Ir$$

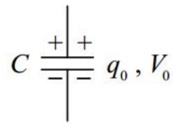
Note that, in the second line, I used the definition of resistance ($V = IR$) in the form $V_r = Ir$, to replace V_r with Ir.

We have been consistent, in this book, with the convention that a double subscript such as AB can be read "A to B" meaning, in the case at hand, that V_{AB} is the sum of the potential changes from A to B (rather than the other way around), in other words, that V_{AB} is how much higher the electric potential at point B is than the electric potential at point A. Still, there are some books out there that take V_{AB} (all by itself) to mean the voltage of A with respect to B (which is the negative of what we mean by it). So, for folks that may have used a different convention than you use, it is a good idea to diagrammatically define exactly what you mean by V_{AB}. Putting a voltmeter, labeled to indicate that it reads V_{AB}, and labeled to indicate which terminal is its "+" terminal and which is its "−" terminal is a good way to do this.

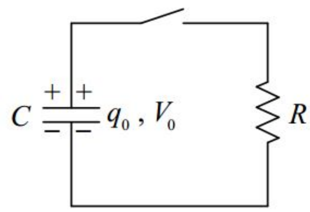

Battery
$\mathcal{E}$
r
A
B
I
V_r
V_{AB}
R

B13: RC Circuit

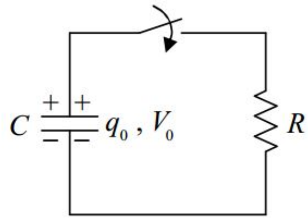
Suppose you connect a capacitor across a battery, and wait until the capacitor is charged to the extent that the voltage across the capacitor is equal to the EMF *V_0\) of the battery. Further suppose that you remove the capacitor from the battery. You now have a capacitor with voltage V_0 and charge q_o, where $q_o = CV_0$.



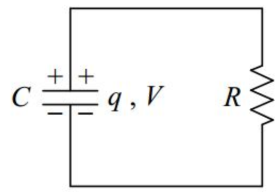
The capacitor is said to be charged. Now suppose that you connect the capacitor in series with an open switch and a resistor as depicted below.



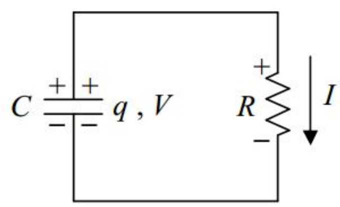
The capacitor remains charged as long as the switch remains open. Now suppose that, at a clock reading we shall call time zero, you close the switch.



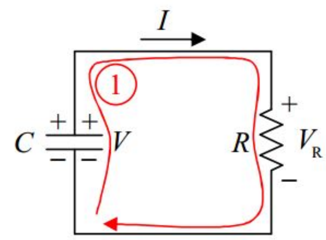
From time 0 on, the circuit is:



The potential across the resistor is now the same as the potential across the capacitor. This results in current through the resistor:



Positive charge flows from the upper plate of the capacitor, down through the resistor to the lower plate of the capacitor. The capacitor is said to be discharging. As the charge on the capacitor decreases; according to $q = CV$, which can be written $V = q/C$, the voltage across the capacitor decreases. But, as is clear from the diagram, the voltage across the capacitor is the voltage across the resistor. What we are saying is that the voltage across the resistor decreases. According to $V = IR$, which can be written as $I = V/R$, this means that the current through the resistor decreases. So, the capacitor continues to discharge at an ever decreasing rate. Eventually, the charge on the capacitor decreases to a negligible value, essentially zero, and the current dies down to a negligible value, essentially zero. Of interest is how the various quantities, the voltage across both circuit elements, the charge on the capacitor, and the current through the resistor depend on the time t. Let's apply the loop rule to the circuit while the capacitor is discharging:



$$KVL\ 1 + V - V_R = 0$$

Using $q = CV$ expressed as $V = \frac{q}{C}$ and $V_R = IR$, we obtain

$$\frac{q}{C} - IR = 0$$

I is the charge flow rate through the resistor, which is equivalent to the rate at which charge is being depleted from the capacitor (since the charge flowing through the resistor comes from the capacitor). Thus I is the negative of the rate of change of the charge on the capacitor:

$$I = -\frac{dq}{dt}$$

Substituting this ($I = -\frac{dq}{dt}$) into our loop rule equation ($\frac{q}{C} - IR = 0$) yields:

$$\frac{q}{C} + \frac{dq}{dt} R = 0$$

$$\frac{dq}{dt} = -\frac{1}{RC} q$$

Thus $q(t)$ is a function whose derivative with respect to time is itself, times the constant $-\frac{1}{RC}$. The function is essentially its own derivative. This brings the exponential function e^t to mind. The way to get that constant ($-\frac{1}{RC}$) to appear when we take the derivative of $q(t)$ with respect to t is to include it in the exponent. Try $q(t) = q_o\, e^{-\frac{1}{RC} t}$. Now, when you apply the chain rule for the function of a function you get $\frac{dq}{dt} = -\frac{1}{RC} q$ which is just what we wanted. Let's check the units. R was defined as $\frac{V}{I}$ meaning the ohm is a volt per ampere. And C was defined as $\frac{q}{V}$ meaning that the farad is a coulomb per volt. So the units of the product RC are:

$$[RC] = \frac{V}{A} \frac{\text{coulombs}}{V} = \frac{\text{coulombs}}{A} = \frac{\text{coulombs}}{\text{coulombs/s}} = s$$

So the exponent in $e^{-\frac{1}{RC} t}$ is unitless. That works. We can't raise e to something that has units. Now, about that q_o out front in $q = q_o e^{-\frac{1}{RC} t}$. The exponential evaluates to a unitless quantity. So we need to put the qo there to get units of charge. If you plug the value 0 in for the time in $q = q_o e^{-\frac{1}{RC} t}$ you get $q = q_o$. Thus, q_o is the initial value of the charge on the capacitor. One final point: The product RC is called the "RC time constant." The symbol τ is often used to represent that time constant. In other words,

$$\tau = RC \tag{B13.1}$$

where τ is also referred to as the RC time constant. In terms of τ, our expression for q becomes:

$$q = q_o e^{-\frac{t}{\tau}}$$

which we copy here for your convenience:

$$q = q_o e^{-\frac{t}{\tau}}$$

Note that when $t = \tau$, we have

$$q = q_o e^{-1}$$

$$q = \frac{1}{e} q_o$$

$\frac{1}{e}$ is $.368$ so τ is the time it takes for q to become 36.8 of its original value.

With our expression for q in hand, it is easy to get the expression for the voltage across the capacitor (which is the same as the voltage across the resistor, $V_C = V_R$) which we have been calling V. Substituting our expression $q = q_o e^{-\frac{1}{RC}t}$ into the defining equation for capacitance $q = CV$ solved for V,

$$V = \frac{q}{C}$$

yields:

$$V = \frac{q_o}{C} e^{-\frac{1}{RC}t}$$

But if q_o is the charge on the capacitor at time 0, then $q_o = CV_0$ where V_0 is the voltage across the capacitor at time 0 or:

$$\frac{q_o}{C} = V_o.$$

Substituting V_o for $\frac{q_o}{C}$ in $V = \frac{q_o}{C} e^{-\frac{1}{RC}t}$ above yields:

$$V = V_o e^{-\frac{t}{RC}} \tag{B13.2}$$

for both the voltage across the capacitor and the voltage across the resistor. From, the defining equation for resistance:

$$V = IR,$$

we can write:

$$I = \frac{V}{R}$$

Substituting our expression $V_o e^{-\frac{t}{RC}}$ in for V turns this equation $\left(I = \frac{V}{R}\right)$ into:

$$I = \frac{V_o e^{-\frac{t}{RC}}}{R}$$

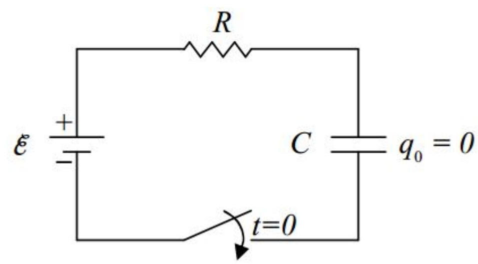
But, $\frac{V_o}{R}$ is just I_0 (from $V_0 = I_0 R$ solved for I_0), the current at the time 0, so:

$$I = I_0 e^{-\frac{t}{RC}} \tag{B13.3}$$

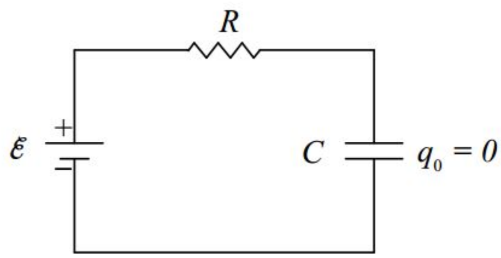
Summarizing, we note that all three of the quantities, V, I, and q decrease exponentially with time.

Charging Circuit

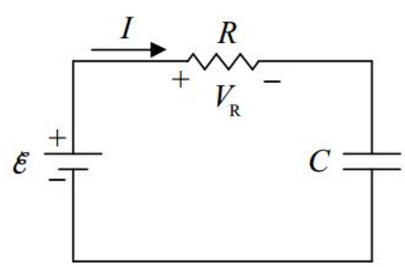
Consider the following circuit, containing an initially-uncharged capacitor, and



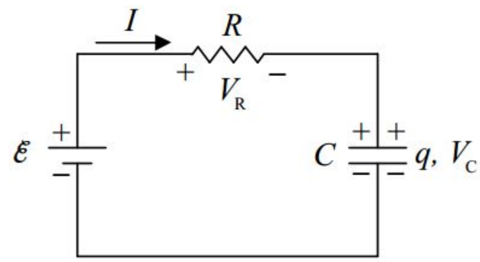
annotated to indicate that the switch is closed at time 0 at which point the circuit becomes:



Let's think about what will happen as time elapses. With no charge on the capacitor, the voltage across it is zero, meaning the potential of the right terminal of the resistor is the same as the potential of the lower-potential terminal of the seat of EMF. Since the left end of the resistor is connected to the higher-potential terminal of the seat of EMF, this means that at time 0, the voltage across the resistor is equivalent to the EMF ε of the seat of EMF. Thus, there will be a rightward current through the resistor.



The positive charge flowing through the resistor has to come from someplace. Where does it come from? Answer: The bottom plate of the capacitor. Also, charge can't flow through an ideal capacitor. So where does it go? It piles up on the top plate of the capacitor.



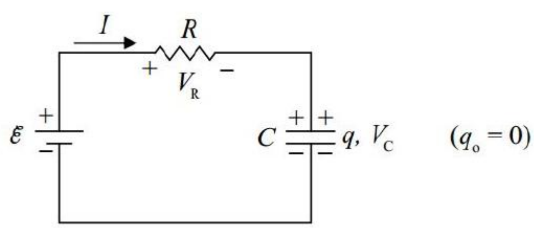
The capacitor is becoming charged. As it does, the voltage across the capacitor increases, meaning the potential of the right terminal of the resistor (relative to the potential of the lowerpotential terminal of the seat of EMF) increases. The potential of the left terminal of the resistor remains constant, as dictated by the seat of EMF. This means that the voltage across the resistor continually decreases. This, in turn; from $V_R = IR$, written as $I = V_R/R$, means that the current continually decreases. This occurs until there is so much charge on the capacitor that $V_c = \varepsilon$, meaning that $V_R = 0$ so $I = 0$.

Recapping our conceptual discussion:

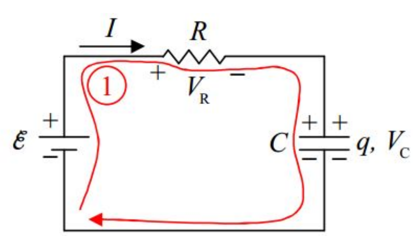
At time 0, we close the switch:

- The charge on the capacitor starts off at 0 and builds up to $q = C\varepsilon$ where ε is the EMF voltage.
- The capacitor voltage starts off at 0 and builds up to the EMF voltage ε.
- The current starts off at $I_o = \dfrac{\varepsilon}{R}$ and decreases to 0.

Okay, we have a qualitative understanding of what happens. Let's see if we can obtain formulas for V_R, I, V_C, and q as functions of time. Here's the circuit:



We apply the loop rule:



$$KVL1 + \varepsilon - V_R + V_C = 0$$

and the definitions of resistance and capacitance:

$$V_R = IR$$

$$q = CV_C$$

$$V_C = \frac{q}{C}$$

to obtain:

$$\varepsilon - IR - \frac{q}{C} = 0$$

$$IR + \frac{q}{C} = \varepsilon$$

Then we use the fact that the current is equal to the rate at which charge is building up on the capacitor, $I = \frac{dq}{dt}$, to get:

$$\frac{dq}{dt}R + \frac{q}{C} = \varepsilon$$

$$\frac{dq}{dt} + \frac{q}{RC} = \frac{\varepsilon}{R}$$

This is interesting. This is the same equation that we had before, except that we have the constant ε/R on the right instead of 0.

For this equation, I'm simply going to provide and discuss the solution, rather than show you how to solve the differential equation. The charge function of time that solves this equation is:

$$q = C\varepsilon(1 - e^{-\frac{t}{RC}})$$

Please substitute it into the differential equation ($\frac{dq}{dt} + \frac{q}{RC} = \frac{\varepsilon}{R}$) and verify that it leads to an identity.

Now let's check to make sure that $q = C\varepsilon(1 - e^{-\frac{t}{RC}})$ is consistent with our conceptual understanding. At time zero $(t = 0)$, our expression $q(t) = C\varepsilon(1 - e^{-\frac{t}{RC}})$ evaluates to:

$$\begin{aligned} q(0) &= C\varepsilon(1 - e^{-\frac{0}{RC}}) \\ &= C\varepsilon(1 - e^{0}) \\ &= C\varepsilon(1 - 1) \\ q(0) &= 0 \end{aligned}$$

Excellent. This is consistent with the fact that the capacitor starts out uncharged.

Now, what does our charge function $q(t) = C\varepsilon(1 - e^{-\frac{t}{RC}})$ say about what happens to the charge of the capacitor in the limit as t goes to infinity?

$$\begin{aligned} \lim_{t\to\infty} q(t) &= \lim_{t\to\infty} C\varepsilon(1 - e^{-\frac{t}{RC}}) \\ &= C\varepsilon \lim_{x\to\infty} (1 - e^{-x}) \\ &= C\varepsilon \lim_{x\to\infty} (1 - \frac{1}{e^{x}}) \\ &= C\varepsilon \lim_{y\to\infty} (1 - \frac{1}{y}) \\ &= C\varepsilon\, (1 - \lim_{y\to\infty} \frac{1}{y}) \\ &= C\varepsilon(1 - 0) \\ \lim_{t\to\infty} q(t) &= C\varepsilon \end{aligned}$$

Well, this makes sense. Our conceptual understanding was that the capacitor would keep charging until the voltage across the capacitor was equal to the voltage across the seat of EMF. From the definition of capacitance, when the capacitor voltage is ε, its charge is indeed $C\varepsilon$. The formula yields the expected result for $\lim_{t\to\infty} q(t)$.

Once we have $q(t)$ it is pretty easy to get the other circuit quantities. For instance, from the definition of capacitance:

$$q = CV_c,$$

we have $V_c = q/C$ which, with $1 = C\varepsilon(1 - e^{-\frac{t}{RC}})$ evaluates to:

$$V_c = \varepsilon(1 - e^{-\frac{t}{RC}}) \tag{B13.4}$$

Our original loop equation read:

$$\varepsilon - V_R - V_c = 0$$

So:

$$V_R = \varepsilon - V_c$$

which, with $V_c = \varepsilon(1 - e^{-\frac{t}{RC}})$ can be written as:

$$V_R = \varepsilon - \varepsilon(1 - e^{-\frac{t}{RC}})$$

$$V_R = \varepsilon - \varepsilon + \varepsilon e^{-\frac{t}{RC}}$$

$$V_R = \varepsilon e^{-\frac{t}{RC}}$$

From our definition of resistance:

$$V_R = IR$$

$$I = \frac{V_R}{R}$$

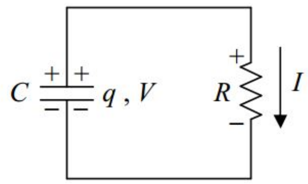
with $V_R = \varepsilon e^{-\frac{t}{RC}}$, this can be expressed as:

$$I = \frac{\varepsilon}{R} e^{-\frac{t}{RC}}$$

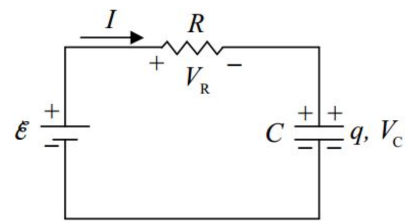
At time 0, this evaluates to ε/R meaning that ε/R can be interpreted as the current at time 0 allowing us to write our function $I(t)$ as

$$I = I_o e^{-\frac{t}{RC}}$$

Our formula has the current starting out at its maximum value and decreasing exponentially with time, as anticipated based on our conceptual understanding of the circuit. Note that this is the same formula that we got for the current in the discharging-capacitor circuit. In both cases, the current dies off exponentially. The reasons differ, but the effect ($I = I_o e^{-\frac{t}{RC}}$) is the same:



In the discharging-capacitor circuit, the current dies off because the capacitor runs out of charge.



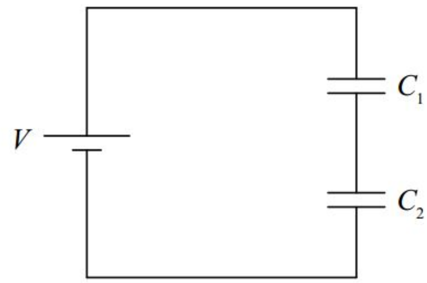
In the charging-capacitor circuit, the current dies off because the capacitor voltage, which counteracts the EMF, builds up to ℰ as the capacitor charges.

B14: Capacitors in Series & Parallel

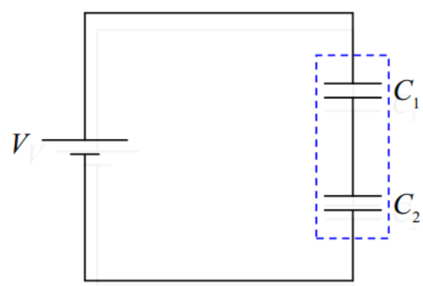
The method of ever-simpler circuits that we used for circuits with more than one resistor can also be used for circuits having more than one capacitor. The idea is to replace a combination circuit element consisting of more than one capacitor with a single equivalent capacitor. The equivalent capacitor should be equivalent in the sense that, with the same potential across it, it will have the same charge as the combination circuit element.

Capacitors in Series

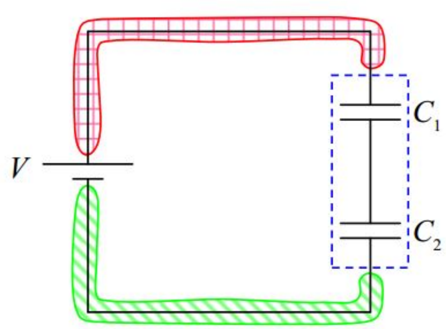
Let's start with a case in which the combination circuit element consists of two capacitors in series with each other:



We consider the two capacitors to be a two-terminal combination circuit element:



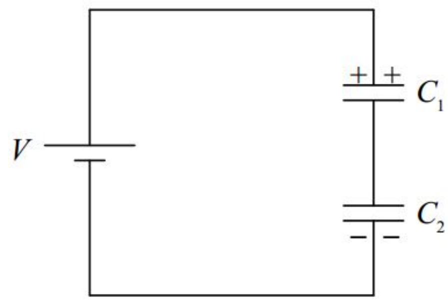
The voltage across the combination circuit element is clearly the EMF voltage V since, for both the seat of EMF and the combination circuit element, we're talking about the potential difference between the same two conductors:



The voltage across each individual capacitor is, however, not known.

But consider this: After that last wire is connected in the circuit, the charging process (which takes essentially no time at all) can be understood to proceed as follows (where, for ease of understanding, we describe things that occur simultaneously as if they occurred sequentially):

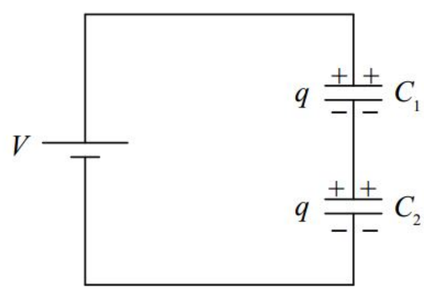
The seat of EMF pulls some positive charge from the bottom plate of the lower capacitor and pushes it onto the top plate of the upper capacitor.



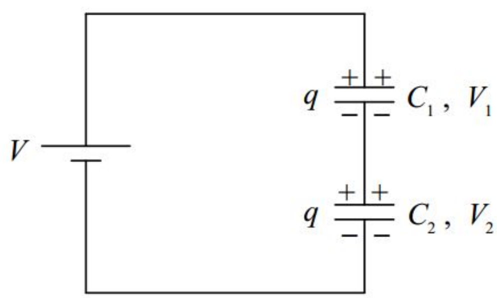
The key point about this movement of charge is that the amount of positive charge on the top plate of the upper capacitor is exactly equal to the amount of negative charge on the bottom plate of the lower capacitor (because that's where the positive charge came from!)

Now, the positive charge on the upper plate of the top capacitor repels the positive charge (remember, every neutral object consists of huge amounts of both kinds of charge, and, in our positive-charge-carrier convention, the positive charges are free to move) on the bottom plate of the upper capacitor and that charge has a conducting path to the top plate of the lower capacitor, to which it (the positive charge) is attracted by the negative charge on the bottom plate of the lower capacitor.

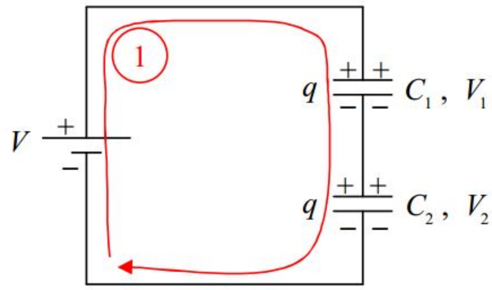
The final result is that both capacitors have one and the same charge q:



which in turn causes capacitor C_1 to have voltage $V_1 = \frac{q}{C_1}$ and capacitor C_2 to have voltage $V_2 = \frac{q}{C_2}$.



By the loop rule,



KVL 1

$$V - V_1 - V_2 = 0$$

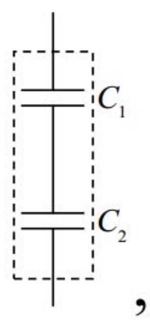
$$V = V_1 + V_2$$

$$V = \frac{q}{C_1} + \frac{q}{C_2}$$

$$V = q(\frac{1}{C_1} + \frac{1}{C_2})$$

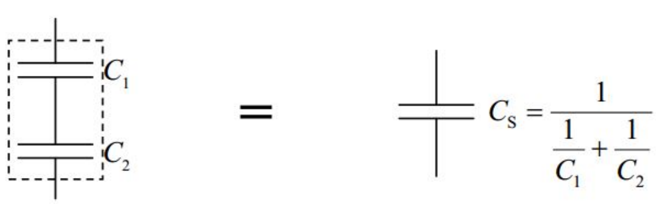
$$q = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2}} V$$

So, what we’re saying is, that when you put a voltage V across the two-terminal circuit element



,

an amount of charge $q = \frac{1}{\frac{1}{C_1}+\frac{1}{C_2}} V$ is moved from the bottom terminal of the combination circuit element, around the circuit, to the top terminal. Then charge stops moving. Recall that we defined the capacitance of a capacitor to be the ratio $\frac{q}{V}$ of the charge on the capacitor to the corresponding voltage across the capacitor. $\frac{q}{V}$ for our two-terminal combination circuit element is thus the equivalent capacitance of the two terminal circuit element. Solving $q = \frac{1}{\frac{1}{C_1}+\frac{1}{C_2}} V$ for the ratio $\frac{q}{V}$ yields $\frac{q}{V} = \frac{1}{\frac{1}{C_1}+\frac{1}{C_2}}$ so our equivalent capacitance for two capacitors in series is $C_s = \frac{1}{\frac{1}{C_1}+\frac{1}{C_2}}$



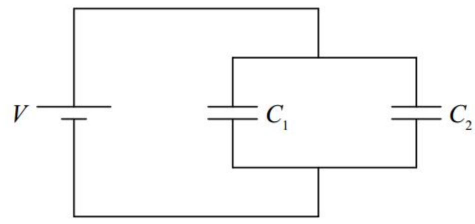
By logical induction, we can extend this argument to cover any number of capacitors in series with each other, obtaining:

$$C_s = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \dots} \tag{B14.1}$$

As far as making things easy to remember, it's just too bad the way things work out sometimes. This expression is mathematically identical to the expression for resistors in parallel. But, this expression is for capacitors in series.

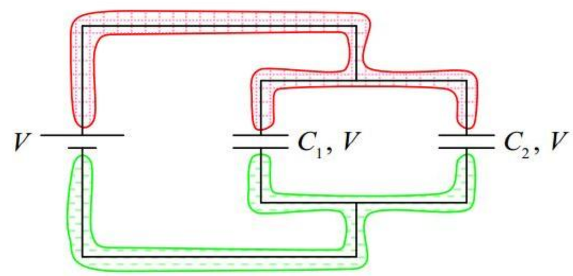
Capacitors in Parallel

Suppose we put a voltage V across a combination circuit element consisting of a pair of capacitors in parallel with each other:

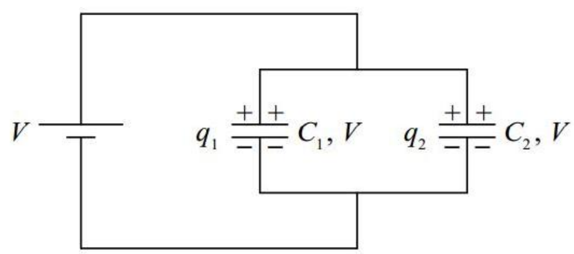


Capacitors in Parallel

It is clear from the diagram that the voltage across each capacitor is just the EMF V since the voltage across every component in the circuit is the potential difference between the same two conductors.



So what happens (almost instantaneously) when we make that final connection? Answer: The seat of EMF pulls charge off the bottom plates of the two capacitors and pushes it onto the top plates until the charge on C_1 is $q_1 = C_1 V$ and the charge on C_2 is $q_2 = C_2 V$.



To do that, the seat of EMF has to move a total charge of

$$q = q_1 + q_2$$

$$q = C_1 V + C_2 V$$

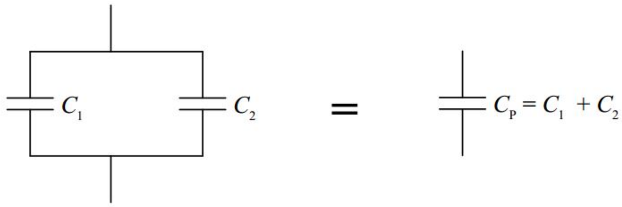
$$q = (C_1 + C_2) V$$

Solving the last equation, $q = (C_1 + C_2) V$, for the equivalent capacitance C_P, defined as q/V, yields:

$$\frac{q}{V} = C_1 + C_2$$

$$C_p = C_1 + C_2$$

In other words:



So, the equivalent capacitance of capacitors in parallel is simply the sum of the individual capacitances. (This is the way resistors in series combine.) By means of inductive reasoning, the result can be extended to any number of capacitors, yielding:

$$C_P = C_1 + C_2 + C_3 + \ldots \tag{B14.2}$$

Concluding Remarks

The facts that the voltage is the same for capacitors in parallel and the charge is the same for capacitors in series are important, but, if you look at these as two more things that you have to commit to memory then you are not going about your study of physics the right way. You need to be able to "see" that the charge on capacitors in series has to be the same because the charge on one capacitor comes from its (originally-neutral) neighbor. You need to be able to "see" that the voltage across capacitors in parallel has to be the same because, for each capacitor, the voltage is the potential difference between the same two conductors.

B15: Magnetic Field Introduction - Effects

We now begin our study of magnetism, and, analogous to the way in which we began our study of electricity, we start by discussing the effect of a given magnetic field without first explaining how such a magnetic field might be caused to exist. We delve into the causes of magnetic fields in subsequent chapters.

A magnetic field is a vector field. That is, it is an infinite set of vectors, one at each point in the region of space where the magnetic field exists. We use the expression "magnetic field" to designate both the infinite set of vectors, and, when one is talking about the magnetic field at a point in space, the one magnetic field vector at that point in space. We use the symbol $\vec{B}$ to represent the magnetic field. The most basic effect of a magnetic field is to exert a torque on an object that has a property known as magnetic dipole moment, and, that finds itself in the magnetic field. A particle or object that has a non-zero value of magnetic dipole moment is called a magnetic dipole. A magnetic dipole is a bar magnet. The value of the magnitude of the magnetic dipole moment of an object is a measure of how strong a bar magnet it is. A magnetic dipole has two ends, known as poles—a north pole and a south pole. Magnetic dipole moment is a property of matter which has direction. We can define the direction, of the magnetic dipole moment of an object, by considering the object to be an arrow whose north pole is the arrowhead and whose south pole is the tail. The direction in which the arrow is pointing is the direction of the magnetic dipole moment of the object. The unit of magnetic dipole moment is the $A \cdot m^2$ (ampere meter-squared). While magnetic compass needles come in a variety of magnetic dipole moments, a representative value for the magnetic dipole moment of a compass needle is $.1 A \cdot m^2$.

Again, the most basic effect of a magnetic field is to exert a torque on a magnetic dipole that finds itself in the magnetic field. The magnetic field vector, at a given point in space, is the maximum possible torque-per-magnetic-dipole-moment-of-would-be-victim that the magnetic field would/will exert on any magnetic dipole (victim) that might find itself at the point in question. I have to say "maximum possible" because the torque exerted on the magnetic dipole depends not only on the magnitude of the magnetic field at the point in space and the magnitude of the magnetic dipole moment of the victim, but it also depends on the orientation of the magnetic dipole relative to the direction of the magnetic field vector. In fact:

$$\vec{\tau} = \vec{\mu} \times \vec{B} \tag{B15.1}$$

where $\vec{\tau}$ is the torque exerted on the magnetic dipole (the bar magnet) by the magnetic field, $\vec{\mu}$ is the magnetic dipole moment of the magnetic dipole (the bar magnet, the victim), and $\vec{B}$ is the magnetic field vector at the location in space at which the magnetic dipole is.

For the cross product of any two vectors, the magnitude of the cross product is the product of the magnitudes of the two vectors, times the sine of the angle the two vectors form when placed tail to tail. In the case of $\vec{\tau} = \vec{\mu} \times \vec{B}$, this means:

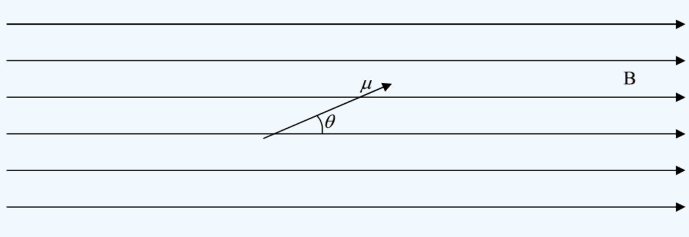
$$\tau = \mu\, B \sin \theta$$

In the SI system of units, torque has units of $N \cdot m$ (newton-meters). For the units on the right side of $\tau = \mu\, B \cos \theta$ to work out to be $N \cdot m$, what with μhaving units of electric dipole moment ($A \cdot m^2$) and $\sin \theta$ having no units at all, B must have units of torque-per-magnetic-dipole-moment, namely, $\frac{N \cdot m}{A \cdot m^2}$. That combination unit is given a name. It is called the tesla, abbreviated T.

$$1T = 1 \frac{N \cdot m}{A \cdot m^2}$$

Example 1

Consider a magnetic dipole having a magnetic dipole moment $\mu = 0.045\, A \cdot m^2$, oriented so that it makes an angle of 23° with the direction of a uniform magnetic field of magnitude $5.0 \times 10^{-5} T$ as depicted below. Find the torque exerted on the magnetic dipole, by the magnetic field.



Recall that the arrowhead represents the north pole of the bar magnet that a magnetic dipole is. The direction of the torque is such that it tends to cause the magnetic dipole to point in the direction of the magnetic field. For the case depicted above, that would be clockwise as viewed from the vantage point of the creator of the diagram. The magnitude of the torque for such a case can be calculated as follows:

$$\begin{aligned} \tau &= \mu B \sin\theta \\ &= (.045 A\cdot m^2)(5.0\times 10^{-5} T)\sin 23^\circ \\ &= 8.8\times 10^{-7} A\cdot m^2\cdot T \end{aligned}$$

Recalling that a tesla is a $\frac{N\cdot m}{A\cdot m^2}$ we have:

$$\begin{aligned} \tau &= 8.8\times 10^{-7} A\cdot m^2\cdot \frac{N\cdot m}{A\cdot m^2} \\ &= 8.8\times 10^{-7} N\cdot m \end{aligned}$$

Thus, the torque on the magnetic dipole is $\tau = 8.8\times 10^{-7} N\cdot m$ clockwise, as viewed from the vantage point of the creator of the diagram.

Example 2

A particle having a magnetic dipole moment $\vec{\mu} = 0.025 A\cdot m^2\hat{i} - 0.035 A\cdot m^2\hat{j} + 0.015 A\cdot m^2\hat{k}$ is at a point in space where the magnetic field $\vec{B} = 2.3mT\hat{i} + 5.3mT\hat{j} - 3.6mT\hat{k}$. Find the torque exerted on the particle by the magnetic field.

Solution

$$\begin{aligned} \vec{\tau} &= \vec{\mu}\times\vec{B} \\ &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 0.025 A\cdot m^2 & -0.035 A\cdot m^2 & 0.015 A\cdot m^2 \\ 0.0023\frac{Nm}{Am^2} & 0.0053\frac{Nm}{Am^2} & -0.0036\frac{Nm}{Am^2} \end{vmatrix} \\ &= \hat{i}\left[(-0.035 Am^2)(-0.0036\frac{Nm}{Am^2}) - (0.015 Am^2)(0.0053\frac{Nm}{Am^2})\right] \\ &+ \hat{j}\left[(0.015 Am^2)(0.0023\frac{Nm}{Am^2}) - (0.025 Am^2)(-0.0036\frac{Nm}{Am^2})\right] \\ &+ \hat{k}\left[(0.025 Am^2)(0.0053\frac{Nm}{Am^2}) - (-0.035 Am^2)(0.0023\frac{Nm}{Am^2})\right] \\ &= 1.2\times 10^{-4} Nm\hat{i} - 1.2\times 10^{-4} Nm\hat{j} + 2.1\times 10^{-4} Nm\hat{k} \end{aligned}$$

The Magnetic Force Exerted Upon a Magnetic Dipole

A uniform magnetic field exerts no force on a bar magnet that is in the magnetic field. You should probably pause here for a moment and let that sink in. A uniform magnetic field exerts no force on a bar magnet that is in that magnetic field.

You have probably had some experience with bar magnets. You know that like poles repel and unlike poles attract. And, from your study of the electric field, you have probably (correctly) hypothesized that in the field point of view, the way we see this is that one bar magnet (call it the source magnet) creates a magnetic field in the region of space around itself, and, that if there is another bar magnet in that region of space, it will be affected by the magnetic field it is in. We have already discussed the fact that the victim bar magnet will experience a torque. But you know, from your experience with bar magnets, that it will also experience a force. How can that be when I just stated that a uniform magnetic field exerts no force on a bar magnet? Yes, of course. The magnetic field of the source magnet must be non-uniform. Enough about the nature of the magnetic field of a bar magnet, I'm supposed to save that for an upcoming chapter. Suffice it to say that it is non-uniform and to focus our attention on the effect of a non-uniform field on a bar magnet that finds itself in that magnetic field.

First of all, a non-uniform magnetic field will exert a torque on a magnetic dipole (a bar magnet) just as before ($\vec{\tau} = \vec{\mu}\times\vec{B}$). But, a non-uniform magnetic field (one for which the magnitude, and/or direction, depends on position) also exerts a force on a magnetic dipole. The force is given by:

$$\vec{F}_B = \nabla(\vec{\mu} \cdot \vec{B}) \tag{B15.2}$$

where

- $\vec{F}_B$ is the force exerted by the magnetic field $\vec{B}$ on a particle having a magnetic dipole moment $\vec{\tau}$
- $\vec{\mu}$ is the magnetic dipole of the "victim", and,
- $\vec{B}$ is the magnetic field at the position in space where the victim finds itself. To evaluate the force, once must know $\vec{B}$ as a function of x, y and z (wherea $\vec{\mu}$ is a constant) .

Note that after you take the gradient of $\vec{\mu} \cdot \vec{B}$, you have to evaluate the result at the values of x, y and z corresponding to the location of the victim.

Just to make sure that you know how to use this equation, please note that if $\vec{\mu}$ and $\vec{B}$ are expressed in $\hat{i}, \hat{j}, \hat{k}$ notation, so that they appear as $\vec{\mu} = \mu_x \hat{i} + \mu_y \hat{j} + \mu_z \hat{k}$ and $\vec{B} = B_x \hat{i} + B_y \hat{j} + B_z \hat{k}$ respectively, then:

$$\begin{aligned} \vec{\mu} \cdot \vec{B} &= (\mu_x \hat{i} + \mu_y \hat{j} + \mu_z \hat{k}) \cdot (B_x \hat{i} + B_y \hat{j} + B_z \hat{k}) \\ &= \mu_x B_x + \mu_y B_y + \mu_z B_z \end{aligned}$$

And the gradient of $\vec{\mu} \cdot \vec{B}$ (which by equation B15.2 is the force we seek) is given by

$$\nabla(\vec{\mu} \cdot \vec{B}) = \frac{\partial(\vec{\mu} \cdot \vec{B})}{\partial x} \hat{i} + \frac{\partial(\vec{\mu} \cdot \vec{B})}{\partial y} \hat{j} + \frac{\partial(\vec{\mu} \cdot \vec{B})}{\partial z} \hat{k}$$

where derivatives in this equation can (using $\vec{\mu} \cdot \vec{B} = \mu_x B_x + \mu_y B_y + \mu_z B_z$ from just above) can be expressed as:

$$\begin{aligned} \frac{\partial(\vec{\mu} \cdot \vec{B})}{\partial x} &= \mu_x \frac{\partial B_x}{\partial x} + \mu_y \frac{\partial B_y}{\partial x} + \mu_z \frac{\partial B_z}{\partial x} \\ \frac{\partial(\vec{\mu} \cdot \vec{B})}{\partial y} &= \mu_x \frac{\partial B_x}{\partial y} + \mu_y \frac{\partial B_y}{\partial y} + \mu_z \frac{\partial B_z}{\partial y} \\ \frac{\partial(\vec{\mu} \cdot \vec{B})}{\partial z} &= \mu_x \frac{\partial B_x}{\partial z} + \mu_y \frac{\partial B_y}{\partial z} + \mu_z \frac{\partial B_z}{\partial z} \end{aligned}$$

where we have taken advantage of the fact that the components of the magnetic dipole moment of the victim are not functions of position. Also note that the derivatives are all partial derivatives. Partial derivatives are the easy kind in the sense that, when, for instance, you take the derivative with respect to x, you are to treat y and z as if they were constants. Finally, it is important to realize that, after you take the derivatives, you have to plug the values of x, y and z corresponding to the location of the magnetic dipole (the victim), into the given expression for the force.

Example 3

There exists, in a region of space, a magnetic field, given in terms of Cartesian unit vectors by:

$$\vec{B} = -5.82 \times 10^{-6} T \cdot m \frac{y}{x^2 + y^2} \hat{i} + 5.82 \times 10^{-6} T \cdot m \frac{x}{x^2 + y^2} \hat{j}$$

A particle is in the region of space where the magnetic field exists. The particle has a magnetic dipole moment given by:

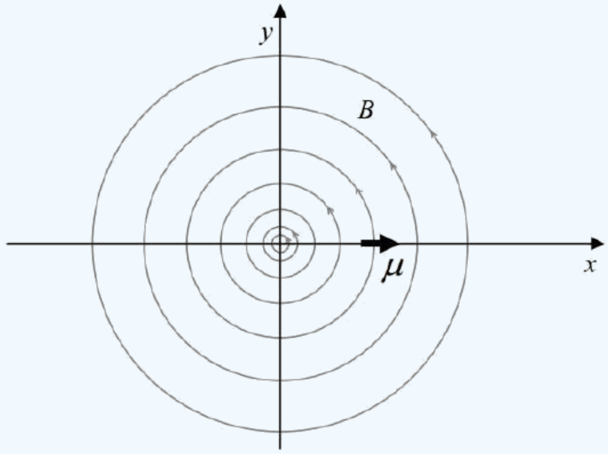
$$\vec{\mu} = .514 A \cdot m^2 \hat{i}$$

The particle is at $(0.110m, 0, 0)$

Find the force exerted on the particle by the magnetic field.

Solution

First, we sketch the configuration:



Substituting the given $\vec{\mu}$ and $\vec{B}$, into our expression for the force yields:

$$\begin{aligned}\vec{F}_B &= \nabla(\vec{\mu}\cdot\vec{B}) \\ &= \nabla\left[(0.514 A\cdot m^2\,\hat{i})\cdot\left(-5.82\times10^{-6} T\cdot m\frac{y}{x^2+y^2}\,\hat{i} + 5.82\times10^{-6} T\cdot m\frac{x}{x^2+y^2}\,\hat{j}\right)\right] \\ &= \nabla\left(-2.00\times10^{-6} A\cdot m^2\cdot T\cdot m\frac{y}{x^2+y^2}\right) \\ &= -2.99\times10^{-6} N\cdot m^2\,\nabla\left[y(x^2+y^2)^{-1}\right] \\ &= -2.99\times10^{-6} N\cdot m^2\left\{\frac{\partial}{\partial x}\left[y(x^2+y^2)^{-1}\right]\hat{i} + \frac{\partial}{\partial y}\left[y(x^2+y^2)^{-1}\right]\hat{j} + \frac{\partial}{\partial z}\left[y(x^2+y^2)^{-1}\right]\hat{k}\right\} \\ &= -2.99\times10^{-6} N\cdot m^2\{[y(-1)(x^2+y^2)^{-2}2x\,\hat{i} + [(x^2+y^2)^{-1} + y(-1)(x^2+y^2)^{-2}2y]\,\hat{j} + 0\hat{k} \\ &= -2.99\times10^{-6} N\cdot m^2\left\{-\frac{2xy}{(x^2+y^2)^2}\,\hat{i} + \left[\frac{1}{x^2+y^2} - \frac{2y^2}{(x^2+y^2)^2}\right]\hat{j}\right.\end{aligned}$$

Recalling that we have to evaluate this expression at the location of the victim, a location that was given as $(0.110m, 0, 0)$,we find that:

$$\begin{aligned}\vec{F}_B &= -2.99\times10^{-6} N\cdot m^2\left\{-\frac{2(0.110m)0}{[(0.110m^2)+0^2]^2}\,\hat{i} + \left[\frac{1}{(0.110m)^2+0^2} - \frac{2(0)^2}{[(0.110m)^2+0^2]^2}\right]\hat{j}\right\} \\ &= -2.47\times10^{-4} N\hat{j}\end{aligned}$$

Characteristics of the Earth's Magnetic Field

We live in a magnetic field produced by the earth. Both its magnitude and its direction are different at different locations on the surface of the earth. Furthermore, at any given location, the earth's magnetic field varies from year to year in both magnitude and direction. Still, on the geographical scale of a college campus, and, on a time scale measured in days, the earth's magnetic field is approximately uniform and constant.

To align your index finger with the magnetic field of the earth on the Saint Anselm College campus, first point in the horizontal direction 15.4° West of North. Then tilt your arm downward so that you are pointing in a direction that is 68.9° below the horizontal. (Yes! Can you believe it? It's mostly downward!) You are now pointing your finger in the direction of the earth's magnetic field. The magnitude of the magnetic field, on the Saint Anselm College campus, is $5.37\times10^{-5}T$. In other words:

The Earth's Magnetic Field on the Saint Anselm College Campus in 2006

Characteristic	Value	**Rate of Change**
Declination	-15.4°	$+0.074^{\circ}/\text{year}$

Characteristic	Value	**Rate of Change**
Inclination (Dip Angle)	68.8°	$-0.096^{\circ}/\text{year}$
Magnitude	$5.36 \times 10^{-5} T$	$-0.012 \times 10^{-5} T/\text{year}$
Horizontal Component	$1.93 \times 10^{-5} T$	$+0.004 \times 10^{-5} T/\text{year}$
Vertical Component	$5.00 \times 10^{-5} T$	$-0.014 \times 10^{-5} T/\text{year}$

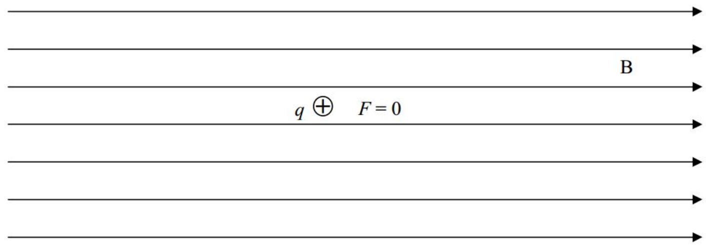
A compass needle is a tiny bar magnet that is constrained to rotate about a vertical axis. The earth's magnetic field exerts a torque on the compass needle that tends to make the compass needle point in the direction of the horizontal component of the earth's magnetic field, a direction we call "magnetic north". Recall that when we talk about which way a bar magnet (such as a compass needle) is pointing, we imagine there to be an arrowhead at its north pole.

B16: Magnetic Field - More Effects

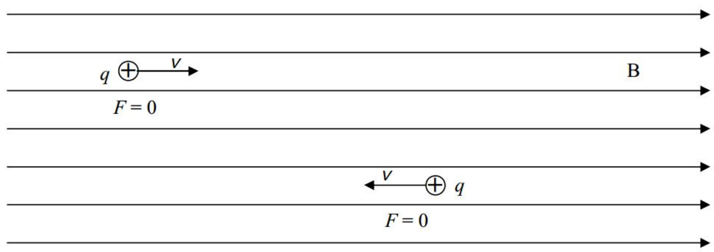
The electric field and the magnetic field are not the same thing. An electric dipole with positive charge on one end and negative charge on the other is not the same thing as a magnetic dipole having a north and a south pole. More specifically: An object can have positive charge but it can't have "northness".

On the other hand, electricity and magnetism are not unrelated. In fact, under certain circumstances, a magnetic field will exert a force on a charged particle that has no magnetic dipole moment. Here we consider the effect of a magnetic field on such a charged particle.

FACT: A magnetic field exerts no force on a charged particle that is at rest in the magnetic field.



FACT: A magnetic field exerts no force on a charged particle that is moving along the line along which the magnetic field, at the location of the particle, lies.

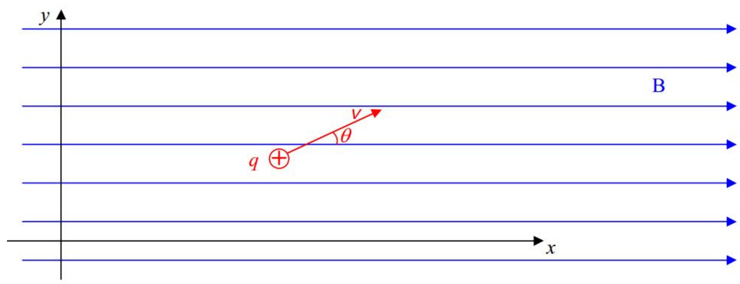


FACT: A magnetic field does exert a force on a charged particle that is in the magnetic field, and, is moving, as long as the velocity of the particle is not along the line, along which, the magnetic field is directed. The force in such a case is given by:

$$\vec{F} = q\vec{v} \times \vec{B} \tag{B16.1}$$

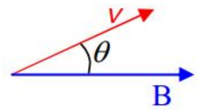
Note that the cross product yields a vector that is perpendicular to each of the multiplicands. Thus the force exerted on a moving charged particle by the magnetic field within which it finds itself, is always perpendicular to both its own velocity, and the magnetic field vector at the particle's location.

Consider a positively-charged particle moving with velocity v at angle θ in the $x-y$ plane of a Cartesian coordinate system in which there is a uniform magnetic field in the $+x$ direction.



To get the magnitude of the cross product $\vec{v} \times \vec{B}$ that appears in $\vec{F} = q\vec{v} \times \vec{B}$ we are supposed to establish the angle that $\vec{v}$ and $\vec{B}$ make with each other when they are placed tail to tail. Then the magnitude $|\vec{v} \times \vec{B}|$ is just the absolute value of the product of the magnitudes of the vectors times the sine of the angle in between them. Let's put the two vectors tail to tail and establish that angle.

Note that the magnetic field as a whole is an infinite set of vectors in the $+x$ direction. So, of course, the magnetic field vector $\vec{B}$, at the location of the particle, is in the $+x$ direction.



Clearly the angle between the two vectors is just the angle θ that was specified in the problem. Hence,

$$|\vec{v} \times \vec{B}| = |vB \sin\theta|,$$

so, starting with our given expression for $\vec{F}$, we have:

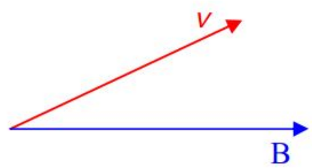
$$\vec{F} = q\vec{v} \times \vec{B}$$

$$|\vec{F}| = |q\vec{v} \times \vec{B}|$$

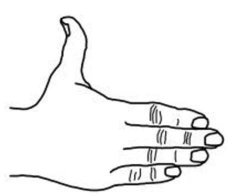
$$|\vec{F}| = |qvB \sin\theta|$$

Okay, now let's talk about the direction of $\vec{F} = q\vec{v} \times \vec{B}$. We get the direction of $\vec{v} \times \vec{B}$ and then we think. The charge q is a scalar. If q is positive, then, when we multiply the vector $\vec{v} \times \vec{B}$ by q (to get $\vec{F}$), we get a vector in the same direction as that of $\vec{v} \times \vec{B}$. So, whatever we get (using the right-hand rule for the cross product) for the direction of $\vec{v} \times \vec{B}$ is the direction of $\vec{F} = q\vec{v} \times \vec{B}$. But, if q is negative, then, when we multiply the vector $\vec{v} \times \vec{B}$ by q (to get $\vec{F}$), we get a vector in opposite direction to that of $\vec{v} \times \vec{B}$. So, once we get the direction of $\vec{v} \times \vec{B}$ by means of the righthand rule for the cross product of two vectors, we have to realize that (because the charge is negative) the direction of $\vec{F} = q\vec{v} \times \vec{B}$ is opposite the direction that we found for $\vec{v} \times \vec{B}$.

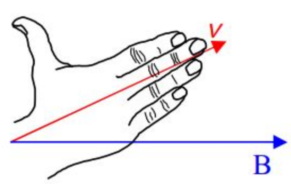
Let's do it. To get the direction of the cross product vector $\vec{v} \times \vec{B}$ (which appears in $\vec{F} = q\vec{v} \times \vec{B}$, draw the vectors $\vec{v}$ and $\vec{B}$ tail to tail.



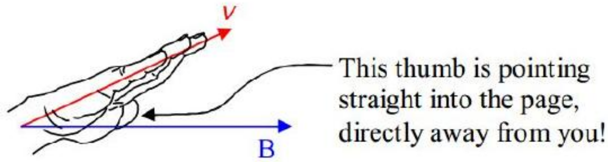
Extend the fingers of your right hand so that they are pointing directly away from your right elbow. Extend your thumb so that it is at right angles to your fingers.



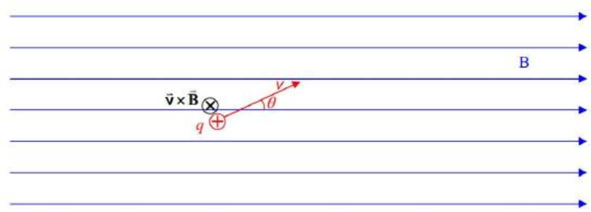
Now, keeping your fingers aligned with your forearm, align your fingers with the first vector appearing in the cross product $\vec{v} \times \vec{B}$, namely $\vec{v}$.



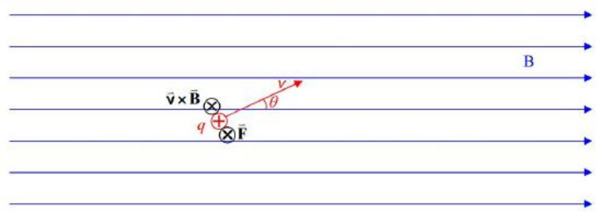
Now rotate your hand, as necessary, about an imaginary axis extending along your forearm and along your middle finger, until your hand is oriented such that, if you were to close your fingers, they would point in the direction of the second vector.



The direction in which your right thumb is now pointing is the direction of $\vec{v} \times \vec{B}$. We depict a vector in that direction by means of an $\times$ with a circle around it. That symbol is supposed to represent the tail feathers of an arrow that is pointing away from you.

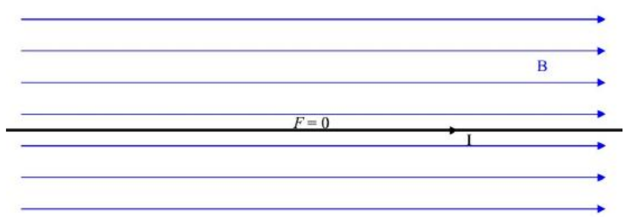


Let's not forget about that q in the expression $\vec{F} = q\vec{v} \times \vec{B}$. In the case at hand, the charged particle under consideration is positive. In other words q is positive. So, $\vec{F} = q\vec{v} \times \vec{B}$ is in the same direction as $\vec{v} \times \vec{B}$.

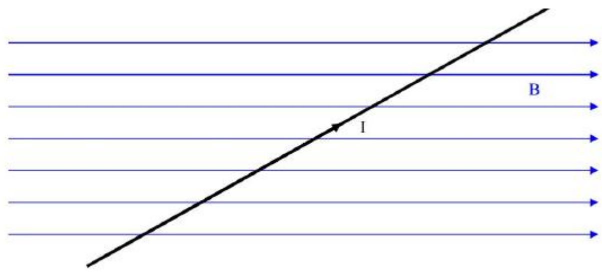


A magnetic field will also interact with a current-carrying conductor. We focus our attention on the case of a straight current-carrying wire segment in a magnetic field:

FACT: Given a straight, current carrying conductor in a magnetic field, the magnetic field exerts no force on the wire segment if the wire segment lies along the line along which the magnetic field is directed. (Note: The circuit used to cause the current in the wire must exist, but, is not shown in the following diagram.)



FACT: A magnetic field exerts a force on a current-carrying wire segment that is in the magnetic field, as long as the wire is not collinear with the magnetic field.



The force exerted on a straight current-carrying wire segment, by the (uniform) magnetic field in which the wire is located, is given by

$$\vec{F} = I\vec{L} \times \vec{B} \tag{B16.2}$$

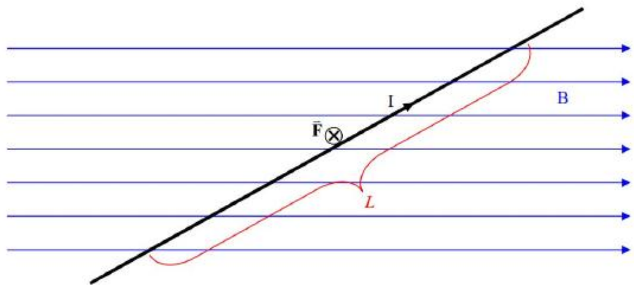
where:

$\vec{F}$ is the force exerted on the wire-segment-with-current by the magnetic field the wire is in,

I is the current in the wire,

$\vec{L}$ is a vector whose magnitude is the length of that segment of the wire which is actually in the magnetic field, and, whose direction is the direction of the current (which depends both on how the wire segment is oriented and how it is connected in the (not-shown) circuit.)

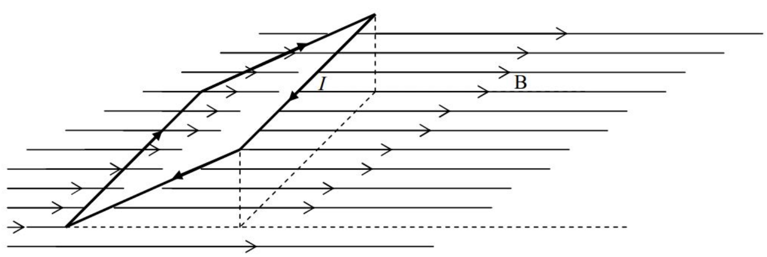
$\vec{B}$ is the magnetic field vector. The magnetic field must be uniform along the entire length of the wire for this formula to apply, so, $\vec{B}$ is the magnetic field vector at each and every point along the length of the wire.



Note that, in the preceding diagram, $\vec{F}$ is directed into the page as determined from $\vec{F} = I\vec{L} \times \vec{B}$ by means of the right-hand rule for the cross product of two vectors.

Effect of a Uniform Magnetic Field on a Current Loop

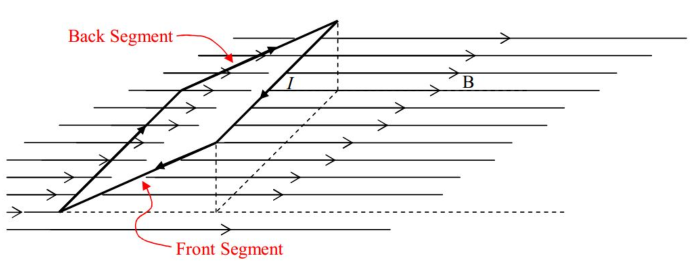
Consider a rectangular loop of wire. Suppose the loop to be in a uniform magnetic field as depicted in the following diagram:



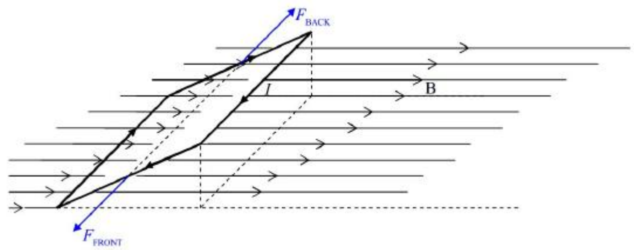
Note that, to keep things simple, we are not showing the circuitry that causes the current in the loop and we are not showing the cause of the magnetic field. Also, the magnetic field exists throughout the region of space in which the loop finds itself. We have not shown the full extent

of either the magnetic field lines depicted, or, the magnetic field itself.

Each segment of the loop has a force exerted on it by the magnetic field the loop is in. Let's consider the front and back segments first:

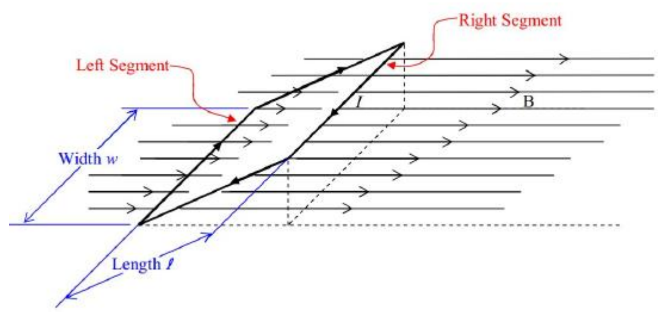


Because both segments have the same length, both segments make the same angle with the same magnetic field, and both segments have the same current; the force $\vec{F} = I\vec{L} \times \vec{B}$ will be of the same magnitude in each. (If you write the magnitude as $F = ILB\sin\theta$, you know the magnitudes are the same as long as you know that for any angle θ, $\sin(\theta) = \sin(180^\circ - \theta)$.) Using the right-hand rule for the cross product to get the direction, we find that each force is directed perpendicular to the segment upon which it acts, and, away from the center of the rectangle:

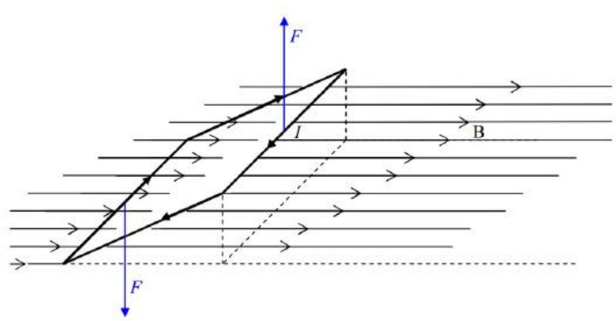


The two forces, F_{FRONT} and F_{BACK} are equal in magnitude, collinear, and opposite in direction. About the only effect they could have would be to stretch the loop. Assuming the material of the loop is rigid enough not to stretch, the net effect of the two forces is no effect at all. So, we can forget about them and focus our attention on the left and right segments in the diagram.

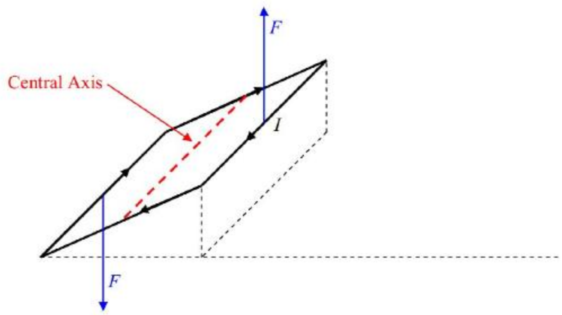
Both the left segment and the right segment are at right angles to the magnetic field. They are also of the same length and carry the same current. For each, the magnitude of $\vec{F} = I\vec{L} \times \vec{B}$ is just IwB where w is the width of the loop and hence the length of both the left segment and the right segment.



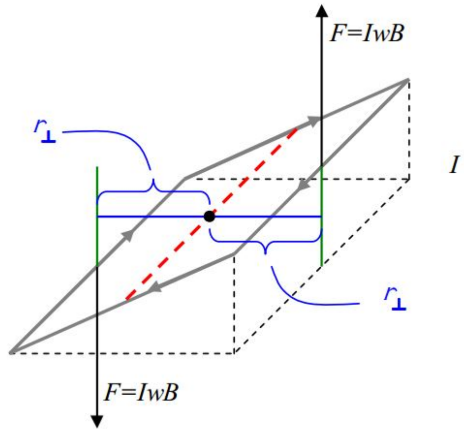
Using the right-hand rule for the cross product of two vectors, applied to the expression $\vec{F} = I\vec{L} \times \vec{B}$ for the force exerted on a wire segment by a magnetic field, we find that the force $F = IwB$ on the right segment is upward and the force $F = IwB$ on the left segment is downward.



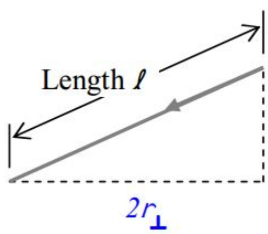
The two forces are equal (both have magnitude $F = IwB$) and opposite in direction, but, they are not collinear. As such, they will exert a net torque on the loop. We can calculate the torque about the central axis:



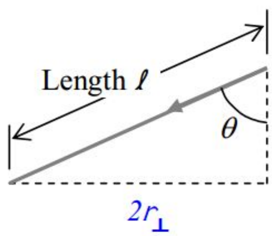
by extending the lines of action of the forces and identifying the moment arms:



The torque provided by each force is $r_\perp F$. Both torques are counterclockwise as viewed in the diagram. Since they are both in the same direction, the magnitude of the sum of the torques is just the sum of the magnitudes of the two torques, meaning that the magnitude of the total torque is just $\tau = 2r_\perp F$. We can get an expression for $2r_\perp$ by recognizing, in the diagram, that $2r_\perp$ is just the distance across the bottom of the triangle in the front of the diagram:



and defining the angle θ, in the diagram, to be the angle between the plane of the loop and the vertical.



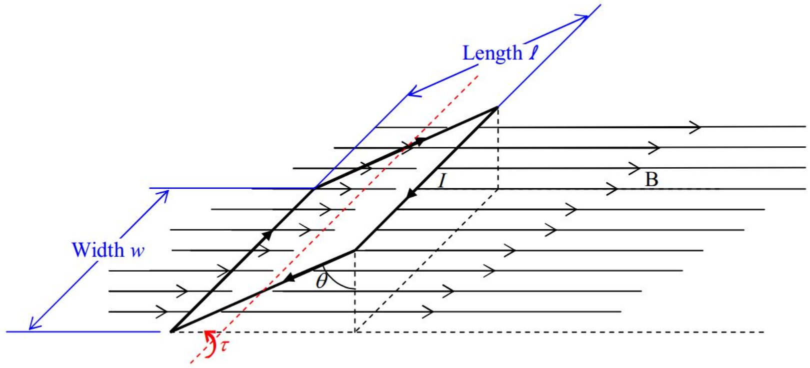
From the diagram, it is clear that $2r_\perp = l \sin\theta$.

Thus the magnetic field exerts a torque of magnitude

$$\tau = r_\perp F$$

$$\tau = [l(\sin\theta)](IwB)$$

on the current loop.



The expression for the torque can be written more concisely by first reordering the multiplicands so that the expression appears as

$$\tau = IlwB\sin\theta$$

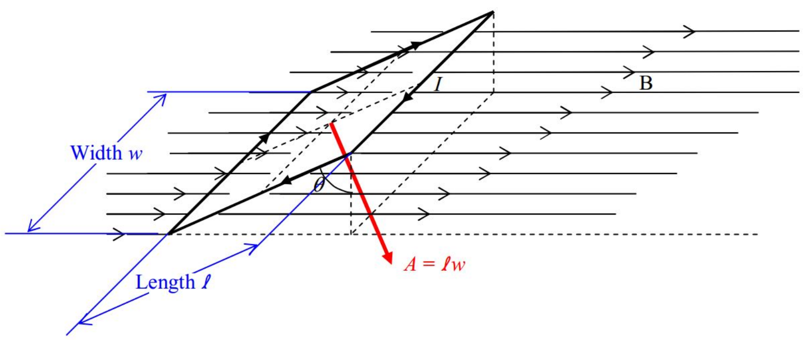
and then recognizing that the product lw is just the area A of the loop. Replacing lw with A yields:

$$\tau = I\ AB\sin\theta$$

Torque is something that has direction, and, you might recognize that $\sin\theta$ appearing in the preceding expression as something that can result from a cross product. Indeed, if we define an area vector to have a magnitude equal to the area of the loop,

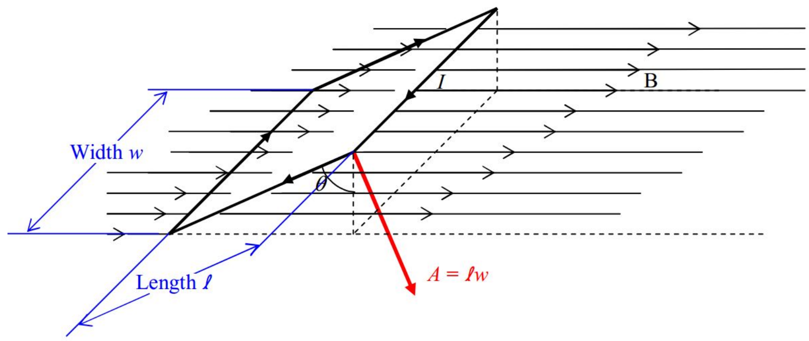
$$|\vec{A}| = lw$$

and, a direction perpendicular to the plane of the loop,

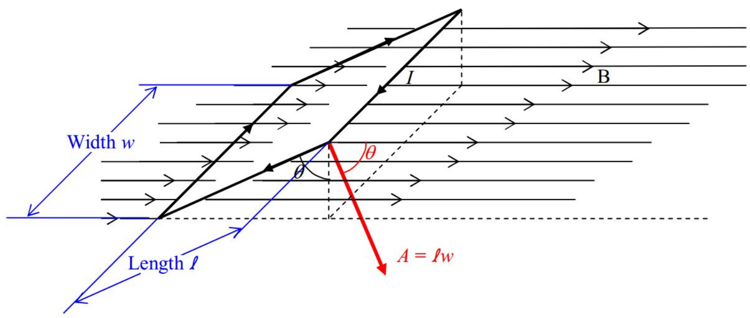


we can write the torque as a cross product. First note that the area vector as I have defined it in words to this point, could point in the exact opposite direction to the one depicted in the diagram. If, however, we additionally stipulate that the area vector is directed in accord with the righthand rule for something curly something straight, with the loop current being the something curly and the area vector the something straight (and we do so stipulate), then the direction of the area vector is uniquely determined to be the direction depicted in the diagram.

Now, if we slide that area vector over to the right front corner of the loop,



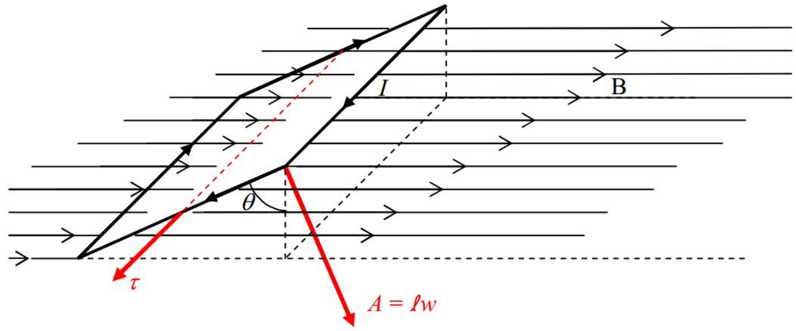
it becomes more evident (you may have already noticed it) that the angle between the area vector $\vec{A}$ and the magnetic field vector $\vec{B}$, is the same θ defined earlier and depicted in the diagram just above.



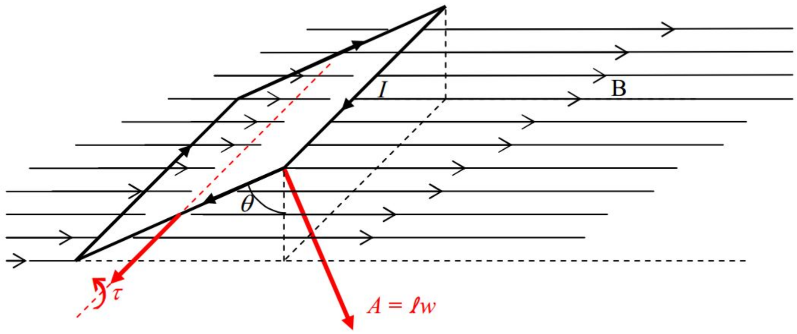
This allows us to write our expression for the torque $\tau = IAB\sin\theta$ counterclockwise as viewed in the diagram, as:

$$\vec{\tau} = I\vec{A} \times \vec{B}$$

Check it out. The magnitude of the cross product $|\vec{A} \times \vec{B}|$ is just $AB\sin\theta$, meaning that our new expression yields the same magnitude $\tau = IAB\sin\theta$ for the torque as we had before. Furthermore, the right-hand rule for the cross product of two vectors yields the torque direction depicted in the following diagram.



Recalling that the sense of rotation associated with an axial vector is determined by the righthand rule for something curly, something straight; we point the thumb of our cupped right hand in the direction of the torque vector and note that our fingers curl around counterclockwise, as viewed in the diagram.



Okay, we're almost there. So far, we have the fact that if you put a loop of wire carrying a current I in it, in a uniform magnetic field $\vec{B}$, with the loop oriented such that the area vector $\vec{A}$ of the current loop makes an angle θ with the magnetic field vector, then, the magnetic field exerts a torque

$$\vec{\tau} = I\vec{A} \times \vec{B}$$

on the loop.

This is identical to what happens to a magnetic dipole when you put it in a uniform magnetic field. It experiences a torque $\vec{\tau} = \vec{\mu} \times \vec{B}$. In fact, if we identify the product $I\vec{A}$ as the magnetic dipole moment of the current loop, then the expressions for the torque are completely identical:

$$\vec{\tau} = \vec{\mu} \times \vec{B} \tag{B16.3}$$

where:

- $\vec{\tau}$ is the torque exerted on the victim. The victim can be either a particle that has an inherent magnetic dipole moment, or, a current loop.
- $\vec{\mu}$ is the magnetic dipole moment of the victim. If the victim is a particle, $\vec{\mu}$ is simply the magnitude and direction of the inherent magnetic dipole moment of the particle. If the victim is a current loop, then $\vec{\mu} = I\vec{A}$ where I is the current in the loop and $\vec{A}$ is the area vector of the loop, a vector whose magnitude is the area of the loop and whose direction is the direction in which your right thumb points when you curl the fingers of your right hand around the loop in the direction of the current. (See the discussion below for the case in which the victim is actually a coil of wire rather than a single loop.)
- $\vec{B}$ is the magnetic field vector at the location of the victim.

A single loop of wire can be thought of as a coil of wire that is wrapped around once. If the wire is wrapped around N times, rather than once, then the coil is said to have N turns or N windings. Each winding makes a contribution of $I\vec{A}$ to the magnetic dipole moment of the current loop. The contribution from all the loops is in one and the same direction. So, the magnetic moment of a current-carrying coil of wire is:

$$\vec{\mu} = NI\vec{A} \tag{B16.4}$$

where:

- $\vec{\mu}$ is the magnetic moment of the coil of wire.
- N is the number of times the wire was wrapped around to form the coil. N is called the number of windings. N is also known as the number of turns.
- I is the current in the coil. The coil consists of one long wire wrapped around many times, so, there is only one current in the wire. We call that one current the current in the coil.
- $\vec{A}$ is the area vector of the loop or coil. Its magnitude is the area of the plane shape whose perimeter is the loop or coil. Its direction is the direction your extended right thumb would point if you curled the fingers of your right hand around the loop in the direction of the current.

Some Generalizations Regarding the Effect of a Uniform Magnetic Field on a Current Loop

We investigated the effect of a uniform magnetic field on a current loop. A magnetic field will exert a torque on a current loop whether or not the magnetic field is uniform. Since a current loop has some spatial extent (it is not a point particle), using a single value-plus-direction for $\vec{B}$ in $\vec{\tau} = \vec{\mu} \times \vec{B}$ will yield an approximation to the torque. It is a good approximation as long as the magnetic field is close to being uniform in the region of space occupied by the coil.

We investigated the case of a rectangular loop. The result for the torque exerted on the current-carrying loop or coil is valid for any plane loop or coil, whether it is circular, oval, or rectangular.

B17: Magnetic Field: Causes

This chapter is about magnetism but let's think back to our introduction to charge for a moment. We talked about the electric field before saying much about what caused it. We said the electric field exerts a force on a particle that has charge. Later we found out that charged particles play not only the role of "victim" to the electric field but, that charged particles cause electric fields to exist.

Now we have been talking about the magnetic field. We have said that the magnetic field exerts a torque on a particle that has magnetic dipole moment. You might guess that a particle that has magnetic dipole moment would cause a magnetic field. You'd be right! A particle that has the physical property known as magnetic dipole moment causes a magnetic field to exist in the region of space around it. A magnetic field can be caused to exist by a particle having magnetic dipole moment or a distribution of particles having magnetic dipole moment.

The magnetic field at point P, an empty point in space in the vicinity of a particle that has a magnetic dipole moment, due to that particle-with-magnetic-dipole-moment, is given by

$$\vec{B}=\frac{\mu_o}{4\pi}\frac{3(\vec{\mu}\cdot\hat{r})\hat{r}-\vec{\mu}}{r^3} \tag{B17.1}$$

where

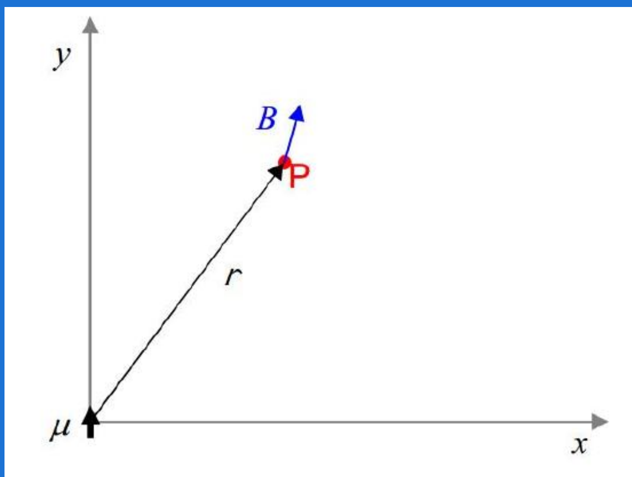
- $\mu_o=4\pi\times 10^{-7}\,\frac{T\cdot m}{A}$ is a universal constant which goes by the name of "the magnetic permeability of free space." This value is to be taken as exact. (Do not treat the "4" as a value known to only one significant digit.)
- $\vec{B}$ is the magnetic field vector at point P, where P is an empty point in space a distance r away from the particle-with-magnetic-dipole-moment that is causing $\vec{B}$.
- $\vec{\mu}$ is the magnetic dipole moment of the particle that is causing the magnetic field.
- $\hat{r}$ is a unit vector in the direction "from the particle, toward point P". Defining $\vec{r}$ to be the position vector of point P relative to the location of the particle-with-magnetic-dipole-moment, $\vec{r}=r\hat{r}$ so $\hat{r}=\frac{\vec{r}}{r}$.
- r is the distance that point P is from the particle-with-magnetic-dipole-moment.

A particle-with-magnetic-dipole-moment is called a magnetic dipole. Note that the magnetic field due to a magnetic dipole dies off like $\frac{1}{r^3}$.

A particle is at the origin of a Cartesian coordinate system. The magnetic dipole moment of the particle is $1.0\,A\cdot m^2\,\hat{j}$. Find the magnetic field vector, due to the particle, at $(3.0cm, 4.0cm)$.

Solution

I'm going to start with a diagram of the configuration.



Note that I do not know the direction of B in advance, so, I have drawn $\vec{B}$ on the diagram in a fairly arbitrary direction. I did want to put $\vec{B}$ on there to make it more evident that we are dealing with the magnetic field at point P, caused by the particle at the origin. Also, intentionally drew $\vec{B}$ in a direction other than that of $\vec{r}$, to avoid conveying the false impression that $\vec{B}$ is necessarily in the direction of $\vec{r}$. (At some points, it is, but those points are the exception. In general, $\vec{B}$ is not in the same direction as $\vec{r}$. As we shall soon see, for the case at hand, it turns out that $\vec{B}$ is not in the same direction as $\vec{r}$.)

Given $x = 0.030m$ and $y = 0.040m$, the position vector, for point P is $\vec{r} = 0.030m\,\hat{i} + 0.040m\,\hat{j}$. The magnitude of $\vec{r}$ is given by:

$$r = \sqrt{x^2 + y^2} \quad r = \sqrt{(.030m)^2 + (.040m)^2} \quad r = .050m$$

The unit vector $\hat{r}$ is thus given by:

$$\hat{r} = \frac{\vec{r}}{r}$$

$$\hat{r} = \frac{0.030m\,\hat{i} + 0.040m\,\hat{j}}{0.050m} \quad \hat{r} = 0.60\hat{i} + 0.80\hat{j}$$

Substituting what we have into our expression for $\vec{B}$ we find:

$$\vec{B} = \frac{\mu_o}{4\pi}\frac{3(\mu\cdot\hat{r})\hat{r} - \vec{\mu}}{r^3}$$

$$\vec{B} = \frac{4\pi\times 10^{-7}T\cdot m/A}{4\pi}\,\frac{3[(1.0A\cdot m^2\hat{j})\cdot(0.60\hat{i} + 0.80\hat{j})](0.60\hat{i} + 0.80\hat{j}) - 1.0A\cdot m^2\hat{j}}{(.050m)^3}$$

$$\vec{B} = 1\times 10^{-7}\frac{T\cdot m}{A}\,\frac{3[0.80A\cdot m^2](0.60\hat{i} + 0.80\hat{j}) - 1.0A\cdot m^2\hat{j}}{(.050m)^3}$$

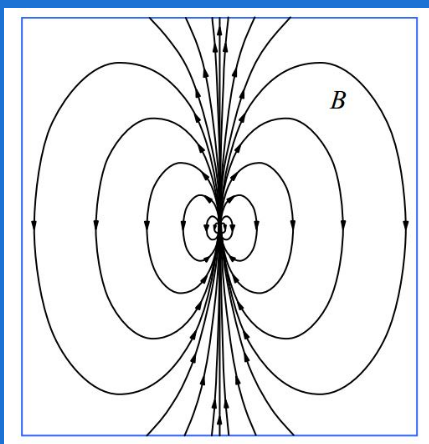
$$\vec{B}=1\times 10^{-7}\frac{T\cdot m}{A}\frac{1.44A\cdot m^2\hat{i}+1.92A\cdot m^2\hat{j}-1.0A\cdot m^2\hat{j}}{(.050m)^3}$$

$$\vec{B}=1\times 10^{-7}\frac{T\cdot m}{A}\frac{1.44A\cdot m^2\hat{i}+0.92A\cdot m^2\hat{j}}{(.050m)^3}$$

$$\vec{B}=1\times 10^{-7}\frac{T\cdot m}{A}(11520\frac{A}{m}\hat{i}+7360\frac{A}{m}\hat{j})$$

$$\vec{B}=1.152mT\hat{i}+.736mT\hat{j}$$

So ends our solution to the sample problem. Here's a magnetic field diagram of the magnetic field due to a particle that has a magnetic dipole moment.



The Magnetic Field Due to a Loop or Coil

We discovered in the last chapter that, as a victim to a magnetic field, a current loop or a currentcarrying coil behaves as if it were a particle with a magnetic dipole moment

$$\vec{\mu}=NI\vec{A}$$

where:

- $\vec{\mu}$ is the magnetic moment of the coil of wire.
- N is the number of windings, a.k.a. the number of turns. ($N=1$ in the case of a loop.)
- I is the current in the coil.
- $\vec{A}$ is the area vector of the loop or coil. Its magnitude is the area of the plane shape whose perimeter is the loop or coil. Its direction is the direction your extended right thumb would point if you curled the fingers of your right hand around the loop in the direction of the current.

You might guess that if a coil of wire responds to a magnetic field as if it were a particle with a magnetic dipole moment, then perhaps it can also behave as a source of magnetic field lines and create the same kind of magnetic field that a particle with a magnetic dipole moment produces. Indeed it does. As compared to a particle like the electron that has a magnetic dipole moment but itself has no extent in space, a loop or coil of wire does have extent in space. The magnetic field very near the loop or coil is more complicated than a dipole field, but, at points whose distance from the loop or coil are large compared to the diameter of the coil, the magnetic field of the loop or coil is the dipole magnetic field

$$\vec{B}=\frac{\mu_o}{4\pi}\frac{3(\vec{\mu}\cdot\hat{r})\hat{r}-\vec{\mu}}{r^3}$$

In the case of a loop or coil, the $\vec{\mu}$ that appears in this equation is $\vec{\mu}=NI\vec{A}$.

A Bar Magnet

An atom is made of a nucleus containing neutrons and protons; and; electrons in orbit about the nucleus. Each of these elementary particles has a magnetic moment. The magnetic moment of the electron is $9.28 \times 10^{-24} A \cdot m^2$, the magnetic moment of the proton is $1.41 \times 10^{-26} A \cdot m^2$, and, the magnetic moment of the neutron is $9.66 \times 10^{-27} A \cdot m^2$. When these particles combine to form atoms, they each contribute to the magnetic field of the atom. In addition to these contributions to the magnetic field, the protons move in loops within the nucleus and the electrons move in loops about the nucleus. A charged particle that is moving in a loop is a current loop and such current loops contribute to the overall magnetic field of the atom. In many atoms the various contributions to the magnetic field cancel each other out in such a manner that the overall magnetic field is essentially zero. In some atoms, such as iron, cobalt, and neodymium, the various contributions to the magnetic field do not cancel out. In such cases, the observed total magnetic field of the atom is a dipole magnetic field, and, the atom behaves as a magnetic dipole. Substances consisting of such atoms are referred to as ferromagnetic materials.

Consider an iron rod or bar that is not a magnet. The bar was formed from molten iron. As the iron cooled, seed crystals formed at various locations within the iron. At the start of crystallization, the iron atoms forming the seed crystal tend to align with each other, south pole to north pole. The magnetic field of the seed crystal causes neighboring iron atoms to align with the seed crystal magnetic dipole moment so that when they crystallize and become part of the growing crystal they also align south pole to north pole. The contributions of the atoms making up the crystal to the magnetic field of the crystal tend to add together constructively to form a relatively large magnetic field. There is a multitude of sites at which crystals begin to form and at each site, in the absence of an external magnetic field, the seed crystal is aligned in a random direction. As the crystals grow, they collectively form a multitude of microscopic bar magnets. When the iron bar is completely solidified it consists of a multitude of microscopic bar magnets called domains. Because they are aligned in random directions, their magnetic fields cancel each other out. Put the iron rod or bar in a magnetic field and the magnetic field will cause the microscopic bar magnets, the domains, in the iron to line up with each other to an extent that depends on the strength of the magnetic field. This turns the iron rod or bar into a magnet. Remove the rod or bar from the magnetic field and local forces on the domains cause them to revert back toward their original orientations. They do not achieve their original orientations and the iron remains at least weakly magnetized, an effect known as hysteresis.

Getting back to the cooling process, if we allow the molten iron to crystallize within an external magnetic field, the seed crystals, will all tend to line up with the external magnetic field, and hence, with each other. When the iron is completely solidified, you have a permanent magnet.

So a bar magnet consists of a bunch of microscopic bar magnets which themselves consist of a bunch of atoms each of which has a magnetic dipole moment because it consists of particles that each have a magnetic dipole moment and in some cases have charge and move in a loop within the atom.

The magnetic field of a bar magnet is thus the superposition (vector sum at each point in space) of a whole lot of magnetic dipole fields. As such, at distances large compared to the length of the magnet, the magnetic field of a bar magnet is a magnetic dipole field. As such, we can assign, based on measurements, a magnetic dipole vector $\vec{\mu}$ to the bar magnet as a whole, and compute its magnetic field (valid for distances large compared to the length of the magnet) as

$$\vec{B} = \frac{\mu_o}{4\pi} \frac{3(\vec{\mu} \cdot \hat{r})\hat{r} - \vec{\mu}}{r^3}$$

The Dipole-Dipole Force

The magnetic field produced by one bar magnet will exert a torque on another bar magnet. Because the magnetic field due to a magnetic dipole is non-uniform (you can see in

$$\vec{B} = \frac{\mu_o}{4\pi} \frac{3(\vec{\mu} \cdot \hat{r})\hat{r} - \vec{\mu}}{r^3}$$

that it dies off like $\frac{1}{r^3}$), it also exerts a force on another bar magnet.

We are now in a position to say something quantitative about the force that one bar magnet exerts on another. Consider an object that is at the origin of a Cartesian coordinate system. Suppose that object to have a magnetic dipole moment given by $\vec{\mu}_1 = \mu_1 \hat{i}$. Clearly we're talking about a magnet pointing (treating the magnet as an arrow with its head at the north pole of the magnet) in the $+x$ direction. Let's find the force that that magnet would exert on another one at $(x, 0, 0)$ given that the magnetic dipole moment of the second magnet is $\vec{\mu}_2 = -\mu_2 \hat{i}$. The second magnet is pointing back toward the origin, so we are talking about two magnets whose north poles are facing each other. Knowing that like poles repel, you should be able to anticipate that the second magnet will experience a force in the $+x$ direction. The magnetic field produced by the first magnet is given (for any point in space, as long as the distance to that point, from the origin, is large compared to the size of the magnet) by

$$\vec{B} = \frac{\mu_o}{4\pi}\frac{3(\vec{\mu}\cdot\hat{r})\hat{r}-\vec{\mu}}{r^3}$$

$$\vec{B}_1 = \frac{\mu_o}{4\pi}\frac{3(\mu_1\hat{i}\cdot\hat{r})\hat{r}-\mu_1\hat{i}}{r^3}$$

$$\vec{B}_1 = \frac{\mu_o}{4\pi}\Big(\frac{3(\mu_1\hat{i}\cdot\vec{r})\vec{r}}{r^5}-\frac{\mu_1\hat{i}}{r^3}\Big)$$

The force on the second particle is given by:

$$\vec{F} = \nabla(\vec{\mu}_2\cdot\vec{B}_1)$$

evaluated at the position of magnet 2, namely at $(x, 0, 0)$. Substituting the given $\vec{\mu}_2 = -\mu_2\hat{i}$ in for the magnetic dipole of particle 2, and, the expression just above for $\vec{B}_1$, we obtain:

$$\vec{F} = \nabla\Big\{-\mu_2\hat{i}\cdot\Big[\frac{\mu_o}{4\pi}\Big(\frac{3(\mu_1\cdot\vec{r})\vec{r}}{r^5}-\frac{\mu_1\hat{i}}{r^3}\Big)\Big]\Big\}$$

$$\vec{F} = -\frac{\mu_o}{4\pi}\mu_1\mu_2\nabla\Big\{\hat{i}\cdot\Big[\Big(\frac{3(\hat{i}\cdot\vec{r})\vec{r}}{r^5}-\frac{\hat{i}}{r^3}\Big)\Big]\Big\}$$

Now, if you substitute $\vec{r} = x\hat{i}+y\hat{j}+z\hat{k}$ and $r = \sqrt{x^2+y^2+z^2}$, take the gradient, and then (after taking the gradient) evaluate the result at $(x, 0, 0)$, you find that

$$\vec{F} = \frac{3\mu_o}{2\pi}\frac{\mu_1\mu_2}{x^4}\hat{i}$$

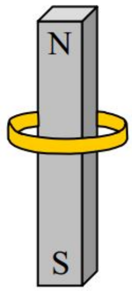
So, when like poles are facing each other, two magnets repel each other with a force that dies off like $\frac{1}{r^4}$ where r is the distance between the magnets (measure it center to center), the x in the case that we investigated.

B18: Faraday's Law and Lenz's Law

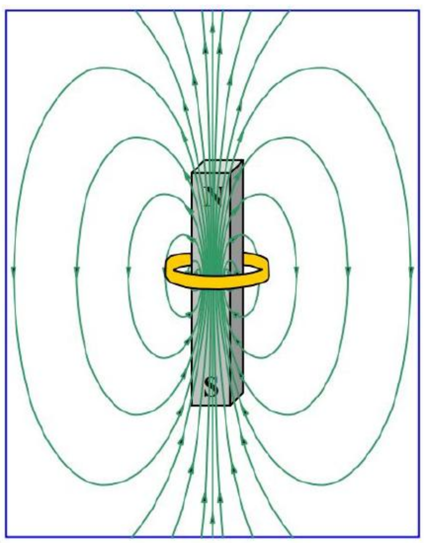
Do you remember Archimedes's Principle? We were able to say something simple, specific, and useful about a complicated phenomenon. The gross idea was that a submerged object being pressed upon on every surface element in contact with fluid, by the fluid, experiences a net upward force because the pressure in a fluid is greater at greater depth. The infinite sum, over all the surface area elements of the object in contact with the fluid, of the force of magnitude pressure-times-the area, and direction normal to and into the area element, resulted in an upward force that we called the buoyant force. The thing is, we were able to prove that the buoyant force is equal in magnitude to the weight of that amount of fluid that would be where the object is if the object wasn't there. Thus we can arrive at a value for the buoyant force without having to even think about the vector integration of pressure-related force that causes it.

We are about to encounter another complicated phenomenon which can be characterized in a fruitful fashion by a relatively simple rule. I'm going to convey the idea to you by means of a few specific processes, and then sum it up by stating the simple rule.

Consider a gold ring and a bar magnet in the hands of a person. The person is holding the ring so that it encircles the bar magnet. She is holding the magnet, north end up.

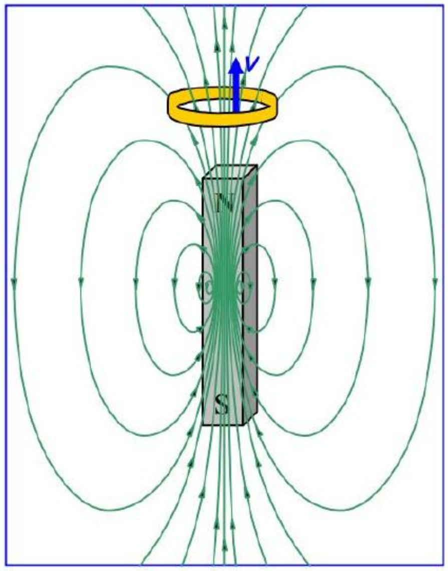


There is a magnetic field, due to the bar magnet, within the bar magnet, and in the region of space around it.

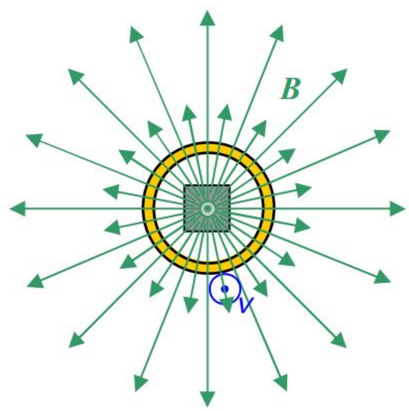


It is important to note that the magnetic field lines are most densely packed inside the bar magnet.

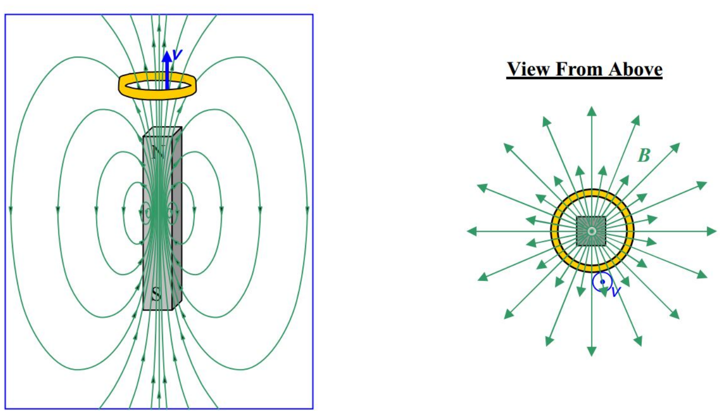
Now suppose that the person, holding the magnet at rest in one hand, moves the loop upward. I want to focus on what is going on while she is moving it upward. As she moves the loop upward, she is moving it roughly along the direction of the magnetic file lines, but, and this is actually the important part, that loop will also be "cutting through" some magnetic field lines. Consider an instant in time when the loop is above the magnet, and moving upward:



From above, the scene looks like:



where it is important to realize that none of those magnetic field lines begin on the magnet or end at the tip of the arrow depicted, rather, they extend out of the magnet toward us, flower out and over, back down away from us, and then they loop around to enter the south pole of the magnet from which they extend back up through the magnet toward us. In fact, no magnetic field line ever begins or ends anywhere. They all form closed loops. This is a manifestation of the fact that there is no such thing as magnetic charge. (There are no magnetic monopoles.)



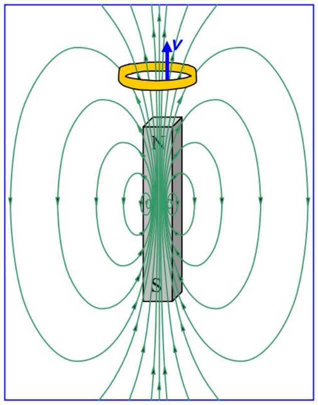
Here's where we're going with this: The motion of the ring relative to the magnet is going to cause a current in the ring. Here's how: The ring is neutral, but, it is chock full of charged particles that are free to move around within the gold. [I'm going to discuss it in our positive charge carrier model but you can verify that you get the same result if the charge carriers are negative (recalling that our current is in the direction opposite that in which negative charge carriers are moving.)] Pick any short segment of the ring and get the direction of the force exerted on the charge carriers of that segment using $\vec{F} = q\vec{v} \times \vec{B}$ and the right-hand rule for the cross product of two vectors. In the view from above, all we can see is the horizontal component of the magnetic field vectors in the vicinity of the moving ring but that's just dandy; the vertical component, being parallel to the ring's velocity (and hence parallel to the velocity of the charge in the ring), makes no contribution to $\vec{v} \times \vec{B}$. Now, pick your segment of the ring. Make your fingers point away from your elbow, and, in the direction of the first vector (the velocity vector) in $\vec{v} \times \vec{B}$, namely, "out of the page". Now, keeping your fingers pointing both away from your elbow, and, out of the page, rotate your forearm as necessary so that your palm is facing in the direction of $\vec{B}$ (at the location of the segment you are working on), meaning that if you were to close your fingers, they would point in the direction of $\vec{B}$. Your extended thumb is now pointing in the direction of the force exerted on the positive charge carriers in the ring segment you chose. No matter what ring segment you pick, the force is always in that direction which tends to push the positive charge carriers counterclockwise around the ring! The result is a counterclockwise (as viewed from above) current in the ring.

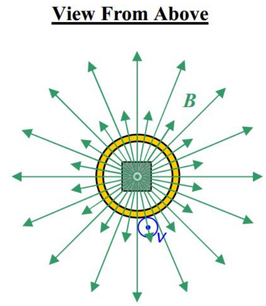
Suppose that, starting with the ring encircling the magnet, the person who was holding the ring and the magnet moved the magnet downward rather than moving the ring upward. She holds the ring stationary, and moves the magnet. I said earlier that a charged particle at rest in a magnetic field has no force exerted on it by the magnetic field. But we were talking about stationary magnetic fields at the time. Now we are talking about the magnetic field of a magnet that is moving. Since the magnet responsible for it is moving, the magnetic field itself must be moving. Will that result in a force on the charges in the ring (and hence a current in the ring)? This brings us to a consideration of relative motion. To us, the two processes (person moves ring upward while holding magnet at rest, vs. person moves magnet downward while holding ring at rest) are different. But that is just because we are so used to viewing things from the earth's reference frame. Have you ever been riding along a highway and had the sense that you were at rest and the lampposts on the side of the road were moving past you at high speed. That is a valid viewpoint. Relative to your reference frame of the car, the lampposts are indeed moving and the car is a valid reference frame. Suppose we view the magnet moving downward through a ring situation from a platform that is moving downward at the same speed as the magnet. In that reference frame, the magnet is at rest. If for instance, as we, while seated on the platform, see the magnet at eye level, it remains at eye level. But the ring, as viewed from the platform reference frame is moving upward. So in the platform reference frame, we have, in the new processes (which in the room reference frame is a magnet moving downward through and away from a ring) the same situation that we had in the room frame for the original process (which in the room reference frame is a ring, originally encircling a stationary magnet, moving upward). Thus in the platform reference frame, we must have the same result for the new process that we had for the original process in the room frame, namely, a counterclockwise (as viewed from above) current in the ring. The current in the ring doesn't depend on what reference frame we view the ring from. Hence, we can conclude that the magnet moving downward through the stationary ring at speed v results in the same current as we have when the ring moves upward at the same speed v relative to the stationary magnet.

When the person holding the magnet and the ring moved the ring upward, there was a current in the ring. Now we have established that if, instead of moving the ring, she moves the magnet downward at the same speed, she will get the same current in the ring. Based on what caused that current, the $\vec{F} = q\vec{v} \times \vec{B}$ force on the charged particles in the ring, you can surmise that the current will depend on things like the velocity of the ring relative to the magnet, the strength of the magnetic field, and the relative orientation of the velocity vector and the magnetic field. It has probably occurred to you that the current also depends on the resistance of the ring.

Michael Faraday came up with a very fruitful way of looking at the phenomenon we are discussing and I will convey his idea to you by means of the example we have been working with.

Looking at the diagrams of that ring moving relative to the magnet again,

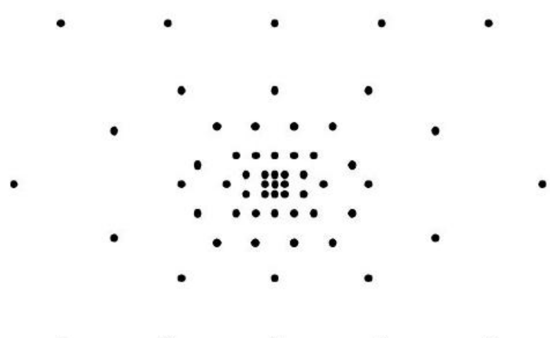




we can describe what's happening by saying that the ring is "cutting through" magnetic field lines (or, equivalently, by saying that the magnetic field lines are "cutting through" the ring). What Faraday recognized was that, in conceptual terms, by the ring cutting through magnetic field lines (or vice versa depending on your point of view), what was happening was, that the number of magnetic field lines encircled by the loop was changing. In the diagrams above, each time the ring "cuts through" one more field line, the number of field lines encircled by the loop decreases by one. The rate at which the ring "cuts through" magnetic field lines (or the magnetic field lines cut through the ring) is determined by the same things that determine the force on the charged particles making up the ring (relative speed between ring and magnetic field, strength of magnetic field, relative orientation of velocity of ring and magnetic field) such that, the greater the rate at which the ring "cuts through" magnetic field lines (or the greater the rate at which magnetic field lines cut through the ring), the greater the force on the charged particles and hence the greater the current. Faraday expressed this in a manner that is easier to analyze. He said that the current is determined by the rate at which the number of magnetic field lines encircled by the loop is changing. In fact, Faraday was able to write this statement in equation form. Before I show you that, I have to be a lot more specific about what I mean by "the number of magnetic field lines."

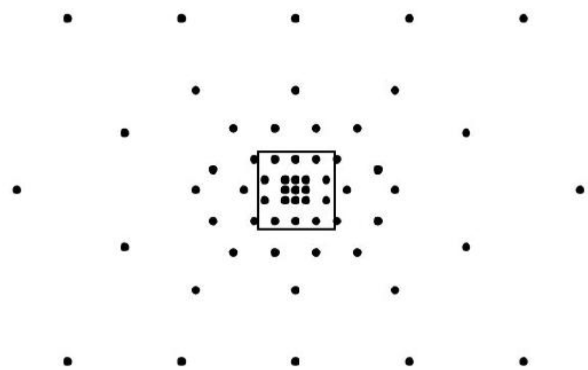
I'm going to call the statement I have just attributed to Faraday, the conceptual form of Faraday's Law. In other words, Faradays Law, in conceptual form is: A changing number of magnetic field lines through a closed loop or coil causes a current in that loop or coil, and, the faster the number is changing, the greater the current.

Our field line concept is essentially a diagrammatic scheme used to convey some information about the direction and the relative strength of a field. We have used it both for the electric field and the magnetic field. What I say here about the number of field lines can be applied to both, but, since we are presently concerned with the magnetic field, I will talk about it in terms of the magnetic field. Conceptually, the number of field lines encircled by a loop is going to depend on how closely packed the field lines are, how big the loop is, and to what degree the loop is oriented "face-on" to the field lines. (Clearly, if the loop is oriented edge-on to the field lines, it will encircle none of them.) Now, diagrammatically, how closely packed the field lines are is representative of how strong the magnetic field is. The more closely-packed the field lines, the greater the value of B. Imagine that someone has created a beautiful, three-dimensional, magnetic field diagram. Now if you view the field lines end-on, e.g. such that the magnetic field lines are directed right at you, and depict a cross section of "what you see" in a two-dimensional diagram, you would get something like this.



This is a graphical representation of the magnitude of that component of the magnetic field which is directed straight at you.

Suppose the scale of the diagram to be given by $(1\mu T \cdot m^2)$n where n is the magnetic field line density, the number-of-magnetic-field-lines-per-area, directed through the plane represented by the page, straight at you. Let's use a square, one centimeter on a side, to sample the field at a position near the center,



I count 19 field lines that are clearly in the square centimeter and four that are touching it, I'm going to count two of those four for an estimated 21 field lines in one square centimeter. Thus, in that region,

$$n = \frac{21 \text{ lines}}{(1 \times 10^{-2} m)^2}$$

$$n = 2100 \frac{\text{lines}}{\text{m}^2}$$

Using the given scale factor,

$$B = (1.0\mu T \cdot m^2)n$$

$$B = (1.0\mu T \cdot m^2) 2100 \frac{\text{lines}}{\text{m}^2}$$

$$B = 2.1 mT$$

Let's make it more clear what the number of lines represents by replacing n with $\frac{\text{Number of Lines}}{A}$ and solving the expression $V = (1.0\mu T)n$ for the number of lines.

$$B = (1.0\mu T \cdot m^2) \frac{\text{Number of Lines}}{A}$$

$$\text{Number of Lines} = \frac{BA}{1.0\mu T \cdot m^2}$$

So the number of lines through a loop encircling a plane region of area A is proportional to BA, with the constant of proportionality being the reciprocal of our scale factor for the field diagram. The simple product BA is really only good if the magnetic field lines are "hitting" the area encircled by the loop "head on," and, if the magnetic field is single-valued over the whole area. We can take care of the "which way the loop is facing" issue by replacing BA with $\vec{B} \cdot \vec{A}$ where $\vec{A}$, the area vector, is a vector whose magnitude is the area of the plane region encircled by the loop and whose direction is perpendicular to the plane of the loop. There are actually two directions that are perpendicular to the loop. One is the opposite of the other. In practice, one picks one of the two directions arbitrarily, but, picking a direction for the area vector establishes a positive direction for the current around the loop. The positive direction for the current is the direction around the loop that makes the current direction and the area vector direction, together, conform to the right-hand rule for something curly something straight. We take care of the possible variation of the magnetic field over the region enclosed by the loop, by cutting that plane region up into an infinite number of infinitesimal area elements dA, calculating $\vec{B} \cdot \overrightarrow{dA}$ for each area element, and adding up all the results. The final result is the integral $\int \vec{B} \cdot \overrightarrow{dA}$. You won't be held responsible for using the calculus algorithms for analyzing such an integral, but, you are responsible for knowing what $\int \vec{B} \cdot \overrightarrow{dA}$ means. It is the infinite sum you get when you subdivide the area enclosed by the loop up into an infinite number of infinitesimal area elements, and, for each area element, dot the magnetic field vector at the location of that area element into the area vector of that area element, and add up all the resulting dot products. You also need to know that, in the special case of a magnetic field that is constant in both magnitude and direction over the entire area enclosed by the loop, $\int \vec{B} \cdot \overrightarrow{dA}$ is just $\vec{B} \cdot \vec{A}$.

Using a generic "constant" for the reciprocal of the field diagram scale factor yields

$$\text{Number of Lines} = (constant) \int \vec{B} \cdot \overrightarrow{dA}$$

for the number of field lines encircled by the loop or coil. The quantity $\int \vec{B} \cdot \overrightarrow{dA}$ is called the magnetic flux through the plane region enclosed by the loop. Note that the flux is directly proportional to the number of magnetic field lines through the loop.

The magnetic flux is given the name Φ_B (the Greek letter upper case phi).

$$\Phi_B = \int \vec{B} \cdot \overrightarrow{dA}$$

The expression yields $T \cdot m^2$ as the units of magnetic flux. This combination of units is given a name, the Weber, abbreviated Wb.

$$1 Wb = T \cdot m^2$$

Faraday's Law, the one that says that the current induced in a loop or coil is proportional to the rate of change in the number of magnetic field lines encircled by the loop or coil, can be written in terms of the flux as:

$$I = -\frac{N}{R} \frac{d\Phi_B}{dt}$$

where:

N is the number of windings or turns making up the closed coil of wire. $N = 1$ for a single loop.

R is the resistance of the loop or coil.

$\frac{d\Phi_B}{dt}$ is the rate of change in the flux through the loop.

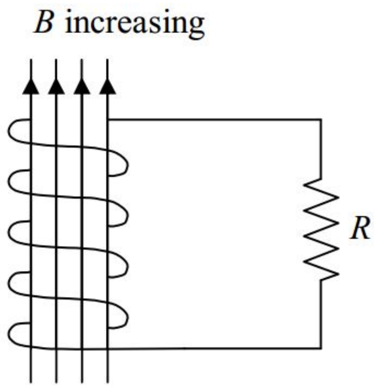
The derivative of a function with respect to time is often abbreviated as the function itself with a dot over it. In other words,

$$\dot{\Phi}_B = \frac{d\Phi_B}{dt}$$

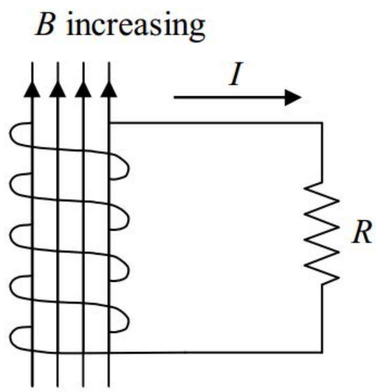
Using this notation in our expression for the current in Faradays Law of induction we have:

$$I = -\frac{N}{R} \dot{\Phi}_B$$

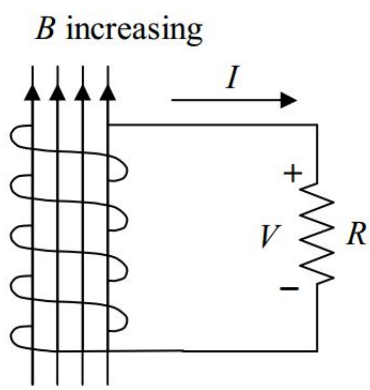
Faraday's Law is usually expressed in terms of an EMF rather than a current. I'm going to use a specific case study to develop the idea which is of general applicability. Consider a coil of ideally-conducting wire in series with a resistor. For closure of the loop, the resistor is to be considered part of the loop (and hence is the resistance of the loop), but, we have a negligible number of magnetic field lines cutting through the resistor itself. Suppose there to be an increasing magnetic flux directed upward through the coil.



By Faraday's Law of Induction, there will be a current $I=-\frac{N}{R}\dot{\Phi}_B$ induced in the coil. The charge will flow around and around the coil, out the top of the coil and down through the resistor.



But, for a resistor to have a current in it, there must be a potential difference $V=IR$ between the terminals of the resistor.



Recognizing that, in the case at hand, the I in $V=IR$ is the $I=-\frac{N}{R}\dot{\Phi}_B$ resulting from the changing magnetic flux through the coil, we have

$$V=\left(-\frac{N}{R}\dot{\Phi}_B\right)R$$

which we can write as

$$V=-N\dot{\phi}_B$$

Where there is a voltage across a resistor, there is an electric field in the resistor. What exactly causes that electric field? The answer is, the changing flux through the coil. More specifically, it is the magnetic field lines cutting through the coil as they must be doing to cause a change in the number of field lines through the coil. The field lines through the coil causes a force on the charge carriers in the coil. In our positive charge carrier model, this causes positive charge carriers in the coil all to surge toward the top of the resistor, leaving an absence of same on the bottom of the resistor. It only takes a minuscule amount of charge to cause an appreciable electric field in the resistor. A dynamic equilibrium is reached in which the changing magnetic field force on the charged particle becomes unable to push any more charge to the top terminal of the resistor than is forced through the resistor by the electric field in the resistor. The changing magnetic field can't push more charge there because of the repulsion of the charge that is already there. The changing magnetic field force in the coil maintains the potential difference across the resistor in spite of the fact that charge carriers keep "falling" through the resistor. This should sound familiar. A seat of EMF does the same thing. It maintains a constant potential difference between two conductors (such as the terminals of the resistor in the case at hand). The coil with the changing flux through it is acting like a seat of EMF. One says that the changing flux induces an EMF in the coil, calls that Faraday's Law of Induction, and writes:

$$\varepsilon = -N\dot{\Phi}_B$$

where:

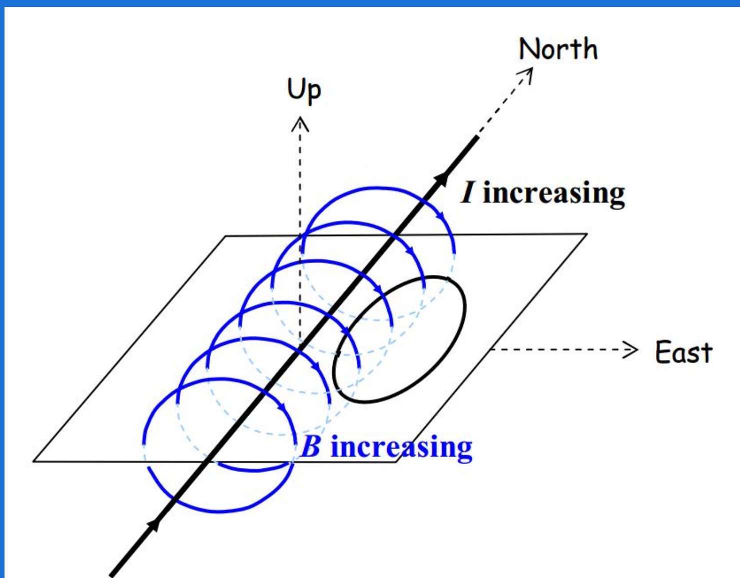
- ε is the EMF induced in the loop.
- N is the number of windings or turns making up the coil of wire.
- $\dot{\Phi}_B$ is the rate of change in the flux through the loop.

B19: Induction, Transformers, and Generators

In this chapter we provide examples chosen to further familiarize you with Faraday's Law of Induction and Lenz's Law. The last example is the generator, the device used in the world's power plants to convert mechanical energy into electrical energy.

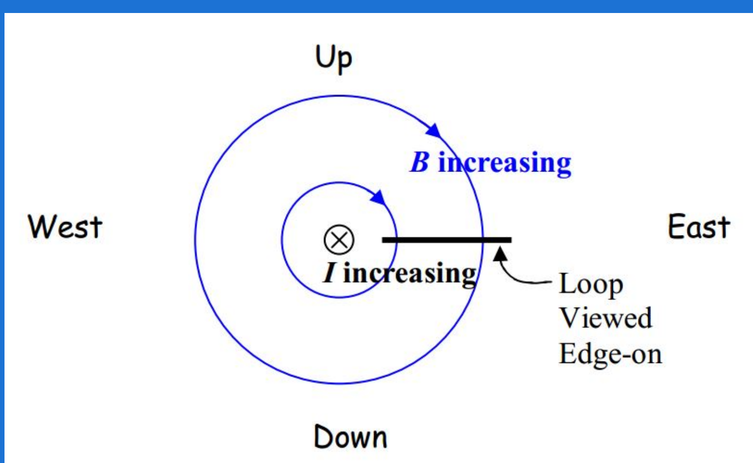
A straight wire carries a current due northward. Due east of the straight wire, at the same elevation as the straight wire, is a horizontal loop of wire. The current in the straight wire is increasing. Which way is the current, induced in the loop by the changing magnetic field of the straight wire, directed around the loop?

I'm going to draw the given situation from a few different viewpoints, just to help you get used to visualizing this kind of situation. As viewed from above-and-to-the-southeast, the configuration (aside from the fact that magnetic field lines are invisible) appears as:



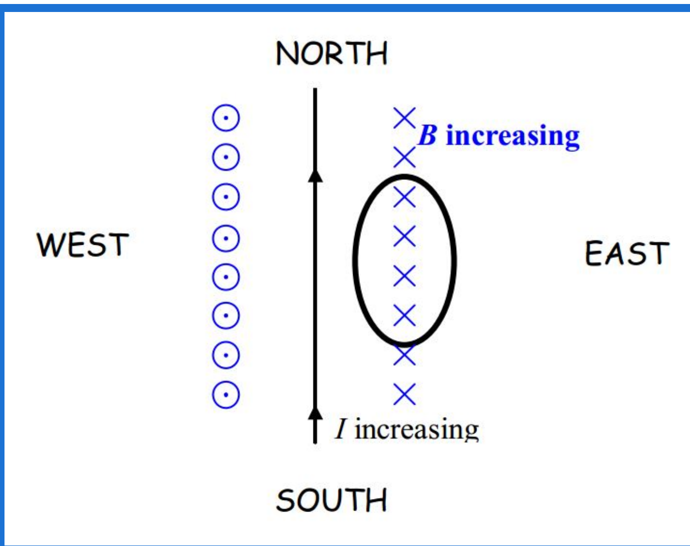
where I included a sheet of paper in the diagram to help you visualize things.

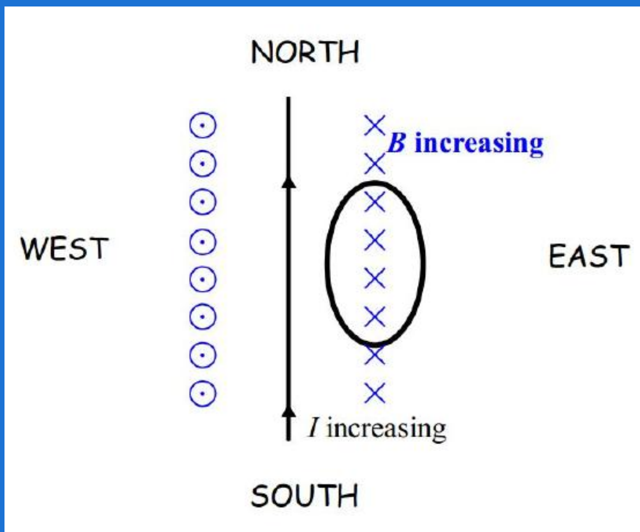
Here's a view of the same configuration from the south, looking due north:



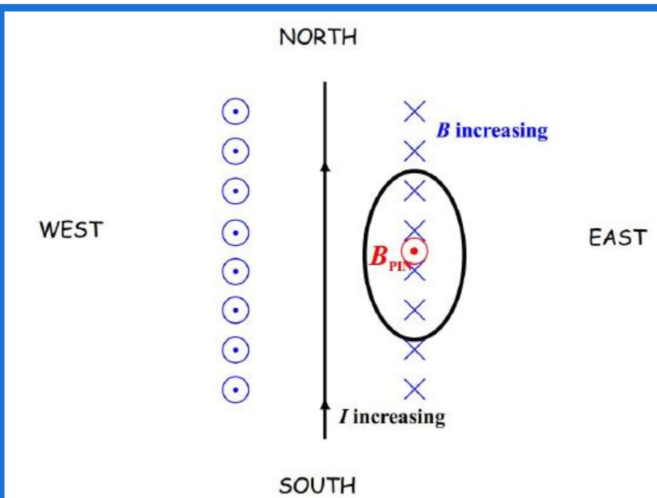
Both diagrams make it clear that we have an increasing number of downward-directed magnetic field lines through the loop. It is important to keep in mind that a field diagram is a diagrammatic manner of conveying information about an infinite set of vectors. There is no such thing as a curved vector. A vector is always directed along a straight line. The magnetic field vector is tangent to the magnetic field lines characterizing that vector. At the location of the loop, every magnetic field vector depicted in the diagram above is straight downward. While it is okay to say that we have an increasing number of magnetic field lines directed downward through the loop, please keep in mind that the field lines characterize vectors.

In presenting my solution to the example question, "What is the direction of the current induced in a horizontal loop that is due east of a straight wire carrying an increasing current due north?" I wouldn't draw either one of the diagrams above. The first one takes too long to draw and there is no good way to show the direction of the current in the loop in the second one. The view from above is the most convenient one:

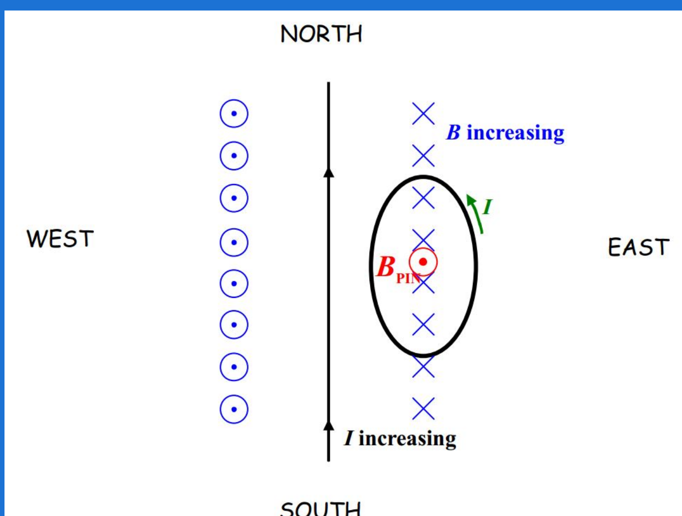




In this view (in which the downward direction is into the page) it is easy to see that what we have is an increasing number of downward-directed magnetic field lines through the loop (more specifically, through the region enclosed by the loop). In its futile attempt to keep the number of magnetic field lines directed downward through the loop the same as what it was, $\vec{B}_{PIN}$ must be directed upward in order to cancel out the newly-appearing downward-directed magnetic field lines. [Recall the sequence: The changing number of magnetic field lines induces (by Faraday's Law) a current in the loop. That current produces (by Ampere's Law) a magnetic field ($\vec{B}_{PIN}$) of its own. Lenz's Law relates the end product ($\vec{B}_{PIN}$) to the original change (increasing number of downward-through-the-loop magnetic field lines).]

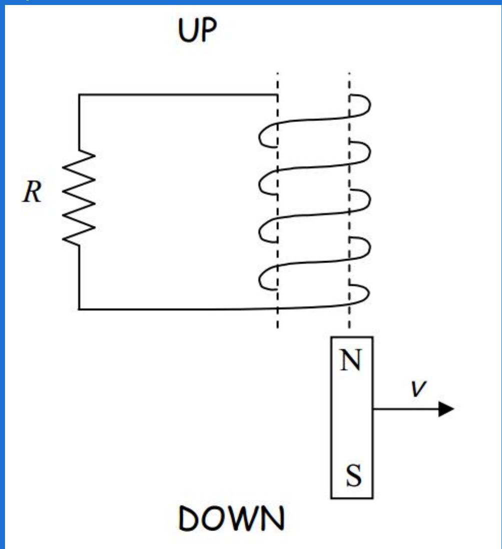


That's interesting. We know the direction of the magnetic field produced by the induced current before we even known the direction of the induced current itself. So, what must the direction of the induced current be in order to produce an upward-directed magnetic field ($\vec{B}_{PIN}$)? Well, by the right-hand rule for something curly something straight, the current must be counterclockwise, as viewed from above.

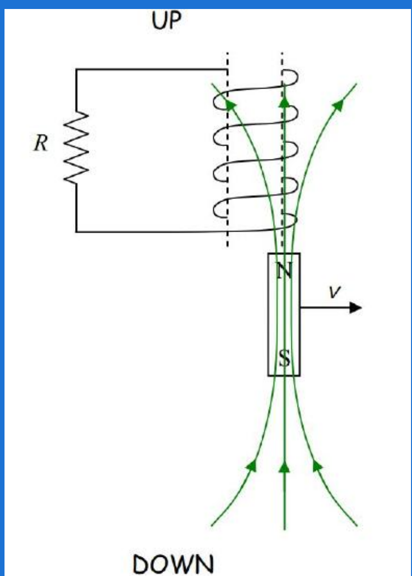


Hey. That's the answer to the question. We're done with that example. Here's another one:

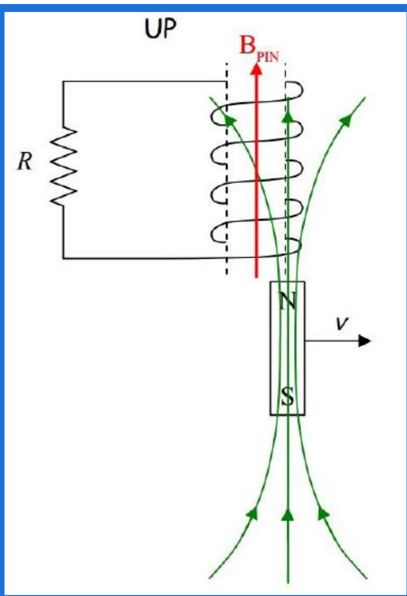
A person is moving a bar magnet, aligned north pole up, out from under a coil of wire, as depicted below. What is the direction of the current in the resistor?



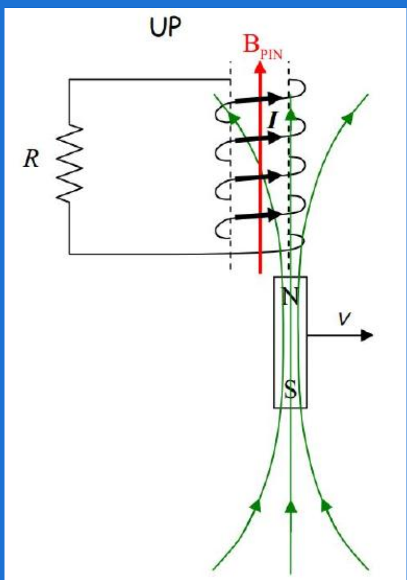
The magnetic field of the bar magnet extends upward through the coil.



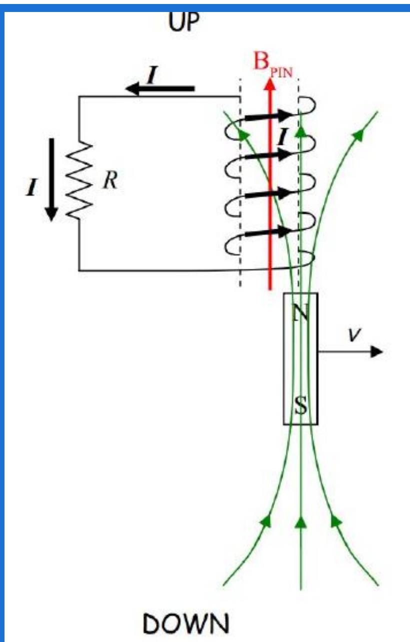
As the magnet moves out from under the coil, it takes its magnetic field with it. So, as regards the coil, what we have is a decreasing number of upward-directed magnetic field lines through the coil. By Faraday's Law, this induces a current in the coil. By Ampere's Law, the current produces a magnetic field, $\vec{B}_{PIN}$. By Lenz's Law $\vec{B}_{PIN}$ is upward, to make up for the departing upward-directed magnetic field lines through the coil.



So, what is the direction of the current that is causing $\vec{B}_{PIN}$? The right-hand rule will tell us that. Point the thumb of your cupped right hand in the direction of $\vec{B}_{PIN}$. Your fingers will then be curled around (counterclockwise as viewed from above) in the direction of the current.

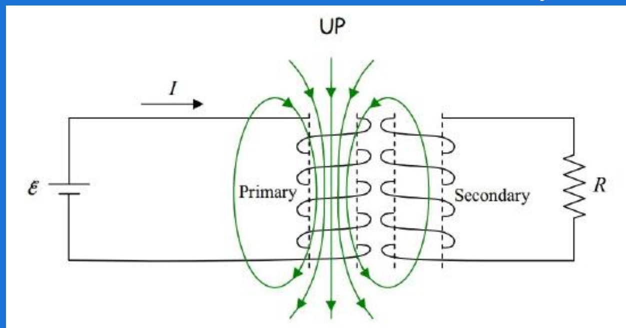


Because of the way the coil is wound, such a current will be directed out the top of the coil downward through the resistor.

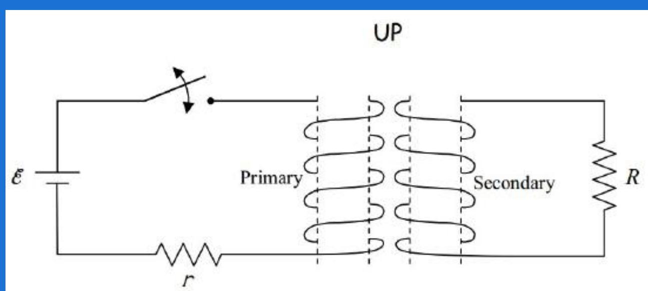


That's the answer to the question posed in the example. (What is the direction of the current in the resistor?)

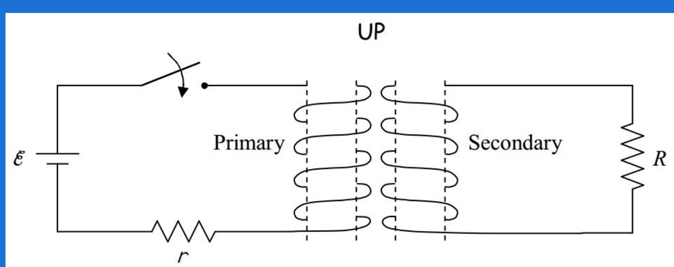
When you put two coils of wire near each other, such that when you create a magnetic field by using a seat of EMF to cause a current in one coil, that magnetic field extends through the region encircled by the other coil, you create a transformer. Let's call the coil in which you initially cause the current, the primary coil, and the other one, the secondary coil.



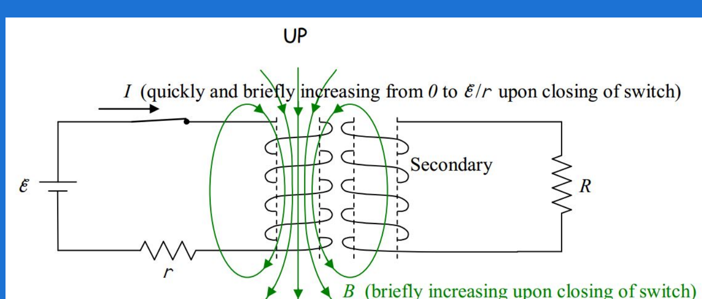
If you cause the current in the primary coil to be changing, then the magnetic field produced by that coil is changing. Thus, the flux through the secondary coil is changing, and, by Faraday's Law of Induction, a current will be induced in the secondary coil. One way to cause the current in the primary coil to be changing would be to put a switch in the primary circuit (the circuit in which the primary coil is wired) and to repeatedly open and close it.



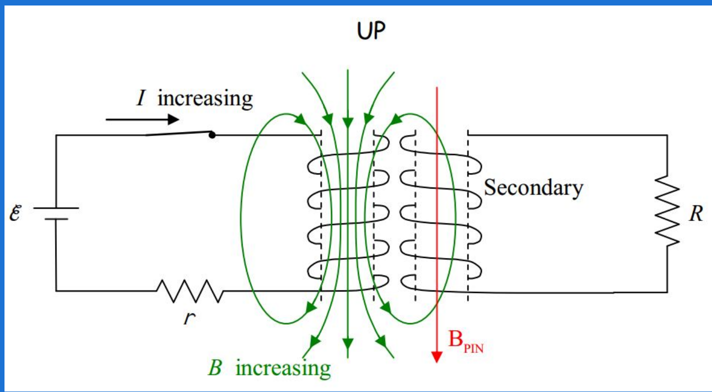
Okay, enough preamble, here's the question: What is the direction of the transient current induced in the circuit above when the switch is closed?



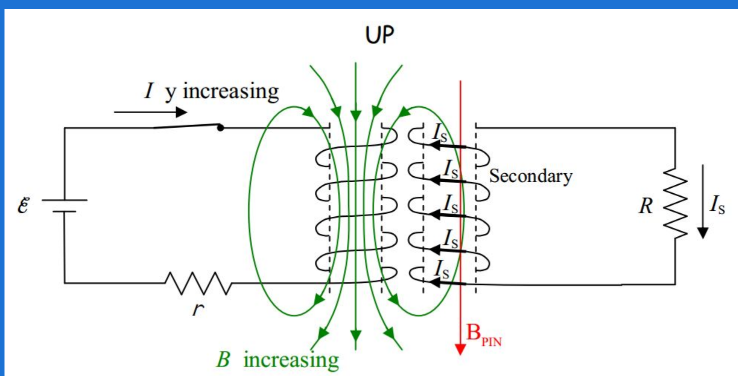
Upon closing the switch, the current in the primary circuit very quickly builds up to ε/r. While the time that it takes for the current to build up to ε/r is very short, it is during this time interval that the current is changing. Hence, it is on this time interval that we must focus our attention in order to answer the question about the direction of the transient current in resistor R in the secondary circuit. The current in the primary causes a magnetic field. Because the current is increasing, the magnetic field vector at each point in space is increasing in magnitude.



The increasing magnetic field causes upward-directed magnetic field lines in the region encircled by the secondary coil. There were no magnetic field lines through that coil before the switch was closed, so clearly, what we have here is an increasing number of upward-directed magnetic field lines through the secondary coil. By Faraday's Law this will induce a current in the coil. By Ampere's law, the current induced in the secondary will produce a magnetic field of its own, one that I like to call $\vec{B}_{PIN}$ for "The Magnetic field Produced by the Induced Current." By Lenz's Law, $\vec{B}_{PIN}$ must be downward to cancel out some of the newly-appearing upward-directed magnetic field lines through the secondary. (I hope it is clear that what I call the magnetic field lines through the secondary, are the magnetic field lines passing through the region encircled by the secondary coil.)



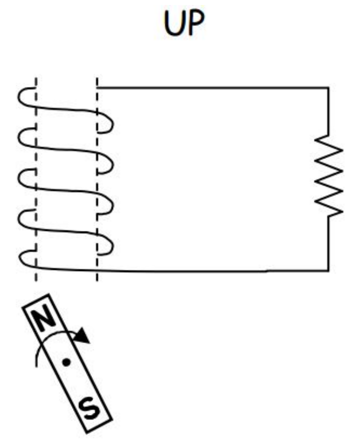
Okay. Now the question is, which way must the current be directed around the coil in order to create the downward-directed magnetic field $\vec{B}_{PIN}$ that we have deduced it does create. As usual, the right-hand rule for something curly something straight reveals the answer. We point the thumb of the cupped right hand in the direction of $\vec{B}_{PIN}$ and cannot fail to note that the fingers curl around in a direction that can best be described as "clockwise as viewed from above."



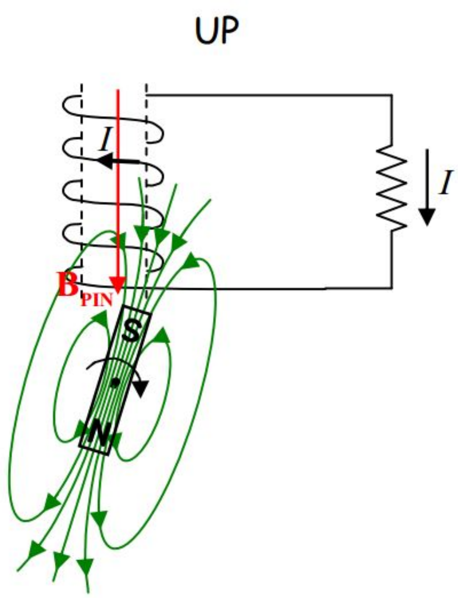
Because of the way the secondary coil is wound, such a current will be directed out of the secondary at the top of the coil and downward through resistor R. This is the answer to the question posed in the example.

An Electric Generator

Consider a magnet that is caused to rotate in the vicinity of a coil of wire as depicted below.

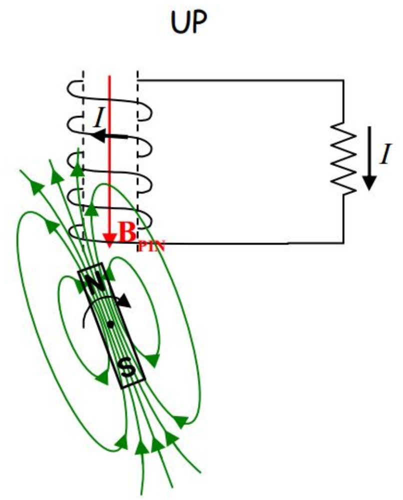


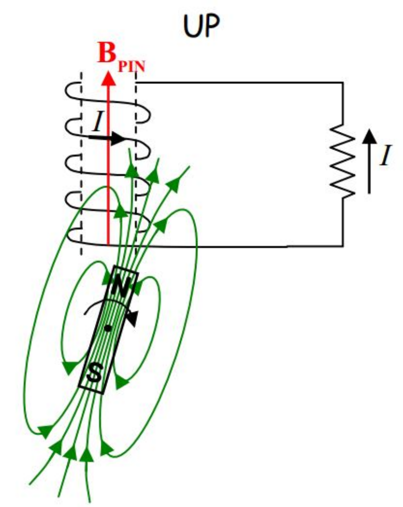
As a result of the rotating magnet, the number and direction of the magnetic field lines through the coil is continually changing. This induces a current in the coil, which, as it turns out, is also changing. Check it out in the case of magnet that is, from our viewpoint, rotating clockwise. In the orientation of the rotating magnet depicted here:

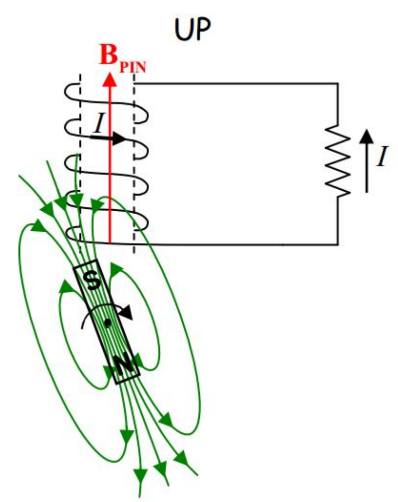


as the magnet rotates, the number of its magnetic field lines extending downward through the coil is decreasing. In accord with Faraday's Law, this induces a current in the coil which, in accord with Ampere's Law, produces a magnetic field of its own. By Lenz's Law, the field ($\vec{B}_{PIN}$) produced by the induced current must be downward to make up for the loss of downwarddirected magnetic field lines through the coil. To produce $\vec{B}_{PIN}$ downward, the induced current must be clockwise, as viewed from above. Based on the way the wire is wrapped and the coil is connected in the circuit, a current that is clockwise as viewed from above, in the coil, is directed out of the coil at the top of the coil and downward through the resistor.

In the following diagrams we show the magnet in each of several successive orientations. Keep in mind that someone or something is spinning the magnet by mechanical means. You can assume for instance that a person is turning the magnet with her hand. As the magnet turns the number of magnetic field lines is changing in a specific manner for each of the orientations depicted. You the reader are asked to apply Lenz's Law and the Right Hand Rule for Something Curley, Something Straight to verify that the current (caused by the spinning magnet) through the resistor is in the direction depicted:





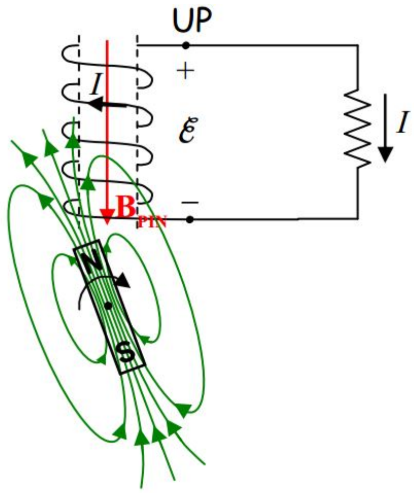


As the magnet continues to rotate clockwise, the next orientation it achieves is our starting point and the process repeats itself over and over again.

Recapping and extrapolating, the current through the resistor in the series of diagrams above, is:

downward, downward, upward, upward, downward, downward, upward, upward, …

For half of each rotation, the current is downward, and for the other half of each rotation, the current is upward. In quantifying this behavior, one focuses on the EMF induced in the coil:

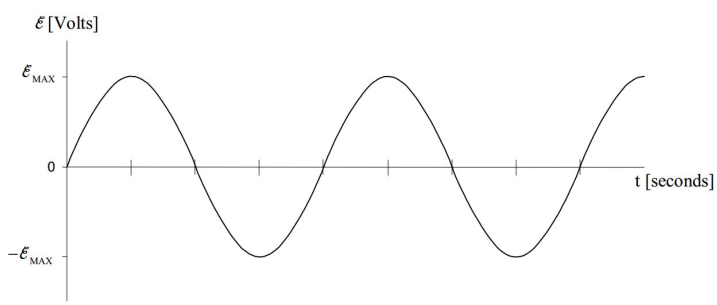


The EMF across the coil varies sinusoidally with time as:

$$\varepsilon=\varepsilon_{MAX}\sin(2\pi ft) \tag{B19.1}$$

where:

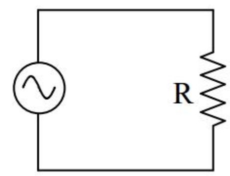
- ε which stands for EMF, is the time-varying electric potential difference between the terminals of a coil in close proximity to a magnet that is rotating relative to the coil as depicted in the diagrams above. This potential difference is caused to exist, and to vary the way it does, by the changing magnetic flux through the coil.
- ε_{MAX} is the maximum value of the EMF of the coil.
- f is the frequency of oscillations of the EMF across the coil. It is exactly equal to the rotation rate of the magnet expressed in rotations per second, a unit that is equivalent to hertz.



The device that we have been discussing (coil-plus-rotating magnet) is called a generator, or more specifically, an electric generator. A generator is a seat of EMF that causes there to be a potential difference between its terminals that varies sinusoidally with time. The schematic representation of such a time-varying seat of EMF is:

It takes work to spin the magnet. The magnetic field caused by the current induced in the coil exerts a torque on the magnet that always tends to slow it down. So, to keep the magnet spinning, one must continually exert a torque on the magnet in the direction in which it is spinning. The generator is the main component of any electrical power plant. It converts mechanical energy to electrical energy. The kind of power plant you are dealing with is determined by what your power company uses to spin the magnet. If moving water is used to spin the magnet, we call the power plant a hydroelectric plant. If a steam turbine is used to spin the magnet, then the power plant is designated by its method of heating and vaporizing water. For instance, if one heats and vaporizes the water by means of burning coal, one calls the power plant a coal-fired power plant. If one heats and vaporizes the water by means of a nuclear reactor, one calls the power plant a nuclear power plant.

Consider a “device which causes a potential difference between its terminals that varies sinusoidally with time” in a simple circuit:

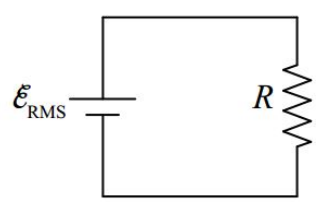


The time-varying seat of EMF causes a potential difference across the resistor, in this simple circuit, equal, at any instant in time, to the voltage across the time-varying seat of EMF. As a result, there is a current in the resistor. The current is given by $I=\frac{V}{R}$, our defining equation for resistance, solved for the current I. Because the algebraic sign of the potential difference across the resistor is continually alternating, the direction of the current in the resistor is continually alternating. Such a current is called an alternating current (AC). It has become traditional to use the abbreviation AC) to the extent that we do so in a redundant fashion, often referring to an alternating current as an AC current. (When we need to distinguish it from AC, we call the “oneway” kind of current that, say, a battery causes in a circuit, direct current, abbreviated DC.)

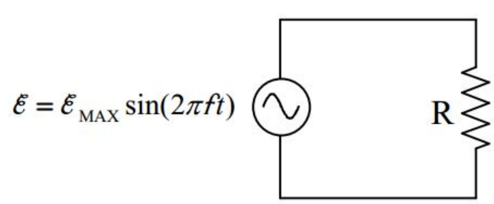
A device that causes current in a resistor, whether that current is alternating or not, is delivering energy to the resistor at a rate that we call power. The power delivered to a resistor can be expressed as $P=IV$ where I is the current through the resistor and V is the voltage across the resistor. Using the defining equation of resistance, $V=IR$, the power can be expressed as $P=I^2R$. A “device which causes a potential difference between its terminals that varies sinusoidally with time”, what I have been referring to as a “time-varying seat of EMF” is typically referred to as an AC power source. An AC power source is typically referred to in terms of the frequency of oscillations, and, the voltage that a DC power source, an ordinary seat of EMF, would have to maintain across its terminals to cause the same average power in any resistor that might be connected across the terminals of the AC power source. The voltage in question is typically referred to as ε_{RMS} or V_{RMS} where the reasoning behind the name of the subscript will become evident shortly.

Since the power delivered by an ordinary seat of EMF is a constant, its average power is the value it always has.

Here’s the fictitious circuit



that would cause the same resistor power as the AC power source in question. The average power (which is just the power in the case of a DC circuit) is given by $P_{AVG}=I\varepsilon_{RMS}$, which, by means of our defining equation of resistance solved for I, $I=V/R$, (where the voltage across the resistor is, by inspection, ε_{RMS}) can be written $P_{AVG}=\frac{\varepsilon_{RMS}^2}{R}$. So far, this is old stuff, with an unexplained name for the EMF voltage. Now let’s consider the AC circuit:



The power is $P=\frac{\varepsilon^2}{R}=\frac{[\varepsilon_{MAX}\sin(2\pi ft)]^2}{R}=\frac{\varepsilon_{MAX}^2[\sin(2\pi ft)]^2}{R}$. The average value of the square of the sine function is $\frac{1}{2}$. So the average power is $P_{AVG}=\frac{1}{2}\frac{\varepsilon_{MAX}^2}{R}$. Combining this with our expression $P_{AVG}=\frac{\varepsilon_{RMS}^2}{R}$ from above yields:

$$\frac{\varepsilon_{RMS}^2}{R} = \frac{1}{2}\frac{\varepsilon_{MAX}^2}{R}$$

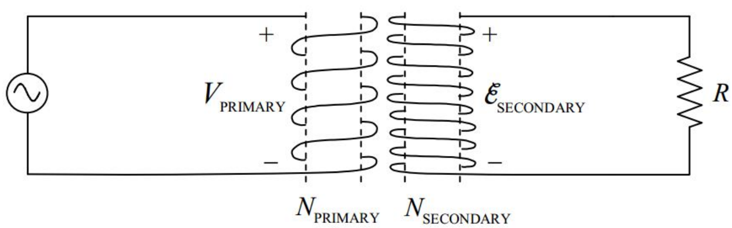
$$\varepsilon_{RMS} = \sqrt{\frac{1}{2}}\varepsilon_{MAX} \tag{B19.2}$$

Now we are in a position to explain why we called the equivalent EMF, ε_{RMS}. In our expression $P_{AVG} = \frac{1}{2}\frac{\varepsilon_{MAX}^2}{R}$, we can consider $\frac{\varepsilon_{MAX}^2}{2}$ to be the average value of the square of our timevarying EMF $\varepsilon = \varepsilon_{MAX} = \sin(2\pi ft)$. Another name for "average" is "mean" so we can consider $\frac{\varepsilon_{MAX}^2}{2}$ to be the mean value of ε^2. On the right side of our expression for our equivalent EMF, $\varepsilon_{RMS} = \frac{1}{\sqrt{2}}\varepsilon_{MAX}$, we have the square root of $\frac{\varepsilon_{MAX}^2}{2}$, that is, we have the square root of the mean of the square of the EMF ε. And indeed the subscript "RMS" stands for "root mean squared." RMS values are convenient for circuits consisting of resistors and AC power sources in that, one can analyze such circuits using RMS values the same way one analyzes DC circuits.

More on the Transformer

When the primary coil of a transformer is driven by an AC power source, it creates a magnetic field which varies sinusoidally in such a manner as to cause a sinusoidal EMF, of the same frequency as the source, to be induced in the secondary coil. The RMS value of the EMF induced in the secondary coil is directly proportional to the RMS value of the sinusoidal potential difference imposed across the primary. The constant of proportionality is the ratio of the number of turns in the secondary to the number of turns in the primary.

$$\varepsilon_{SECONDARY} = \frac{N_{SECONDARY}}{N_{PRIMARY}}V_{PRIMARY} \tag{B19.3}$$



When the number of windings in the secondary coil is greater than the number of windings in the primary coil, the transformer is said to be a step-up transformer and the secondary voltage is greater than the primary voltage. When the number of windings in the secondary coil is less than the number of windings in the primary coil, the transformer is said to be a step-down transformer and the secondary voltage is less than the primary voltage.

The Electrical Power in Your House

When you plug your toaster into a wall outlet, you bring the prongs of the plug into contact with two conductors between which there is a time-varying potential difference characterized as 115 volts 60 Hz AC. The 60 Hz is the frequency of oscillations of the potential difference resulting from a magnet completing 60 rotations per second, back at the power plant. A step-up transformer is used near the power plant to step the power plant output up to a high voltage. Transmission lines at a very high potential, with respect to each other, provide a conducting path to a transformer near your home where the voltage is stepped down. Power lines at a much lower potential provide the conducting path to the wires in your home. 115 volts is the RMS value of the potential difference between the two conductors in each pair of slots in your wall outlets. Since $\varepsilon_{RMS} = \frac{1}{\sqrt{2}}\varepsilon_{MAX}$, we have $\varepsilon_{MAX} = \sqrt{2}\,\varepsilon_{RMS}$, so $\varepsilon_{MAX} = \sqrt{2}(115 \text{ volts})$, or $\varepsilon_{MAX} = 163$ volts. Thus,

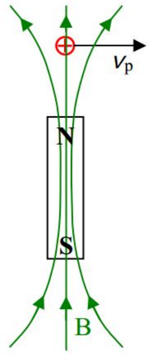
$$\varepsilon = (163 \text{ volts}) \sin[2\pi(60 Hz)t]$$

which can be written as,

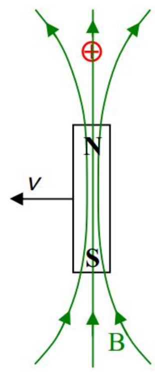
$$\varepsilon = (163 \text{ volts}) \sin[(377 \frac{\text{rad}}{s}t]$$

B20: Faraday’s Law and Maxwell’s Extension to Ampere’s Law

Consider the case of a charged particle that is moving in the vicinity of a moving bar magnet as depicted in the following diagram:



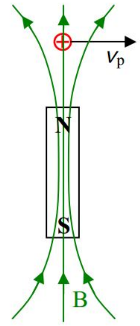
When we view the situation from the reference frame of the magnet, what we see (as depicted just above) is a charged particle moving in a stationary magnetic field. We have already studied the fact that a magnetic field exerts a force $\vec{F}=q\vec{v}_p\times\vec{B}$ on a charged particle moving in that magnetic field. Now let’s look at the same phenomenon from the point of view of the charged particle:



(where $\vec{\mathrm{v}}=-\vec{\mathrm{v}}_p$).

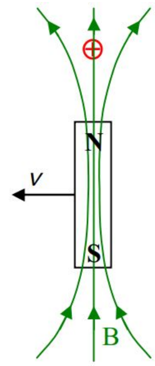
Surely we aren’t going to change the force exerted on the charged particle by the magnetic field of the magnet just by looking at the situation from a different reference frame. In fact we’ve already addressed this issue. What I said was that it is the relative motion between the magnet and the charged particle that matters. Whether the charged particle is moving through magnetic field lines, or the magnetic field lines, due to their motion, are moving sideways through the particle, the particle experiences a force. Now here’s the new viewpoint on this situation: What we say is, that the moving magnetic field doesn’t really exert a force on the stationary charged particle, but rather, that by moving sideways through the point at which the particle is located, the magnetic field creates an electric field at that location, and it is the electric field that exerts the force on the charged particle. In this viewpoint, we have, at the location of the stationary charged particle, an electric field that is exerting a force on the particle, and a magnetic field that is exerting no force on the particle. At this stage it might seem that it would be necessary to designate the magnetic field as some special kind of magnetic field that doesn’t exert a force on a charged particle despite the relative velocity between the charged particle and the magnetic field. Instead, what we actually do is to characterize the magnetic field as being at rest relative to the charged particle.

So, as viewed from the reference frame in which the magnet is at rest:



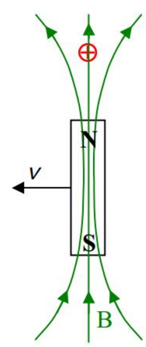
the particle experiences a force $\vec{F}$ directed out of the page in the diagram above due to its motion through the magnetic field.

And, as viewed from the reference frame in which the charged particle is at rest:



the particle finds itself in a stationary magnetic field but experiences the same force $\vec{F}$ because it also finds itself in an electric field directed out of the page.

So we have two models for explaining the force on the stationary charged particle in the case
depicted by:



In model 1 we simply say that in terms of the Lorentz Force $\vec{F}=q\vec{v}\times\vec{B}$, what matters is the relative velocity between the particle and the magnetic field and to calculate the force we identify the velocity $\vec{v}_p$ of the particle relative to the magnetic field as being rightward at magnitude $v_p=v$ in the diagram above so $\vec{F}=q\vec{v}\times\vec{B}$ (where q is the charge of the particle). In model 2 we say that the apparent motion of the magnetic field "causes" there to be an electric field and a stationary magnet field so the particle experiences a force $\vec{F}=q\vec{E}$. Of course we are using two different models to characterize the same force. In order for both models to give the same result we must have:

$$\vec{E}=\vec{v}_p\times\vec{B} \tag{B20.1}$$

where:

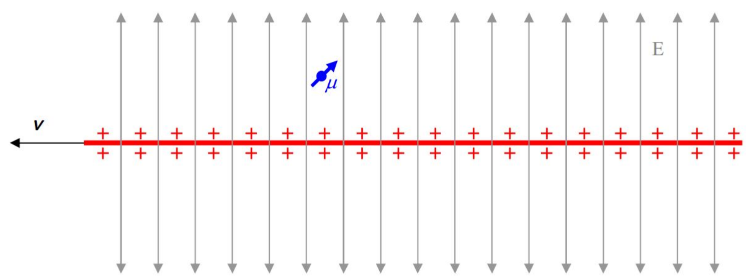
$\vec{E}$ is the electric field at an empty point in space due to the motion of that point relative to a magnetic field vector that exists at that point in space,

$\vec{v}_p$ is the velocity of the empty point in space relative to the magnetic field vector, and

$\vec{B}$ is the magnetic field vector.

Physicists have found model 2 to be more fruitful, especially when attempting to explain magnetic waves. The idea that a magnetic field in apparent sideways motion through a point in space "causes" there to be an electric field at that point in space, is referred to as Faraday's Law of Induction. Our mnemonic for Faraday's Law of Induction is: "A changing magnetic field causes an electric field."

The acceleration experienced by a charged particle in the vicinity of a magnet, when the charged particle is moving relative to the magnet represents an experimental result that we have characterized in terms of the model described in the preceding part of this chapter. The model is useful in that it can be used to predict the outcome of, and provide explanations regarding, related physical processes. Another experimental result is that a particle that has a magnetic dipole moment and is moving in an electric field with a velocity that is neither parallel nor antiparallel to the electric field, does (except for two special magnetic dipole moment directions) experience angular acceleration. We interpret this to mean that the particle experiences a torque. Recalling that a particle with a magnetic dipole moment that is at rest in an electric field experiences no torque, but one that is at rest in a magnetic field does indeed experience a torque (as long as the magnetic dipole moment and the magnetic field it is in are not parallel or antiparallel to each other), you might think that we can model the fact that a particle with a magnetic dipole moment experiences a torque when it is moving relative to an electric field, by defining a magnetic field "caused" by the apparent motion of the electric field relative to the particle. You would be right. To build such a model, we consider a charged particle that is moving in an electric field produced by a long line of charge that is uniformly distributed along the line. We start by depicting the situation in the reference frame in which the particle is at rest and the line of charge is moving:



Note that we have two different ways of accounting for the magnetic field due to the moving line of charge, at the location of the particle with a magnetic dipole moment. The moving line of charge is a current so we can think of the magnetic field as being caused by the current.

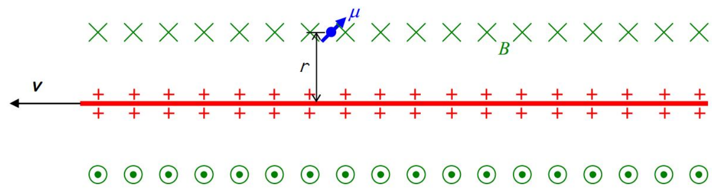
The other option is to view the magnetic field as being caused by the electric field lines moving sideways through the particle. There is, however, only one magnetic field, so, the two different ways of accounting for it must yield the same result. We are going to arrive at an expression for the magnetic field due to the motion of an electric field by forcing the two different ways of accounting for the magnetic field to be consistent with each other. First, we'll simply use Ampere's law to determine the magnetic field at the location of the particle. Let's define the linear charge density (the charge per length) of the line of charge to be λ and the distance that the particle is from the line of charge to be r. Suppose that in an amount of time dt the line of charge moves a distance dx. Then the amount of charge passing a fixed point on the line along which the charge is moving, in time dt, would be λdx. Dividing the latter by dt yields $\lambda dx/dt$ which can be expressed as λv and is just the rate at which charge is flowing past the fixed point, that is, it is the current I. In other words, the moving line of charge is a current $I = \lambda v$. Back in chapter 17 we gave the experimental result for the magnetic field due to a long straight wire carrying current I in the form of an equation that we called "Ampere's Law." It was equation 17-2; it read:

$$B = \frac{\mu_o}{2\pi}\frac{I}{r}$$

and it applies here. substituting $I=\lambda v$ into this expression for B yields

$$B=\frac{\mu_o}{2\pi}\frac{\lambda v}{r} \tag{B20.2}$$

By the right hand rule for something curly something straight we know that the magnetic field is directed into the page at the location of the particle that has a magnetic dipole moment, as depicted in the following diagram:



Now let's work on obtaining an expression for the same magnetic field from the viewpoint that it is the electric field moving sideways through the location of the particle that causes the magnetic field. First we need an expression for the electric field due to the line of charge, at the location of the particle, that is, at a distance r from the line of charge. The way to get that is to consider the line of charge as consisting of an infinite number of bits of charged material, each of which is a segment of infinitesimal length dx of the line of charge. Since the line of charge has a linear charge density λ, this means that each of the infinitesimal segments dx has charge λdx. To get the electric field at the location of the particle that has a magnetic dipole moment, all we have to do is to add up all the contributions to the electric field at the location of the particle, due to all the infinitesimal segments of charged material making up the line of charge. Each contribution is given by Coulomb's Law for the Electric Field. The difficulty is that there are an infinite number of contributions. You will be doing such calculations when you study chapter 30 of this textbook. At this stage, we simply provide the result for the electric field due to an infinitely long line of charge having a constant value of linear charge density λ:

$$E=\frac{\lambda}{2\pi r\epsilon_o}$$

Multiplying both sides by ϵ_o yields

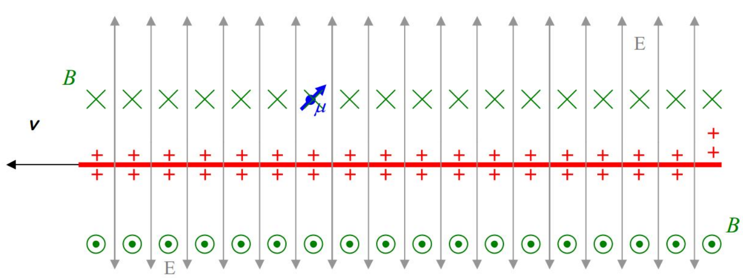
$$\epsilon_o E=\frac{\lambda}{2\pi r}$$

The expression on the right side of this equation appears in equation $ref20-2$, $B=\frac{\mu_o}{2\pi}\frac{\lambda v}{r}$. Substituting $\epsilon_o E$ for $\frac{\lambda}{2\pi r}$ where the latter appears in equation B20.2 yields:

$$B=\mu_o\epsilon_o E v$$

This represents the magnitude of the magnetic field that is experienced by a particle when it is moving with speed $v_p=v$ relative to an electric field $\vec{E}$ when the velocity is perpendicular to $\vec{E}$. Experimentally we find that a particle with a magnetic dipole moment experiences no torque (and hence no magnetic field) if its velocity is parallel or antiparallel to the electric field $\vec{E}$. As such, we can make our result more general (not only good for the case when the velocity is perpendicular to the electric field) if we write, $E_\perp$ in place of E.

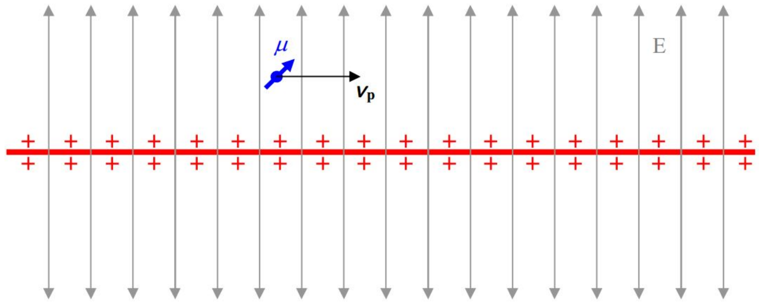
$$B=\mu_o\epsilon_o E_\perp v$$



Starting with the preceding equation, we can bundle both the magnitude and the direction (as determined from Ampere's Law and the right hand rule when we treat the moving line of charge as a current, and as depicted in the diagram above) of the magnetic field into one equation by writing:

$$\vec{B} = \mu_o \epsilon_o \vec{v} \times \vec{E}$$

We can express $\vec{B}$ in terms of the velocity $\vec{v}_p$ of the particle relative to the line of charge



(instead of the velocity $\vec{v}$ of the line of charge relative to the particle) just by recognizing that

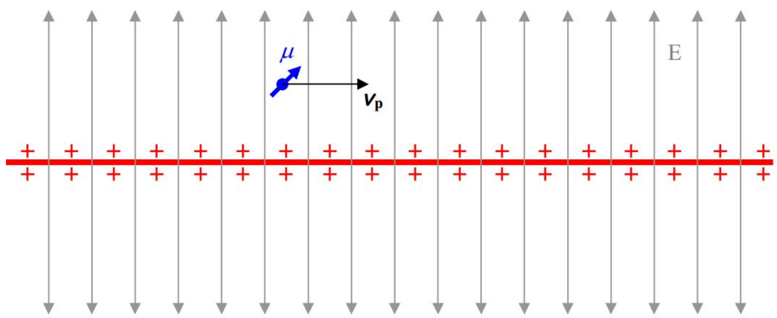
$$\vec{v}_p = -\vec{v}$$

Substituting this expression (\vec{v}_p=-\vec{v}) into our expression for the magnetic field ($\vec{B} = \mu_o \epsilon_o \vec{v} \times \vec{E}$) yields:

$$\vec{B} = -\mu_o \epsilon_o \vec{v}_P \times \vec{E}$$

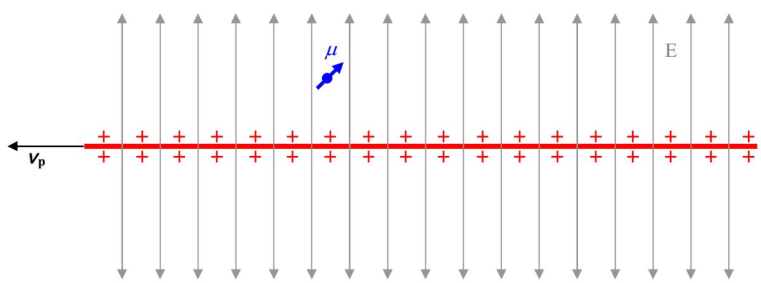
In this model, where we account for the torque experienced by a particle that has a magnetic dipole moment when that particle is moving in an electric field, by defining a magnetic field $\vec{B} = -\mu_o \epsilon_o \vec{v}_P \times \vec{E}$ which depends both on the velocity of the particle relative to the electric field and the electric field itself, the electric field itself is considered to exert no torque on the charged particle. At this stage it might seem that it would be necessary to designate the electric field as some special kind of electric field that doesn't exert a torque on a charged particle despite the relative velocity between the charged particle and the electric field. Instead, what we actually do is to characterize the electric field as being at rest relative to the charged particle.

So, as viewed from the reference frame in which the line of charge is at rest:



the particle that has a magnetic dipole moment experiences a torque due to its motion through the electric field.

And, as viewed from the reference frame in which the particle is at rest:

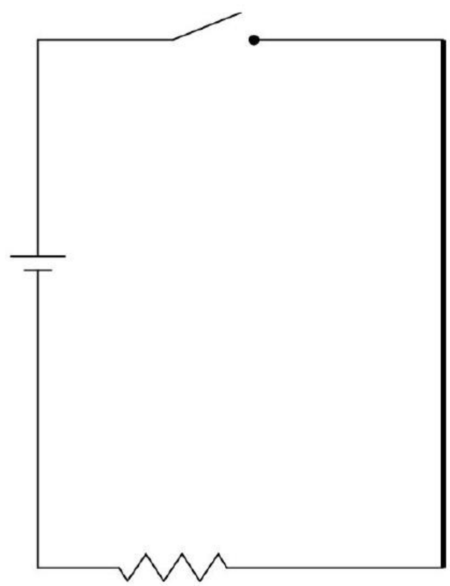


the particle that has a magnetic dipole moment finds itself in a stationary electric field but experiences the same torque because it also finds itself in a magnetic field directed, in the diagram above, into the page. One way of saying what is going on here is to say that, loosely speaking: A changing electric field "causes" a magnetic field. The phenomenon of a changing electric field "causing" a magnetic field is referred to as Maxwell's Extension to Ampere's Law.

So far, in this chapter we have addressed two major points: A magnetic field moving sideways through a point in space causes there to be an electric field at that point in space, and, an electric field moving sideways through a point in space causes there to be a magnetic field at that point in space. In the remainder of this chapter we find that putting these two facts together yields something interesting.

Expressing what we have found in terms of the point of view in which point P is fixed and the field is moving through point P with speed $\vec{v} = -\vec{v}_P$, we have: a magnetic field vector $\vec{B}$ moving with velocity $\vec{v}$ transversely through a point in space will "cause" an electric field $\vec{E} = -\vec{v} \times \vec{B}$ at that point in space; and; an electric field vector moving with velocity $\vec{v}$ transversely through a point in space will "cause" a magnetic field $\vec{B} = \mu_o \epsilon_o \vec{v} \times \vec{E}$ at that point in space. The word "cause" is in quotes because there is never any time delay. A more precise way of putting it would be to say that whenever we have a magnetic field vector moving transversely through a point in space, there exists, simultaneously, an electric field $\vec{E} = -\vec{v} \times \vec{B}$ at that point in space, and whenever we have an electric field vector moving transversely through a point in space there exists, simultaneously, a magnetic field $\vec{B} = \mu_o \epsilon_o \vec{v} \times \vec{E}$ at that point in space.

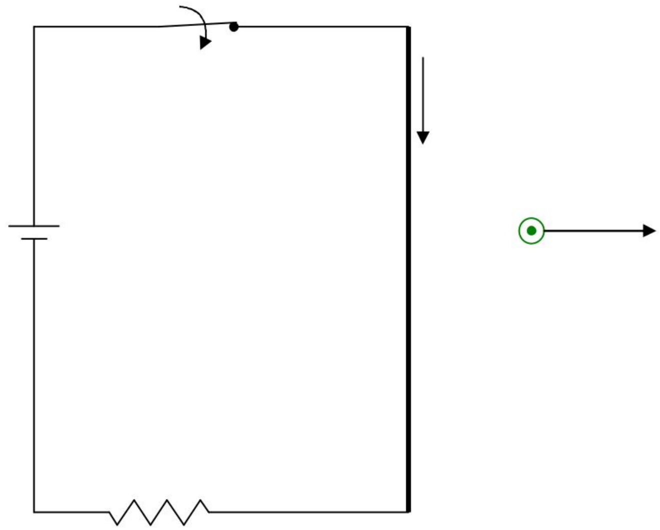
Consider the following circuit. Assume that we are looking down on the circuit from above, meaning that into the page is downward, and out of the page is upward.



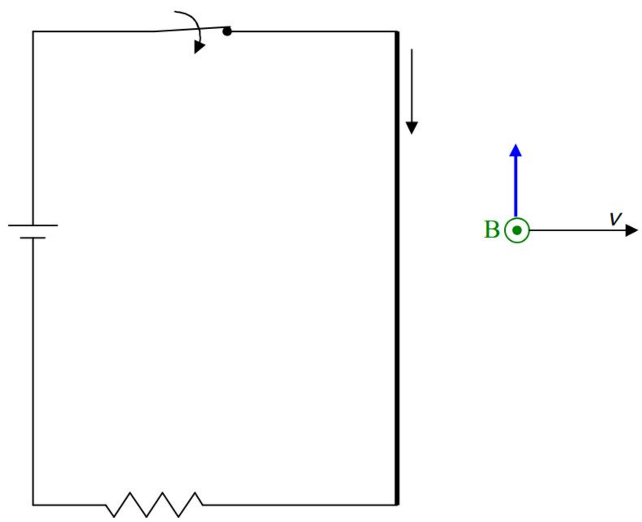
I want you to focus your attention on the rightmost wire of that circuit. As soon as someone closes that switch we are going to get a current through that wire and that current is going to produce a magnetic field. By means of the right-hand rule for something curly something straight, with the current being the something straight, and our knowledge that straight currents cause magnetic fields that make loops around the current, we can deduce that there will be an upward-directed (pointing out of the page) magnetic field at points to the right of the wire. In steady state, we understand that the upward-directed magnetic field vectors will be everywhere to the right of the wire with the magnitude of the magnetic field vector being smaller the greater the distance the point in question is from the wire. Now the question is, how long does it take for the magnetic field to become established at some point a specified distance to the right of the wire? Does the magnetic field appear instantly at every point to the right of the wire or does it take time? James Clerk Maxwell decided to explore the possibility that it takes time, in other words, that the magnetic field develops in the vicinity of the wire and moves outward with a finite velocity.

Here I want to talk about the leading edge of the magnetic field, the expanding boundary within which the magnetic field already exists, and outside of which, the magnetic field does not yet exist. With each passing infinitesimal time interval another infinitesimal layer is added to the region within which the magnetic field exists. While this is more a case of magnetic field vectors growing sideways through space, the effect of the motion of the leading edge through space is the same, at the growing boundary, as magnetic field vectors moving through space. As such, I am going to refer to this magnetic field growth as motion of the magnetic field through space.

To keep the drawing uncluttered I'm going to show just one of the infinite number of magnetic field vectors moving rightward at some unknown velocity (and it is this velocity that I am curious about) as the magnetic field due to the wire becomes established in the universe.



Again, what I'm saying is that, as the magnetic field builds up, what we have, are rightwardmoving upward (pointing out of the page, toward you) magnetic field lines due to the current that just began. Well, as a magnetic field vector moves through whatever location it is moving through, it "causes" an electric field $\vec{E}=-\vec{v}\times\vec{B}$.



At any point P through which the magnetic field vector passes, an electric field exists consistent with $\vec{E}=-\vec{v}$\time$\vec{B}$. What this amounts to is that we have both a magnetic field and an electric field moving rightward through space. But we said that an electric field moving transversely through space "causes" a magnetic field. More specifically we said that it is always accompanied by a magnetic field given by $\vec{B}=\mu_o\epsilon_o\vec{v}\times\vec{E}$. Now we've argued around in a circle. The current "causes" the magnetic field and its movement through space "causes" an electric field whose movement through space "causes" the magnetic field. Again, the word "causes" here should really be interpreted as "exists simultaneously with." Still, we have two explanations for the existence of one and the same magnetic field and the two explanations must be consistent with each other. For that to be the case, if we take our expression for the magnetic field "caused" by the motion of the electric field,

$$\vec{B}=\mu_o\epsilon_o\vec{v}\times\vec{E}$$

and substitute into it, our expression $\vec{E} = -\vec{v} \times \vec{B}$ for the electric field "caused" by the motion of the magnetic field, we must obtain the same $\vec{B}$ that, in this circular argument, is "causing" itself. Let's try it. Substituting $\vec{E} = -\vec{v} \times \vec{B}$ into $\vec{B} = \mu_o \epsilon_o \vec{v} \times \vec{E}$, we obtain:

$$\vec{B} = -\mu_o \epsilon_o \vec{v} \times (\vec{v} \times \vec{B})$$

All right. Noting that $\vec{v}$ is perpendicular to both $\vec{B}$ and $\vec{v} \times \vec{B}$, meaning that the magnitude of the cross product, in each case, is just the product of the magnitudes of the multiplicand vectors, we obtain:

$$\vec{B} = \mu_o \epsilon_o v^2 \vec{B}$$

which I copy here for your convenience:

$$\vec{B} = \mu_o \epsilon_o v^2 \vec{B}$$

Again, it is one and the same $\vec{B}$ on both sides, so, the only way this equation can be true is if $\mu_o \epsilon_o v^2$ is exactly equal to 1. Let's see where that leads us:

$$\mu_o \epsilon_o v^2 = 1$$

$$v^2 = \frac{1}{\mu_o \epsilon_0}$$

$$v = \frac{1}{\sqrt{\mu_o \epsilon_o}}$$

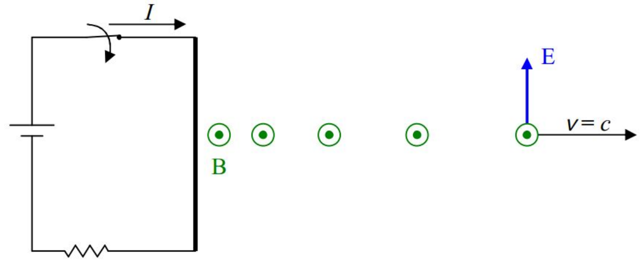
$$v = \frac{1}{\sqrt{\left(4\pi \times 10^{-7} \frac{T \cdot m}{A}\right) 8.85 \times 10^{-12} \frac{C^2}{N \cdot m^2}}}$$

$$v = 3.00 \times 10^8 \frac{m}{s}$$

Wow! That's the speed of light! When James Clerk Maxwell found out that electric and magnetic fields propagate through space at the (already known) speed of light he realized that light is electromagnetic waves.

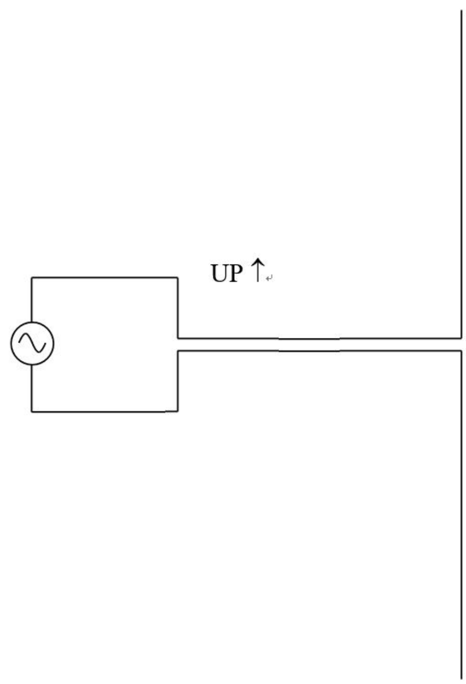
B21: The Nature of Electromagnetic Waves

When we left off talking about the following circuit:



we had recently closed the switch and the wire was creating a magnetic field which was expanding outward. The boundary between that part of the universe in which the magnetic field is already established and that part of the universe in which the magnetic field has not yet been established is moving outward at the speed of light, $c = 3 \times 10^8 m/s$. Between that boundary and the wire we have a region in which there exists a steady unmoving magnetic field. Note that it was the act of creating the current that caused the magnetic field "edge" that is moving at the speed of light. In changing from a no-current situation to one in which there was current in the wire, charged particles in the wire went from no net velocity in the along-the-wire direction to a net velocity along the wire, meaning, that the charged particles were accelerated. In other words, accelerated charged particles cause light. We can also cause light by means of the angular acceleration of particles having a magnetic dipole moment, but, the short answer to the question about what causes light, is, accelerated charged particles.

Here's a simple circuit that one might use to intentionally cause light:



The vertical arrangement of wires on the right is referred to as a dipole antenna. As the AC power source alternately causes charge to surge upward in both parts of the antenna, and then downward, the dipole antenna creates electric and magnetic fields that oscillate sinusoidally in both time and space. The fields propagate through space away from the antenna at the speed of light.

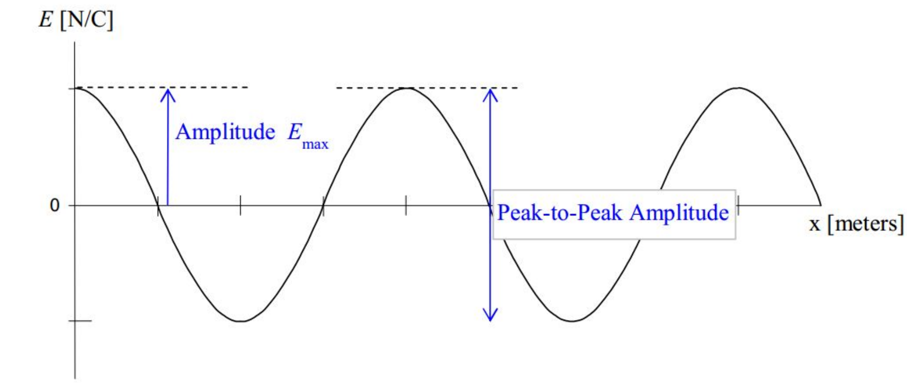
The charged particles oscillating up and down in the antenna causes waves of electric and magnetic fields known as light. The frequency of the waves is the same as the frequency of oscillations of the particles which is determined by the frequency of the power source. The speed of the waves is the speed of light $c = 3 \times 10^8 m/s$, because the waves are light. For any kind of wave, the frequency, wavelength, and wave speed are related by:

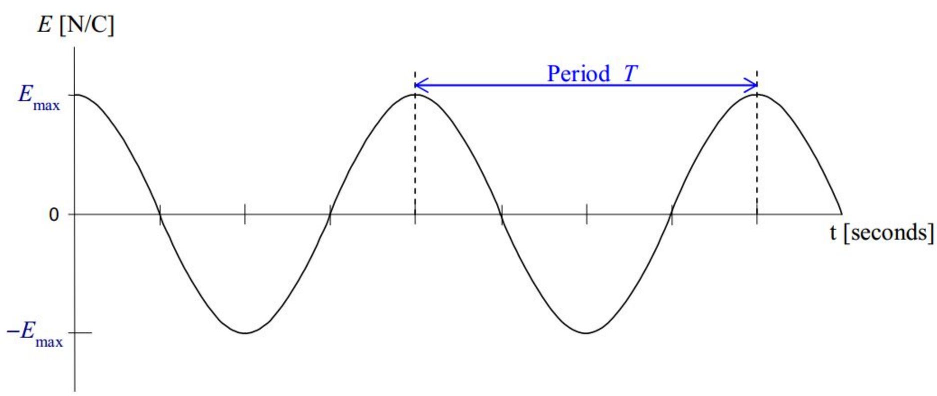
$$v=\lambda f$$

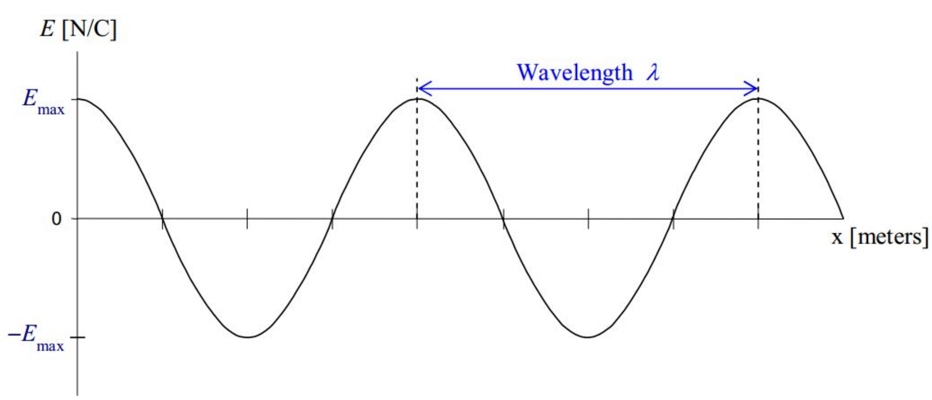
which, in the case of light reads:

$$c=\lambda f$$

Here's a quick pictorial review of some properties of waves. In the case of light, we have electric and magnetic fields oscillating in synchronization with each other. It is customary to characterize the waves in terms of the electric field. I'll do that here, but, one should keep in mind that the magnetic field oscillates and moves in the same manner that the electric field does, but, at right angles to the electric field.







The intensity of a wave is proportional to the square of its amplitude, so, in the case of light:

$$I\propto (E_{MAX})^2$$

The frequency of light is determined by the frequency of oscillations of the charged particles constituting the source of the light. How we categorize light depends on the frequency of the light. In order of increasing frequency, we refer to light as: radio waves, microwaves, infrared radiation, visible light, ultraviolet light, X rays, and gamma rays. They are all the same thing— electric and magnetic fields that are oscillating in time and space. I am using the word light in a generic sense. It refers to waves of any one of these various frequencies of oscillations of electric and magnetic fields. In this context, if I want to talk about light whose frequency falls in the range to which our eyes are sensitive, I refer to it as visible light. Another name for light is electromagnetic radiation. The entire set of the different frequencies of light is referred to as the electromagnetic spectrum. The following table indicates the way in which humans categorize the various frequencies of light in the electromagnetic spectrum. While I do give definite values, boundaries separating one frequency from the next are not well defined and hence, should be treated as approximate values.

Kind of Light	***Frequency***	***Wavelength***
Radio Waves	$< 300\,MHz$	$> 1\,m$
Microwaves	$300 - 750000\,MHz$	$.4\,mm - 1\,m$
Infrared	$750\,GHz - 430\,GHz$	$700\,nm - .4\,mm$
Visible	$430 - 750\,THz$	$400 - 700\,nm$
Ultraviolet	$750 - 6000\,THz$	$5 - 400\,nm$
X rays	$6000 - 50000000\,THz$	$.006 - 5\,nm$
Gamma Rays	$> 50000000\,THz$	$< .006\,nm$

Note that the visible regime is but a tiny slice of the overall electromagnetic spectrum. Within it, red light is the long-wavelength, low-frequency visible light, and, blue/violet light is the short-wavelength high-frequency visible light. **AM** radio stations broadcast in the kHz range and **FM** stations broadcast in the MHz range. For instance, setting your **AM** dial to 100 makes your radio sensitive to radio waves of frequency 100 kHz and wavelength 3000 m. Setting your **FM** dial to 100 makes your radio sensitive to radio waves of frequency 100 MHz and wavelength 3 m.

We call the superposition of the changing electric and magnetic field vectors, with other changing electric and magnetic field vectors, interference. Many of the phenomena involving light are understood in terms of interference.

When light interacts with matter, its electric field exerts forces on the charged particles that make up matter. The direction of the force exerted on a charged particle is the same direction as the electric field if the particle is positive, and in the opposite direction if it is negative. The magnetic field exerts a torque on the magnetic-dipole-possessing particles that make up matter. Because so many of the observable effects associated with the interaction of visible light have to do with the force exerted on charged particles by the electric field, it has become customary to talk about the interaction of light with matter in terms of the interaction of the electric field with matter. I will follow that custom. Please keep in mind that the magnetic field, always at right angles to the electric field in light, is also present. As a result of the force exerted on the charged particles by the electric field of which the light consists, the charged particles accelerate, and, as a result, produce their own electric and magnetic fields. Because there is no time delay between the exertion of the force and the resulting acceleration, the newly produced electric and magnetic field vectors superpose with the very electric and magnetic field vectors causing the acceleration. Because the mass of an electron is approximately $1/2000$ of the mass of a proton, the acceleration experienced by an electron is 2000 times greater than that experienced by a proton subject to the same force. Hence, the interaction of light with matter, can often be explained in terms of the interaction of light with the electrons in matter.

How the electrons in matter interact with light, is largely determined by the degree to which the electrons are bound in the matter. As a rather bizarre example of how a large number of complicated interactions can combine to form a simple total effect, the mix of attractive and repulsive $1/r^2$ Coulomb forces exerted on the electron in a solid material by the protons and electrons on all sides of it, results in a net force on the electron that is well modeled by the force that would be exerted on a particle "tied" to its equilibrium position by a spring. Hence the electron acts like a "mass on a spring." As such, it can undergo simple harmonic motion like a mass on the end of a spring. The way light interacts with the electrons can thus be said to depend on the frequency of the light and the force constant of the effective spring. If we limit our discussion to visible light, the degree to which the electrons are bound (the spring constant) determines how the light interacts with the matter. In the case of what we would consider opaque light-absorbing matter such as flat black paint, the electron accelerations result in destructive interference of the incoming light with the light produced by the electrons. Light doesn't go through, nor is much reflected off the material. In the case of shiny metal surfaces, the electrons that the light interacts with are virtually free. The light emitted by these electrons as a result of the acceleration caused by the light, interferes constructively with the light in a very specific backward direction and destructively in forward directions. Hence, the light does not get through the metal, but, it is reflected in a mirror-like manner referred to as specular reflection. In the case of a transparent medium such as glass, the light given off by the electrons interferes with the incoming light in such a manner as to cause constructive interference in specific forward and backward directions. But, the constructive interference in the forward direction is such that the pattern of electric and magnetic waves formed by all the interference taken as a whole, moves more slowly through the glass than light moves through vacuum. We say that the speed of light in a transparent medium is less than the speed of light in vacuum. The ratio of the speed of light in vacuum to the speed of light in a transparent medium is called the index of refraction, n, of that transparent medium.

$$n = \frac{c}{v}$$

where:

n is the index of refraction of a transparent medium,

$c = 3.00 \times 10^8\ m/s$ is the speed of light in vacuum, and,

v is the speed of light in the transparent medium.

Because the speed of light never exceeds the-speed-of-light-in-vacuum, the index of refraction is always greater than or equal to 1 (equal when the medium is, or behaves as, vacuum). Some values for the index of refraction of light for a few transparent media are:

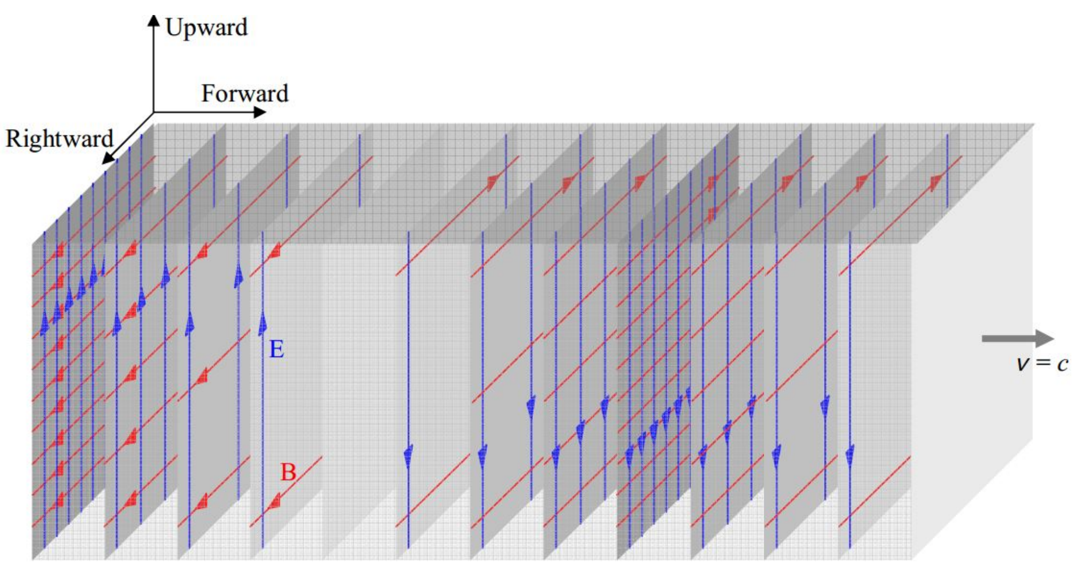
Medium	***Index of Refraction***
Vacuum	1
Air	1.00
Water	1.33
Glass (Depends on the kind of glass. Here is one typical value).	1.5

The kind of interaction of light with matter with which we are most familiar is called diffuse reflection. It is the light that is diffusely reflecting off a person that enters your eyes when you are looking at that person. The electron motion produces light that interferes destructively with the incoming light in the forward direction (the direction in which the incoming light is traveling), so, essentially none gets through but, for a particular frequency range, for all backward directions, very little destructive interference occurs. When the object is illuminated by a mix of all visible frequencies (white light), the frequency of the reflected light depends on the force constant of the effective spring that is binding the electrons to the material of which they are a part. The frequency reflected (in all backward directions) corresponds to what we call the color of the object.

B22: Huygens's Principle and 2-Slit Interference

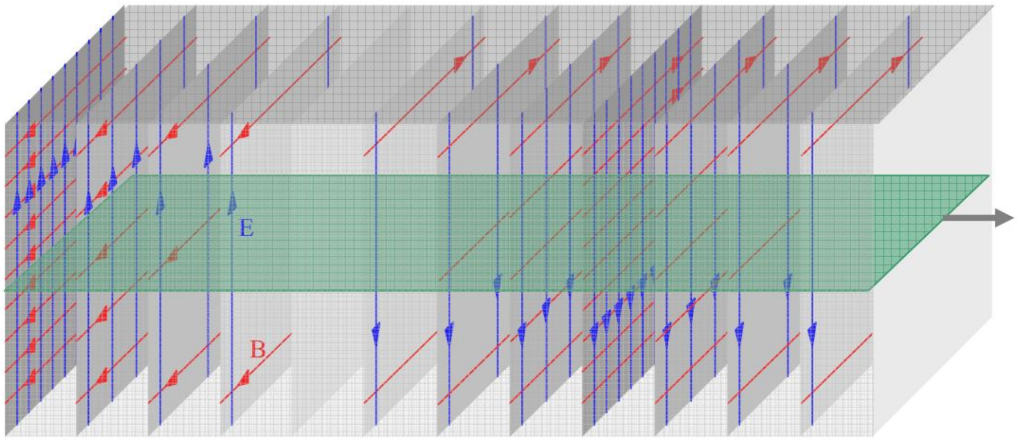
Consider a professor standing in front of the room holding one end of a piece of rope that extends, except for sag, horizontally away from her in what we'll call the forward direction. She asks, "What causes sinusoidal waves?" You say, "Something oscillating." "Correct," she replies. Then she starts moving her hand up and down and, right before your eyes, waves appear in the rope. For purposes of discussion, we will consider the waves only before any of them reach the other end, so we are dealing with traveling waves, not standing waves. "What, specifically, is causing these waves," she asks, while pointing, with her other hand, at the waves in the rope. You answer that it is her hand oscillating up and down that is causing the waves and again you are right. Now suppose you focus your attention on a point in the rope, call it point P, somewhat forward of her hand. Like all points in the rope where the wave is, that point is simply oscillating up and down. At points forward of point P, the rope is behaving just as if the professor were holding the rope at point P and moving her hand up and down the same way that point P is actually moving up and down. Someone studying only those parts of the rope forward of point P would have no way of knowing that the professor is actually holding onto the rope at a point further back and that point P is simply undergoing its part of the wave motion caused by the professors hand at the end of the rope. For points forward of point P, things are the same as if point P were the source of the waves. For predicting wave behavior forward of point P, we can treat point P, an oscillating bit of the rope, as if it were the source of the waves. This idea that you can treat one point in a wave medium as if it were the source of the waves forward of it, is called Huygens's Principle. Here, we have discussed it in terms of a one dimensional medium, the rope. When we go to more than one dimension, we can do the same kind of thing, but we have more than one point in the wave medium contributing to the wave behavior at forward points. In the case of light, that which is oscillating are the electric field and the magnetic field. For smooth regular light waves traveling in a forward direction, if we know enough about the electric and magnetic fields at all points on some imaginary surface through which all the light is passing, we can determine what the light waves will be like forward of that surface by treating all points on the surface as if they were point sources of electromagnetic waves. For any point forward of the surface, we just have to (vectorially) add up all the contributions to the electric and magnetic fields at that one point, from all the "point sources" on the imaginary surface. In this chapter and the next, we use this Huygens's Principle idea in a few simple cases (e.g. when, except for two points on the kind of surface just mentioned, all the light is blocked at the surface so you only have two "point sources" contributing to the electric and magnetic fields at points forward of the surface—you could create such a configuration by putting aluminum foil on the surface and poking two tiny holes in the aluminum foil) to arrive at some fairly general predictions (in equation form) regarding the behavior of light. The first has to do with a phenomenon called two-slit interference. In applying the Huygens's principle idea to this case we use a surface that coincides with a wave front. In diagrams, we have some fairly abstract ways of representing wave fronts which we share with you by means of a series of diagrams of wave fronts, proceeding from less abstract to more abstract.

Here's one way of depicting a portion of a beam of light traveling forward through space, at an instant in time:

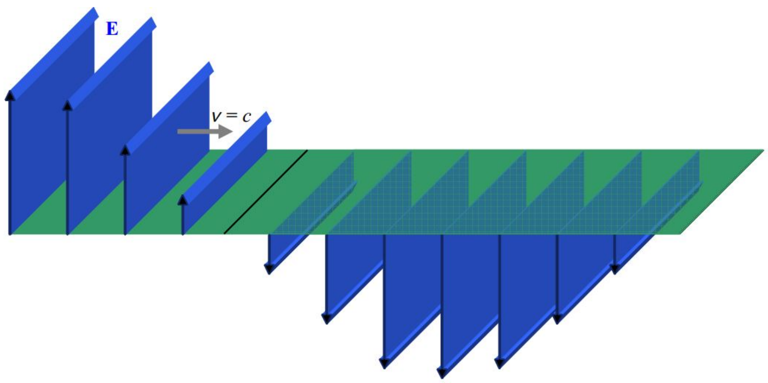


Each "sheet" in the diagram characterizes the electric (vertical arrows) and magnetic (horizontal arrows) fields at the instant in time depicted. On each sheet, we use the field diagram convention with which you are already familiar—the stronger the field, the more densely packed the field lines in the diagram.

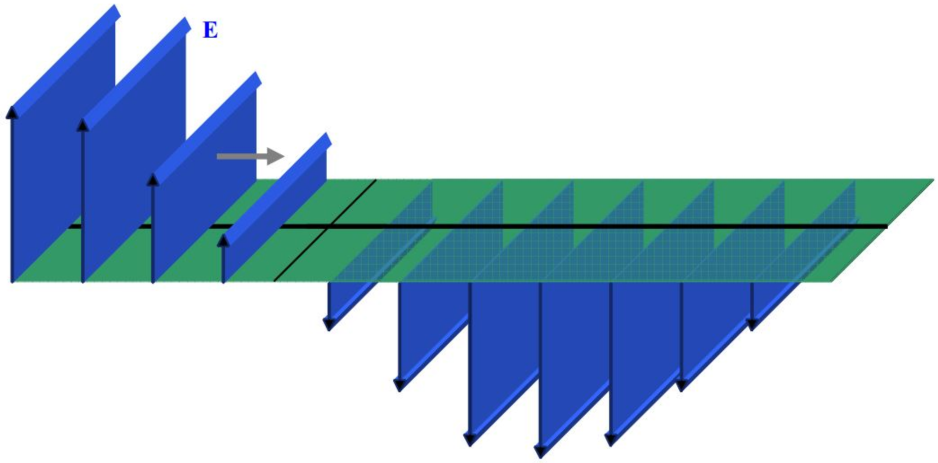
Consider the one thin horizontal slice of the beam depicted in this diagram:



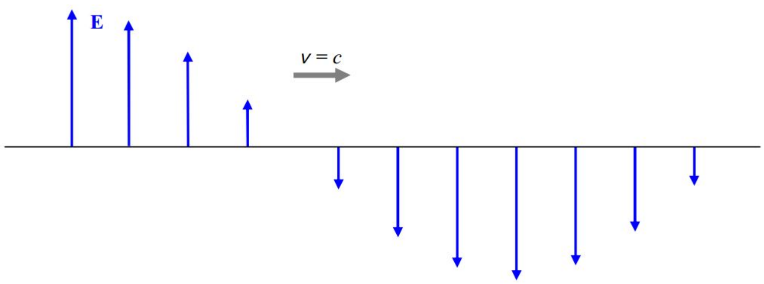
Here we depict the electric field on that one thin horizontal region:



This is a simpler diagram with less information which is, perhaps, easier to interpret, but, also, perhaps, easier to misinterpret. It may for instance, be more readily apparent to you that what we are depicting (as we were in the other diagram) is just $\frac{3}{4}$ of a wavelength of the wave. On the other hand, the diagram is more abstract—the length of the electric field vectors does not represent extent through space, but rather, the magnitude of the electric field at the tail of the arrow. As mentioned, the set of locations of the tails of the arrows, namely the horizontal plane, is the only place the diagram is giving information. It is generally assumed that the electric field has the same pattern for some distance above and below the plane on which it is specified in the diagram. Note the absence of the magnetic field. It is up to the reader to know that; as part of the light, there is a magnetic field wherever there is an electric field, and that, the greater the electric field, the greater the magnetic field, and that the magnetic field is perpendicular both to the direction in which the light is going, and to the electric field. (Recall that the direction of the magnetic field is such that the vector $\vec{E} \times \vec{B}$ is in the same direction as the velocity of the wave.)

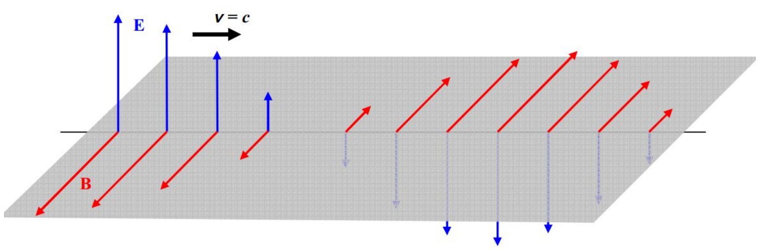


An example of an electric field depiction, on a single line along the direction of travel, would be:

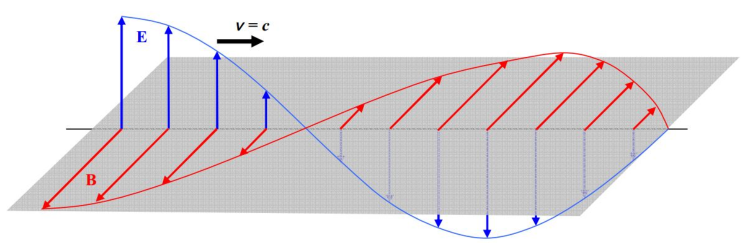


It is important to keep in mind that the arrows are characterizing the electric field at points on the one line, and, it's the length of the arrow that indicates the strength of the electric field at the tail of the arrow, not the spacing between arrows.

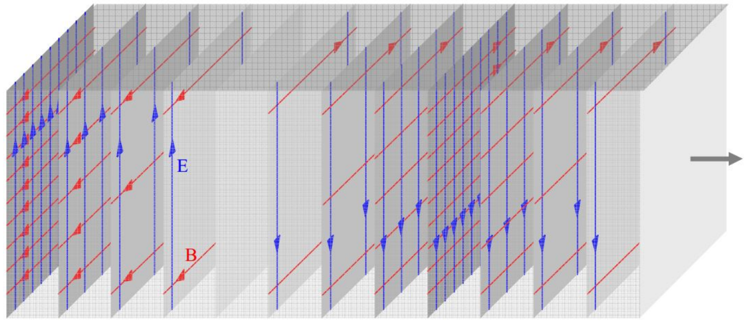
The simplicity of this field-on-a-line diagram allows for the inclusion of the magnetic field vectors in the same diagram:



If one connects the tips of the arrows in this kind of diagram, the meaning of those Electric Field vs. Position sinusoidal curves presented in the last chapter becomes more evident:

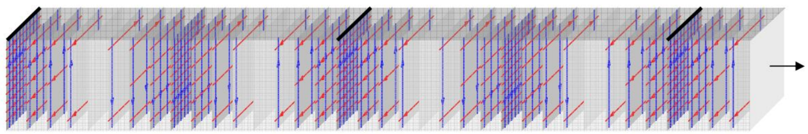


Huygens' Principle involves wavefronts. A wavefront is the part of a wave which is at a surface that is everywhere perpendicular to the direction in which the wave is traveling. If such surfaces are planes, the wave is called a plane wave. The kind of wave we have been depicting is a plane wave. The set of fields on any one of the gray "sheets" on in the diagram:

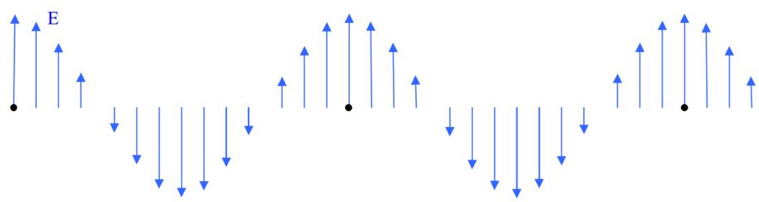


is a part of a wavefront. It is customary to focus our attention on wavefronts at which the electric and magnetic fields are a maximum in one direction. The rearmost sheet in the diagram above represents such a wavefront.

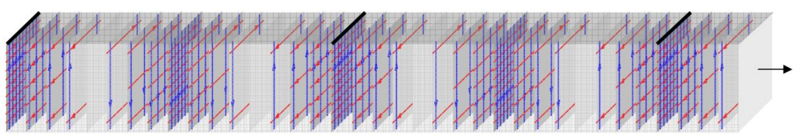
In the following diagram, you see black lines on the top of each sheet representing maximum-upward-directed-electric-field wavefronts.



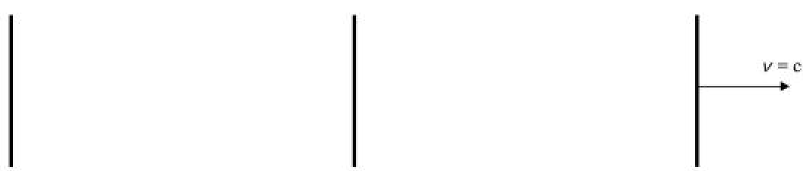
And, in the following, each such wavefront is marked with a black dot:



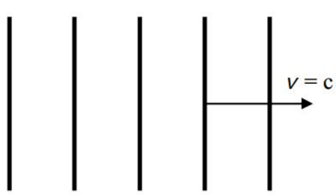
A common method of depicting wavefronts corresponds to a view from above, of the preceding electric field sheet diagram which I copy here:



Such a wavefront diagram appears as:



More commonly, you'll see more of them packed closer together. The idea is that the wavefronts look like a bird's eye view of waves in the ocean.

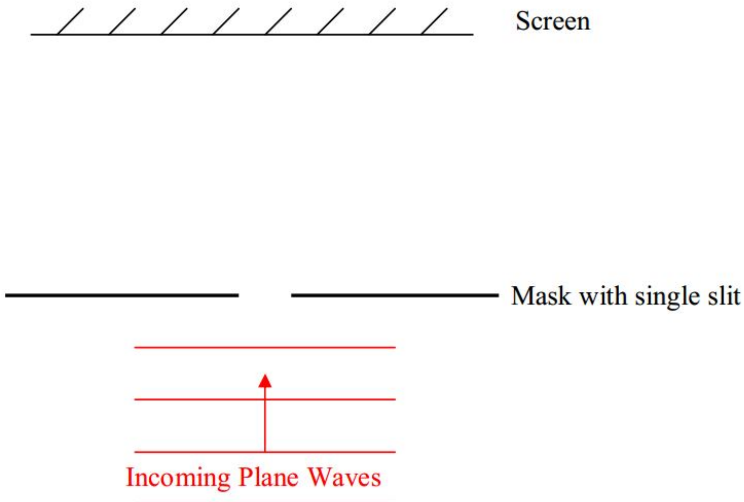


Note that the distance between adjacent maximum-field wavefronts, as depicted here, is one wavelength.

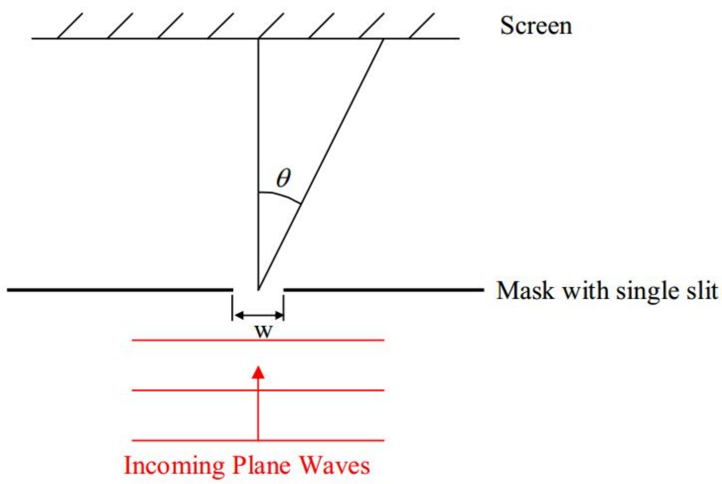
At this point we are ready to use the Huygens's Principle idea, the notion of a wavefront, and our understanding of the way physicists depict wavefronts diagrammatically, to explore the phenomenon of two slit interference.

B23: Single-Slit Diffraction

Single-slit diffraction is another interference phenomenon. If, instead of creating a mask with two slits, we create a mask with one slit, and then illuminate it, we find, under certain conditions, that we again get a pattern of light and dark bands. It is not the same pattern that you get for two-slit interference, but, it's quite different from the single bright line in the straight-ahead direction that you might expect. Here's how it comes about. Firstly, here's the setup:

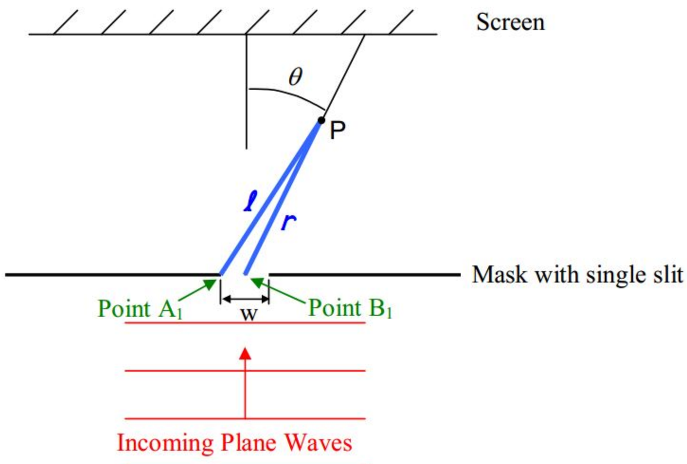


Again, we get a bright fringe in the straight-ahead position on the screen. From there, working out to either side, we get bands that alternate between dark and bright. The first maximum to the right or left of the central maximum is not nearly as bright as the central maximum. And each maximum after that is less bright than the maximum preceding it. As far as the analysis goes, I want to start with the minima. Consider an imaginary line extending out from the midpoint of the slit all the way to the screen.



So now the question is, "Under what conditions will there be completely destructive interference along a line such as the one depicted to be at angle θ above?" To get at the answer, we first divide the slit in half. I'm going to enlarge the mask so that you can see what I mean.

Now I imagine dividing side A up into an infinite number of pieces and side B up the same way. When the slit is illuminated by the light, each piece becomes a point source. Consider the first point source (counting from the left) on side A and the first point source (again counting from the left) on side B. These two point sources are a distance $w/2$ apart, where w is the width of the slit. If the light from these two point sources (which are in phase with each other because they are really both part of the same incoming plane wave), interferes completely destructively, at some angle θ with respect to the straight-ahead direction, then the light from the second point source on side A and the second point source on side B will also interfere with each other completely destructively because these two point sources are also $w/2$ apart. The same goes for the third-from-the-left point sources on both sides, the fourth, the fifth, and so on, ad infinitum. So, all we need is to establish the condition that makes the light from the leftmost point source on side A (overall, the leftmost point of the slit) interfere completely destructively with the leftmost point source on side B (overall, essentially the midpoint of the slit). So, consider any point P on a proposed line of minima.

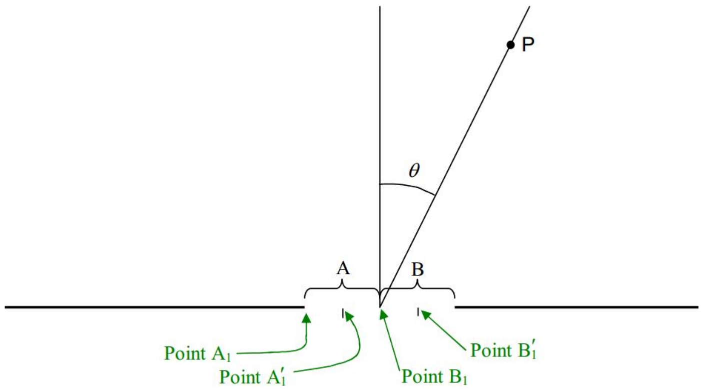


The distance between the two point sources is $w/2$. From the analysis done for the case of two-slit interference, we know that this results in a path difference $\ell - r = \frac{w}{2}\sin\theta$. And, as you know, the condition for completely destructive interference is that the path difference is half a wavelength, or, any integer number of wavelengths plus a half a wavelength. So, we have a minimum along any angle θ (less than 90°) such that:

$$(m+\frac{1}{2})\lambda = \frac{w}{2}\sin\theta \quad (m=0,1,2,\dots)$$

Now we turn our attention to the question of diffraction maxima. I should warn you that this analysis takes an unexpected turn. We do the exact same thing that we did to locate the minima, except that we set the path difference $\ell - r$ equal to an integer number of wavelengths (instead of a half a wavelength plus an integer number of wavelengths). This means that the path to point P from the leftmost point on side A (position A_1) of the slit, differs by an integer number of wavelengths, from the path to point P from the leftmost point on side B (position B_1). This will also hold true for the path from A_2 vs. the path from B_2. It will hold true for the path from A_3 vs. the path from B_3 as well. Indeed, it will hold true for any pair of corresponding points, one from side A and one point from side B. So, at point P, we have maximally constructive interference for every pair of corresponding points along the width of the slit. There is, however, a problem. While, for any pair of points, the oscillations at P will be maximal; that just means that P is at an angle that will make the amplitude of the oscillations of the electric field due to the pair of points maximal. But the electric field due to the pair of points will still be oscillating, e.g. from max up, to 0, to max down, to 0, and back to max up. And, these oscillations will not be in synchronization with the maximal oscillations due to other pairs of points. So, the grand total will not necessarily correspond to an intensity maximum. The big difference between this case and the minima case is that, in contrast to the time varying maximal oscillations just discussed, when a pair of contributions results in an electric field amplitude of zero, the electric field due to the pair is always zero. It is constant at zero. And, when every pair in an infinite sum of pairs contributes zero to the sum, at every instant in time, the result is zero. In fact, in our attempt to locate the angles at which maxima will occur, we have actually found some more minima. We can see this if we sum the contributions in a different order.

Consider the following diagram in which each half of the slit has itself been divided up into two parts:



If the path difference between "A_1 to P" and "B_1 to P", is one wavelength, then the path difference between " A_1 to P" and " A_1 'to P" must be half a wavelength. This yields completely destructive interference. Likewise for the path difference between " B_1 to P" and " B_1 'to P". So, for each half of the slit (with each half itself being divided in half) we can do the same kind of pair-wise sum that we did for the whole slit before. And, we get the same result—an infinite number of zero contributions to the electric field at P. All we have really done is to treat each half of the slit the way we treated the original slit. For the entire slit we found

$$(m+\frac{1}{2})\lambda = \frac{w}{2}\sin\theta \quad (m=0,1,2,\ldots)$$

Here we get the same result but with w itself replaced by $w/2$ (since we are dealing with half the slit at a time.) So now we have:

$$(m+\frac{1}{2})\lambda = \frac{w}{4}\sin\theta \quad (m=0,1,2,\ldots)$$

Let's abandon our search for maxima, at least for now, and see where we are in terms of our search for minima. From our consideration of the entire slit divided into two parts, we have $(m+\frac{1}{2})\lambda = \frac{w}{2}\sin\theta$ which can be written as $(2m+1)\lambda = w\sin\theta$ meaning that we have a minimum when:

$$w\sin\theta = 1\lambda, 3\lambda, 5\lambda, \ldots$$

From our consideration of each half divided into two parts (for a total of four parts) we have $(m+\frac{1}{2})\lambda = \frac{w}{4}\sin\theta$ which can be written $(4m+2)\lambda = w\sin\theta$ meaning that we have a minimum when:

$$w\sin\theta = 2\lambda, 6\lambda, 10\lambda, 14\lambda, \ldots$$

If we cut each of the four parts of the slit in half so we have four pairs of two parts, each $\frac{w}{8}$ in width, we find minima at $(m+\frac{1}{2})\lambda = \frac{w}{8}\sin\theta$ which can be written $(8m+4)\lambda = w\sin\theta$ meaning that we have a minimum when:

$$w\sin\theta = 4\lambda, 12\lambda, 20\lambda, 28\lambda, \ldots$$

If we continue this process of splitting each part of the slit in two and finding the minima for each adjacent pair, ad infinitum, we eventually find that we get a minimum when $w\sin\theta$ is equal to any integer number of wavelengths.

$$w\sin\theta = 1\lambda, 2\lambda, 3\lambda, 4\lambda, \ldots$$

a result which write as

$$m\lambda = w\sin\theta \quad (m=1,2,3,\ldots) \tag{B23.1}$$

We still haven't found any maxima. The only analytical way to determine the angles at which maxima occur is to do a full-fledged derivation of the intensity of the light as a function of position, and then mathematically solve for the maxima. While, this is not really as hard as it sounds, let's save that for an optics course and suffice it to say that, experimentally, we find maxima approximately midway between the minima. This includes the straight-ahead (0°) direction except that the straight-ahead maximum, a.k.a. the central maximum, is exactly midway between its neighboring minima.

Conditions Under Which Single-Slit Diffraction and Two-Slit Interference Occurs

To see the kinds of interference patterns that we have been talking about in this and the preceding chapter, certain conditions need to be met. For instance, in order to see one set of bright fringes in the two-slit interference experiment we need monochromatic light. Translated literally, from the Latin, monochromatic means one-color. Monochromatic light is single-frequency light. Strictly monochromatic light is an idealization. In practice, light that is classified as being essentially monochromatic, actually consists of an infinite set of frequencies that are all very close to the nominal frequency of the light. We refer to the set of frequencies as a band of frequencies. If all the frequencies in the set are indeed very close to the nominal frequency of the light, we refer to the light as narrow-band radiation.

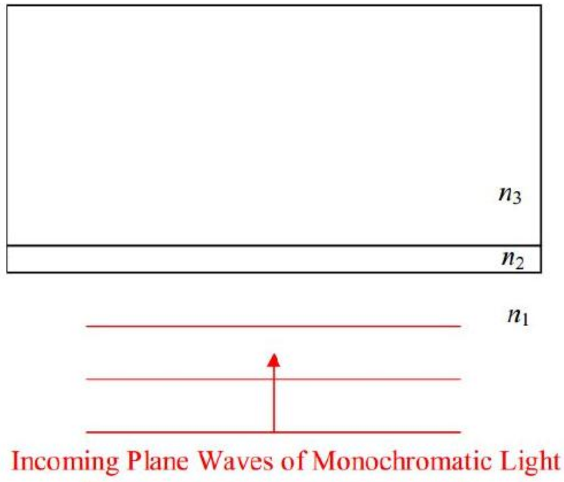
If you illuminate a single or double slit with light consisting of several discrete (individual) wavelengths of light, you get a mix of several interference/diffraction patterns. If you illuminate a single or double slit with a continuum of different frequencies, you find that minima from light of one wavelength are "filled in" by maxima and/or intermediate-amplitude oscillations of light of other wavelengths. Depending on the slit width and (in the case of two-slit interference) slit separation, and the wavelengths of the light, you may see a spectrum of colors on the screen.

In order for the kind of interference that we have been talking about to occur, the light must be coherent. The light must be temporally coherent (coherent with respect to time). While it applies to any part of a wave, I am going to talk about it in terms of crests. In temporally coherent light, one wave crest is part of the same wave that the preceding wave crest is a part of. In light having a great deal of temporal coherence this holds true for thousands of crests in a row. In light with very little temporal coherence this may hold true for only one or two crests in a row. Another way of stating it is to say that light that consists of a bunch of little wave pulses is temporally incoherent and light that consists of long continuous waves is temporally coherent. The long continuous wave can be called a "wave train". In terms of wave trains, light that is temporally incoherent consists of lots of short wave trains, whereas light that is temporally coherent consists of a relatively small number of long wave trains. The kinds of interference we have been talking about involve one part of a wave passing through a slit or slits in a mask, interfering with another part of the same wave passing through the same mask at a later time. In order for the latter to indeed be part of the same wave, the light must consist of long wave trains, in other words, it must be temporally coherent. Now, if the wave crest following a given wave crest is not part of the same wave, the distance from one wave crest to the next will be different for different crests. This means the wavelengths are different and hence the frequencies are different. Thus the light is not monochromatic. Under the opposite circumstances, the light is monochromatic. So, monochromatic light is temporally coherent.

The other condition is that the light must be spatially coherent. In the context of light that is normally incident on plane masks, this means that the wavefronts must be plane and they must have extent transverse to the direction in which the light is traveling. In the case of two-slit interference for instance, spatial coherence means that the light at one slit really is in phase with the light at the other slit. In the case of single-slit diffraction, spatial coherence means that light passing through the right half of the slit is in phase with light passing through the left half of the slit.

B24: Thin Film Interference

As the name and context imply, thin-film interference is another interference phenomenon involving light. Here's the picture, as viewed from above:

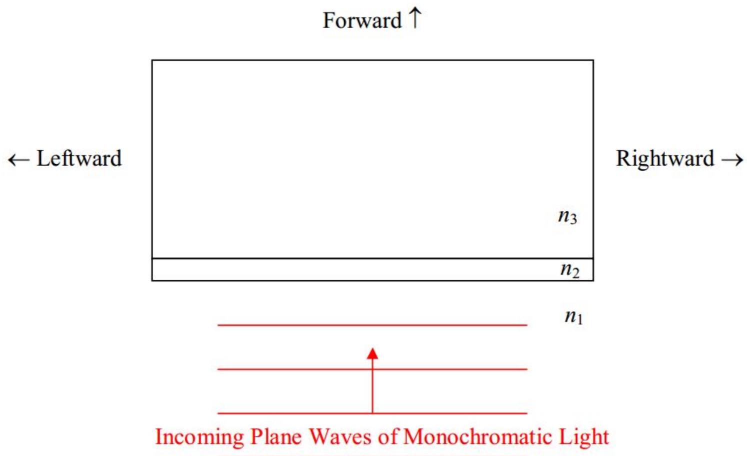


Involved are three transparent media: medium 1, medium 2, and medium 3, of index of refraction n_1, n_2, and n_3, respectively. (In general, a medium is a substance, but, in this context, vacuum is also considered a medium. The index of refraction n of a medium is the ratio of the speed of light in vacuum to the speed of light in that medium.) The phenomenon occurs whether or not $n_1 = n_3$, but, n_2 must be different from n_1 and n_3. Medium 2 is the "thin film." For thin-film interference to occur, the thickness of medium 2 must be on the order of the wavelength of the light. (The actual maximum thickness for which thin-film interference can occur depends on the coherence of the light.)

Here's the deal: Under most circumstances, when light encounters a smooth interface between two transparent media, some of the light goes through (transmitted light) and some of the light bounces off (reflected light). In the thin-film arrangement of three transparent media depicted above, for certain thicknesses of the thin film (medium 2) all the light can be reflected, and, for certain other thicknesses, all the light can be transmitted. You see this phenomenon when looking at soap bubbles, and sometimes when looking at puddles in the road (when there is a thin layer of oil on top of the water). Humans take advantage of the phenomenon by putting a thin coating of a transparent substance on lenses such as camera lenses and binocular lenses, a layer of just the right thickness for maximum transmission.

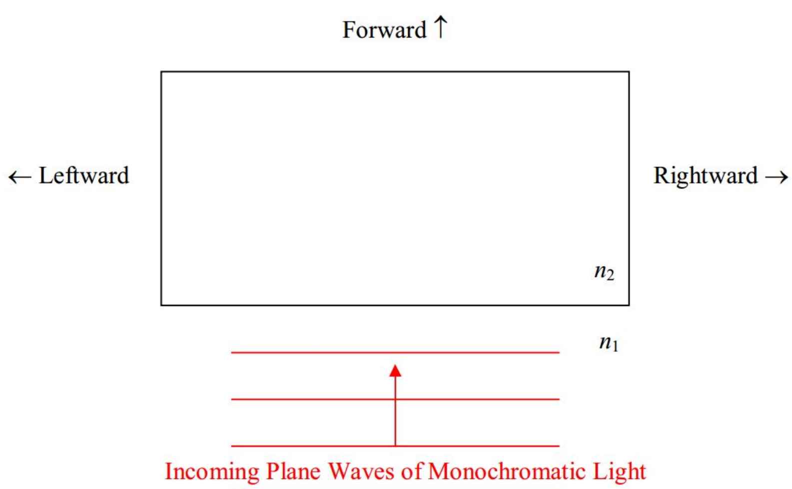
Based on the situations in which it occurs, it should be clear that we do not need monochromatic light to make thin-film interference happen. However, I am going to discuss it in terms of monochromatic light to get the idea across. Once you understand it in terms of monochromatic light, you can apply it to white light (a mixture of all the visible frequencies) to answer questions such as, "What wavelength of incoming white light will experience maximum reflection?" The answer helps one understand the rainbow of colors you might see on the surface of a puddle in broad daylight. You put a clear layer of gasoline on top of a clear puddle of water and thin-film interference results in maximal constructive interference of the reflected light, at certain wavelengths.

Based on your experience with soap bubbles and puddle surfaces, you know that the light does not have to be normally incident upon the interface between transparent media in order for thin-film interference to occur. However, the analysis is easier for the case of normal incidence, so, in this chapter, I am going to limit our analysis to the case of normal incidence.



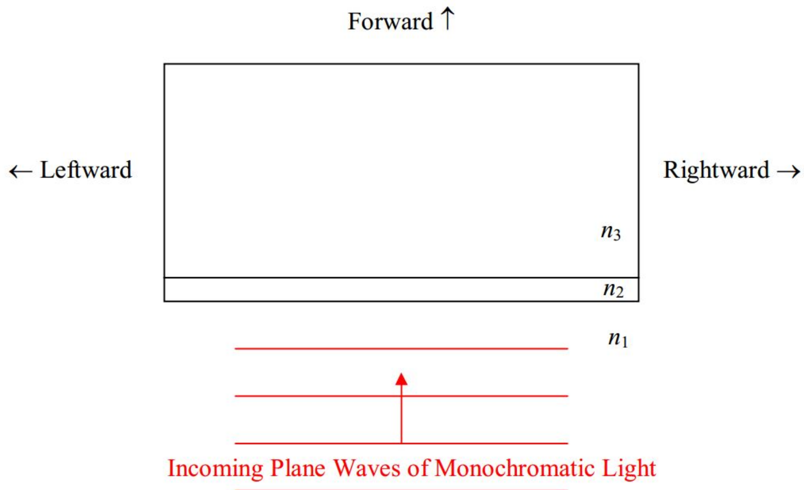
Here's the gross idea: In going through a thin transparent film, light encounters two interfaces, the n_1 abutting n_2 interface, and, the n_2 abutting n_3 interface. At each interface, some of the light gets through and some is reflected. We can say all that we need to say about thin-film interference, just by talking about the reflected light. The thing is, light reflected off the second interface interferes with light reflected off the first interface. The reflected light can be thought of as being from two sources at two different locations, one source being the n_1 abutting n_2 interface, and the other being the n_2 abutting n_3 interface. But, there is a fixed phase difference between the light from the two sources because the light was originally part of one and the same source of incoming light. The light reflected from the second interface travels farther, to arrive back at the same backward position, than the light reflected from the first interface does. If you figure, that, when that extra distance is one wavelength, the interference of reflected light is constructive, and that, when that extra distance is half a wavelength, the interference of reflected light is destructive, then you've got the right idea, but, there are two "complications" that need to be taken into account.

The first complication has to do with phase reversal upon reflection. Consider a single interface (forget about the thin film for a moment) between two transparent medium. Assume light to be incident upon the interface. Call the index of refraction of the medium in which the light is initially traveling, n_1, and call the index of refraction of the medium in which the transmitted light travels, n_2. Experimentally we find that if $n_2 > n_1$, then the reflected light is phase reversed, but, that if $n_2 < n_1$, the reflected light experiences no phase change at all.



Regarding what we mean by phase reversal: Think of a crest of a wave hitting the interface. More specifically, let the electric field be oscillating along the vertical (into and out of the page in the diagram) so that at the instant under consideration, we have a forward-moving maximum upward-directed electric field vector at the location of the interface. An infinitesimal time dt later, we will find a forward-moving maximum-upward electric field vector at a point v_2dt forward of the interface. If $n_2 > n_1$ (phase reversal condition met), then, at the same instant in time (dt after the forward-moving maximum upward-directed electric field vector hits the interface) we have a maximum downward-directed electric field vector, traveling backward, at a point v_1 dt behind the interface. This is what we mean by phase reversal. An incoming electric field vector pointing in one direction, bounces off the interface as an electric field vector pointing in the opposite direction. If there is no phase reversal, then, at the specified instant in time, we would have a maximum upward-directed electric field vector, traveling backward, at a point$v_1 dt$ behind the interface. With no phase reversal, an electric field vector pointing in one direction, bounces off the interface as an electric field vector pointing in the same direction.

Now back to the thin-film setup:



Recall that to get back to some specified point in space, light reflecting off the second interface (between medium 2 and medium 3) travels farther than the light reflecting off the first interface (between medium 1 and medium 2). Before, we hypothesized that if the path difference was a half a wavelength, the light from the two "sources" would interfere destructively, but, that if it was a full wavelength, the interference would be constructive. Now, if there is no phase reversal from either surface (because $n_2 < n_1$ and $n_3 < n2$), or, if there is phase reversal from both surfaces (because $n_2 > n_1$ and $n_3 > n_2$) then our original hypothesis is still viable. But, if we have phase reversal at one of the interfaces but not the other ($n_2 > n_1$ but $n_3 < n_2$, or, $n_3 > n_2$ but $n_2 < n_1$), then the situation is reversed. A path difference of one wavelength would result in a crest interfering with a "crest that upon reflection turned into a trough" meaning that a path difference of one wavelength would result in destructive interference. And, a path difference of half a wavelength would result in a crest interfering with a "trough that upon reflection turned into a crest" meaning that a path difference of half a wavelength would result in constructive interference. Okay, we have addressed the phase reversal issue. We have one more complication to deal with. The thing is, the light that bounces off the second interface, not only travels a greater distance, but, it travels at a different speed while it is traveling that extra distance because it is in a different medium. Let's see how this complication plays out.

The phenomenon holds true for every part of the wave. I focus the attention on crests, just because I find them easier to keep track of. For now, I also want to focus attention on the no-phase-reversal constructive interference case. Consider an instant when a crest of the forward traveling incoming wave hits the first interface. The crest of the transmitted wave travels through the interface, proceeds through the second medium at speed $v_2 = c/n_2$, bounces off the interface with the third medium, and travels back through the second medium, completing its round trip (of distance two times the thickness of the second medium) through the second medium in time:

$$t_2 = \frac{2(\text{thickness})}{v_2}$$

Now, while that is going on, the next crest from the incoming wave is traveling forward at speed $v_1 = c/n_1$. It arrives at the same interface (between medium 1 and medium 2) at time

$$t_1 = \frac{\lambda_1}{v_1}$$

where λ_1 is the wavelength of the light while it is traveling in medium 1. (Remember, the source establishes the frequency of the light and that never changes, but, from $v = \lambda f$, the wavelength depends on the speed of the wave in the medium in which the light is traveling.) For constructive interference (under no phase-reversal conditions), we must have

$$t_1 = t_2$$

which, from the expressions for t_1 and t_2 above, can be written as:

$$\frac{\lambda_1}{v_1} = \frac{2(\text{thickness})}{v_2}$$

Substituting $v_1 = \lambda_1 f$ and $v_2 = \lambda_2 f$ yields:

$$\frac{\lambda_1}{\lambda_1 f} = \frac{2(\text{thickness})}{\lambda_2 f}$$

$$\lambda_2 = 2(\text{thickness})$$

I'm going to leave the result in this form because twice the thickness of the thin film is the path difference. So the equation is saying that, under no-phase-reversal conditions, there will be constructive interference of the light reflected from the two interfaces, when the wavelength that the light has in the material of which the thin film consists, is equal to the path difference. Of course, if the path difference is $2\lambda_2, 3\lambda_2, 4\lambda_2$, etc. we will also get constructive interference. We can write this as:

$$m\lambda_2 = 2(\text{thickness}) \quad (m = 1, 2, 3, \dots) \tag{B24.1}$$

where:

> m is an integer,
>
> λ_2 is the in-the-thin-film wavelength of the light, and,
>
> $thickness$ is the thickness of the thin film.

This condition is also appropriate for the case of maximal constructive interference when phase reversal occurs at both interfaces. But, this condition yields completely destructive interference of reflected light when there is phase reversal at one interface but not the other.

The in-the-film wavelength λ_2 of the light can be expressed most simply in terms of the wavelength λ_1 of the light in the medium in which it is originally traveling, by setting the two expressions for the frequency equal to each other. From $v_1 = \lambda_1 f$ we have $f = v_1/\lambda_1$. but from $n_1 = c/v_1$ we have $v_1 = c/n_1$. Replacing v_1 in $f = v_1/\lambda_1$ with c/n_1 yields $f = \frac{c}{n_1\lambda_1}$. In a similar manner, we find that f can be expressed as $f = \frac{c}{n_2\lambda_2}$. Setting the two expressions for f equal to each other yields:

$$\frac{c}{n_1\lambda_1} = \frac{c}{n_2\lambda_2}$$

which can be written as:

$$\lambda_2 = \frac{n_1}{n_2}\lambda_1 \tag{B24.2}$$

To get a maximum when we have phase reversal at one, and only one, interface, we need the path difference (twice the thickness) to be half a wavelength-in-the-film,

$$\frac{1}{2}\lambda_2 = 2(\text{thickness})$$

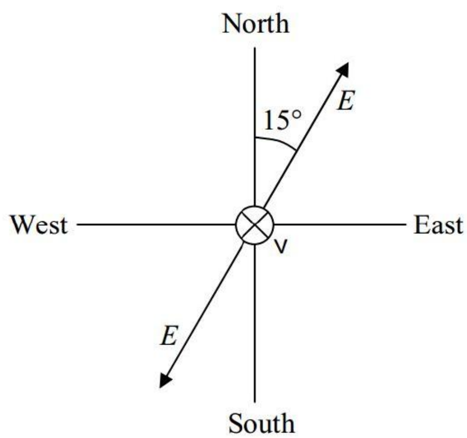
or that, plus, an integer number of full wavelengths-in-the-film:

$$(m + \frac{1}{2})\lambda_2 = 2(\text{thickness}) \quad (m = 1, 2, 3, \dots) \tag{B24.3}$$

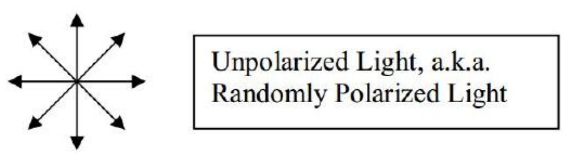
This is also the condition for completely destructive interference for the case of no phase reversal, or, phase reversal at both interfaces. This is exactly what we want for a camera lens that is to be used in air ($n_1 = 1.00$). Consider a clear plastic medium of index of refraction $n_2 = 1.3$. Now consider the wavelength that light from the middle of the visible spectrum (green light) would have in that medium. Put a coating of the plastic, one quarter as thick as that wavelength is long, on a lens made of glass having an index of refraction $n_3 = 1.5$. (Note that we have phase reversal at both interfaces.) With that coating, the lens reflects none of the light of the specified wavelength that is normally incident on the lens (and a reduced amount of light of nearby wavelengths). That is, it transmits more of the light than it would without the coating. This is the desired effect.

B25: Polarization

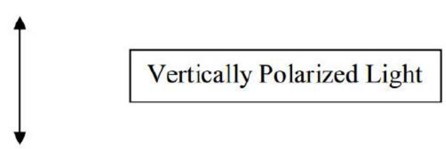
The polarization direction of light refers to the two directions or one of the two directions in which the electric field is oscillating. For the case of completely polarized light there are always two directions that could be called the polarization direction. If a single direction is specified, then that direction, and the exact opposite direction, are both the directions of polarization. Still, specifying one direction completely specifies the direction of polarization. For instance for light that is traveling straight downward near the surface of the earth, if the polarization direction is said to be a compass heading of 15°, that unambiguously means that the electric field oscillates so that it is at times pointing in the direction with a compass heading of 15°, and at times pointing in the direction with a compass heading of 195° (15° west of south).



Randomly polarized light, a.k.a. unpolarized light, has electric field oscillations in each and every direction perpendicular to the direction in which the light is traveling. Such light is often depicted, as viewed from behind, (where forward is the direction in which the light is traveling) as:

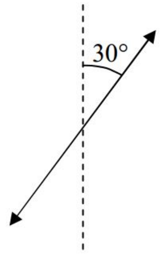


Vertically polarized light traveling horizontally away from you is typically depicted as:

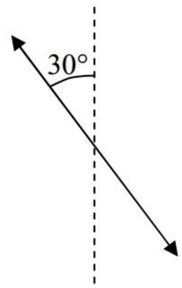


where the direction in which the light is traveling is “into the page” and upward is “toward the top of the page.” At a particular position through which the light is traveling, starting at an instant when the electric field vector at that position is upward and maximum, the electric field will decrease to zero, then be downward and increasing, reach a maximum downward, then be downward and decreasing, become zero, then be upward and increasing, then reach a maximum upward, and repeat, continually. The diagram depicting the polarization indicates the directions that the electric field does point, at some time during its oscillations. It in no way is meant to imply that the electric field is pointing in two directions at the same time.

Light that is traveling horizontally away from you that is polarized at 30° with respect to the vertical could be either:

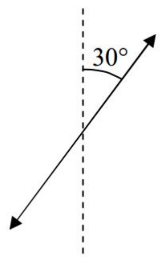


or



If you encounter such an ambiguous specification of polarization in a problem statement then the answer is the same for either case, so, it doesn't matter which of the two possible polarization directions you pick. Pick one arbitrarily and work with it.

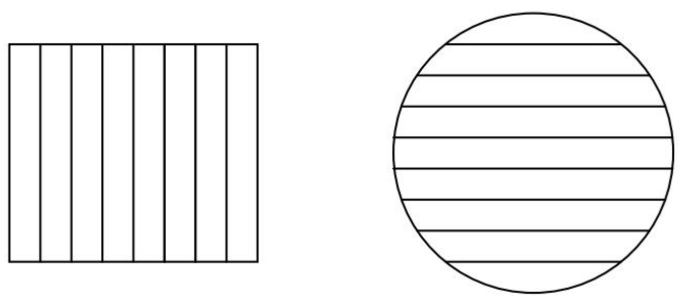
Light that is traveling horizontally away from you and is polarized, from your point of view, at 30° clockwise from the vertical is, however, unambiguously:



Polarizers

A plastic material is manufactured in the form of flat sheets that polarize light that travels through them. A sample of such a flat sheet is called a polarizer. In use, one typically causes light to travel toward a polarizer along a direction that is perpendicular to the polarizer. In other words, one causes the light to be normally incident upon the polarizer.

Schematically, one typically depicts a polarizer by means of a rectangle or a circle filled with parallel line segments.



The orientation of the lines is referred to as the polarization direction of the polarizer. The effect of a polarizer is to transmit light that is polarized in the same direction as that of the polarizer, and to block (absorb or reflect) light that is polarized at right angles to the direction of the polarizer.

The polarization direction of the rectangular polarizer depicted above is vertical. So, it lets vertically-polarized light through and blocks horizontally-polarized light.

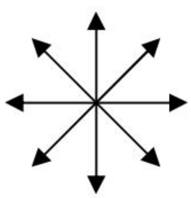
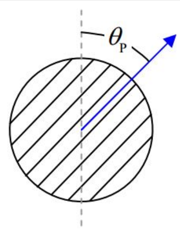
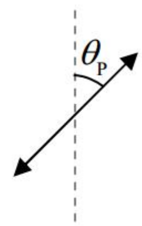
The polarization direction of the circle-shaped sample of polarizing material depicted above is horizontal. So, it lets horizontally-polarized light through and blocks vertically-polarized light.

When unpolarized light (a.k.a. randomly-polarized light) is normally incident on any polarizer, half the light gets through. So, if the intensity of the incoming light is I_0, then the intensity of the light that gets through, call it I_1, is given by:

$$I_1 = \frac{1}{2} I_0 \tag{B25.1}$$

In completely unpolarized light, the electric field vectors are oscillating in every direction that is perpendicular to the direction in which the light is traveling. But all the electric field vectors are, as the name implies, vectors. As such, we can break every single one of them up into a component along the direction of polarization of the polarizer and a component that is perpendicular to the polarization direction of the polarizer. A polarizer will let every component that is along the direction of polarization of the polarizer through, and block every component that is perpendicular to the polarization direction. In completely unpolarized light, no matter what the direction of polarization of the polarizer is, if you break up all the electric field vectors into components parallel to and perpendicular to the polarizer's polarization direction, and add all the parallel components together, and then separately add all the perpendicular components together, the two results will have the same magnitude. This means that we can view completely unpolarized light as being made up of two halves: half polarized parallel to the polarizer's polarization direction, and half polarized perpendicular to the polarizer's polarization direction. The half that is polarized parallel to the polarizer's polarization direction gets through the polarizer, and the other half doesn't.

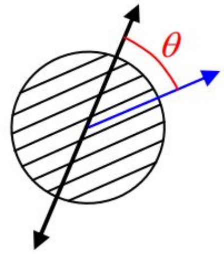
Unpolarized Light Traveling Directly Away From You

 When completely unpolarized light of intensity $I_0...$	 ... is normally incident on a polarizer whose polarization direction makes an angle θ_p with the vertical...	 ... the light that gets through is polarized in the polarizer's direction of polarization, and, has an intensity $I_1 = \frac{1}{2} I_0$.

Note how the effect of a polarizer on the intensity of normally-incident unpolarized light does not depend on the orientation of the polarizer. You get the same intensity $I_1 = \frac{1}{2} I_0$ of light getting through the polarizer, no matter what the direction of polarization of the polarizer is.

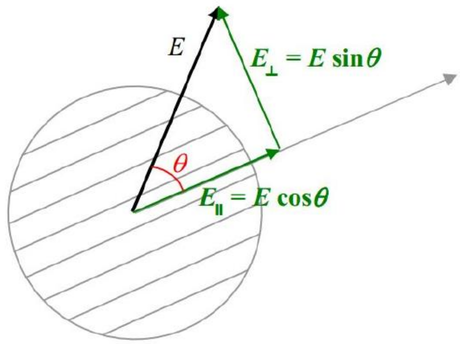
Now suppose that we have some light that, for whatever reason, is already polarized. When polarized light is normally incident on a polarizer, the intensity of the light that gets through does depend on the direction of polarization of the polarizer (relative to that of the incoming light). Suppose for instance, that the incoming light,

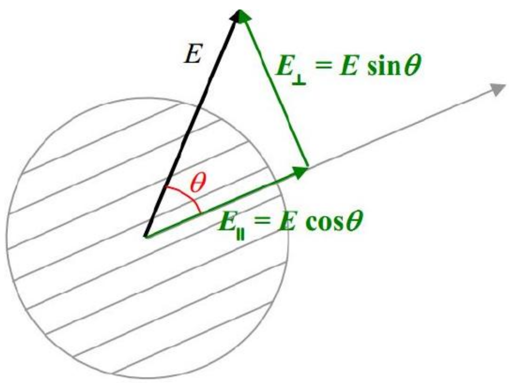
is polarized at an angle θ with respect to the polarization direction of a polarizer upon which the light is normally incident:



Before it hits the polarizer, the light's electric-field-oscillations-amplitude vector,

can be broken up into a component parallel to the polarizer's polarization direction and a component perpendicular to the polarizer's polarization direction.





The parallel component $E_{||}$ = E cos \theta\) gets through the polarizer, the perpendicular component does not.

Now the intensity of polarized light is proportional to the square of the amplitude of the oscillations of the electric field. So, we can express the intensity of the incoming light as

$$I_o = (constant)E^2$$

and the intensity of the light that gets through as:

$$I_1 = (constant)E_{||}^2$$

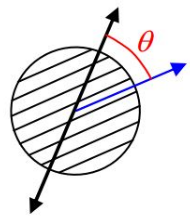
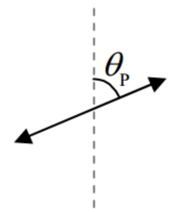
$$I_1 = (constant)(E\cos\theta)^2$$

$$I_1 = (constant)E^2(\cos\theta)^2$$

$$I_1 = I_o (\cos\theta)^2 \tag{B25.2}$$

Summarizing:

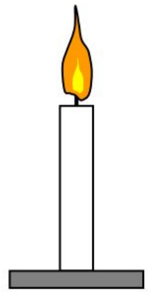
Polarized Light Traveling Directly Away From You

 When polarized light of intensity $I_0 \ldots$	 ... is normally incident on a polarizer whose polarization direction makes an angle θ with the polarization direction of the light...	 ... the light that gets through is polarized in the polarizer's direction of polarization, and, has an intensity $I_1 = I_0 (\cos\theta)^2$.

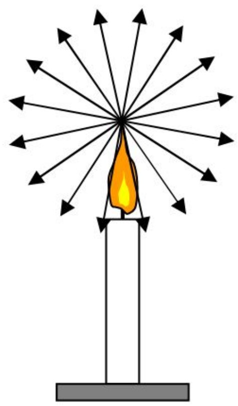
B26: Geometric Optics, Reflection

We now turn to a branch of optics referred to as geometric optics and also referred to as ray optics. It applies in cases where the dimensions of the objects (and apertures) with which the light interacts are so large as to render diffraction effects negligible. In geometric optics we treat light as being made up of an infinite set of narrow beams of light, called light rays, or simply rays, traveling through vacuum or transparent media along straight line paths. Where a ray of light encounters the surface of a mirror, or the interface between the transparent medium in which it (the light) is traveling and another transparent medium, the ray makes an abrupt change in direction, after which, it travels along a new straight line path.

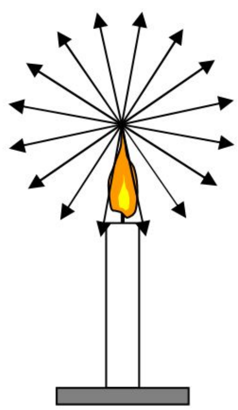
In the geometric optics model of light, we see light emitted by sources of light because the light enters our eyes. Consider for instance, a candle.



Every point of the flame of the candle emits rays of light in every direction.



While the preceding diagram conveys the idea in the statement preceding the diagram, the diagram is not the complete picture. To get a more complete picture of what's going on, what I want you to do is to look at the diagram provided, form a picture of it in your mind, and, to the picture in your mind, add the following embellishments:

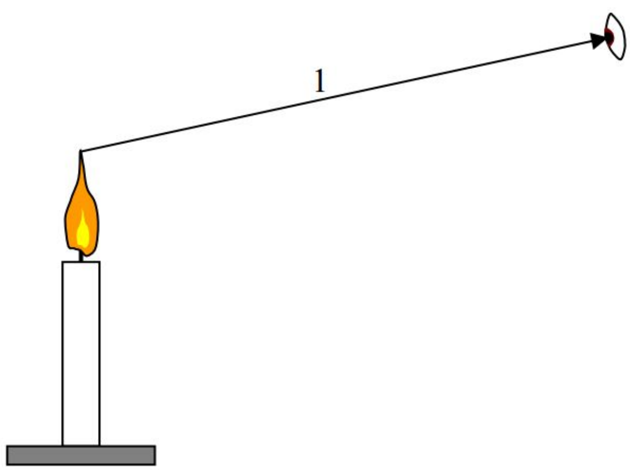


1. First off, I need you to imagine it to be a real candle extending in three dimensions. Our set of rays depicted as arrows whose tips are all on a circle becomes a set of rays depicted as arrows whose tips all end on a sphere. Thus, in addition to rays going

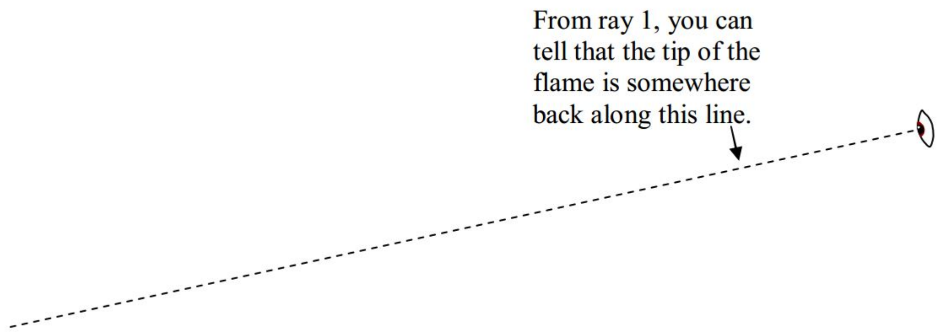
(at various angles) upward, downward and to the sides, you've got rays proceeding (at various angles) away from you and toward you.

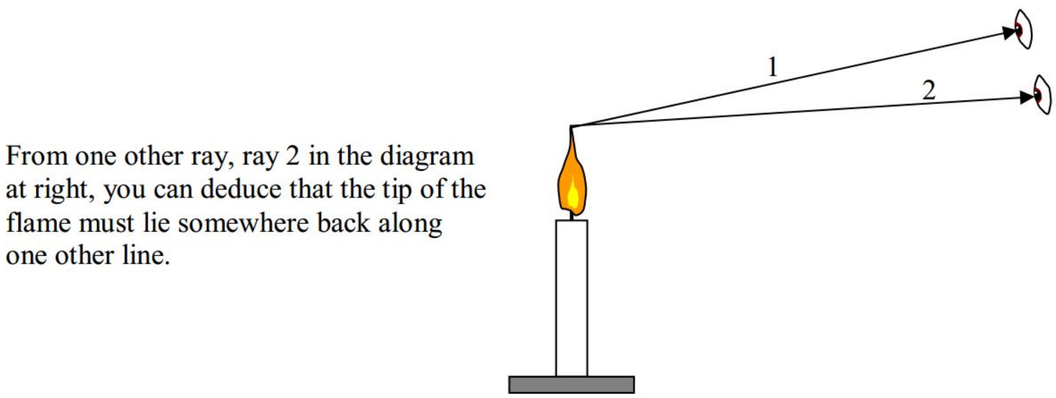
2. Now I need you to add more rays to the picture in your mind. I included 16 rays in the diagram. In three dimensions, you should have about 120 rays in the picture in your mind. I need you to bump that up to infinity.
3. In the original diagram, I showed rays coming only from the tip of the flame. At this point, we have an infinite number of rays coming from the tip of the flame. I need you to picture that to be the case for each point of the flame, not just the tip of the flame. In the interest of simplicity, in the picture in your mind, let the flame of the candle be an opaque solid rather than gaseous, so that we can treat all our rays as coming from points on the surface of the flame. Neglect any rays that are in any way directed into the flame itself (don't include them in the picture in your mind). Upon completion of this step, you should have, in the picture in your mind, an infinite number of rays coming from each of the infinite number of points making up the surface of the flame.
4. For this next part, we need to establish the setting. I'm concerned that you might be reading this in a room in which lit candles are forbidden. If so, please relocate the candle in the picture in your mind to the dining room table in your home, or, replace the candle with a fake electric-powered candle such as you might see in a home around Christmastime. Now I need you to extend each of the rays in the picture in your mind all the way out to the point where they bump into something. Please end each ray at the point where it bumps into something. (A ray that bumps into a non-shiny surface, bounces off in all directions [diffuse reflection]. Thus, each ray that bumps into a non-shiny surface creates an infinite set of rays coming from the point of impact. A ray bumping into perfectly shiny surfaces continues as a single ray in one particular, new, direction [specular reflection]. To avoid clutter, let's omit all the reflected rays from the picture in your mind.)

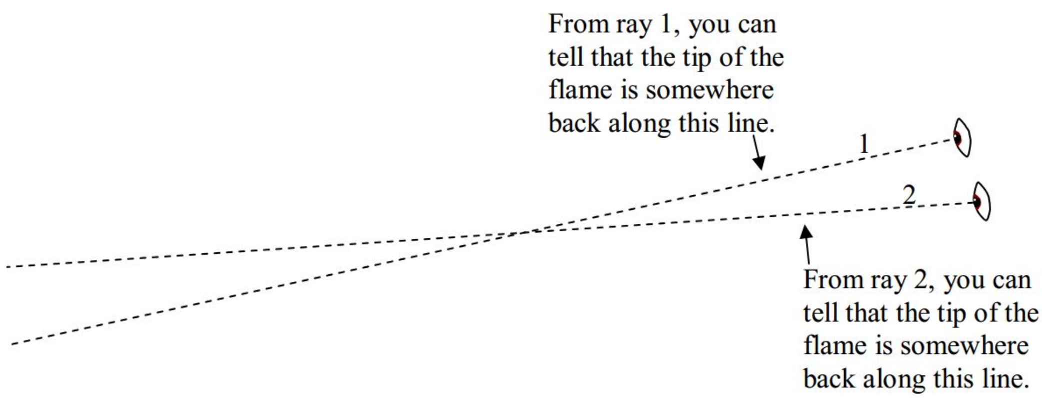
If you have carried out steps 1-4 above, then you have the picture, in your mind, of the geometric optics model of the light given off by a light-emitting object. When you are in a room with a candle such as the one we have been discussing, you can tell where it is (in what direction and how far away—you might not be able to give very accurate values, but you can tell where it is) by looking at it. When you look at it, an infinite number of rays, from each part of the surface of the flame, are entering your eyes. What is amazing is how few rays you need to determine where, for instance, the tip of the flame is. Of the infinite number of rays available to you, you only need two! Consider what you can find out from a single ray entering your eye:



From just one of the infinite number of rays, you can deduce the direction that the tip of the flame is in, relative to you. In other words, you can say that the tip of the flame lies somewhere on the line segment that both contains the ray that enters your eye, and, that ends at the location of your eye.



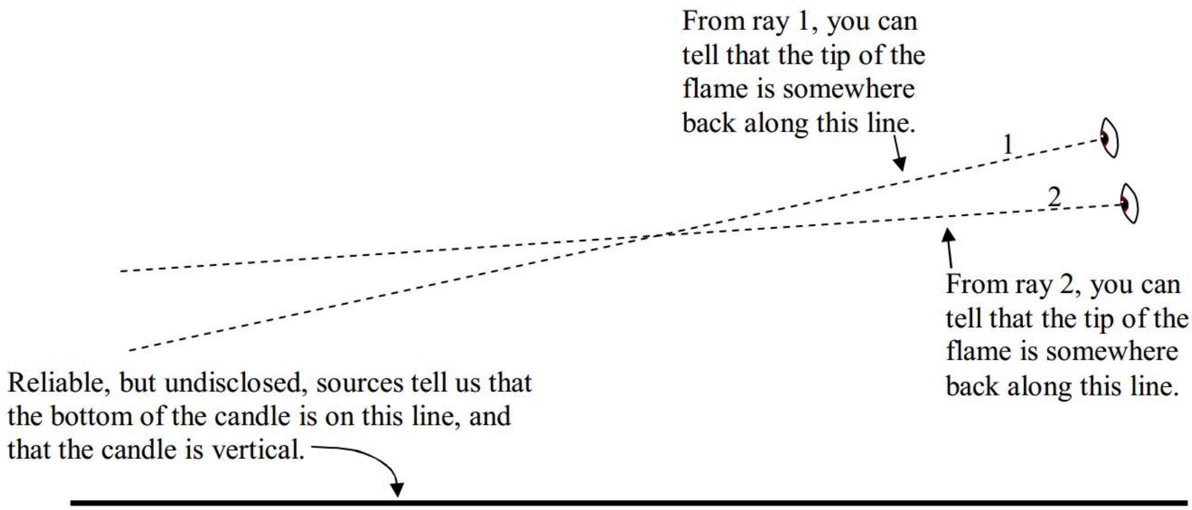




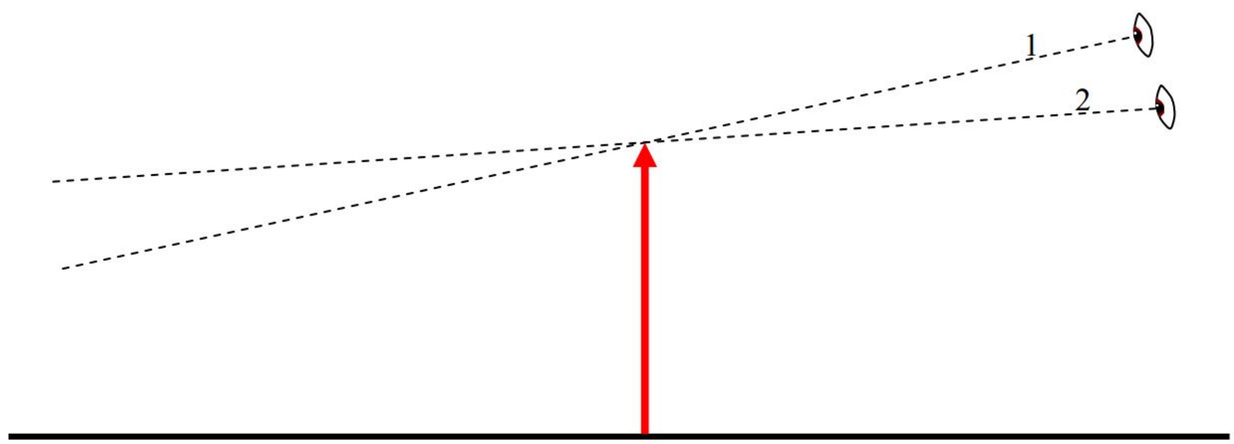
There is only one point in space that is both “somewhere back along line 1” and “somewhere back along line 2.” That one point is, of course, the point where the two lines cross. The eyebrain system is an amazing system. When you look at something, your eye-brain system automatically carries out the “trace back and find the intersection” process to determine how far away that something is. Again, you might not be able to tell me how many centimeters away the candle, for instance, is, but you must know how far away it is because you would know about how hard to throw something to hit the candle.

This business of tracing rays back to see where they come from is known as ray tracing and is what geometric optics is all about.

At this point I want to return our attention to the candle to provide you with a little bit more insight into the practice of ray tracing. Suppose that when you were determining the location of the tip of the flame of the candle, you already had some additional information about the candle. For instance, assume: You know that the rays are coming from the upper extremity of the candle; you know that the bottom of the candle is on the plane of the surface of your dining room table; and you know that the candle is vertical. We’ll also assume that the candle is so skinny that we are not interested in its horizontal extent in space, so, we can think of it as a skinny line segment with a top (the tip of the candle) and a bottom, the point on the candle that is at table level. The intersection of the plane of the table surface with the plane of the two rays is a line, and, based on the information we have, the bottom of the candle is on that line.



Taken together with the information gleaned from the rays, we can draw in the entire (skinny) candle, on our diagram, and from the diagram, determine such things as the candle's height, position, and orientation (whether it is upside down [inverted] or right side up [erect]). In adding the candle to the diagram, I am going to draw it as an arrow. Besides the fact that it is conventional to draw objects in ray tracing diagrams as arrows, we use an arrow to represent the candle to avoid conveying the impression that, from the limited facts we have at our disposal, we have been able to learn more about the candle (diameter, flame height, etc) than is possible. (We can only determine the height, position, and orientation.)

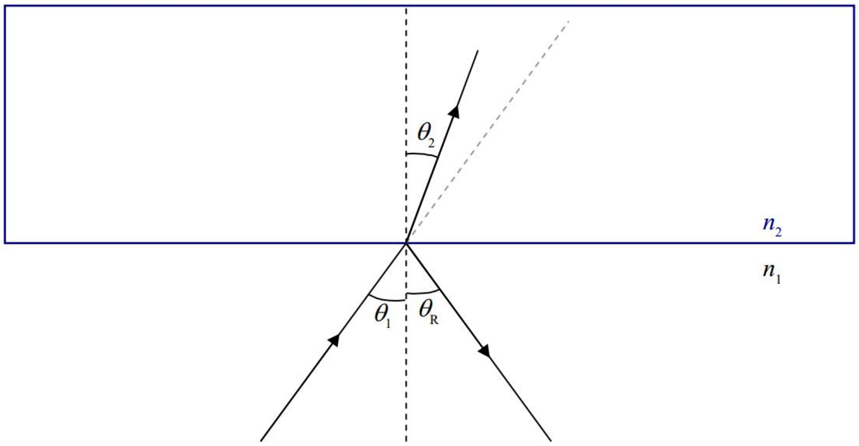


The trace-back method for locating the tip of the candle flame works for any two rays, from among the infinite number of rays emitted by the tip of the candle flame. All the rays come from the same point and they all travel along different straight line paths. As such, the rays are said to diverge from the tip of the candle flame. The trace-back method allows us to determine the point from which the rays are diverging.

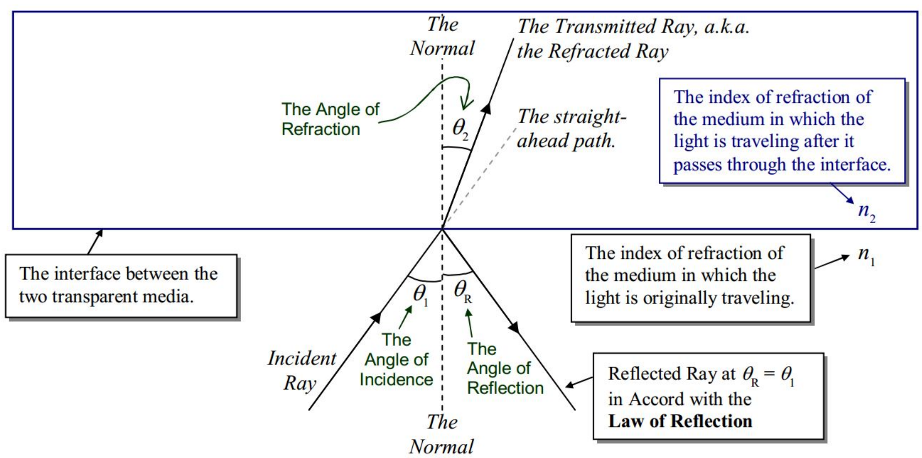
By means of lenses and mirrors, we can redirect rays of light, infinite numbers of them at a time, in such a manner as to fool the eye-brain system that is using the trace-back method into perceiving the point from which the rays are diverging as being someplace other than where the object is. To do so, one simply has to redirect the rays so that they are diverging from someplace other than their point of origin. The point, other than their point of origin, from which the rays diverge (because of the redirection of rays by mirrors and/or lenses), is called the image of the point on the object from which the light actually originates.

B27: Refraction, Dispersion, Internal Reflection

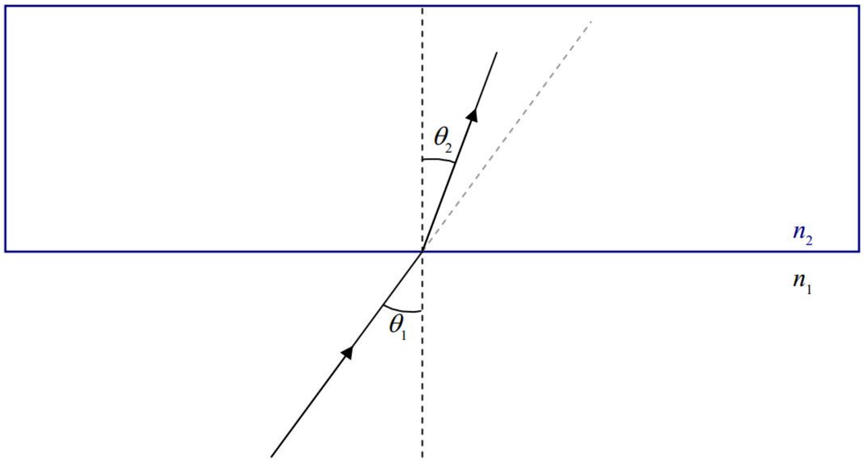
When we talked about thin film interference, we said that when light encounters a smooth interface between two transparent media, some of the light gets through, and some bounces off. There we limited the discussion to the case of normal incidence. (Recall that normal means perpendicular to and normal incidence is the case where the direction in which the light is traveling is perpendicular to the interface.) Now we consider the case in which light shining on the smooth interface between two transparent media, is not normally incident upon the interface. Here's a "clean" depiction of what I'm talking about:



and here's one that's all cluttered up with labels providing terminology that you need to know:



As in the case of normal incidence, some of the light is reflected and some of it is transmitted through the interface. Here we focus our attention on the light that gets through.



Experimentally we find that the light that gets through travels along a different straight line path than the one along which the incoming ray travels. As such, the transmitted ray makes an angle θ_2 with the normal that is different from the angle θ_1 that the incident ray makes with the normal.

The adoption of a new path by the transmitted ray, at the interface between two transparent media is referred to as refraction. The transmitted ray is typically referred to as the refracted ray, and the angle θ_2 that the refracted ray makes with the normal is called the angle of refraction. Experimentally, we find that the angle of refraction θ_2 is related to the angle of incidence θ_1 by Snell's Law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{B27.1}$$

where:

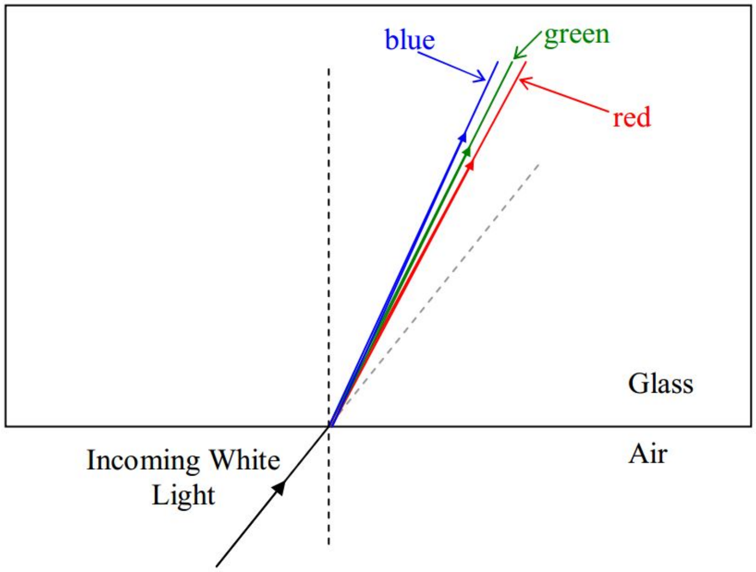
- n_1 is the index of refraction of the first medium, the medium in which the light is traveling before it gets to the interface,
- θ_1 is the angle that the incident ray (the ray in the first medium) makes with the normal,
- n_2 is the index of refraction of the second medium, the medium in which the light is traveling after it goes through the interface, and,
- θ_2 is the angle that the refracted ray (the ray in the second medium) makes with the normal.

Dispersion

On each side of the equation form of Snell's law we have an index of refraction. The index of refraction has the same meaning as it did when we talked about it in the context of thin film interference. It applies to a given medium. It is the ratio of the speed of light in that medium to the speed of light in vacuum. At that time, I mentioned that different materials have different indices of refraction, and in fact, provided you with the following table:

Medium	Index of Refraction
Vacuum	1
Air	1.00
Water	1.33
Glass(Depends on the kind of glass. Here is one typical value.)	1.5

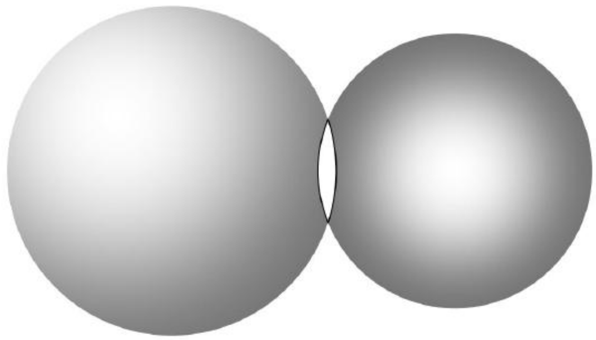
What I didn't mention back then is that there is a slight dependence of the index of refraction on the wavelength of the visible light, such that, the shorter the wavelength of the light, the greater the index of refraction. For instance, a particular kind of glass might have an index of refraction of 1.49 for light of wavelength 695 nm (red light), but an index of refraction that is greater than that for shorter wavelengths, including an index of refraction of 1.51 for light of wavelength 405 nm (blue light). The effect in the case of a ray of white light traveling in air and encountering an interface between air and glass is to cause the different wavelengths of the light making up the white light to refract at different angles.



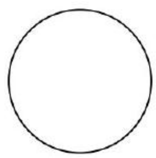
This phenomena of white light being separated into its constituent wavelengths because of the dependence of the index of refraction on wavelength, is called dispersion.

B28: Thin Lenses - Ray Tracing

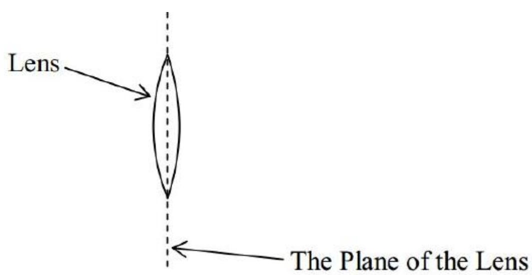
A lens is a piece of transparent material whose surfaces have been shaped so that, when the lens is in another transparent material (call it medium 0), light traveling in medium 0, upon passing through the lens, is redirected to create an image of the light source. Medium 0 is typically air, and lenses are typically made of glass or plastic. In this chapter we focus on a particular class of lenses, a class known as thin spherical lenses. Each surface of a thin spherical lens is a tiny fraction of a spherical surface. For instance, consider the two spheres:



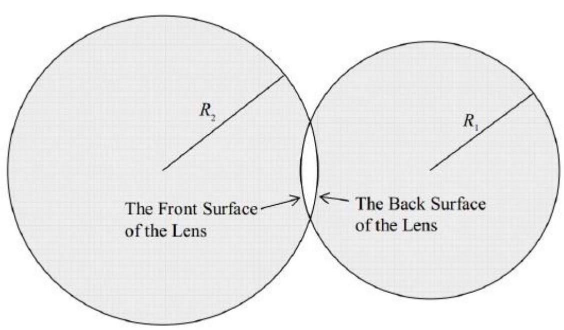
A piece of glass in the shape of the intersection of these two spherical volumes would be a thin spherical lens. The intersection of two spherical surfaces is a circle. That circle would be the rim of the lens. Viewed face on, the outline of a thin spherical lens is a circle.



The plane in which that circle lies is called the plane of the lens. Viewing the lens edge-on, the plane of the lens looks like a line.



Each surface of a thin spherical lens has a radius of curvature. The radius of curvature of a surface of a thin spherical lens is the radius of the sphere of which that surface is a part. Designating one surface of the lens as the front surface of the lens and one surface as the back surface, in the following diagram:



we can identify R_1 as the radius of curvature of the front surface of the lens and R_2 as the radius of curvature of the back surface of the lens.

The defining characteristic of a lens is a quantity called the focal length of the lens. At this point, I'm going to tell you how you can calculate a value for the focal length of a lens, based on the physical characteristics of the lens, before I even tell you what focal length means. (Don't worry, though, we'll get to the definition soon.) The **lens-maker's equation** gives the reciprocal of the focal length in terms of the physical characteristics of the lens (and the medium in which the lens finds itself):

The Lens-Maker's Equation:

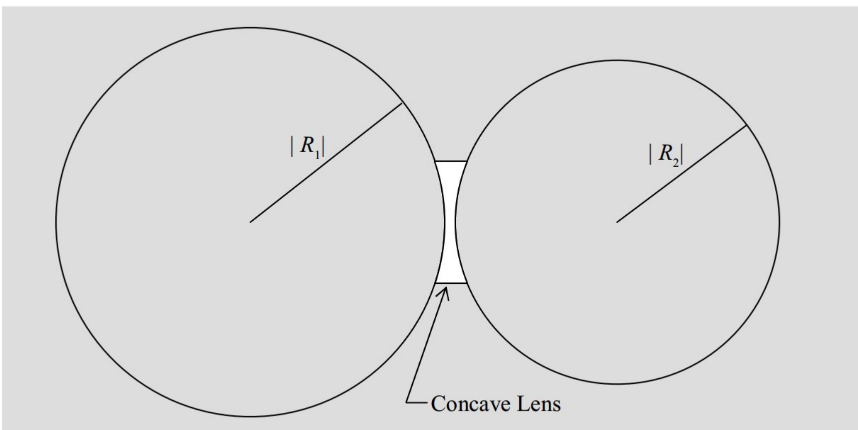
$$\frac{1}{f}=(n-n_0)\left(\frac{1}{R_1}+\frac{1}{R_2}\right) \tag{B28.1}$$

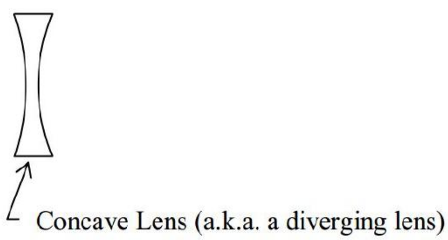
where:

- f is the focal length of the lens,
- n is the index of refraction of the material of which the lens is made,
- n_o is the index of refraction of the medium surrounding the lens (n_o is typically 1.00 because the medium surrounding the lens is typically air),
- R_1 is the radius of curvature of one of the surfaces of the lens, and,
- R_2 is the radius of curvature of the other surface of the lens.

Before we move on from the lens-maker's equation, I need to tell you about an algebraic sign convention for the R values. There are two kinds of spherical lens surfaces. One is the "curved out" kind possessed by any lens that is the intersection of two spheres. (This is the kind of lens that we have been talking about.) Such a lens is referred to as a convex lens (a.k.a. a converging lens) and each ("curved out") surface is referred to as a convex surface. The radius of curvature R for a convex surface is, by convention, positive.

The other kind of lens surface is part of a sphere that does not enclose the lens itself. Such a surface is said to be "curved in" and is called a concave surface.

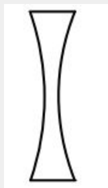




By convention, the absolute value of R for a concave surface is still the radius of the sphere whose surface coincides with that of the lens. But, the quantity R contains additional information in the form of a minus sign used to designate the fact that the surface of the lens is concave. R is still called the radius of curvature of the surface of the lens despite the fact that there is no such thing as a sphere whose radius is actually negative.

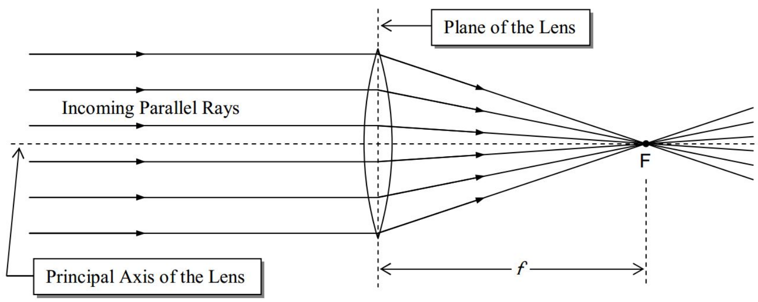
Summarizing, our convention for the radius of curvature of the surface of a lens is:

Surface of Lens	Algebraic Sign of Radius of Curvature R

Surface of Lens	Algebraic Sign of Radius of Curvature R
Convex 	+
Concave 	−

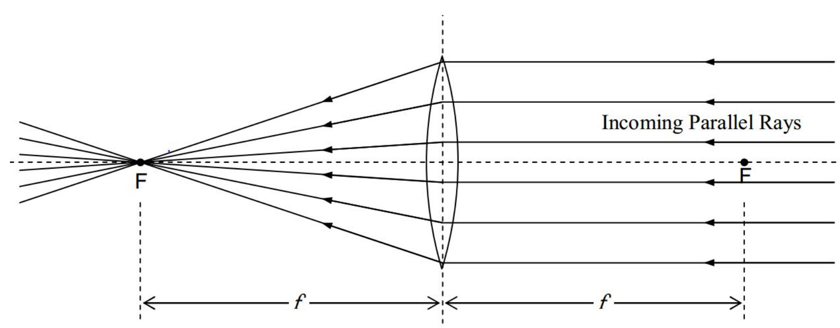
So, what does a lens do? It refracts light at both surfaces. What's special about a lens is the effect that it has on an infinite set of rays, collectively. We can characterize the operational effect of a lens in terms of the effect that it has on incoming rays that are all parallel to the principal axis of the lens. (The principal axis of a lens is an imaginary line that is perpendicular to the plane of the lens and passes through the center of the lens.) A converging lens causes all

such rays to pass through a single point on the other side of the lens. That point is the focal point F of the lens. Its distance from the lens is called the focal length f of the lens.

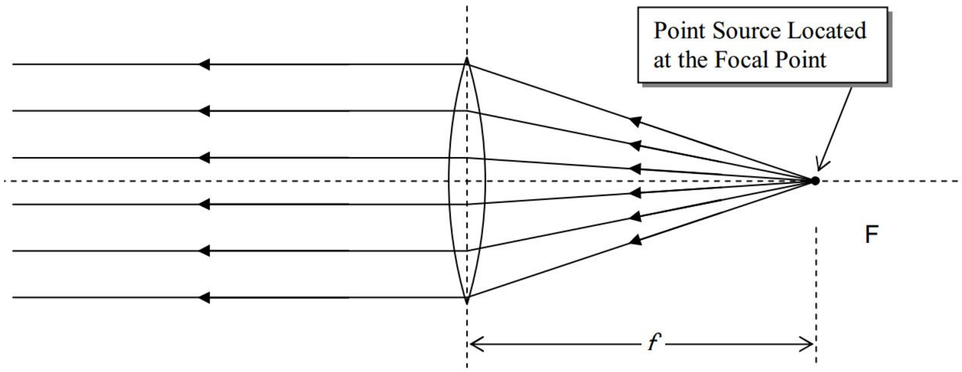


Note that in the diagram, we show the rays of light undergoing an abrupt change in direction at the plane of the lens. This is called the thin lens approximation and we will be using it in all our dealings with lenses. You know that the light is refracted twice in passing through a lens, once at the interface where it enters the lens medium, and again where it exits the lens medium. The two refractions together cause the incoming rays to travel in the directions in which they do travel. The thin lens approximation treats the pair of refractions as a single path change occurring at the plane of the lens. The thin lens approximation is good as long as the thickness of the lens is small compared to the focal length, the object distance, and the image distance.

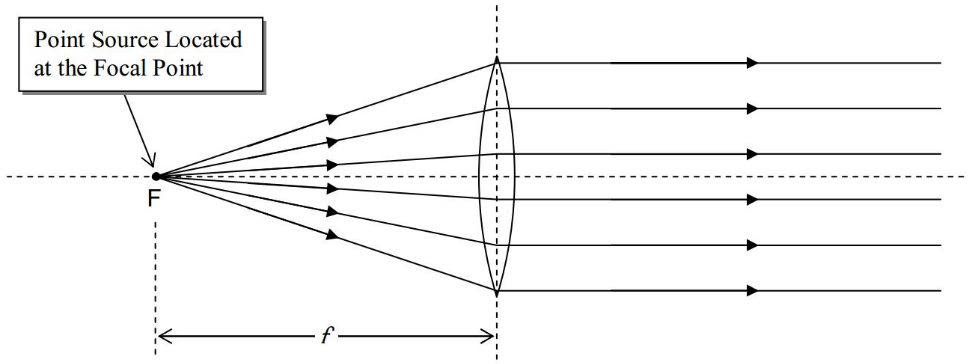
Rays parallel to the principal axis of the lens that enter the lens from the opposite direction (opposite the direction of the rays discussed above) will also be caused to converge to a focal point on the other side of the lens. The two focal points are one and the same distance f from the plane of the lens.



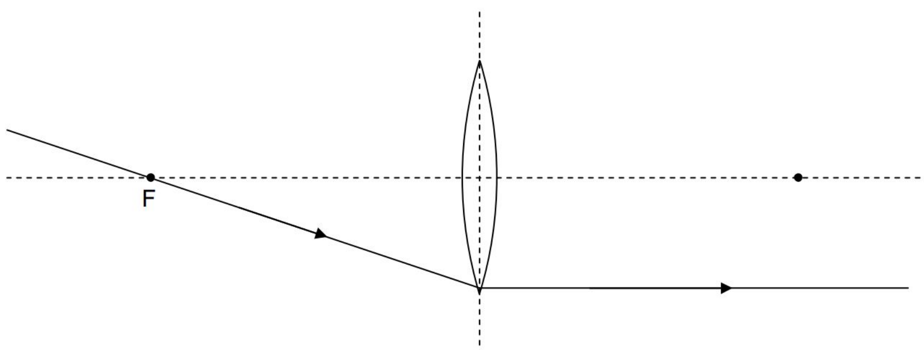
The two phenomena discussed above are reversible in the sense that rays of light coming from a point source, at either focal point, will result in parallel rays on the other side of the lens. Here we show that situation for the case of a point source at one of the focal points:



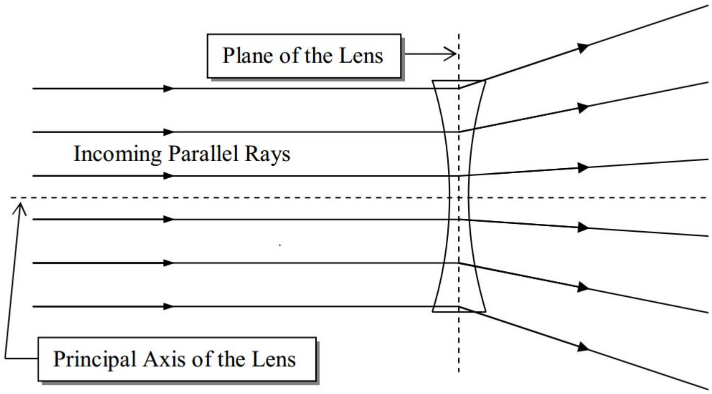
and here we show it for the case of a point source at the other focal point.



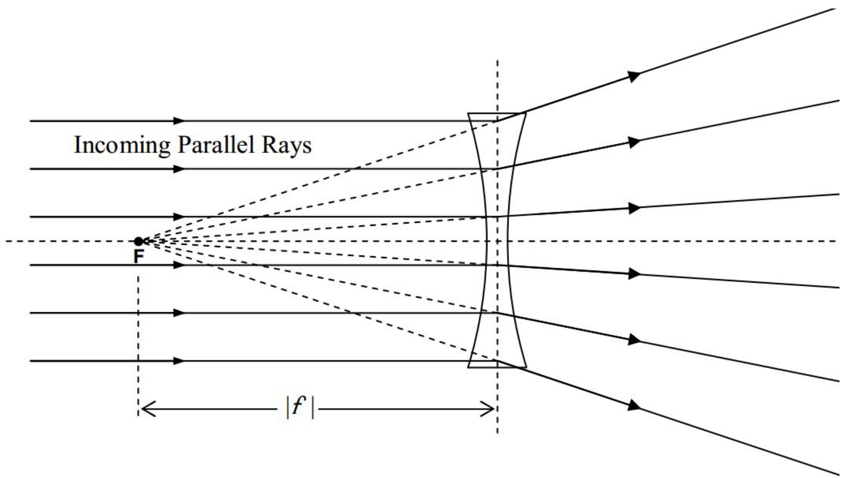
The important thing about this is that, any ray that passes through the focal point on its way to the lens is, after passing through the lens, going to be parallel to the principal axis of the lens.



In the case of a diverging lens, incoming parallel rays are caused to diverge:

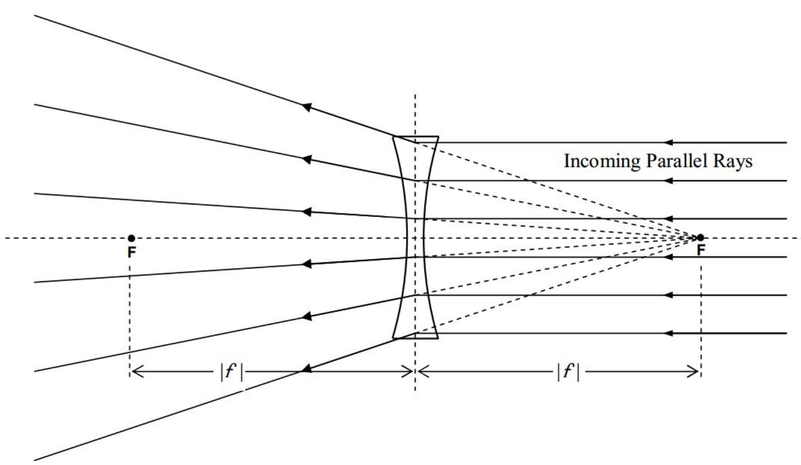


so that they travel along lines which trace-back shows,

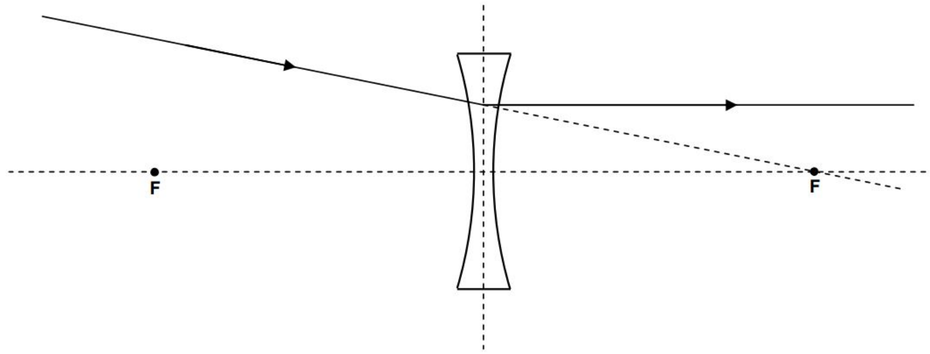


all pass through one and the same point. That is, on passing through the lens, the once-parallel rays diverge as if they originated from a point. That point is known as the focal point of the diverging lens. The distance from the plane of the lens to the focal point is the magnitude of the focal length of the lens. But, by convention, the focal length of a diverging lens is negative. In other words, the focal length of a diverging lens is the negative of the distance from the plane of the lens to the focal point.

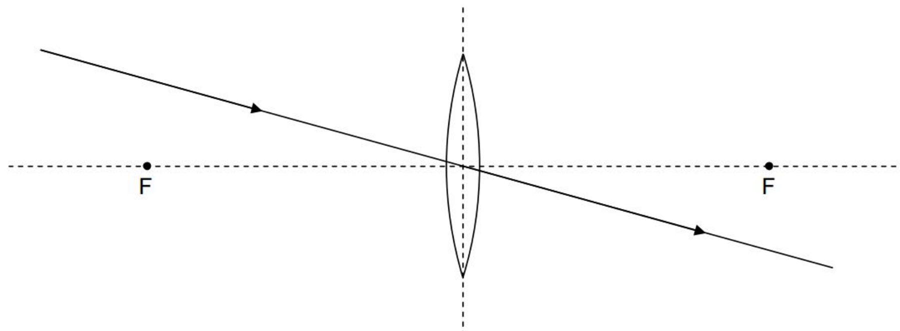
As in the case of the converging lens, there is another focal point on the other side of the lens, the same distance from the plane of the lens as the focal point discussed above:

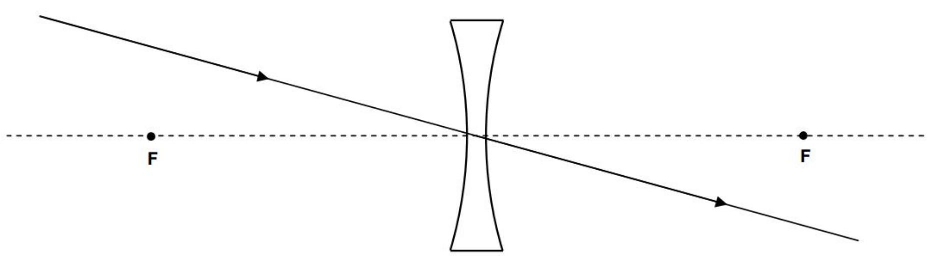


This effect is reversible in that any ray that is traveling through space on one side of the lens, and is headed directly toward the focal point on the other side of the lens, will, upon passing through the lens, become parallel to the principal axis of the lens.



Our plan here is to use the facts about what a lens does to incoming rays of light that are parallel to the principal axis of a lens or are heading directly toward or away from a focal point, to determine where a lens will form an image of an object. Before we do that, I need to tell you one more thing about both kinds of thin spherical lenses. This last fact is a reminder that our whole discussion is an approximation that hinges on the fact that the lenses we are dealing with are indeed thin. Here's the new fact: Any ray that is headed directly toward the center of a lens goes straight through. The justification is that at the center of the lens, the two surfaces of the lens are parallel. So, to the extent that they are parallel in a small region about the center of the lens, it is as if the light is passing through a thin piece of plate glass (or any transparent medium shaped like plate glass.) When light in air, is incident at some angle of incidence other than 0°, on plate glass, after it gets through both air/glass interfaces, the ray is parallel to the incoming ray. The amount by which the outgoing ray is shifted sideways, relative to the incoming ray, depends on how thick the plate is—the thinner the plate, the closer the outgoing ray is to being collinear with the incoming ray. In the thin lens approximation, we treat the outgoing ray as being exactly collinear with the incoming ray.

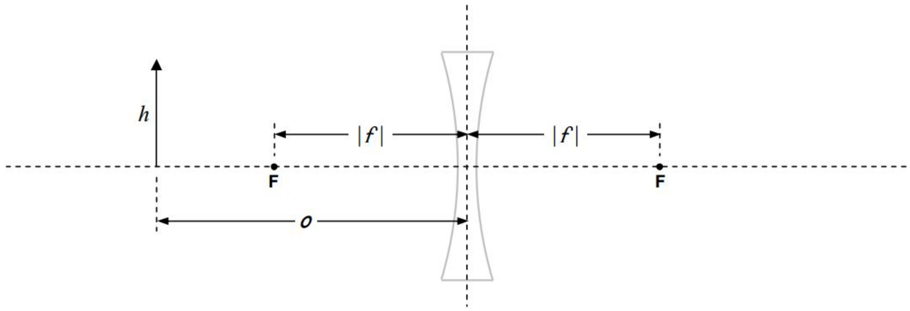




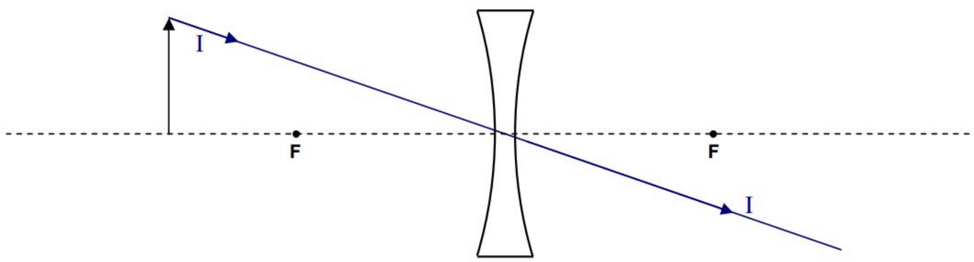
Using Ray Tracing Diagrams

Given an object of height h, the object position o, and the focal length f of the lens with respect to which the object position is given, you need to be able to diagrammatically determine: where the image of that object will be formed by the lens, how big the image is, whether the image is erect (right side up) or inverted (upside down), and whether the image is real or virtual (these terms will be defined soon). Here's how you do that for the case of a diverging lens of specified focal length for which the object distance $o > |f|$:

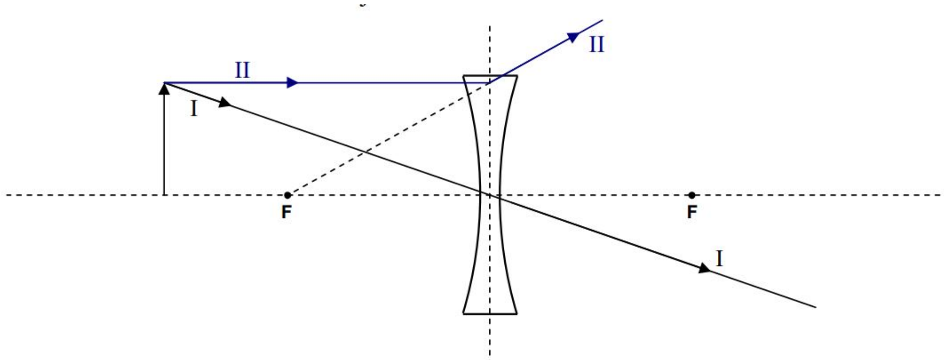
Draw the plane of the lens and the principal axis of the lens. Draw the lens, but think of it as an icon, just telling you what kind of lens you are dealing with. As you proceed with the diagram be careful not to show rays changing direction at the surface of your icon. Also, make sure you draw a diverging lens if the focal length is negative. Measure off the distance |f | to both sides of the plane of the lens and draw the focal points. Measure off the object distance o from the plane of the lens, and, the height h of the object. Draw in the object.



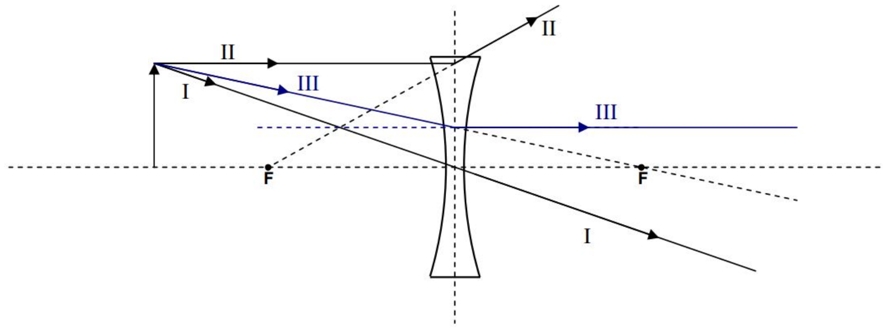
We determine the position of the image of the tip of the arrow by means of three principal rays. The three principal rays are rays on which the effect of the lens is easy to determine based on our understanding of what a lens does to incoming rays that are traveling toward the center of the lens, incoming rays that are traveling toward or away from a focal point, and incoming rays that are traveling directly toward the center of the lens. Let’s start with the easy one, Principal Ray I. It leaves the tip of the arrow and heads directly toward the center of the lens. It goes straight through.



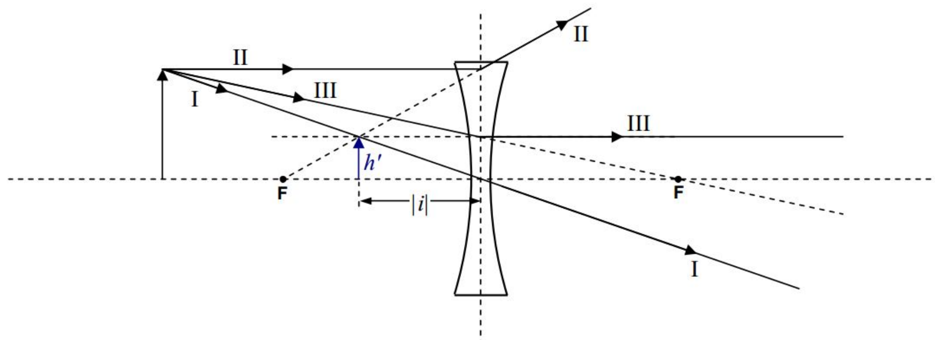
Next comes Principal Ray II. It comes in parallel to the principal axis of the lens, and, at the plane of the lens, jumps on a diverging line, which, if traced back, passes through the focal point on the same side of the lens as the object. Note the need for trace-back.



In the case of a diverging lens, Principal Ray III is the ray that, as it approaches the lens, is headed straight for the focal point on the other side of the lens. At the plane of the lens, Principal Ray III jumps onto a path that is parallel to the principal axis of the lens.



Note that, after passing through the lens, all three rays are diverging from each other. Trace-back yields the apparent point of origin of the rays, the image of the tip of the arrow. It is at the location where the three lines cross. (In practice, using a ruler and pencil, due to human error, the lines will cross at three different points. Consider these to be the vertices of a triangle and draw the tip of the arrow at what you judge to be the geometric center of the triangle.) Having located the image of the tip of the arrow, draw the shaft of the image of the arrow, showing that it extends from point of intersection, to the principal axis of the lens, and, that it is perpendicular to the principal axis of the lens.

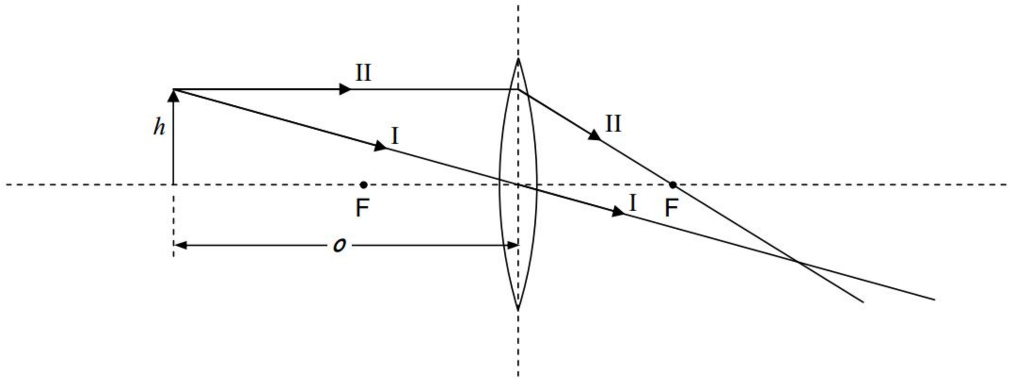


Measurements with a ruler yield the image height h' and the magnitude of the image distance $|i|$. The image is said to be a virtual image. A virtual image of a point, is a point from which rays appear to come, as determined by trace-back, but, through which the rays do not all, actually pass. By convention, the image distance is negative when the image is on the same side of the lens as the object. A negative image distance also signifies a virtual image. Note that the image is erect. By convention, an erect image has a positive image height h'. The magnification M is given by:

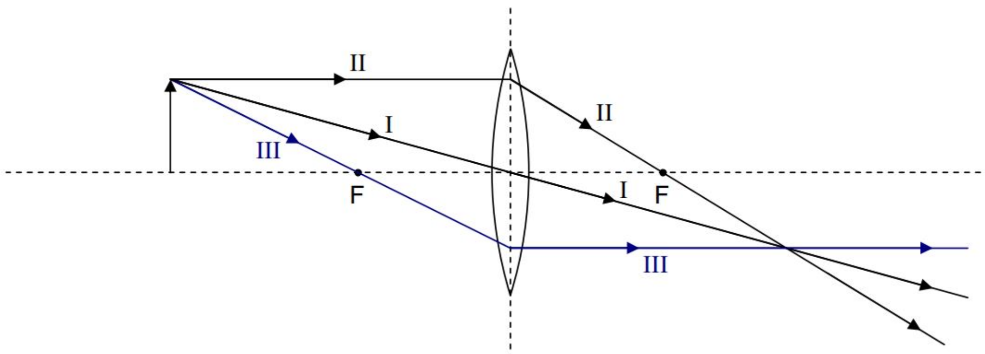
$$M=\frac{h'}{h}$$

By convention, a positive value of M means the image is erect (right side up).

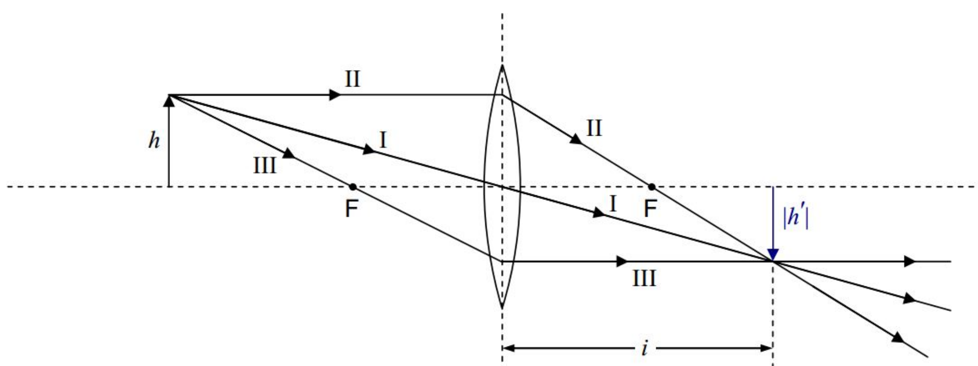
For the case of a converging lens, Principal Ray I is identical to the corresponding ray for the diverging lens. It starts out headed straight for the center of the lens, and, it goes straight through. Principal Ray II starts out the same way Principal Ray II did for the diverging lens—it comes in parallel to the principal axis of the lens—but, starting at the plane of the lens, rather than diverging, it is caused to converge to the extent that it passes through the focal point on the other side of the lens.



Principal Ray III, for a converging lens (with the object farther from the lens than the focal point is), passes through the focal point on the same side of the lens (the side of the lens the object is on) and then, when it gets to the plane of the lens, comes out parallel to the principal axis of the lens.



If you position yourself so that the rays, having passed through the lens, are coming at you, and, you are far enough away from the lens, you will again see the rays diverging from a point. But this time, all the rays actually go through that point. That is, the lens converges the rays to a point, and they don't start diverging again until after they pass through that point. That point is the image of the tip of the arrow. It is a real image. You can tell because if you trace back the lines the rays are traveling along, you come to a point through which all the rays actually travel. Identifying the crossing point as the tip of the arrow, we draw the shaft and head of the arrow.

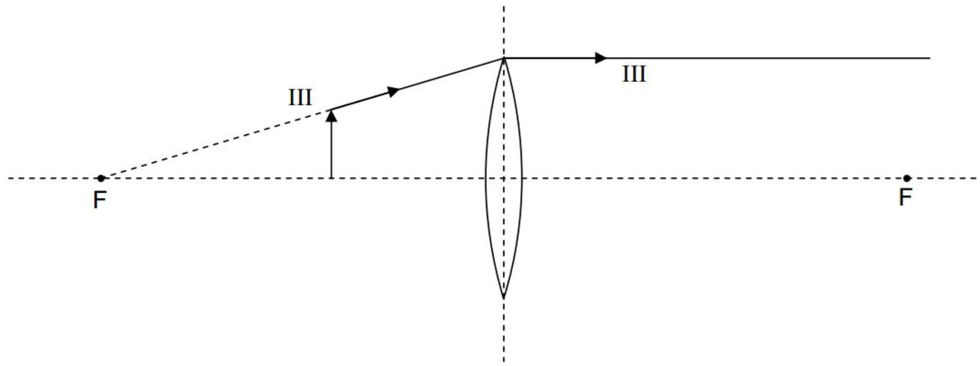


This time, the image is inverted. We can measure the length of the image and the distance of the image from the plane of the mirror. By convention, the image height is negative when the image is inverted, and, the image distance is positive when the image is on the side of the lens opposite that of the object. The magnification M is again given by

$$M = \frac{h'}{h} \tag{B28.2}$$

which, with h' being negative, turns out to be negative itself. This is consistent with the convention that a negative magnification means the image is inverted.

Principal Ray III is different for the converging lens when the object is closer to the plane of the lens than the focal point is:



Principal Ray III, like every principal ray, starts at the tip of the object and travels toward the plane of the lens. In the case at hand, on its way to the plane of lens, Principal Ray III travels along a line that, if traced back, passes through the focal point on the same side of the lens as the object.

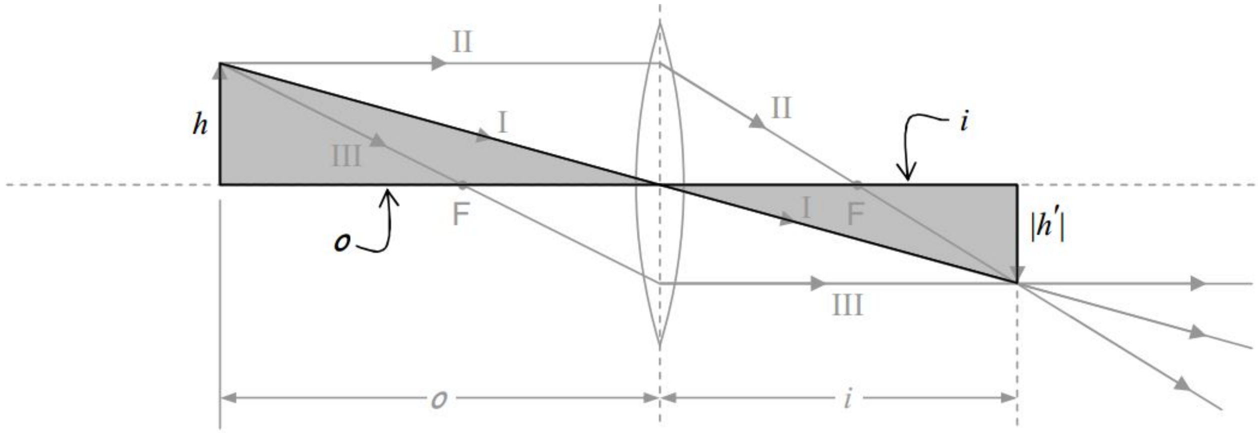
This concludes our discussion of the determination of image features and position by means of ray tracing. In closing this chapter, I summarize the algebraic sign conventions, in the form of a table:

Physical Quantity	Symbol	Sign Convention
focal length	f	+ for converging lens - for diverging lens
image distance	i	+ for real image (on opposite side of lens as object) - for virtual image (on same side of lens as object)
image height	h'	+ for erect image - for inverted image
magnification	M	+ for erect image - for inverted image

B29: Thin Lenses - Lens Equation, Optical Power

From the thin lens ray-tracing methods developed in the last chapter, we can derive algebraic expressions relating quantities such as object distance, focal length, image distance, and magnification.

Consider for instance the case of a converging lens with an object more distant from the plane of the lens than the focal point is. Here's the diagram from the last chapter. In this copy, I have shaded two triangles in order to call your attention to them. Also, I have labeled the sides of those two triangles with their lengths.



By inspection, the two shaded triangles are similar to each other. As such, the ratios of corresponding sides are equal. Thus:

$$\frac{|h'|}{h} = \frac{i}{o}$$

Recall the conventions stated in the last chapter:

Physical Quantity	Symbol	Sign Convention
focal length	f	+ for converging lens - for diverging lens
image distance	i	+ for real image - for virtual image
image height	h'	+ for erect image - for inverted image
magnification	M	+ for erect image - for inverted image

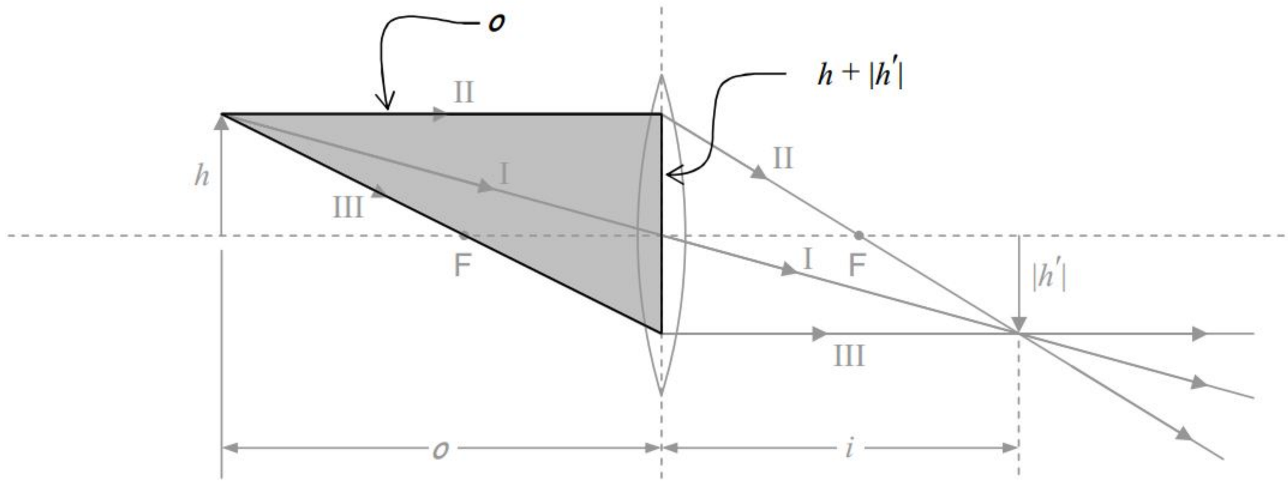
In the case at hand, we have an inverted image, so h' is negative, so $|h'| = -h'$. Thus, the equation $\frac{|h'|}{h} = \frac{i}{o}$ can be written as $\frac{-h'}{h} = \frac{i}{o}$, or, as

$$\frac{h'}{h} = -\frac{i}{o}$$

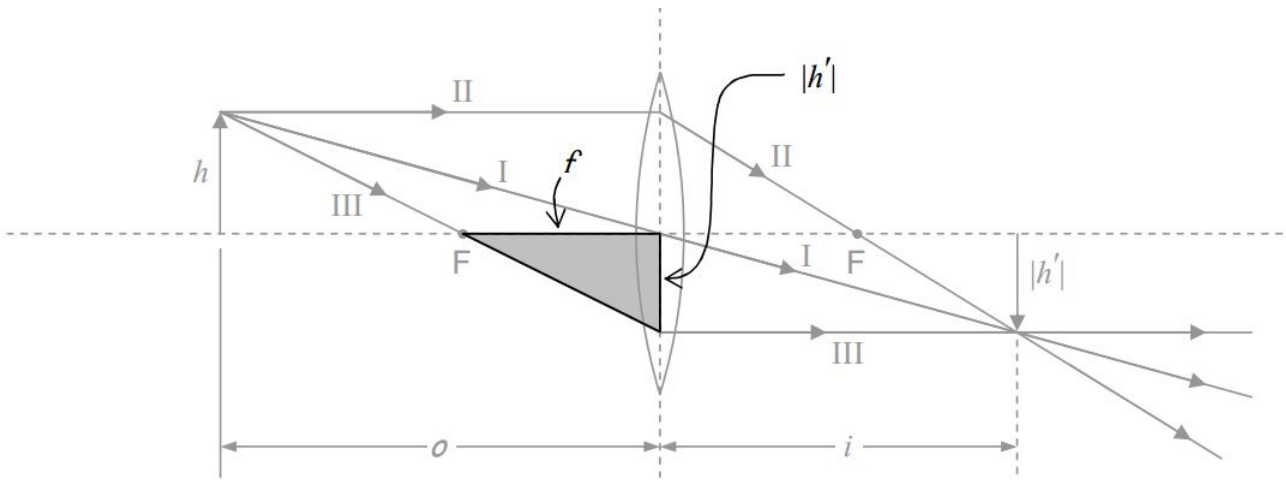
But $\frac{h'}{h}$ is, by definition, the magnification. Thus, we can write the magnification as:

$$M = -\frac{i}{o} \tag{B29.1}$$

Here's another copy of the same diagram with another triangle shaded.



By inspection, that shaded triangle is similar to the triangle that is shaded in the following copy of the same diagram:



Using the fact that the ratios of corresponding sides of similar triangles are equal, we set the ratio of the two top sides (one from each triangle) equal to the ratio of the two right sides:

$$\frac{o}{f} = \frac{h + |h'|}{|h'|}$$

Again, since the image is upside down, h' is negative so $|h'| = -h'$. Thus

$$\frac{o}{f} = \frac{h - h'}{-h'}$$

$$\frac{o}{f} = 1 - \frac{h}{h'}$$

From our first pair of similar triangles we found that $\frac{h'}{h} = -\frac{i}{o}$ which can be written $\frac{h}{h'} = -\frac{o}{i}$ Substituting this into the expression $\frac{o}{f} = 1 - \frac{h}{h'}$ which we just found, we have

$$\frac{o}{f} = 1 - \left(-\frac{o}{i}\right)$$

Dividing both sides by o and simplifying yields:

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i} \tag{B29.2}$$

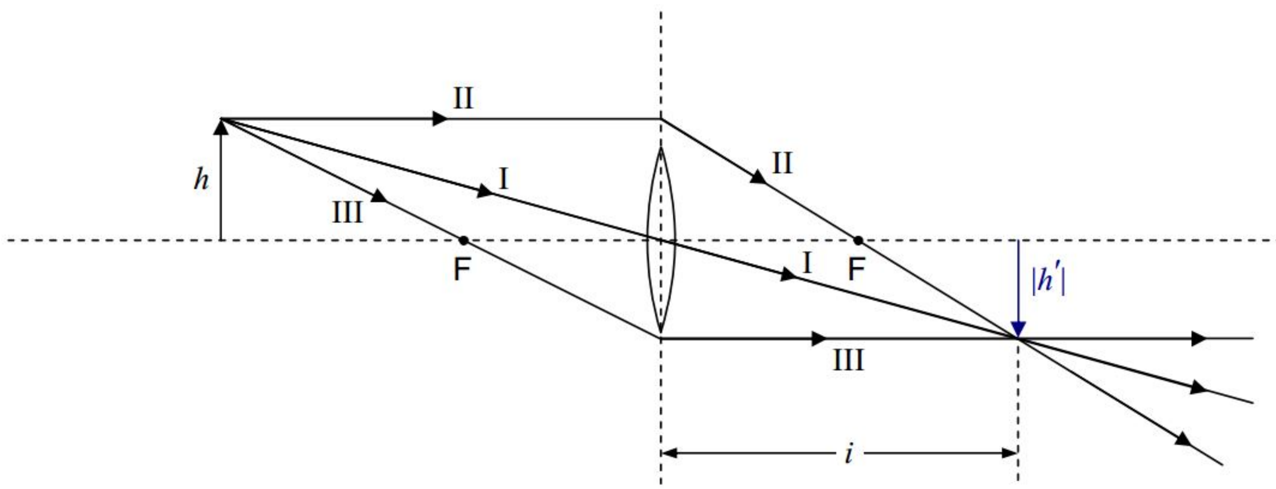
This equation is referred to as the lens equation. Together with our definition of the magnification $M = \frac{h'}{h}$, the expression we derived for the magnification $M = -\frac{i}{o}$, and our conventions:

Physical Quantity	Symbol	Sign Convention
focal length	f	+ for converging lens - for diverging lens
image distance	i	+ for real image - for virtual image
image height	h'	+ for erect image - for inverted image
magnification	M	+ for erect image - for inverted image

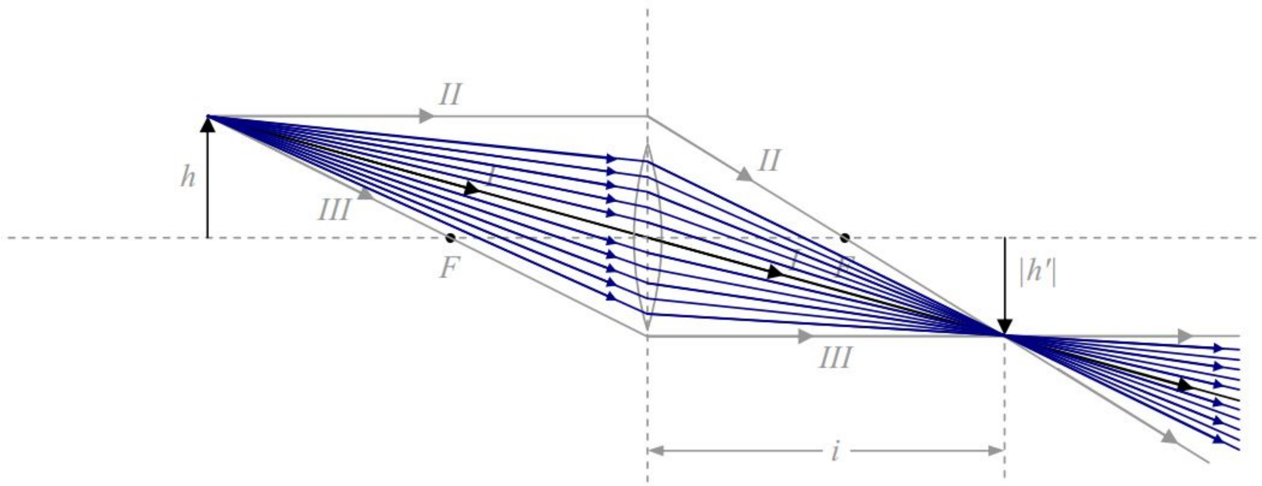
The lens equation tells us everything we need to know about the image of an object that is a known distance from the plane of a thin lens of known focal length. While we have derived it for the case of an object that is a distance greater than the focal length, from a converging lens, it works for all the combinations of lens and object distance for which the thin lens approximation is good. (The thin lens approximation is good as long as i, o, and f are all large compared to the thickness of the lens.) In each case, we derive the lens equation (it always turns out to be the same equation), by drawing the ray tracing diagram and analyzing the similar triangles that appear in it.

An Important Conceptual Point

(We mentioned this in the last chapter but it warrants further attention.) An infinite set of rays contributes to any given point of an image formed by a lens. Consider for instance the case of an object at a greater distance than the focal length from a thin spherical convex (converging) lens. Further, because it's easy to specify, we will consider the image of the tip of the (arrow) object. We have been using the principal rays to locate the image, as in the following diagram:



in which I have intentionally used a small lens icon to remind you that, in using the principal ray diagram to locate the image, we don't really care whether or not the principal rays actually hit the lens. Let's, for the case at hand, consider the diagram to be a life-size diagram of an actual lens. As, important as they are in helping us identify the location of the image, clearly, for the case at hand, Principal Rays II and III do not actually contribute to the image. Principal Ray I does contribute to the image. Let's draw in some more of the contributors:



The fact that every ray that comes from the tip of the object and hits the lens contributes to the image of the tip of the arrow (and the corresponding fact for each and every point on the object) explains why you can cover up a fraction of the lens (such as half the lens) and still get a complete image (albeit dimmer).

The Power of a Lens

When an ophthalmologist writes a prescription for a spherical lens, she or he will typically write either a value around $-.5$ or $.5$, or, a value around -500 or 500 without units. You might well wonder what quantity the given number is a value for, and what the units should be. The answer to the first question is that the physical quantity is the power of the lens being prescribed. In this context, the power is sometimes called the optical power of the lens. The power of a lens has nothing to do with the rate at which energy is being transformed or transferred but instead represents the assignment of a completely different meaning to the same word. In fact, the power of a lens is, by definition, the reciprocal of the focal length of the lens:

$$P=\frac{1}{f} \tag{B29.3}$$

In that the SI unit of focal length is the meter (m), the unit of optical power is clearly the reciprocal meter which you can write as $\frac{1}{m}$ or m^{-1} in accord with your personal preferences. This unit has been assigned a name. It is called the diopter, abbreviated D. Thus, by definition,

$$1D=\frac{1}{m}$$

Thus, a value of $-.5$ on the ophthalmologist's prescription can be interpreted to mean that what is being prescribed is a lens having a power of -0.5 diopters. The minus sign means that the lens is a concave (diverging) lens. Taking the reciprocal yields:

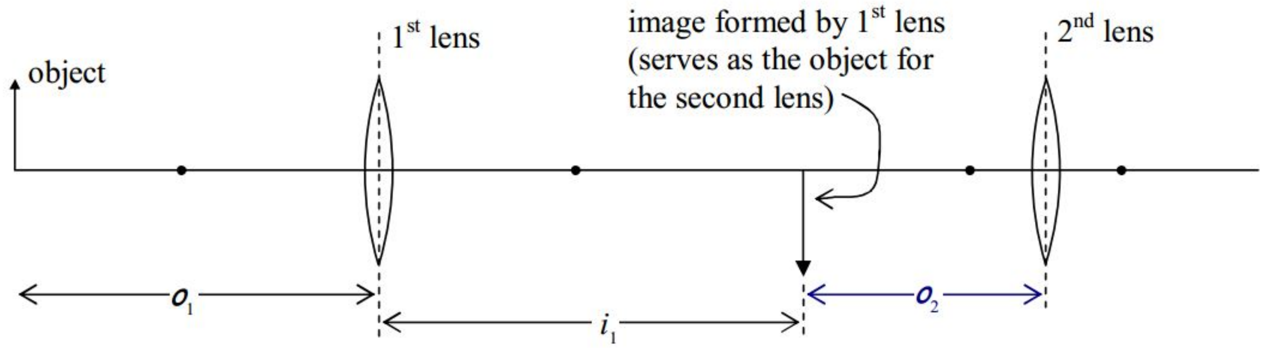
$$f=\frac{1}{P}=\frac{1}{-0.5D}=-2\frac{1}{D}=-2m$$

If you see a number around -500 or 500 on the ophthalmologist's lens prescription, you can assume that the ophthalmologist is giving the power of the lens in units of millidiopters (mD). 500 mD is, of course, equivalent to $.5D$. To avoid confusion, if you are given an optical power in units of mD, convert it to units of diopters before using it to calculate the corresponding focal length.

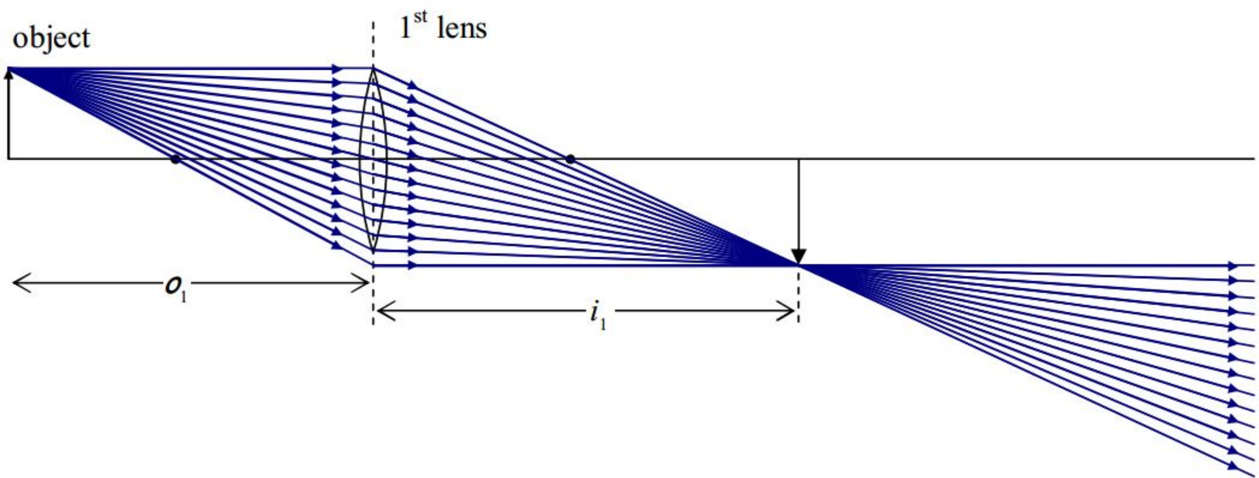
Two-Lens Systems

To calculate the image of a two-lens system, one simply calculates the position of the image for the lens that light from the object hits first, and then uses that image as the object for the second lens. In general, one has to be careful to recognize that for the first lens, the object distance and the image distance are both measured relative to the plane of the first lens. Then, for the second lens, the object distance and the image distance are measured relative to the plane of the second lens. That means that, in general, the object distance for the second lens is not equal in value to the image distance for the first lens.

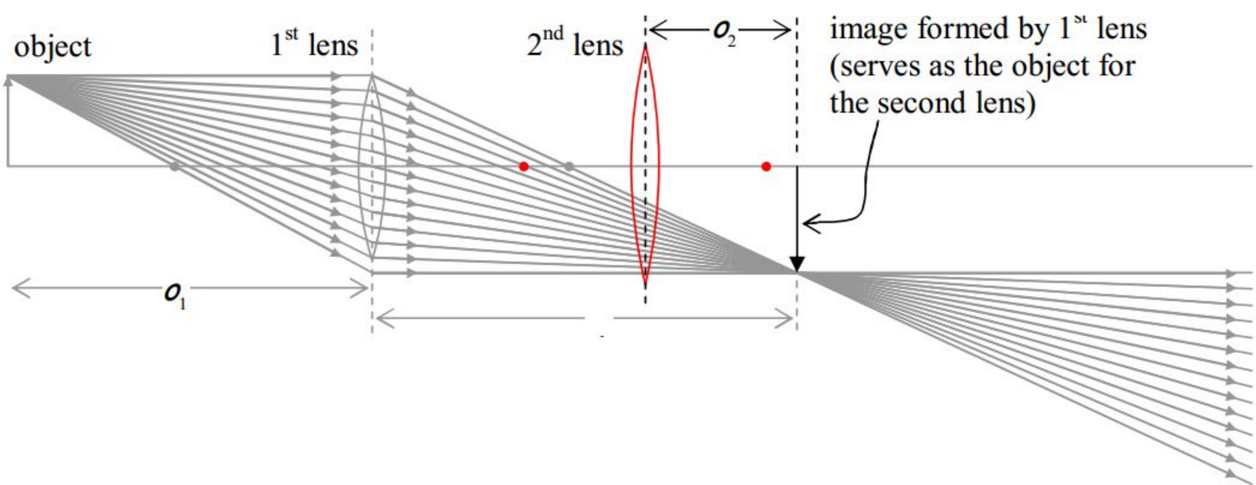
For instance, in the following diagram of two lenses separated by 12cm, if the object is to the left of the first lens, and i_1 turns out to be 8 cm to the right of the first lens,



then o_2, the object distance for the second lens, is 4 cm. A peculiar circumstance arises when the second lens is closer to the first lens than the image formed by the first lens is. Suppose for instance, that we have the image depicted above, formed by the first lens:



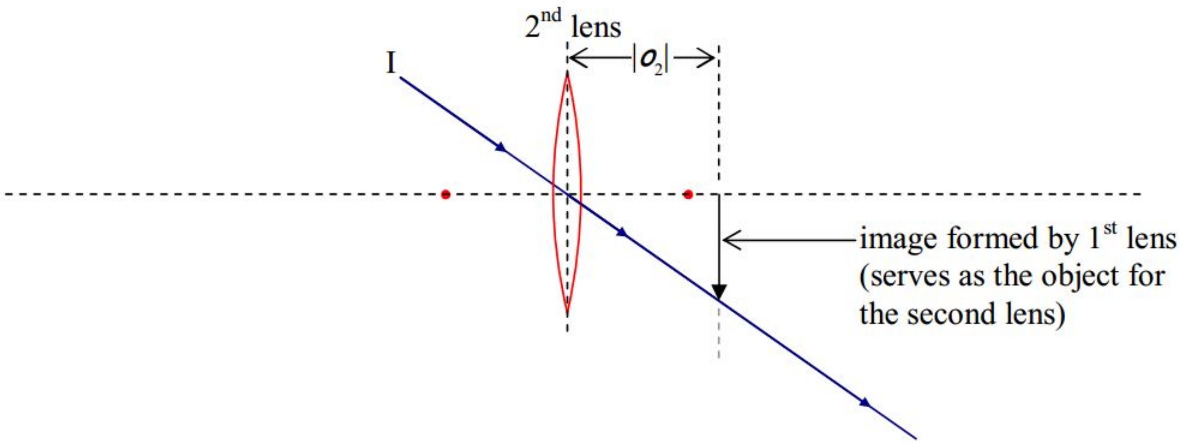
Now suppose that we put a second lens in between the 1st lens and the image.



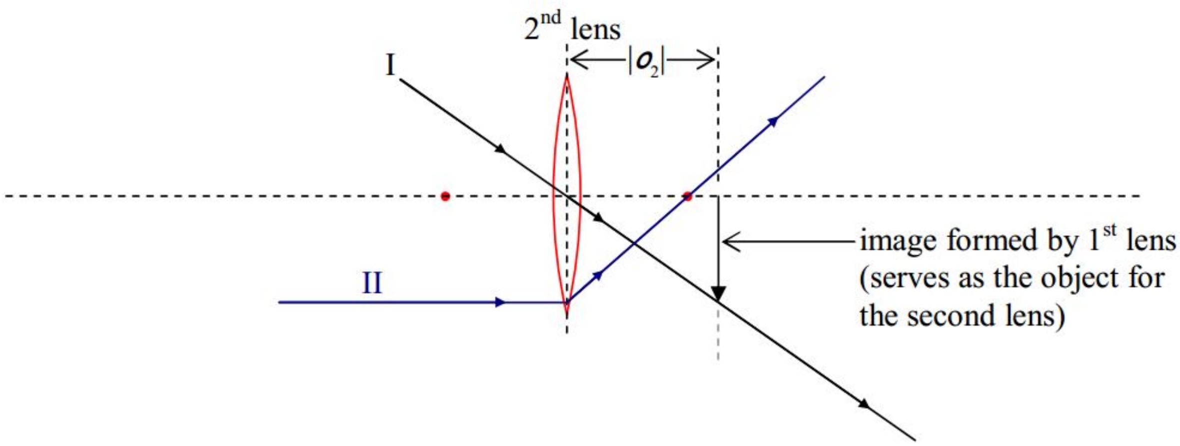
Note that, for the second lens, we have an object to the right of the lens, but, the light associated with that object approaches the object from the left! This can only happen when the object is actually an image formed by another lens. In such a case, we call the object a virtual object. More generally, when an object's light approaches a lens from the side opposite that side to which the object is, the object is considered to be a virtual object, and, the object distance, is, by convention, negative. So, we have one more convention to put in a table for you:

Physical Quantity	Symbol	**Sign Convention**
Object Distance	o	+ for real object (always the case for a physical object) - for virtual object (only possible if "object" is actually the image formed by another lens)

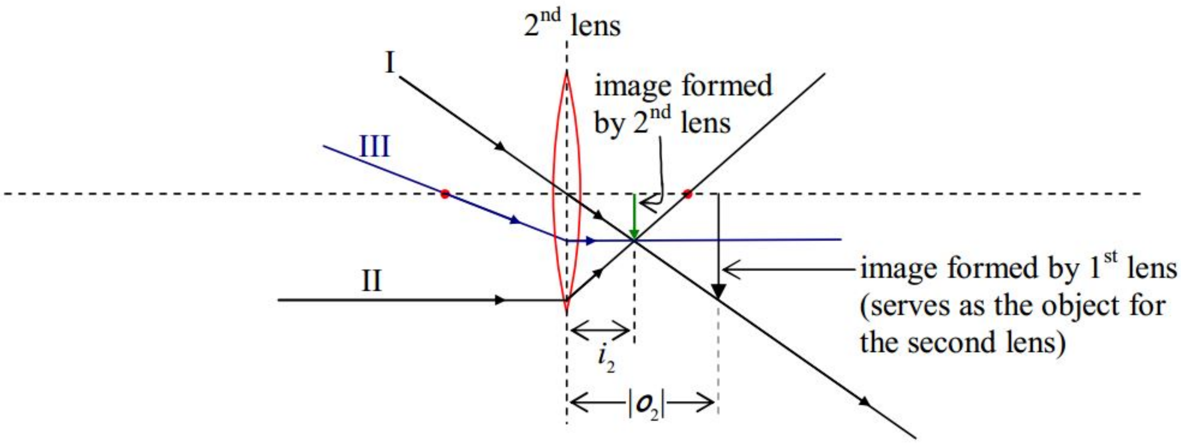
In forming the ray-tracing diagram for the case of the virtual object, we have to remember that every ray coming into the second lens is headed straight for the tip of the arrow that is the virtual object for the second lens. Thus, our Principal Ray I is one that is headed straight toward the tip of the arrow, and, is headed straight toward the center of the lens. It goes straight through.



Principal Ray II is headed straight for the head of the object along a line that is parallel to the principal axis of the lens. At the plane of the lens it jumps onto the straight line path that takes it straight through the focal point on the other side of the lens.

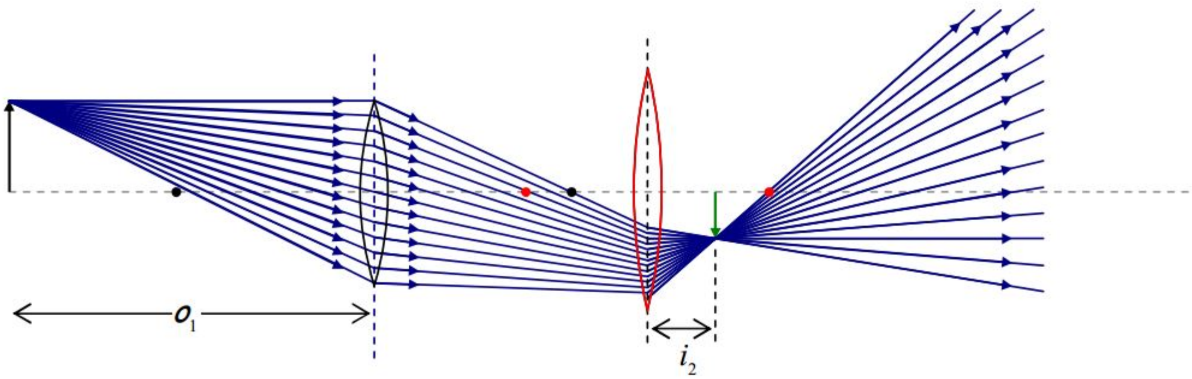


Principal Ray III, is headed straight toward the tip of the virtual object, and, on its way to the lens, it passes through the focal point on the side of the lens from which it approaches the lens. When it hits the plane of the lens, Principal Ray III adopts a path that is parallel to the principal axis of the lens.



Note that, for the case at hand, we get a real image. Relative to the virtual object, the image is not inverted. The virtual object was already upside down. The fact that we can draw a raytracing diagram for the case of a virtual object means that we can identify and analyze similar triangles to establish the relationship between the object distance, the image distance and the focal length of the lens. Doing so, with the convention that the object distance of a virtual object is negative, again yields the lens equation $\frac{1}{f}=\frac{1}{o}+\frac{1}{i}$.

Here's a diagram of the entire two-lens system for the case at hand:



Note that the real image of lens 1 alone is never actually formed, but it was crucial in our determination of the image location, orientation, and size, in the case of the two-lens system.

B30: The Electric Field Due to a Continuous Distribution of Charge on a Line

> *Every integral must include a differential (such as dx, dt, dq, etc.). An integral is an infinite sum of terms. The differential is necessary to make each term infinitesimal (vanishingly small). $\int f(x)dx$ is okay, $\int g(y)dy$ is okay, and $\int h(t)dt$ is okay, but never write $\int f(x)$, never write $\int g(y)$ and never write $\int h(t)$.*

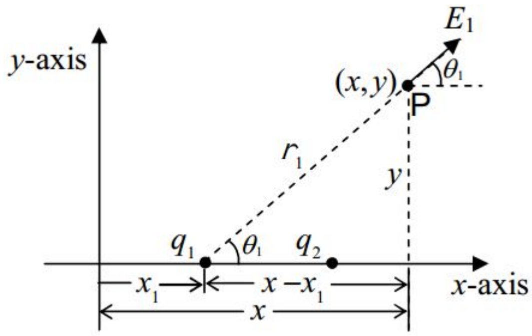
Here we revisit Coulomb's Law for the Electric Field. Recall that Coulomb's Law for the Electric Field gives an expression for the electric field, at an empty point in space, due to a charged particle. You have had practice at finding the electric field at an empty point in space due to a single charged particle and due to several charged particles. In the latter case, you simply calculated the contribution to the electric field at the one empty point in space due to each charged particle, and then added the individual contributions. You were careful to keep in mind that each contribution to the electric field at the empty point in space was an electric field vector, a vector rather than a scalar, hence the individual contributions had to be added like vectors.

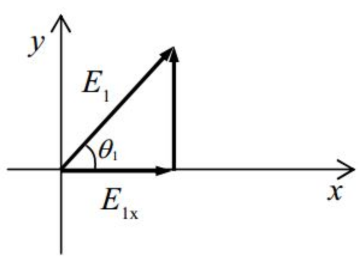
A Review Problem for the Electric Field due to a Discrete Distribution of Charge

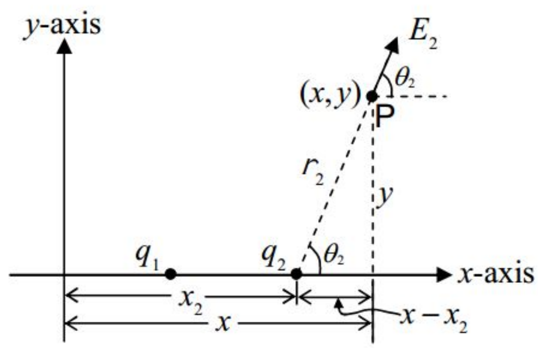
Let's kick this chapter off by doing a review problem. The following example is one of the sort that you learned how to do when you first encountered Coulomb's Law for the Electric Field. You are given a discrete distribution of source charges and asked to find the electric field (in the case at hand, just the x component of the electric field) at an empty point in space.

The example is presented on the next page. Here, a word about one piece of notation used in the solution. The symbol P is used to identify a point in space so that the writer can refer to that point, unambiguously, as "point P." The symbol P in this context does not stand for a variable or a constant. It is just an identification tag. It has no value. It cannot be assigned a value. It does not represent a distance. It just labels a point.

There are two charged particles on the x-axis of a Cartesian coordinate system, q_1 at $x = x_1$ and q_2 at $x = x_2$ where $x_2 > x_1$. Find the x component of the electric field, due to this pair of particles, valid for all points on the x-y plane for which $x > x_2$.







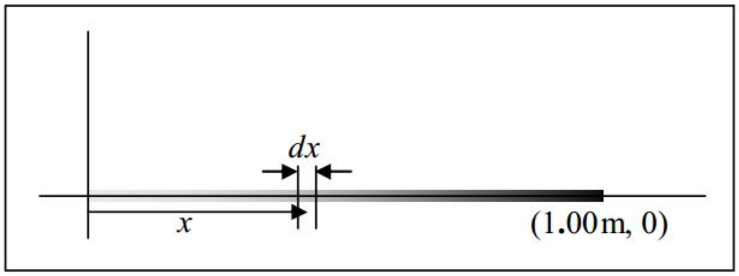
Linear Charge Density

Okay, enough review, now lets consider the case in which we have a continuous distribution of charge along some line segment. In practice, we could be talking about a charged piece of string or thread, a charged thin rod, or even a charged piece of wire. First we need to discuss how one even specifies such a situation. We do so by stating what the linear charge density, the charge per-length, λ is. For now we'll consider the meaning of λ for a few different situations (before we get to the heart of the matter, finding the electric field due to the linear charge distribution). Suppose for instance we have a one-meter string extending from the origin to $x = 1.00m$ along the x axis, and that the linear charge density on that string is given by:

$$\lambda = 1.56 \frac{\mu}{m^2} x.$$

(Just under the equation, we have depicted the linear charge density graphically by drawing a line whose darkness represents the charge density.)

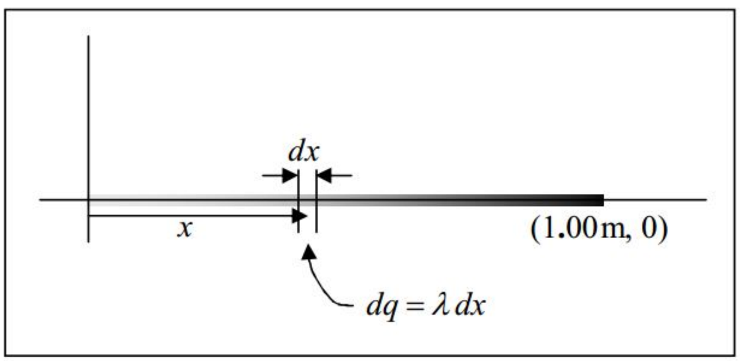
Note that if the value of x is expressed in meters, λ will have units of $\frac{\mu C}{m}$, units of charge-per-length, as it must. Further note that for small values of x, λ is small, and for larger values of x, λ is larger. That means that the charge is more densely packed near the far (relative to the origin) end of the string. To further familiarize ourselves with what λ is, let's calculate the total amount of charge on the string segment. What we'll do is to get an expression for the amount of charge on any infinitesimal length dx of the string, and add up all such amounts of charge for all of the infinitesimal lengths making up the string segment.



The infinitesimal amount of charge dq on the infinitesimal length dx of the string is just the charge per length λ times the length dx of the infinitesimal string segment.

$$dq = \lambda dx$$

Note that you can't take the amount of charge on a finite length (such as $15cm$) of the string to be λ times the length of the segment because λ varies over the length of the segment. In the case of an infinitesimal segment, every part of it is within an infinitesimal distance of the position specified by one and the same value of x. The linear charge density doesn't vary on an infinitesimal segment because x doesn't—the segment is simply too short.



To get the total charge we just have to add up all the dq's. Each dq is specified by its corresponding value of x. To cover all the dq's we have to take into account all the values of x from 0 to 1.00 m. Because each dq is the charge on an infinitesimal length of the line of charge, the sum is going to have an infinite number of terms. An infinite sum of infinitesimal pieces is an integral. When we integrate

$$dq = \lambda dx$$

we get, on the left, the sum of all the infinitesimal pieces of charge making up the whole. By definition, the sum of all the infinitesimal amounts of charge is just the total charge Q (which by the way, is what we are solving for); we don't need the tools of integral calculus to deal with the left side of the equation. Integrating both sides of the equation yields:

$$Q = \int_0^{1.00m} \lambda dx$$

Using the given expression $\lambda = 1.56\frac{\mu}{m^2}x$ we obtain

$$Q = \int_0^{1.00m} 2.56\frac{\mu C}{m^2}x dx = 2.56\frac{\mu C}{m^2}\int_0^{1.00m} x dx = 2.56\frac{\mu C}{m^2}\frac{x^2}{2}\Big|_0^{1.00m} = 2.56\frac{\mu C}{m^2}\left[\frac{(1.00m)^2}{2} - \frac{(0)^2}{2}\right] = 1.28\mu C$$

A few more examples of distributions of charge follow:

For instance, consider charge distributed along the x axis, from $x=0$ to $x=L$ for the case in which the charge density is given by

$$\lambda=\lambda_{MAX}\sin(\pi \text{rad} x/L$$

where λ_{MAX} is a constant having units of charge-per-length, rad stands for the units radians, x is the position variable, and L is the length of the charge distribution. Such a charge distribution has a maximum charge density equal to λ_{MAX} occurring in the middle of the line segment.

Another example would be a case in which charge is distributed on a line segment of length L extending along the y axis from $y=a$ to $y=a+L$ with a being a constant and the charge density given by

$$\lambda=\frac{38\mu C\cdot m}{y^2}$$

In this case the charge on the line is more densely packed in the region closer to the origin. (The smaller y is, the bigger the value of λ, the charge-per-length.)

The simplest case is the one in which the charge is spread out uniformly over the line on which there is charge. In the case of a uniform linear charge distribution, the charge density is the same everywhere on the line of charge. In such a case, the linear charge density λ is simply a constant. Furthermore, in such a simple case, and only in such a simple case, the charge density λ is just the total amount of charge Q divided by the length L of the line along which that charge is uniformly distributed. For instance, suppose you are told that an amount of charge $Q=2.45C$ is uniformly distributed along a thin rod of length $L=0.840m$. Then λ is given by:

$$\lambda=\frac{Q}{L}$$

$$\lambda=\frac{2.45C}{0.840m}$$

$$\lambda=2.92\frac{C}{m}$$

The Electric Field Due to a Continuous Distribution of Charge along a Line

Okay, now we are ready to get down to the nitty-gritty. We are given a continuous distribution of charge along a straight line segment and asked to find the electric field at an empty point in space in the vicinity of the charge distribution. We will consider the case in which both the charge distribution and the empty point in space lie in the x-y plane. The values of the coordinates of the empty point in space are not necessarily specified. We can call them x and y. In solving the problem for a single point in space with unspecified coordinates (x,y), our final answer will have the symbols x and y in it, and our result will actually give the answer for an infinite set of points on the x-y plane.

The plan for solving such a problem is to find the electric field, due to an infinitesimal segment of the charge, at the one empty point in space. We do that for every infinitesimal segment of the charge, and then add up the results to get the total electric field.

Now once we chop up the charge distribution (in our mind, for calculational purposes) into infinitesimal (vanishingly small) pieces, we are going to wind up with an infinite number of pieces and hence an infinite sum when we go to add up the contributions to the electric field at the one single empty point in space due to all the infinitesimal segments of the linear charge distribution. That is to say, the result is going to be an integral.

An important consideration that we must address is the fact that the electric field, due to each element of charge, at the one empty point in space, is a vector. Hence, what we are talking about is an infinite sum of infinitesimal vectors. In general, the vectors being added are all in different directions from each other. (Can you think of a case so special that the infinite set of infinitesimal electric field vectors are all in the same direction as each other? Note that we are considering the general case, not such a special case.) We know better than to simply add the magnitudes of the vectors, infinite sum or not. Vectors that are not all in the same direction as each other, add like vectors, not like numbers. The thing is, however, the x components of all the infinitesimal electric field vectors at the one empty point in space do add like numbers. Likewise for the y components. Thus, if, for each infinitesimal element of the charge distribution, we find, not just the electric field at the empty point in space, but the x component of that electric field, then we can add up all the x components of the electric field at the empty point in space to get the x component of the electric field, due to the entire charge distribution, at the one empty point in space. The sum is still an infinite sum, but this time it is an infinite sum of scalars rather than vectors, and we have the tools for handling that. Of course, if we are asked for the total electric field, we have to repeat the entire procedure to get the y component of the electric field and then combine the two components of the electric field to get the total.

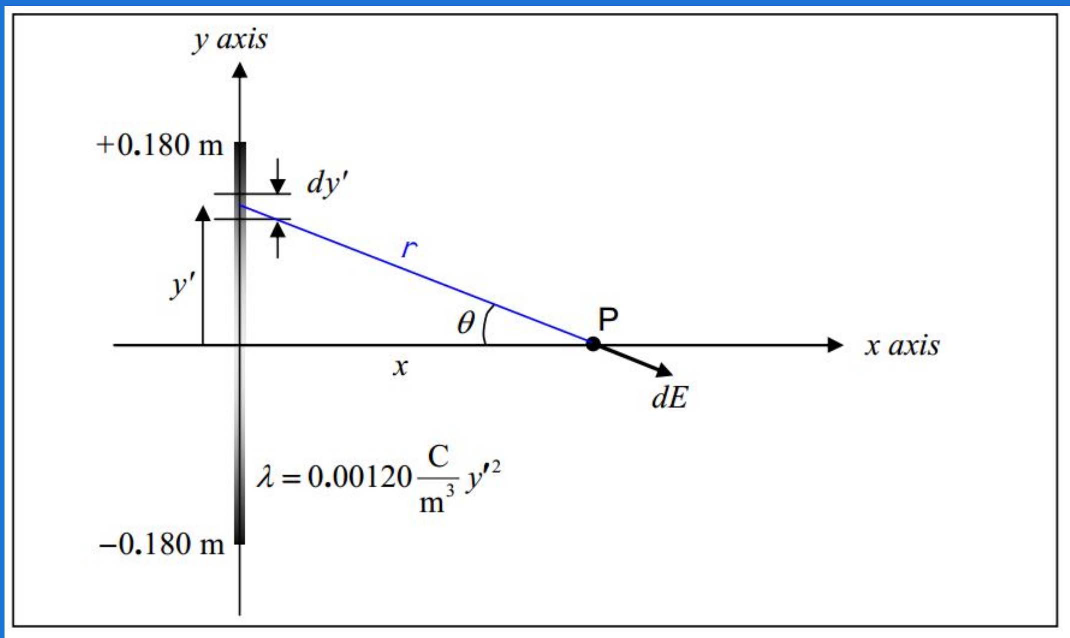
The easy way to do the last step is to use $\hat{\imath}, \hat{\jmath}, \hat{k}$ notation. That is, once we have E_x and E_y, we can simply write:

$$\vec{E} = E_x \hat{\imath} + E_y \hat{\jmath}$$

Find the electric field valid for any point on the positive x axis due a $36.0cm$ long line of charge, lying on the y axis and centered on the origin, for which the charge density is given by

$$\lambda = 0.00120\frac{C}{m^2}y^2$$

As usual, we'll start our solution with a diagram:



Note that we use (and strongly recommend that you use) primed quantities (x', y') to specify a point on the charge distribution and unprimed quantities (x, y) to specify the empty point in space at which we wish to know the electric field. Thus, in the diagram, the infinitesimal segment of the charge distribution is at $(0, y')$ and point P, the point at which we are finding the electric field, is at $(x, 0)$. Also, our expression for the given linear charge density $\lambda = 0.00120\frac{C}{m^2}y^2$ expressed in terms of y' rather than y is:

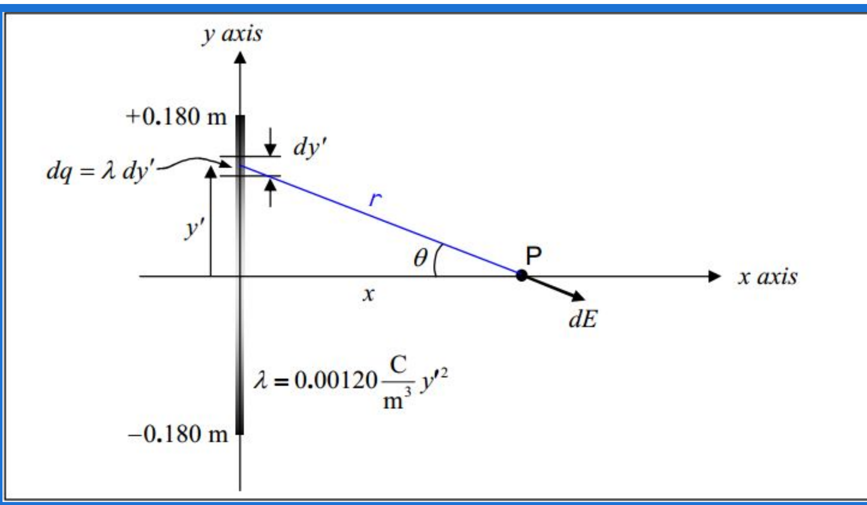
$$\lambda = 0.00120\frac{C}{m^2}y'^2$$

The plan here is to use Coulomb's Law for the Electric Field to get the magnitude of the infinitesimal electric field vector $\overrightarrow{dE}$ at point P due to the infinitesimal amount of charge dq in the infinitesimal segment of length dy'.

$$dE = \frac{kdq}{r^2}$$

The amount of charge dq in the infinitesimal segment dy' of the linear charge distribution is given by

$$dq = \lambda dy'$$



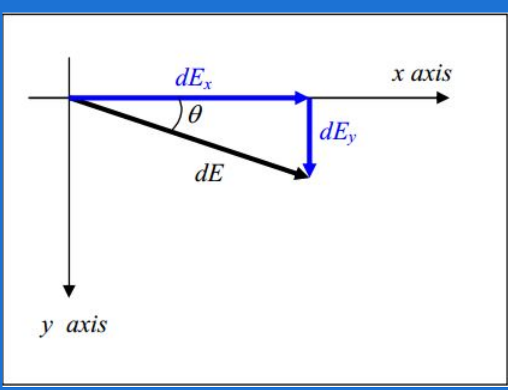
From the diagram, it clear that we can use the Pythagorean theorem to express the distance r that point P is from the infinitesimal amount of charge dq under consideration as:

$r=\sqrt{x^2+y'^2}$ Substituting this and $dq=\lambda dy'$ into our equation for dE ($dE=\frac{kdq}{r^2}$) we obtain

$$dE=\frac{k\lambda dy'}{x^2+y'^2}$$

Recall that our plan is to find E_x, then E_y and then put them together using $\vec{E}=E_x\hat{\imath}+E_y\hat{\jmath}$. So for now, let's get an expression for E_x.

Based on the vector component diagram at right we have $dE_x=dE\cos\theta$



The θ appearing in the diagram at right is the same θ that appears in the diagram above. Based on the plane geometry evident in that diagram (above), we have:

$$\cos\theta=\frac{x}{r}=\frac{x}{\sqrt{x^2+y'^2}}$$

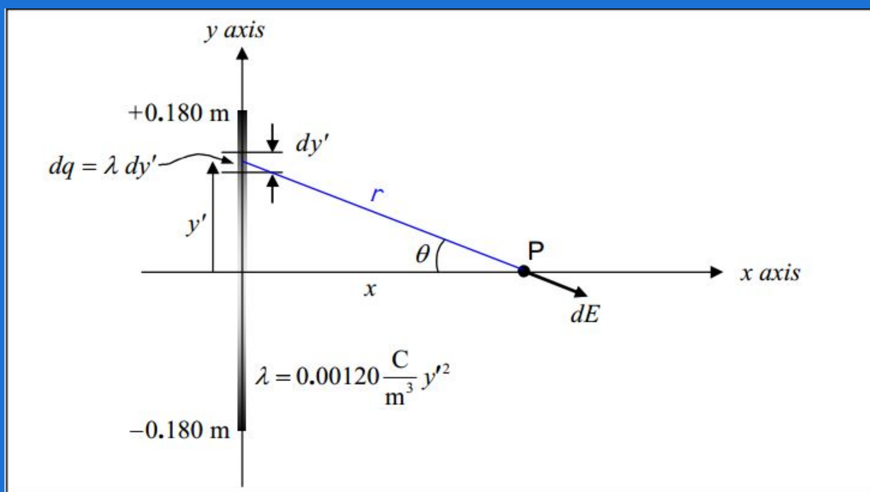
Substituting both this expression for $\cos\theta$($\cos\theta=\frac{x}{\sqrt{x^2+y'^2}}$) and the expression we derived for dE above ($dE=\frac{k\lambda dy'}{(x^2+y'^2)^2}$) into the expression $dE_x=dE\cos\theta$ from the vector component diagram yields:

$$(dE = \frac{k\lambda dy'}{(x^2+y'^2)^{\frac{3}{2}}})$$

Also, let's go ahead and replace λ with the given expression $\lambda = 0.00120\frac{C}{m^3}y'^2$:

$$dE_x = \left(0.00120\frac{C}{m^3}\right)\frac{ky'^2xdy'}{(x^2+y'^2)^{\frac{3}{2}}}$$

Now we have an expression for dE_x that includes only one quantity, namely y', that depends on which bit of the charge distribution is under consideration. Furthermore, although in the diagram



it appears that we picked out a particular infinitesimal line segment dy', in fact, the value of y' needed to establish its position is not specified. That is, we have an equation for dE_x that is good for any infinitesimal segment dy' of the given linear charge distribution. To identify a particular dy' we just have to specify the value of y'. Thus to sum up all the dE_x's we just have to add, to a running total, the dE_x for each of the possible values of y'. Thus we need to integrate the expression for dE_x for all the values of y' from $-0.180m$ to $+0.180m$.

$$\int dE_x = \int_{-0.180m}^{+0.180m}\left(0.00120\frac{C}{m^3}\right)\frac{ky'^2xdy'}{(x^2+y'^2)^{\frac{3}{2}}}$$

Copying that equation here:

$$\int dE_x = \int_{-0.180m}^{+0.180m}\left(0.00120\frac{C}{m^3}\right)\frac{ky'^2xdy'}{(x^2+y'^2)^{\frac{3}{2}}}$$

we note that on the left is the infinite sum of all the contributions to the x component of the electric field due to all the infinitesimal elements of the line of charge. We don't need any special mathematics techniques to evaluate that. The sum of all the parts is the whole. That is, on the left, we have E_x.

The right side, we can evaluate. First, let's factor out the constants:

$$E_x = \left(0.00120\frac{C}{m^3}\right)kx\int_{-0.180m}^{+0.180m}\frac{y'^2dy'}{(x^2+y'^2)^{\frac{3}{2}}}$$

The integral is given on your formula sheet. Carrying out the integration yields:

$$E_x=\left(0.00120\frac{C}{m^3}\right)kx\left[\frac{y'}{\sqrt{x^2+y'^2}}+ln(y'+\sqrt{x^2+y'^2})\right]_{-0.180m}^{+0.180m}$$

$$E_x=\left(.00120\frac{C}{m^3}\right)kx\cdot\left\{\left(\left[\frac{+.018m}{\sqrt{x^2+(+.180m)^2}}+ln\left(+.018m+\sqrt{x^2+(+.018m)^2}\right)\right]-\right.\right.$$

$$\left.\left.\left[\frac{-.180m}{\sqrt{x^2+(-.180)^2}}+ln(-.180m+\sqrt{x^2+(-.180m)^2})\right]\right\}$$

$$E_x=\left(.00120\frac{C}{m^3}\right)kx\cdot\left[\frac{.360m}{\sqrt{x^2+(.180m)^2}}+\ln\frac{\sqrt{x^2+(.180m)^2}+.180m}{\sqrt{x^2+(.180m)^2}-.180m}\right]$$

Substituting the value of the Coulomb constant k from the formula sheet we obtain

$$E_x=\left(.00120\frac{C}{m^3}\right)8.99\times10^9\frac{N\cdot m^2}{C^2}x\cdot\left[\frac{.360m}{\sqrt{x^2+(.180m)^2}}+\ln\frac{\sqrt{x^2+(.180m)^2}+.180m}{\sqrt{x^2+(.180m)^2}-.180m}\right]$$

Finally we have

$$E_x=1.08\times10^7\frac{N}{C\cdot m}\cdot\left[\frac{.360m}{\sqrt{x^2+(.180m)^2}}+\ln\frac{\sqrt{x^2+(.180m)^2}+.180m}{\sqrt{x^2+(.180m)^2}-.180m}\right]$$

It is interesting to note that while the position variable x (which specifies the location of the empty point in space at which the electric field is being calculated) is a constant for purposes of integration (the location of point P does not change as we include the contribution to the electric field at point P of each of the infinitesimal segments making up the charge distribution), an actual value x was never specified. Thus our final result

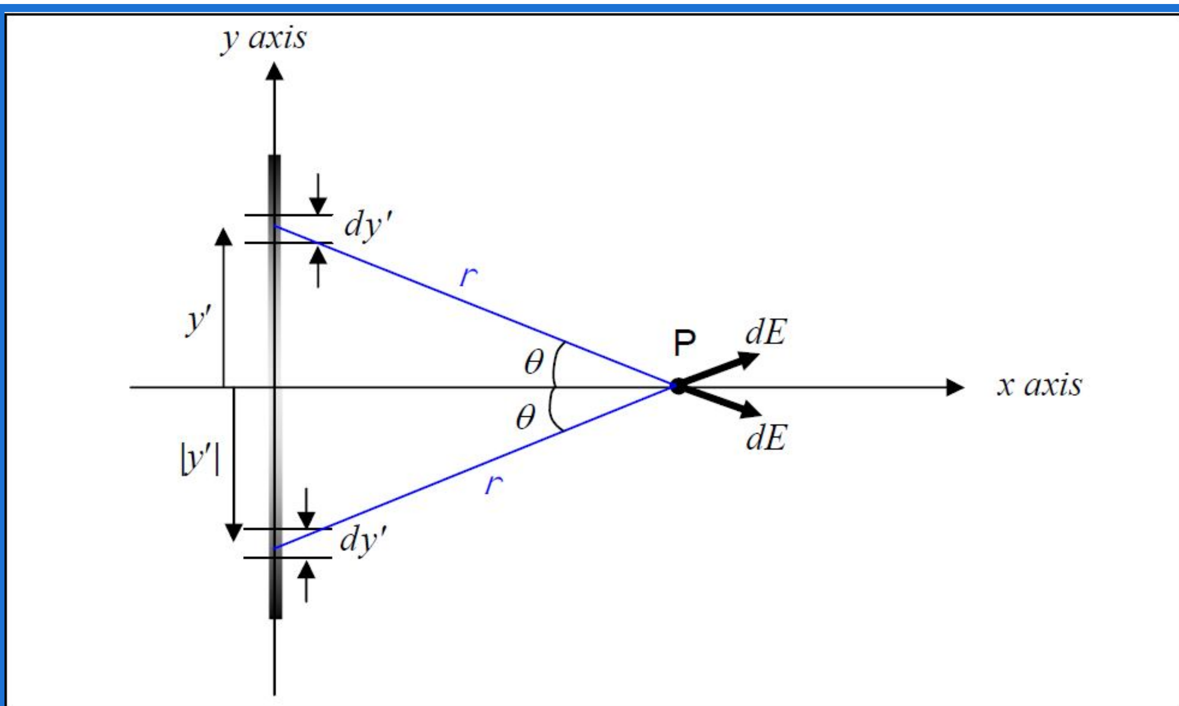
$$E_x=1.08\times10^7\frac{N}{C\cdot m}x\cdot\left[\frac{.360m}{\sqrt{x^2+(.180m)^2}}+\ln\frac{\sqrt{x^2+(.180m)^2}+.180m}{\sqrt{x^2+(.180m)^2}-.180m}\right]$$

for E_x is a function of the position variable x.

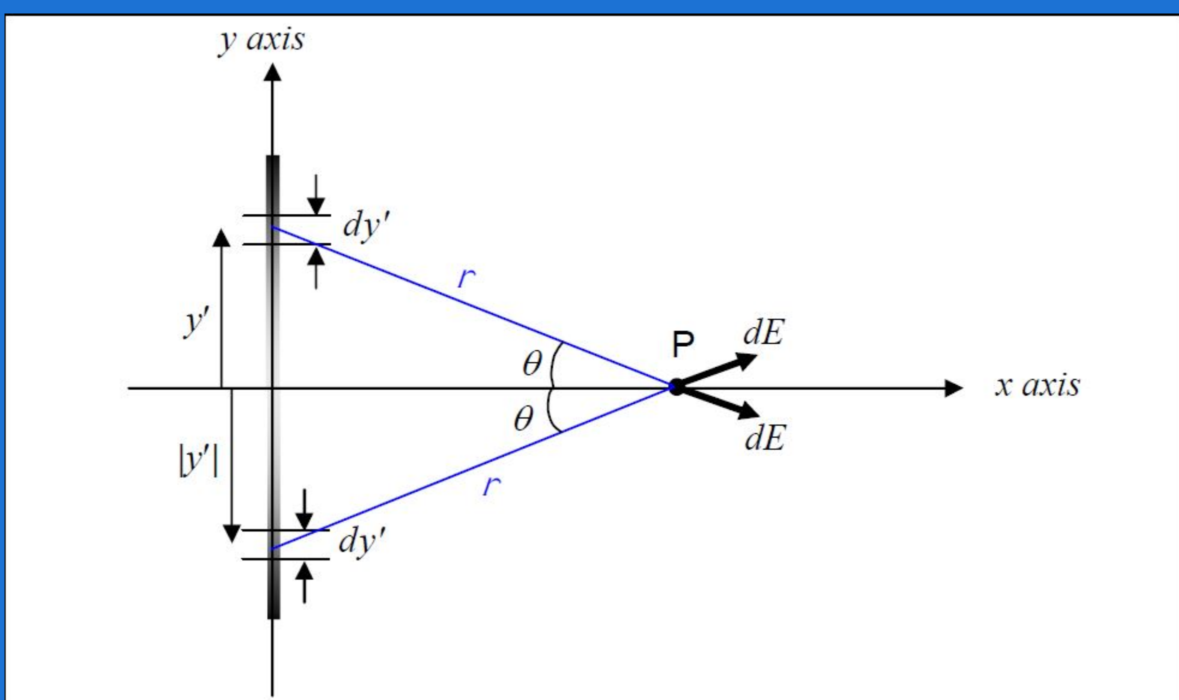
Getting the y-component of the electric field can be done with a lot less work than it took to get E_x if we take advantage of the symmetry of the charge distribution with respect to the x axis. Recall that the charge density λ, for the case at hand, is given by:

$$\lambda=0.00120\frac{C}{m^3}y'^2$$

Because λ is proportional to y'^2, the value of λ is the same at the negative of a specified y' value as it is at the y' value itself. More specifically, the amount of charge in each of the two samesize infinitesimal elements dy' of the charge distribution depicted in the following diagram:



is one and the same value because one element is the same distance below the x axis as the other is above it. This position circumstance also makes the distance r that each element is from point P the same as that of the other, and, it makes the two angles (each of which is labeled θ in the diagram) have one and the same value. Thus the two $\overrightarrow{dE}$ vectors have one and the same magnitude. As a result of the latter two facts (same angle, same magnitude of $\overrightarrow{dE}$), the y components of the two $\overrightarrow{dE}$ vectors cancel each other out. As can be seen in the diagram under consideration:



one is in the $+y$ direction and the other in the $-y$ direction. The y components are "equal and opposite.") In fact, for each and every charge distribution element dy' that is above the x axis and is thus creating a downward contribution to the y component of the electric field at point P, there is an element dy' that is the same distance below the x axis that is creating an upward contribution to the y component of the electric field at point P, canceling the y component of the former. Thus the net sum of all the electric field y components (since they cancel pair-wise) is zero. That is to say that due to the symmetry of the charge distribution with respect to the x axis, $E_y = 0$. Thus,

$$\vec{E} = E_x \hat{i}$$

Using the expression for E_x that we found above, we have, for our final answer:

$$\vec{E} = 1.08 \times 10^7 \frac{N}{C \cdot m} x \left[\frac{.360m}{\sqrt{x^2 + (.180m)^2}} + \ln \frac{\sqrt{x^2 + (.180m)^2} + .180m}{\sqrt{x^2 + (.180m)^2} - .180m} \right] \hat{i}$$

B31: The Electric Potential due to a Continuous Charge Distribution

We have defined electric potential as electric potential-energy-per-charge. Potential energy was defined as the capacity, of an object to do work, possessed by the object because of its position in space. Potential energy is one way of characterizing the effect, or the potential effect, of a force. In the case of electric potential energy, the force in question is the electrostatic force (a.k.a. the Coulomb force)—you know: the repulsive force that two like charges exert on each other, and, the attractive force that two unlike charges exert on each other. The electric potential energy of a charged particle depends on a characteristic of itself, and a characteristic of the point in space at which it finds itself. The characteristic of itself is its charge, and, the characteristic of the point in space is what this chapter is about, the electric potential-energy-per-charge, better known as the electric potential. If we can establish the electric potential-energy-per-charge for each point in space in the vicinity of some source charge, it is easy to determine what the potential energy of a victim charge would be at any such point in space. To do so, we just have to multiply the charge of the victim by the electric potential-energy-per-charge (the electric potential) applicable to the point in space at which the victim is located.

In the next chapter, we exploit the fact that if you know the electric potential throughout a region in space, you can use that knowledge to determine the electric field in that region of space.

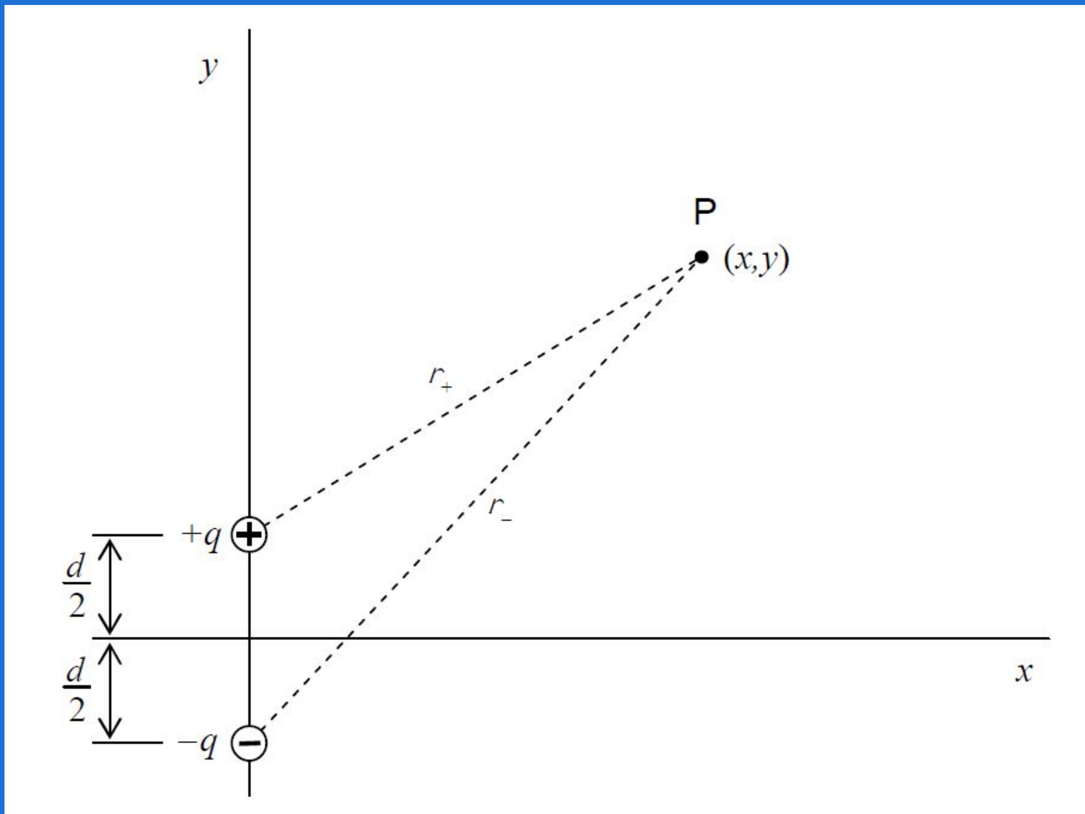
Our purpose of this chapter, is to help you develop your ability to determine the electric potential, as a function of position, in the vicinity of a charge distribution—in particular, in the vicinity of a continuous charge distribution. (Recall that you can think of a continuous charge distribution as some charge that is smeared out over space, whereas a discrete charge distribution is a set of charged particles, with some space between nearest neighbors.)

It's important for you to be able to contrast the electric potential with the electric field. The electric potential is a scalar whereas the electric field is a vector. The electric potential is potential energy-per-charge of the would be victim whereas the electric field is a force-percharge of the would be victim. Hey, that makes this chapter easy compared to the one in which we worked on calculating the electric field due to a continuous charge distribution. It is, in general, easier to calculate a scalar than it is to calculate a vector.

Let's kick things off by doing a review problem involving a discrete distribution of charge. Please solve the following example problem and then check your work against my solution which follows the problem statement.

Find the electric potential on the x-y plane, due to a pair of charges, one of charge $+q$ at $(0, d/2)$ and the other of charge $-q$ at $(0, -d/2)$.

Solution We define a point P to be at some unspecified position (x, y).



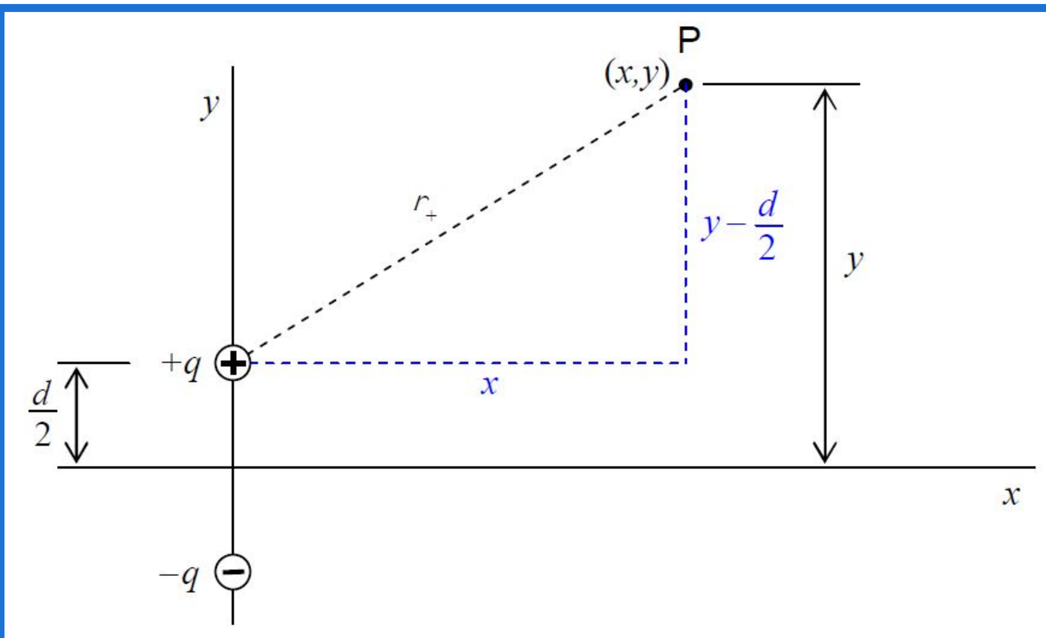
We call the distance from the positive charge to point P, r_+, and, we call the distance from the negative charge to point P, r_-. The electric potential due to a single point charge is given by $\phi = \frac{kq}{r}$. Also, the contributions to the electric potential at one point in space due to more than one point charge simply add like numbers. So, we have:

$$\phi = \phi_1 + \phi_2$$

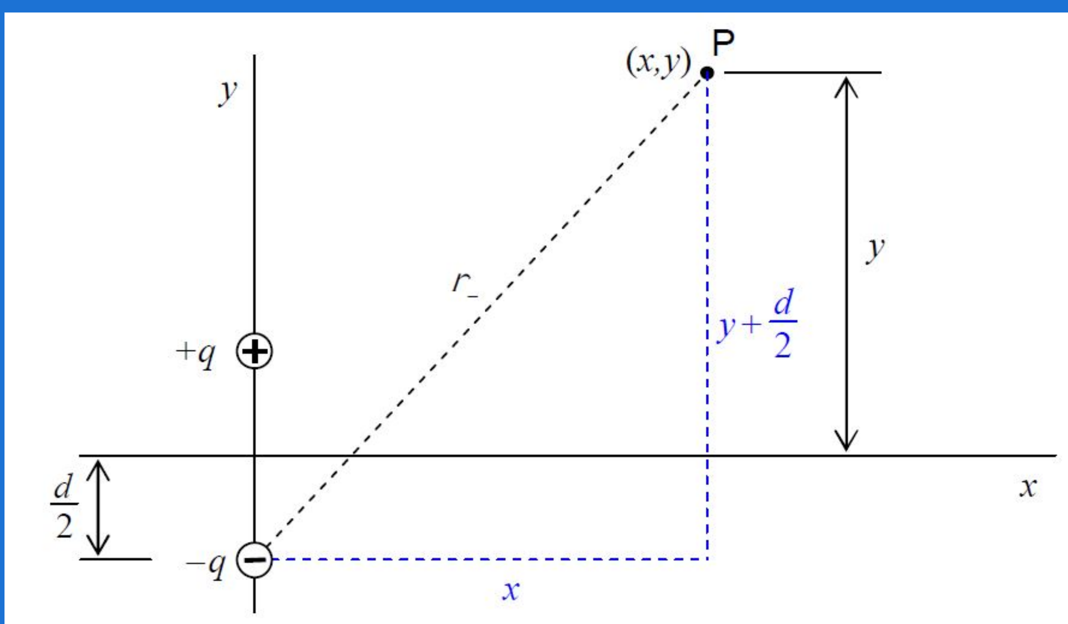
\phi=\frac{kq}{r_{+}}+\frac{k(-q)}{r_{-}} \nonumber

$$\phi = \frac{kq}{r_+} - \frac{kq}{r_-}$$

But, from the diagram:



we can determine that: $r_{+}=\sqrt{x^2+(y-\frac{d}{2})^2}$ and from the diagram:



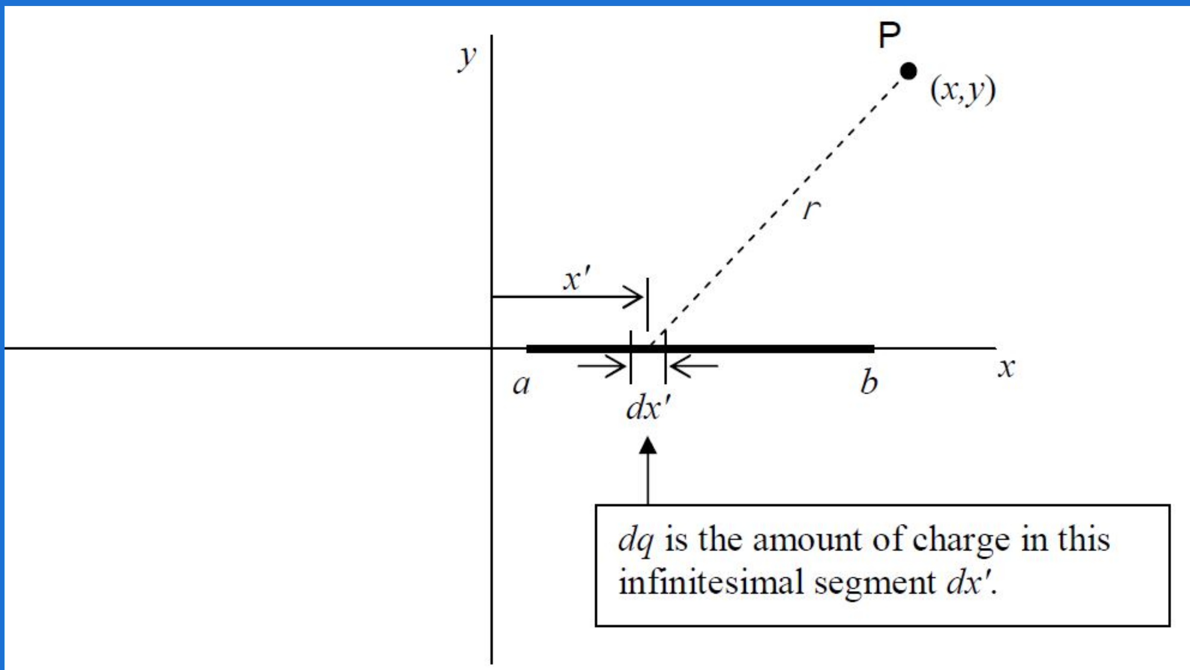
we can see that: \[r_{-}=\sqrt{x^2+(y+\frac{d}{2})^2 \nonumber \]

Plugging both of these results into our expression $\phi=\frac{kq}{r_{+}}-\frac{kq}{r_{-}}$ yields:

$$\phi=\frac{kq}{\sqrt{x^2+(y-\frac{d}{2})^2}}-\frac{kq}{\sqrt{x^2+(y+\frac{d}{2})^2}}$$

That's enough review. Please keep that $\phi=\frac{kq}{r}$ formula in mind as we move on to the new stuff. Also keep in mind the fact that the various contributions to the electric potential at an empty point in space simply add (like numbers/scalars rather than like vectors).

The “new stuff” is the electric potential due to a continuous distribution of charge along a line segment. What we are dealing with is some line segment of charge. It can be anywhere, in any orientation, but for concreteness, let’s consider a line segment of charge on the x axis, say from some $x=a$ to $x=b$ where $a<b$. Furthermore, let’s assume the linear charge density (the chargeper- length) on the line segment to be some function $\lambda(x')$. The idea is to treat the charge distribution as an infinite set of point charges where each point charge may have a different charge value dq depending on where (at what value of x') it is along the line segment.



A particular infinitesimal segment of the line of charge, a length dx' of the line segment, will make a contribution

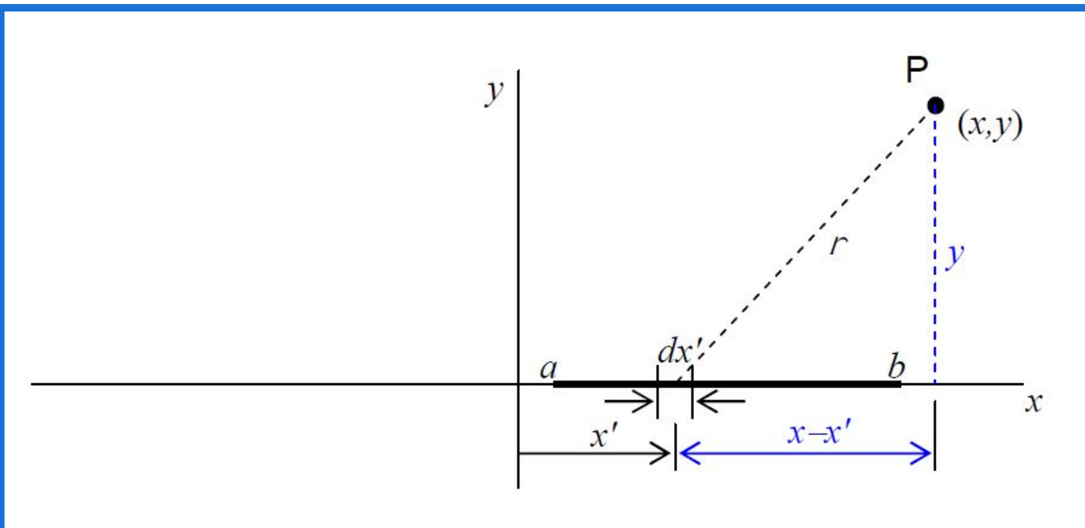
$$d\phi=\frac{kdq}{r}$$

to the electric field at point P.

The amount of charge, dq, in the infinitesimal segment dx′ of the line of charge is just the chargeper-length $\lambda(x')$ (the linear charge density) times the length dx' of the segment. That is to say that $dq=\lambda(x')dx'$. Substituting this into $d\phi=\frac{kdq}{r}$ yields:

$$d\phi=\frac{k\lambda(x')dx'}{r}$$

Applying the Pythagorean theorem to the triangle in the diagram:



tells us that r can be written as $r=\sqrt{(x-x')^2+y^2}$. Substituting this into our expression for dV yields:

\[d\phi=\frac{k\lambda(x')dx'}{\sqrt{(x-x')^2+y^2}}]   Integrating both sides yields:

\[\int d\phi=\int_a^b \frac{k\lambda(x')dx'}{\sqrt{(x-x')^2+y^2}}]

\[\phi=k\int_a^b \frac{\lambda(x')dx'}{\sqrt{(x-x')^2+y^2}}]

This is the electric potential at point P due to the charged line segment on the x axis. Each bit of charge on the line segment is specified by its position variable x'. Thus, in summing the contributions to the electric potential due to each bit of charge, x' is our variable of integration. While its position coordinates have not been specified, but rather, they have been designated x and y, point P is a fixed point in space. Hence, in summing up all the contributions to the electric potential at point P; x and y are to be considered constants. After the integral is done, however, because we never specified values for x and y, the resulting expression for ϕ can be considered to be a function of x and y.

B32: Calculating the Electric Field from the Electric Potential

The plan here is to develop a relation between the electric field and the corresponding electric potential that allows you to calculate the electric field from the electric potential.

The electric field is the force-per-charge associated with empty points in space that have a forceper- charge because they are in the vicinity of a source charge or some source charges. The electric potential is the potential energy-per-charge associated with the same empty points in space. Since the electric field is the force-per-charge, and the electric potential is the potential energy-per-charge, the relation between the electric field and its potential is essentially a special case of the relation between any force and its associated potential energy. So, I'm going to start by developing the more general relation between a force and its potential energy, and then move on to the special case in which the force is the electric field times the charge of the victim and the potential energy is the electric potential times the charge of the victim.

The idea behind potential energy was that it represented an easy way of getting the work done by a force on a particle that moves from point A to point B under the influence of the force. By definition, the work done is the force along the path times the length of the path. If the force along the path varies along the path, then we take the force along the path at a particular point on the path, times the length of an infinitesimal segment of the path at that point, and repeat, for every infinitesimal segment of the path, adding the results as we go along. The final sum is the work. The potential energy idea represents the assignment of a value of potential energy to every point in space so that, rather than do the path integral just discussed, we simply subtract the value of the potential energy at point A from the value of the potential energy at point B. This gives us the change in the potential energy experienced by the particle in moving from point A to point B. Then, the work done is the negative of the change in potential energy. For this to be the case, the assignment of values of potential energy values to points in space must be done just right. For things to work out on a macroscopic level, we must ensure that they are correct at an infinitesimal level. We can do this by setting:

$$\text{Work as Change in Potential Energy} = \text{Work as Force-Along-Path times Path Length}$$

$$-dU = \vec{F} \cdot \overrightarrow{ds}$$

where:

- dU is an infinitesimal change in potential energy,
- $\vec{F}$ is a force, and
- $\overrightarrow{ds}$ is the infinitesimal displacement-along-the-path vector.

In Cartesian unit vector notation, $\overrightarrow{ds}$ can be expressed as $\overrightarrow{ds} = dx\hat{\imath} + dy\hat{\jmath} + dz\hat{k}$, and $\vec{F}$ can be expressed as $\vec{F} = F_x\hat{\imath} + F_y\hat{\jmath} + F_z\hat{k}$. Substituting these two expressions into our expression $-dU = \vec{F} \cdot \overrightarrow{ds}$, we obtain:

$$-dU = (F_x\hat{\imath} + F_y\hat{\jmath} + F_z\hat{k}) \cdot (dx\hat{\imath} + dy\hat{\jmath} + dz\hat{k})$$

$$-dU = F_x dx + F_y dy + F_z dz$$

Now check this out. If we hold y and z constant (in other words, if we consider dy and dz to be zero) then:

$$\underbrace{-dU = F_x dx}_{\text{when y and z are held constant}}$$

Dividing both sides by dx and switching sides yields:

$$\underbrace{F_x = -\frac{dU}{dx}}_{\text{when y and z are held contsant}}$$

That is, if you have the potential energy as a function of x, y, and z; and; you take the negative of the derivative with respect to x while holding y and z constant, you get the x component of the force that is characterized by the potential energy function. Taking the derivative of U with respect to x while holding the other variables constant is called taking the partial derivative of U with respect to x and written

$$\frac{\partial U}{\partial x}$$

Alternatively, one writes

$$\left.\frac{\partial U}{\partial x}\right|_{y,z}$$

to be read, "the partial derivative of U with respect to x holding y and z constant." This latter expression makes it more obvious to the reader just what is being held constant. Rewriting our expression for F_x with the partial derivative notation, we have:

$$F_x = -\frac{\partial U}{\partial x}$$

Returning to our expression $-dU = F_x dx + F_y dy + F_z dz$, if we hold x and z constant we get:

$$F_y = -\frac{\partial U}{\partial y}$$

and, if we hold x and y constant we get,

$$F_z = -\frac{\partial U}{\partial z}$$

Substituting these last three results into the force vector expressed in unit vector notation:

$$\vec{F} = F_x \hat{i} + F_y \hat{j} + F_z \hat{k}$$

yields

$$\vec{F} = -\frac{\partial U}{\partial x}\hat{i} - \frac{\partial U}{\partial y}\hat{j} - \frac{\partial U}{\partial z}\hat{k}$$

which can be written:

$$\vec{F} = -\left(\frac{\partial U}{\partial x}\hat{i} + \frac{\partial U}{\partial y}\hat{j} + \frac{\partial U}{\partial z}\hat{k}\right)$$

Okay, now, this business of:

- taking the partial derivative of U with respect to x and multiplying the result by the unit vector $\hat{i}$ and then,
- taking the partial derivative of U with respect to y and multiplying the result by the unit vector $\hat{j}$ and then,
- taking the partial derivative of U with respect to z and multiplying the result by the unit vector $\hat{k}$, and then,
- adding all three partial-derivative-times-unit-vector quantities up,

is called "taking the gradient of U " and is written ∇U. "Taking the gradient" is something that you do to a scalar function, but, the result is a vector. In terms of our gradient notation, we can write our expression for the force as,

$$\vec{F} = -\nabla U \tag{B32.1}$$

Check this out for the gravitational potential near the surface of the earth. Define a Cartesian coordinate system with, for instance, the origin at sea level, and, with the x-y plane being horizontal and the $+z$ direction being upward. Then, the potential energy of a particle of mass m is given as:

$$U = mgz$$

Now, suppose you knew this to be the potential but you didn't know the force. You can calculate the force using $\vec{F} = -\nabla U$, which, as you know, can be written:

$$\vec{F} = -\left(\frac{\partial U}{\partial x}\hat{i} + \frac{\partial U}{\partial y}\hat{j} + \frac{\partial U}{\partial z}\hat{k}\right)$$

Substituting $U = mgz$ in for U we have

$$\vec{F}=-\left(\frac{\partial}{\partial x}(mgz)\hat{\imath}+\frac{\partial}{\partial y}(mgz)\hat{\jmath}+\frac{\partial}{\partial z}(mgz)\hat{k}\right)$$

Now remember, when we take the partial derivative with respect to x we are supposed to hold y and z constant. (There is no y.) But, if we hold z constant, then the whole thing (mgz) is constant. And, the derivative of a constant, with respect to x, is 0. In other words, $\frac{\partial}{\partial x}(mgz)=0$. Likewise, $\frac{\partial}{\partial y}(mgz)=0$. In fact, the only non zero partial derivative in our expression for the force is $\frac{\partial}{\partial z}(mgz)=mg$. So:

$$\vec{F}=-(0\hat{\imath}+0\hat{\jmath}+mg\hat{k})$$

In other words:

$$\vec{F}=-mg\hat{k}$$

That is to say that, based on the gravitational potential $U=mgz$, the gravitational force is in the $\hat{k}$direction (downward), and, is of magnitude mg. Of course, you knew this in advance, the gravitational force in question is just the weight force. The example was just meant to familiarize you with the gradient operator and the relation between force and potential energy.

Okay, as important as it is that you realize that we are talking about a general relationship between force and potential energy, it is now time to narrow the discussion to the case of the electric force and the electric potential energy, and, from there, to derive a relation between the electric field and electric potential (which is electric potential-energy-per-charge).

Starting with $\vec{F}=-\nabla U$ written out the long way:

$$\vec{F}=-\left(\frac{\partial U}{\partial x}\hat{\imath}+\frac{\partial U}{\partial y}\hat{\jmath}+\frac{\partial U}{\partial z}\hat{k}\right)$$

we apply it to the case of a particle with charge q in an electric field $\vec{E}$ (caused to exist in the region of space in question by some unspecified source charge or distribution of source charge). The electric field exerts a force $\vec{F}=q\vec{E}$ on the particle, and, the particle has electric potential energy $U=q\varphi$ where φ is the electric potential at the point in space at which the charged particle is located. Plugging these into $\vec{F}=-\left(\frac{\partial U}{\partial x}\hat{\imath}+\frac{\partial U}{\partial y}\hat{\jmath}+\frac{\partial U}{\partial z}\hat{k}\right)$ yields:

$$q\vec{E}=-\left(\frac{\partial(q\varphi)}{\partial x}\hat{\imath}+\frac{\partial(q\varphi)}{\partial y}\hat{\jmath}+\frac{\partial(q\varphi)}{\partial z}\hat{k}\right)$$

which I copy here for your convenience:

$$q\vec{E}=-\left(\frac{\partial(q\varphi)}{\partial x}\hat{\imath}+\frac{\partial(q\varphi)}{\partial y}\hat{\jmath}+\frac{\partial(q\varphi)}{\partial z}\hat{k}\right)$$

The q inside each of the partial derivatives is a constant so we can factor it out of each partial derivative.

$$q\vec{E}=-\left(q\frac{\partial\varphi}{\partial x}\hat{\imath}+q\frac{\partial\varphi}{\partial y}\hat{\jmath}+q\frac{\partial\varphi}{\partial z}\hat{k}\right)$$

Then, since q appears in every term, we can factor it out of the sum:

$$q\vec{E}=-q\left(\frac{\partial\varphi}{\partial x}\hat{\imath}+\frac{\partial\varphi}{\partial y}\hat{\jmath}+\frac{\partial\varphi}{\partial z}\hat{k}\right)$$

Dividing both sides by the charge of the victim yields the desired relation between the electric field and the electric potential:

$$\vec{E}=-\left(\frac{\partial\varphi}{\partial x}\hat{\imath}+\frac{\partial\varphi}{\partial y}\hat{\jmath}+\frac{\partial\varphi}{\partial z}\hat{k}\right)$$

We see that the electric field $\vec{E}$ is just the gradient of the electric potential φ. This result can be expressed more concisely by means of the gradient operator as:

$$\vec{E}=-\nabla\varphi \tag{B32.2}$$

In Example 31-1, we found that the electric potential due to a pair of particles, one of charge $+q$ at $(0, d/2)$ and the other of charge $-q$ at $(0, -d/2)$, is given by:

$$\varphi = \frac{kq}{\sqrt{x^2+(y-\frac{d}{2})^2}} - \frac{kq}{\sqrt{x^2+(y+\frac{d}{2})^2}}$$

Such a pair of charges is called an electric dipole. Find the electric field of the dipole, valid for any point on the x axis.

Solution: We can use a symmetry argument and our conceptual understanding of the electric field due to a point charge to deduce that the x component of the electric field has to be zero, and, the y component has to be negative. But, let's use the gradient method to do that, and, to get an expression for the y component of the electric field. I do argue, however that, from our conceptual understanding of the electric field due to a point charge, neither particle's electric field has a z component in the x-y plane, so we are justified in neglecting the z component altogether. As such our gradient operator expression for the electric field

$$\vec{E} = -\nabla\varphi$$

becomes

$$\vec{E} = -\Big(\frac{\partial\varphi}{\partial x}\hat{i} + \frac{\partial\varphi}{\partial y}\hat{j}\Big)$$

Let's work on the $\frac{\partial\varphi}{\partial x}$ part:

$$\frac{\partial\varphi}{\partial x} = \frac{\partial}{\partial x}\Big(\frac{kq}{\sqrt{x^2+(y-\frac{d}{2})^2}} - \frac{kq}{\sqrt{x^2+(y+\frac{d}{2})^2}}\Big)$$

$$\frac{\partial\varphi}{\partial x} = kq\frac{\partial}{\partial x}\Big(\Big[x^2+(y-\frac{d}{2})^2\Big]^{-\frac{1}{2}} - \Big[x^2+(y+\frac{d}{2})^2\Big]^{-\frac{1}{2}}\Big)$$

$$\frac{\partial\varphi}{\partial x} = kq\Big(-\frac{1}{2}\Big[x^2+(y-\frac{d}{2})^2\Big]^{-\frac{3}{2}}2x - \;-\frac{1}{2}\Big[x^2+(y+\frac{d}{2})^2\Big]^{-\frac{3}{2}}2x\Big)$$

$$\frac{\partial\varphi}{\partial x} = kqx\Big(\Big[x^2+(y+\frac{d}{2})^2\Big]^{-\frac{3}{2}} - \Big[x^2+(y-\frac{d}{2})^2\Big]^{-\frac{3}{2}}\Big)$$

$$\frac{\partial\varphi}{\partial x} = \frac{kqx}{\Big[x^2+(y+\frac{d}{2})^2\Big]^{\frac{3}{2}}} - \frac{kqx}{\Big[x^2+(y-\frac{d}{2})^2\Big]^{\frac{3}{2}}}$$

We were asked to find the electric field on the x axis, so, we evaluate this expression at $y=0$:

$$\frac{\partial\varphi}{\partial x} = \frac{kqx}{\Big[x^2+(0+\frac{d}{2})^2\Big]^{\frac{3}{2}}} - \frac{kqx}{\Big[x^2+(0-\frac{d}{2})^2\Big]^{\frac{3}{2}}} \qquad \frac{\partial\varphi}{\partial x}\Big|_{y=0} = 0$$

To continue with our determination of $\vec{E} = -(\frac{\partial\varphi}{\partial x}\hat{i} + \frac{\partial\varphi}{\partial y}\hat{j})$, we next solve for $\frac{\partial\varphi}{\partial y}$.

$$\frac{\partial\varphi}{\partial y} = \frac{\partial}{\partial y}\Big(\frac{kq}{\sqrt{x^2+(y-\frac{d}{2})^2}} - \frac{kq}{\sqrt{x^2+(y+\frac{d}{2})^2}}\Big)$$

$$\frac{\partial\varphi}{\partial y}=kq\frac{\partial}{\partial y}\Big(\Big[x^2+(y-\frac{d}{2})^2\Big]^{-\frac{1}{2}}-\Big[x^2+(y+\frac{d}{2})^2\Big]^{-\frac{1}{2}}\Big)$$

$$\frac{\partial\varphi}{\partial y}=kq\Big(-\frac{1}{2}\Big[x^2+(y-\frac{d}{2})^2\Big]^{-\frac{3}{2}}2(y-\frac{d}{2})-\ -\frac{1}{2}\Big[x^2+(y+\frac{d}{2})^2\Big]^{-\frac{3}{2}}2(y+\frac{d}{2})$$

$$\frac{\partial\varphi}{\partial y}=kq\Big(\Big[x^2+(y+\frac{d}{2})^2\Big]^{-\frac{3}{2}}(y+\frac{d}{2})-\Big[x^2+(y-\frac{d}{2})^2\Big]^{-\frac{3}{2}}(y-\frac{d}{2})\Big)$$

$$\frac{\partial\varphi}{\partial y}=\frac{kq(y+\frac{d}{2})}{\Big[x^2+(y+\frac{d}{2})^2\Big]^{\frac{3}{2}}}-\frac{kq(y-\frac{d}{2})}{\Big[x^2+(y-\frac{d}{2})^2\Big]^{\frac{3}{2}}}$$

Again, we were asked to find the electric field on the x axis, so, we evaluate this expression at $y=0$:

$$\frac{\partial\varphi}{\partial y}\Big|_{y=0}=\frac{kq\Big(0+\frac{d}{2}\Big)}{\Big[x^2+\Big(0+\frac{d}{2}\Big)^2\Big]^{\frac{3}{2}}}-\frac{kq\Big(0-\frac{d}{2}\Big)}{\Big[x^2+\Big(0-\frac{d}{2}\Big)^2\Big]^{\frac{3}{2}}}\quad\frac{\partial\varphi}{\partial y}\Big|_{y=0}=\frac{kqd}{\Big[x^2+\frac{d^2}{4}\Big]^{\frac{3}{2}}}$$

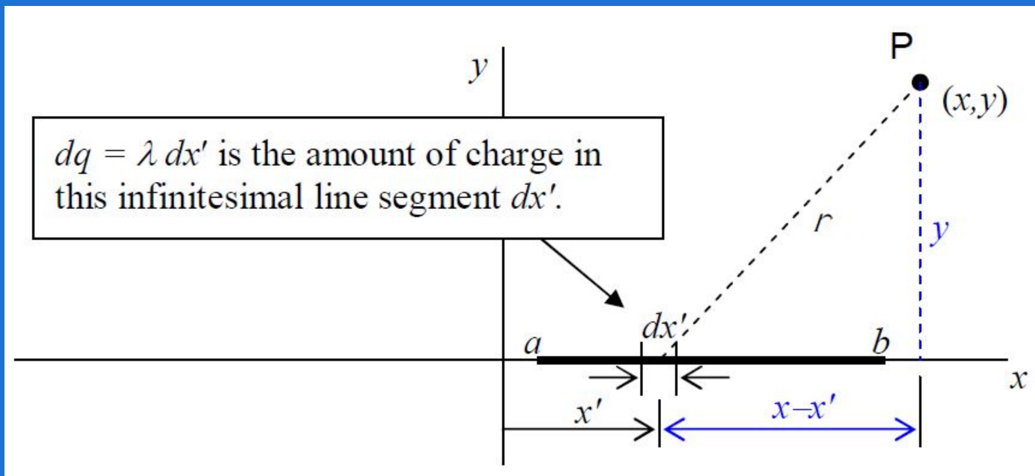
Plugging $\frac{\partial\varphi}{\partial x}\Big|_{y=0}=0$ and $\frac{\partial\varphi}{\partial y}\Big|_{y=0}=\frac{kqd}{\Big[x^2+\frac{d^2}{4}\Big]^{\frac{3}{2}}}$ into $\vec{E}=-\Big(\frac{\partial\varphi}{\partial x}\hat{i}+\frac{\partial\varphi}{\partial y}\hat{j}\Big)$ yields:

$$\vec{E}=-\Big(0\hat{i}+\frac{kqd}{[x^2+\frac{d^2}{4}]^{\frac{3}{2}}}\hat{j}\Big)\quad\vec{E}=-\frac{kqd}{[x^2+\frac{d^2}{4}]^{\frac{3}{2}}}\hat{j}$$

As expected, $\vec{E}$ is in the –y direction. Note that to find the electric field on the x axis, you have to take the derivatives first, and then evaluate at $y=0$.

A line of charge extends along the x axis from $x=a$ to $x=b$. On that line segment, the linear charge density λ is a constant. Find the electric potential as a function of position (x and y) due to that charge distribution on the x-y plane, and then, from the electric potential, determine the electric field on the x axis.

Solution: First, we need to use the methods of chapter 31 to get the potential for the specified charge distribution (a linear charge distribution with a constant linear charge density λ).



$$d\varphi=\frac{k\,dq}{r}$$

$$dq=\lambda dx' \quad \text{and} \quad r=\sqrt{(r-x')^2+y^2} \quad d\varphi=\frac{k\lambda(x')dx'}{\sqrt{(x-x')^2+y^2}} \quad \int d\varphi=\int_a^b\frac{k\lambda dx'}{\sqrt{(x-x')^2+y^2}}$$

$$\varphi=k\lambda\int_a^b\frac{dx'}{\sqrt{(x-x')^2+y^2}}$$

To carry out the integration, we use the variable substitution:

$$u=x-x' \quad du=-dx'\Rightarrow dx'=du$$

Lower Integration Limit: When

$$x'=a, u=x-a$$

Upper Integration Limit: When

$$x'=b, u=x-b$$

Making these substitutions, we obtain:

$$\varphi=k\lambda\int_{x-a}^{x-b}\frac{-du}{\sqrt{u^2+y^2}}$$

which I copy here for your convenience:

$$\varphi=k\lambda\int_{x-a}^{x-b}\frac{-du}{\sqrt{u^2+y^2}}$$

Using the minus sign to interchange the limits of integration, we have:

$$\varphi=k\lambda\int_{x-a}^{x-b}\frac{du}{\sqrt{u^2+y^2}}$$

Using the appropriate integration formula from the formula sheet we obtain:

$$\varphi=k\lambda\ln(u+\sqrt{u^2+y^2})\Big|_{x-b}^{x-a}$$

$$\varphi=k\lambda\left\{\ln[x-a+\sqrt{(x-a)^2+y^2}\,]-\ln\left[x-b+\sqrt{(x-b)^2+y^2}\,\right]\right\}$$

Okay, that's the potential. Now we have to take the gradient of it and evaluate the result at $y=0$ to get the electric field on the x axis. We need to find

$\vec{E} = -\bigtriangledown\varphi$ which, in the absence of any z dependence, can be written as: $\vec{E} = -\Big(\dfrac{\partial\varphi}{\partial x}\hat{i} + \dfrac{\partial\varphi}{\partial y}\hat{j}\Big)$

We start by finding $\frac{\partial\varphi}{\partial x}$:

$$\frac{\partial\varphi}{\partial x} = \frac{\partial}{\partial x}\Big(k\lambda\Big\{\ln\Big[x-a+\sqrt{(x-a)^2+y^2}\,\Big] - \ln\Big[x-b+\sqrt{(x-b)^2+y^2}\,\Big]\Big\}\Big)$$

$$\frac{\partial\varphi}{\partial x} = k\lambda\Big\{\frac{\partial}{\partial x}\ln\Big[x-a+((x-a)^2+y^2)^{\frac{1}{2}}\Big] - \frac{\partial}{\partial x}\ln\Big[x-b+((x-b)^2+y^2)^{\frac{1}{2}}\Big]\Big\}$$

$$\frac{\partial\varphi}{\partial x} = k\lambda\left\{\frac{1+\frac{1}{2}\Big((x-a)^2+y^2\Big)^{-\frac{1}{2}}2(x-a)}{x-a+\Big((x-a)^2+y^2\Big)^{\frac{1}{2}}} - \frac{1+\frac{1}{2}\Big((x-b)^2+y^2\Big)^{-\frac{1}{2}}2(x-b)}{x-b+\Big((x-b)^2+y^2\Big)^{\frac{1}{2}}}\right\}$$

$$\frac{\partial\varphi}{\partial x} = k\lambda\left\{\frac{1+(x-a)\Big((x-a)^2+y^2\Big)^{-\frac{1}{2}}}{x-a+\Big((x-a)^2+y^2\Big)^{\frac{1}{2}}} - \frac{1+(x-b)\Big((x-b)^2+y^2\Big)^{-\frac{1}{2}}}{x-b+\Big((x-b)^2+y^2\Big)^{\frac{1}{2}}}\right\}$$

Evaluating this at $y=0$ yields:

$$\frac{\partial\varphi}{\partial x}\Big|_{y=0} = k\lambda\Big(\frac{1}{x-a} - \frac{1}{x-b}\Big)$$

Now, let's work on getting $\frac{\partial\varphi}{\partial y}$. I'll copy our result for φ from above and then take the partial derivative with respect to y (holding x constant):

$$\varphi = k\lambda\left\{\ln\Big[x-a+\sqrt{(x-a)^2+y^2}\,\Big] - \ln\Big[x-b+\sqrt{(x-b)^2+y^2}\,\Big]\right\}$$

$$\frac{\partial\varphi}{\partial y} = \frac{\partial}{\partial y}\left(k\lambda\Big\{\ln\Big[x-a+\sqrt{(x-a)^2+y^2}\,\Big] - \ln\Big[x-b+\sqrt{(x-b)^2+y^2}\,\Big]\Big\}\right)$$

$$\frac{\partial\varphi}{\partial y} = k\lambda\left\{\frac{\partial}{\partial y}\ln\Big[x-a+\Big((x-a)^2+y^2\Big)^{\frac{1}{2}}\Big] - \frac{\partial}{\partial x}\ln\Big[x-b+\Big((x-b)^2+y^2\Big)^{\frac{1}{2}}\Big]\right\}$$

$$\frac{\partial \varphi}{\partial y}=k\lambda\left\{\frac{\frac{1}{2}\left((x-a)^2+y^2\right)^{-\frac{1}{2}}2y}{x-a+\left((x-a)^2+y^2\right)^{\frac{1}{2}}}-\frac{\frac{1}{2}\left((x-b)^2+y^2\right)^{-\frac{1}{2}}2y}{x-b+\left((x-b)^2+y^2\right)^{\frac{1}{2}}}\right\}$$

$$\frac{\partial \varphi}{\partial y}=k\lambda\left\{\frac{y\left((x-a)^2+y^2\right)^{-\frac{1}{2}}}{x-a+\left((x-a)^2+y^2\right)^{\frac{1}{2}}}-\frac{y\left((x-b)^2+y^2\right)^{-\frac{1}{2}}}{x-b+\left((x-b)^2+y^2\right)^{\frac{1}{2}}}\right\}$$

Evaluating this at $y=0$ yields:

$$\left.\frac{\partial \varphi}{\partial y}\right|_{y=0}=0$$

Plugging $\left.\frac{\partial \varphi}{\partial x}\right|_{y=0}=k\lambda\left(\frac{1}{x-a}-\frac{1}{x-b}\right)$ and $\left.\frac{\partial \varphi}{\partial y}\right|_{y=0}=0$ into $\vec{E}=-\left(\frac{\partial \varphi}{\partial x}\hat{i}+\frac{\partial \varphi}{\partial y}\hat{j}\right)$ yields:

$$\vec{E}=-\left(k\lambda\left(\frac{1}{x-a}-\frac{1}{x-b}\right)\hat{i}+0\hat{j}\right) \quad \vec{E}=k\lambda\left(\frac{1}{x-b}-\frac{1}{x-a}\right)\hat{i}$$

B33: Gauss's Law

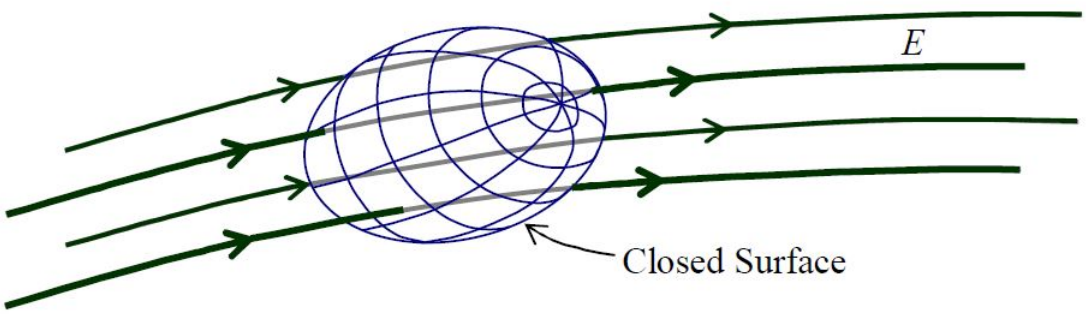
> *When asked to find the electric flux through a closed surface due to a specified non-trivial charge distribution, folks all too often try the immensely complicated approach of finding the electric field everywhere on the surface and doing the integral of $\vec{E}$ dot $\overrightarrow{dA}$ over the surface instead of just dividing the total charge that the surface encloses by ϵ_o.*

Conceptually speaking, Gauss's Law states that the number of electric field lines poking outward through an imaginary closed surface is proportional to the charge enclosed by the surface.

A closed surface is one that divides the universe up into two parts: inside the surface, and, outside the surface. An example would be a soap bubble for which the soap film itself is of negligible thickness. I'm talking about a spheroidal soap bubble floating in air. Imagine one in the shape of a tin can, a closed jar with its lid on, or a closed box. These would also be closed surfaces. To be closed, a surface has to encompass a volume of empty space. A surface in the shape of a flat sheet of paper would not be a closed surface. In the context of Gauss's law, an imaginary closed surface is often referred to as a Gaussian surface.

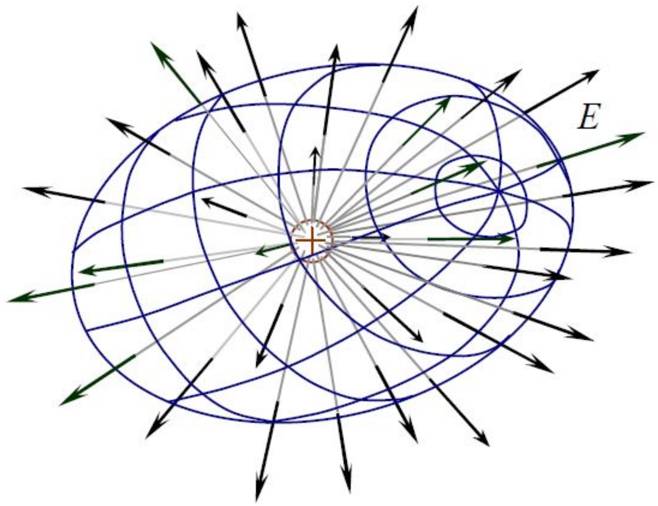
In conceptual terms, if you use Gauss's Law to determine how much charge is in some imaginary closed surface by counting the number of electric field lines poking outward through the surface, you have to consider inward-poking electric field lines as negative outward-poking field lines. Also, if a given electric field line pokes through the surface at more than one location, you have to count each and every penetration of the surface as another field line poking through the surface, adding $+1$ to the tally if it pokes outward through the surface, and -1 to the tally if it pokes inward through the surface.

So for instance, in a situation like:



we have 4 electric field lines poking inward through the surface which, together, count as -4 outward field lines, plus, we have 4 electric field lines poking outward through the surface which together count as $+4$ outward field lines for a total of 0 outward-poking electric field lines through the closed surface. By Gauss's Law, that means that the net charge inside the Gaussian surface is zero.

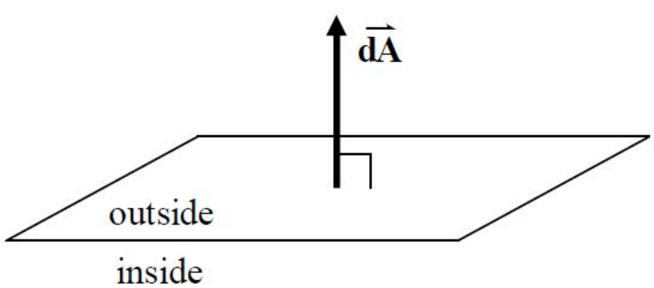
The following diagram might make our conceptual statement of Gauss's Law seem like plain old common sense to you:



The closed surface has the shape of an egg shell. There are 32 electric field lines poking outward through the Gaussian surface (and zero poking inward through it) meaning there must (according to Gauss's Law) be a net positive charge inside the closed surface. Indeed, from your understanding that electric field lines begin, either at positive charges or infinity, and end, either at negative charges or infinity, you could probably deduce our conceptual form of Gauss's Law. If the net number of electric field lines poking out through a closed surface is greater than zero, then you must have more lines beginning inside the surface than you have ending inside the surface, and, since field lines begin at positive charge, that must mean that there is more positive charge inside the surface than there is negative charge.

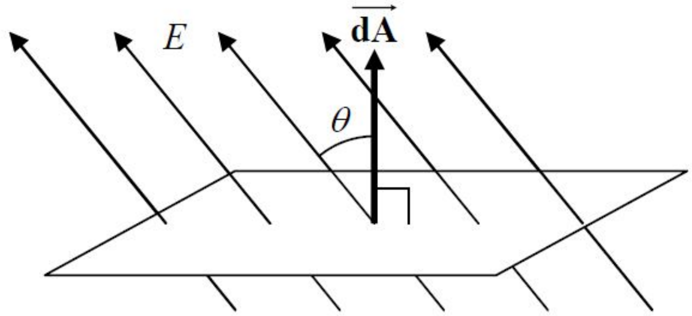
Our conceptual idea of the net number of electric field lines poking outward through a Gaussian surface corresponds to the net outward electric flux Φ_E through the surface.

To write an expression for the infinitesimal amount of outward flux $d\Phi_E$ through an infinitesimal area element dA, we first define an area element vector $\overrightarrow{dA}$ whose magnitude is, of course, just the area dA of the element; and; whose direction is perpendicular to the area element, and, outward. (Recall that a closed surface separates the universe into two parts, an inside part and an outside part. Thus, at any point on the surface, that is to say at the location of any infinitesimal area element on the surface, the direction outward, away from the inside part, is unambiguous.)



In terms of that area element, and, the electric field $\vec{E}$ at the location of the area element, we can write the infinitesimal amount of electric flux $d\Phi_E$ through the area element as:

$$d\Phi_E = \vec{E} \cdot \overrightarrow{dA}$$



Recall that the dot product $\vec{E} \cdot \overrightarrow{dA}$ can be expressed as $EdA\cos\theta$. For a given E and a given amount of area, this yields a maximum value for the case of $\theta = 0^\circ$ (when $\vec{E}$ is parallel to $\overrightarrow{dA}$ meaning that $\vec{E}$ is perpendicular to the surface); zero when $\theta = 90^\circ$ (when $\vec{E}$ is perpendicular to $\overrightarrow{dA}$ meaning that $\vec{E}$ is parallel to the surface); and; a negative value when θ is greater than 90° (with 180° being the greatest value of θ possible, the angle at which $\vec{E}$ is again perpendicular to the surface, but, in this case, into the surface.)

Now, the flux is the quantity that we can think of conceptually as the number of field lines. So, in terms of the flux, Gauss's Law states that the net outward flux through a closed surface is proportional to the amount of charge enclosed by that surface. Indeed, the constant of proportionality has been established to be $\frac{1}{\epsilon_0}$ where ϵ_0 (epsilon zero) is the universal constant known as the electric permittivity of free space. (You've seen ϵ_0 before. At the time, we stated that the Coulomb constant k is often expressed as $\frac{1}{4\pi\epsilon_0}$. Indeed, the identity $k = \frac{1}{4\pi\epsilon_0}$ appears on your formula sheet.) In equation form, Gauss's Law reads:

$$\oint \vec{E} \cdot \overrightarrow{dA} = \frac{Q_{\text{ENCLOSED}}}{\epsilon_0} \tag{B33.1}$$

The circle on the integral sign, combined with the fact that the infinitesimal in the integrand is an area element, means that the integral is over a closed surface. The quantity on the left is the sum of the product $\vec{E} \cdot \overrightarrow{dA}$ for each and every area element dA making up the closed surface. It is the total outward electric flux through the surface.

$$\Phi_E = \oint \vec{E} \cdot \overrightarrow{dA} \tag{B33.2}$$

Using this definition in Gauss's Law allows us to write Gauss's Law in the form:

$$\Phi_E = \frac{Q_{\text{ENCLOSED}}}{\epsilon_0} \tag{B33.3}$$

How You Will be Using Gauss's Law

Gauss's Law is an integral equation. Such an integral equation can also be expressed as a differential equation. We won't be using the differential form, but, because of its existence, the Gauss's Law equation

$$\oint \vec{E} \cdot \overrightarrow{dA} = \frac{Q_{\text{ENCLOSED}}}{\epsilon_0}$$

is referred to as the integral form of Gauss's Law. The integral form of Gauss's Law can be used for several different purposes. In the course for which this book is written, you will be using it in a limited manner consistent with the mathematical prerequisites and co-requisites for the course. Here's how:

1. Gauss's Law in the form $\Phi_E = \frac{Q_{\text{ENCLOSED}}}{\epsilon_0}$ makes it easy to calculate the net outward flux through a closed surface that encloses a known amount of charge Q_{ENCLOSED}. Just divide the amount of charge Q_{ENCLOSED} by ϵ_0 (given on your formula sheet as $\epsilon_0 = 8.85 \times 10^{-12} \frac{C^2}{N \cdot m^2}$ and you have the flux through the closed surface.

2. Given the electric field at all points on a closed surface, one can use the integral form of Gauss's Law to calculate the charge inside the closed surface. This can be used as a check for a case in which the electric field due to a given distribution of charge has been calculated by a means other than Gauss's Law. You will only be expected to do this in cases in which one can treat the closed surface as being made of one or more finite (not vanishingly small) surface pieces on which the electric field is constant over the entire surface piece so that the flux can be calculated algebraically as EA or $EA\cos\theta$. After doing so for each of the finite surface pieces making up the closed surface, you add the results and you have the flux

$$\Phi_E = \oint \vec{E} \cdot \overrightarrow{dA}$$

through the surface. To get the charge enclosed by the surface, you just plug that into
3. In cases involving a symmetric charge distribution, Gauss's Law can be used to calculate the electric field due to the charge distribution. In such cases, the right choice of the Gaussian surface makes E a constant at all points on each of several surface pieces, and in some cases, zero on other surface pieces. In such cases the flux can be expressed as EA and one can simply solve $EA = \frac{Q_{\text{enclosed}}}{\epsilon_o}$ for E and use one's conceptual understanding of the electric field to get the direction of $\vec{E}$. The remainder of this chapter and all of the next will be used to provide examples of the kinds of charge distributions to which you will be expected to be able to apply this method.

Using Gauss's Law to Calculate the Electric Field in the Case of a Charge Distribution Having Spherical Symmetry

A spherically-symmetric charge distribution has a well-defined center. Furthermore, if you rotate a spherically-symmetric charge distribution through any angle, about any axis that passes through the center, you wind up with the exact same charge distribution. A uniform ball of charge is an example of a spherically-symmetric charge distribution. Before we consider that one, however, let's take up the case of the simplest charge distribution of them all, a point charge.

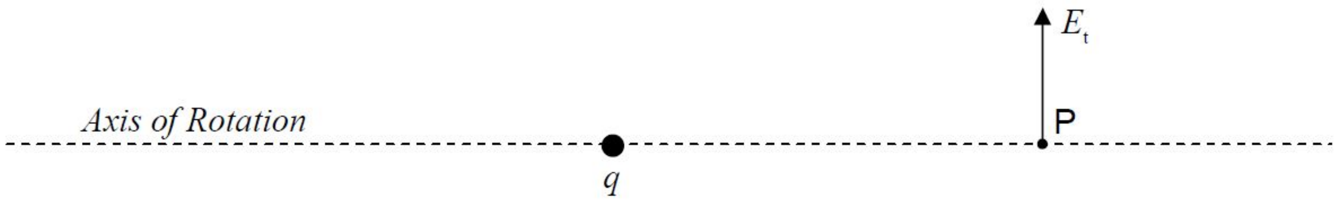
We use the symmetry of the charge distribution to find out as much as we can about the electric field and then we use Gauss's Law to do the rest. Now, when we rotate the charge distribution, we rotate the electric field with it. And, if a rotation of the charge distribution leaves you with the same exact charge distribution, then, it must also leave you with the same electric field.

We first prove that the electric field due to a point charge can have no tangential component by assuming that it does have a tangential component and showing that this leads to a contradiction.

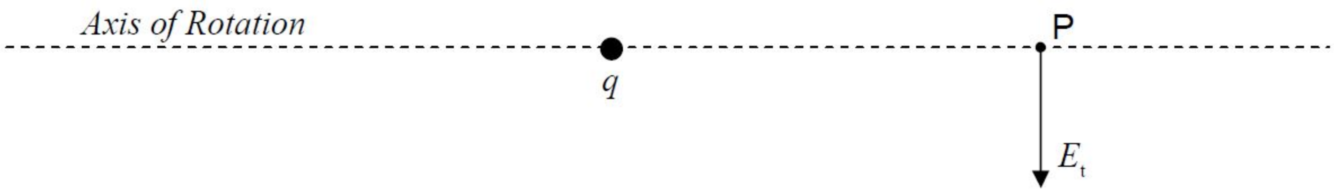
Here's our point charge q, and an assumed tangential component of the electric field at a point P which, from our perspective is to the right of the point charge.

(Note that a radial direction is any direction away from the point charge, and, a tangential direction is perpendicular to the radial direction.)

Now let's decide on a rotation axis for testing whether the electric field is symmetric with respect to rotation. Almost any will do. I choose one that passes through both the point charge, and, point P.



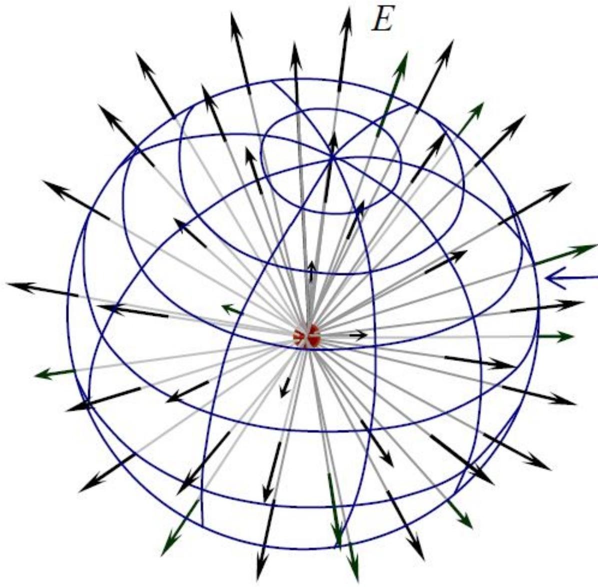
Now, if I rotate the charge, and its associated electric field, through an angle of 180° about that axis, I get:



This is different from the electric field that we started with. It is downward instead of upward. Hence the electric field cannot have the tangential component depicted at point P. Note that the argument does not depend on how far point P is from the point charge; indeed, I never specified the distance. So, no point to the right of our point charge can have an upward component to its electric field. In fact, if I assume the electric field at any point P' in space other than the point at which the charge is, to have a tangential component, then, I can adopt a viewpoint from which point P' appears to be to the right of the charge, and, the electric field appears to be upward. From that viewpoint, I can make the same rotation argument presented above to prove that the tangential component cannot exist. Thus, based on the spherical symmetry of the charge distribution, the electric field due to a point charge has to be strictly radial. Thus, at each point in space, the electric field must be either directly toward the point charge or directly away from it. Furthermore, again from symmetry, if the electric field is directly away from the point charge at one point in space, then it has to be directly away from the point charge at every point in space. Likewise, for the case in which it is directly toward the point charge at one point in space, the electric field has to be directly toward the point charge at every point in space.

We've boiled it down to a 50/50 choice. Let's assume that the electric field is directed away from the point charge at every point in space and use Gauss's Law to calculate the magnitude of the electric field. If the magnitude is positive, then the electric field is indeed directed away from the point charge. If the magnitude turns out to be negative, then the electric field is actually directed toward the point charge.

At this point we need to choose a Gaussian surface. To further exploit the symmetry of the charge distribution, we choose a Gaussian surface with spherical symmetry. More specifically, we choose a spherical shell of radius r, centered on the point charge.



At every point on the shell, the electric field, being radial, has to be perpendicular to the spherical shell. This means that for every area element, the electric field is parallel to our outward-directed area element vector $\overrightarrow{dA}$. This means that the $\vec{E} \cdot \overrightarrow{dA}$ in Gauss's Law,

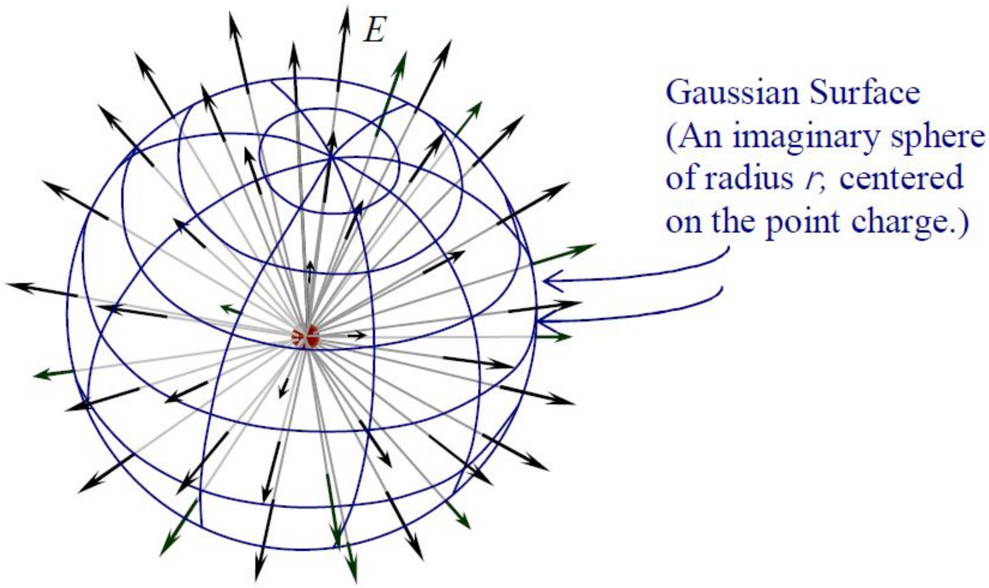
$$\oint \vec{E} \cdot \overrightarrow{dA} = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

evaluates to $E\, dA$. So, for the case at hand, Gauss's Law takes on the form:

$$\oint E dA = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

Furthermore, the magnitude of the electric field has to have the same value at every point on the shell. If it were different at a point P' on the spherical shell than it is at a point P on the spherical shell, then we could rotate the charge distribution about an axis through the point charge in such a manner as to bring the original electric field at point P' to position P. But this would represent a change in the electric field at point P, due to the rotation, in violation of the fact that a point charge has spherical symmetry. Hence, the electric field at any point P' on the Gaussian surface must have the same magnitude as the electric field at point P, which is what I set out to prove. The fact that E is a constant, in the integral, means that we can factor it out of the integral. So, for the case at hand, Gauss's Law takes on the form:

$$E \oint dA = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$



On the preceding page we arrived at $E \oint dA = \frac{Q_{\text{enclosed}}}{\epsilon_o}$.

Now $\oint dA$, the integral of dA over the Gaussian surface is the sum of all the area elements making up the Gaussian surface. That means that it is just the total area of the Gaussian surface. The Gaussian surface, being a sphere of radius r, has area $4\pi r^2$. So now, Gauss's Law for the case at hand looks like:

$$E 4\pi r^2 = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

Okay, we've left that right side alone for long enough. We're talking about a point charge q and our Gaussian surface is a sphere centered on that point charge q, so, the charge enclosed, Q_{enclosed} is obviously q. This yields:

$$E 4\pi r^2 = \frac{q}{\epsilon_o}$$

Solving for E gives us:

$$E = \frac{1}{4\pi\epsilon_o}\frac{q}{r^2}$$

This is positive when the charge q is positive, meaning that the electric field is directed outward, as per our assumption. It is negative when q is negative. So, when the charge q is negative, the electric field is directed inward, toward the charged particle. This expression is, of course, just Coulomb's Law for the electric field. It may look more familiar to you if we write it in terms of the Coulomb constant $k = \frac{1}{4\pi\epsilon_o}$ in which case our result for the outward electric field appears as:

$$E = \frac{kq}{r^2}$$

It's clear that, by means of our first example of Gauss's Law, we have derived something that you already know, the electric field due to a point charge.

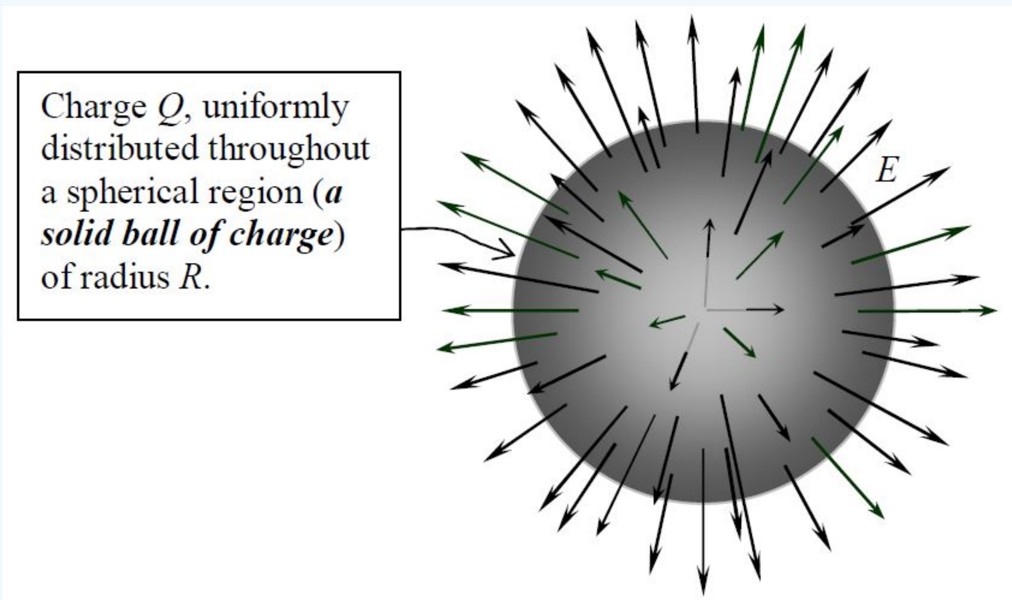
B34: Gauss's Law Example

We finished off the last chapter by using Gauss's Law to find the electric field due to a point charge. It was an example of a charge distribution having spherical symmetry. In this chapter we provide another example involving spherical symmetry.

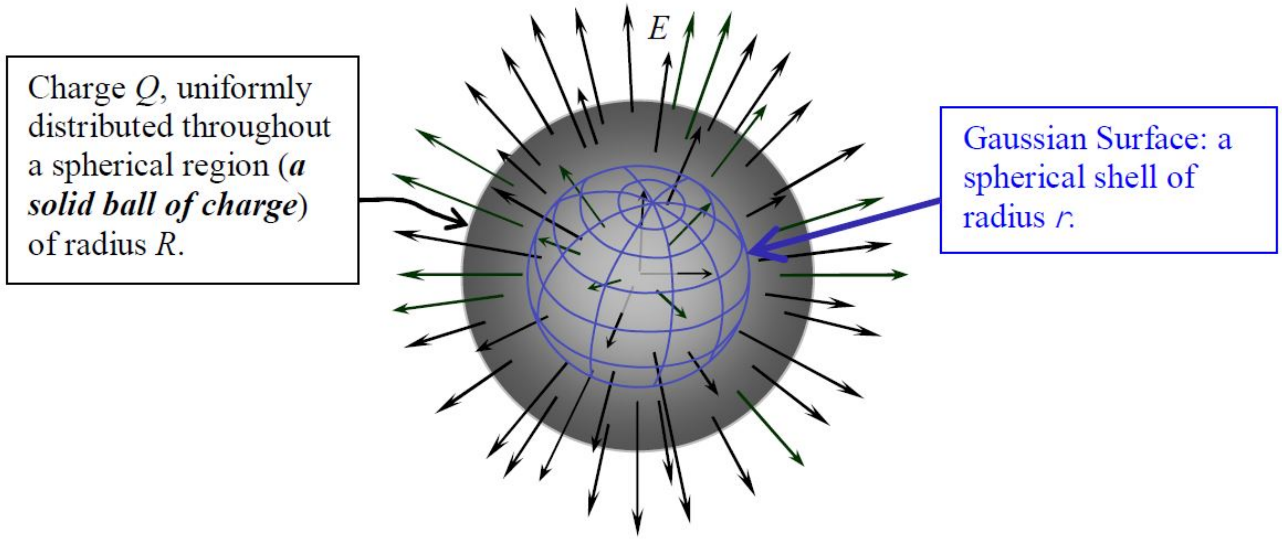
Find the electric field due to a uniform ball of charge of radius R and total charge Q. Express the electric field as a function of r, the distance from the center of the ball.

Solution

Again we have a charge distribution for which a rotation through any angle about any axis passing through the center of the charge distribution results in the exact same charge distribution. Thus, the same symmetry arguments used for the case of the point charge apply here with the result that, the electric field due to the ball of charge has to be strictly radially directed, and, the electric field has one and the same value at every point on any given spherical shell centered on the center of the ball of charge. Again, we assume the electric field to be outward-directed. If it turns out to be inward-directed, we'll simply get a negative value for the magnitude of the outward-directed electric field.



The appropriate Gaussian surface for any spherical charge distribution is a spherical shell centered on the center of the charge distribution.



Okay, let’s go ahead and apply Gauss’s Law.

$$\oint \vec{E} \cdot \overrightarrow{dA} = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

Since the electric field is radial, it is, at all points, perpendicular to the Gaussian Surface. In other words, it is parallel to the area element vector $\overrightarrow{dA}$. This means that the dot product $\vec{E} \cdot \overrightarrow{dA}$ is equal to the product of the magnitudes, EdA. This yields:

$$\oint EdA = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

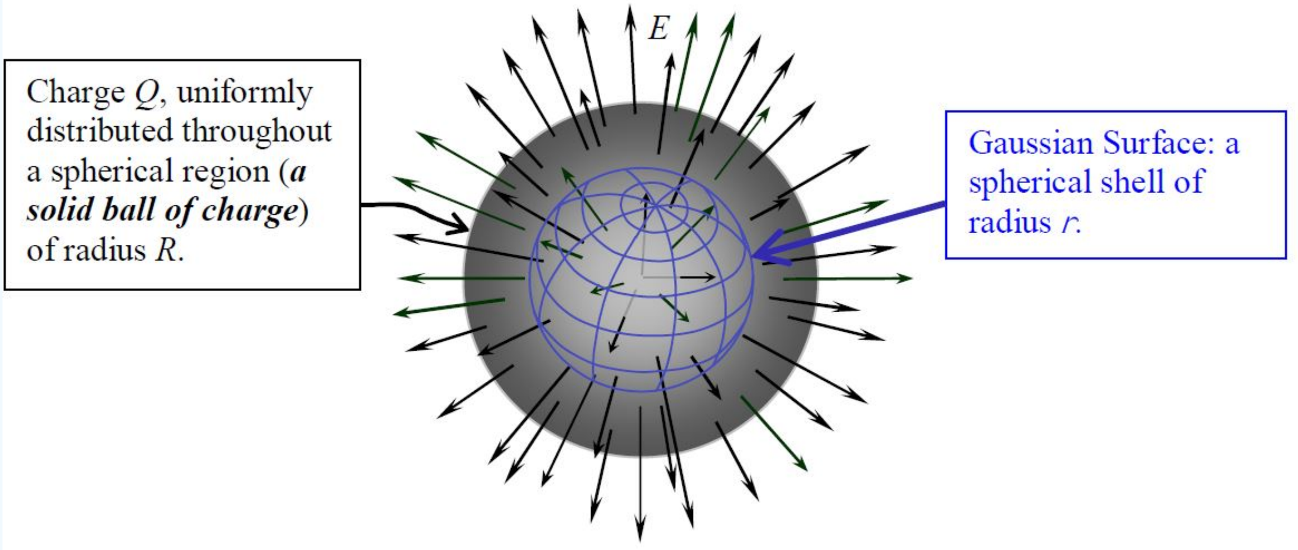
Again, since E has the same value at all points on the Gaussian surface of radius r, each dA in the infinite sum that the integral on the left is, is multiplied by the same value of E. Hence, we can factor the E out of the sum (integral). This yields

$$E \oint dA = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

The integral on the left is just the infinite sum of all the infinitesimal area elements making up the Gaussian surface, our spherical shell of radius r. The sum of all the area elements is, of course, the area of the spherical shell. The area of a sphere is $4\pi r^2$. So,

$$E 4\pi r^2 = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

Now the question is, how much charge is enclosed by our Gaussian surface of radius r?



There are two ways that we can get the value of the charge enclosed. Let's try it both ways and make sure we get one and the same result.

The first way: Because the charge is uniformly distributed throughout the volume, the amount of charge enclosed is directly proportional to the volume enclosed. So, the ratio of the amount of charge enclosed to the total charge, is equal to the ratio of the volume enclosed by the Gaussian surface to the total volume of the ball of charge:

$$\frac{Q_{\text{Enclosed}}}{Q} = \frac{\text{Volume of Gaussian Surface}}{\text{Volume of the Entire Ball of Charge}}$$

$$\frac{Q_{\text{Enclosed}}}{Q} = \frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi R^3}$$

$$Q_{\text{Enclosed}} = \frac{r^3}{R^3}Q$$

The second way: The other way we can look at it is to recognize that for a uniform distribution of charge, the amount of charge enclosed by the Gaussian surface is just the volume charge density, that is, the charge-per-volume ρ, times the volume enclosed.

$$Q_{\text{Enclosed}} = \rho\,(\text{Volume of the Gaussian surface})$$

$$Q_{\text{enclosed}} = \rho\frac{4}{3}\pi r^3$$

In this second method, we again take advantage of the fact that we are dealing with a uniform charge distribution. In a uniform charge distribution, the charge density is just the total charge divided by the total volume. Thus:

$$\rho = \frac{Q}{\text{Volume of Ball of Charge}}$$

$$\rho = \frac{Q}{\frac{4}{3}\pi R^3}$$

Substituting this in to our expression $Q_{\text{enclosed}} = \rho\, 4\pi r^2$ for the charge enclosed by the Gaussian surface yields:

$$Q_{\text{enclosed}} = \frac{Q}{\frac{4}{3}\pi R^3}\frac{4}{3}\pi r^3$$

$$Q_{\text{enclosed}} = \frac{r^3}{R^3}Q$$

which is indeed the same expression that we arrived at in solving for the charge enclosed the first way we talked about.

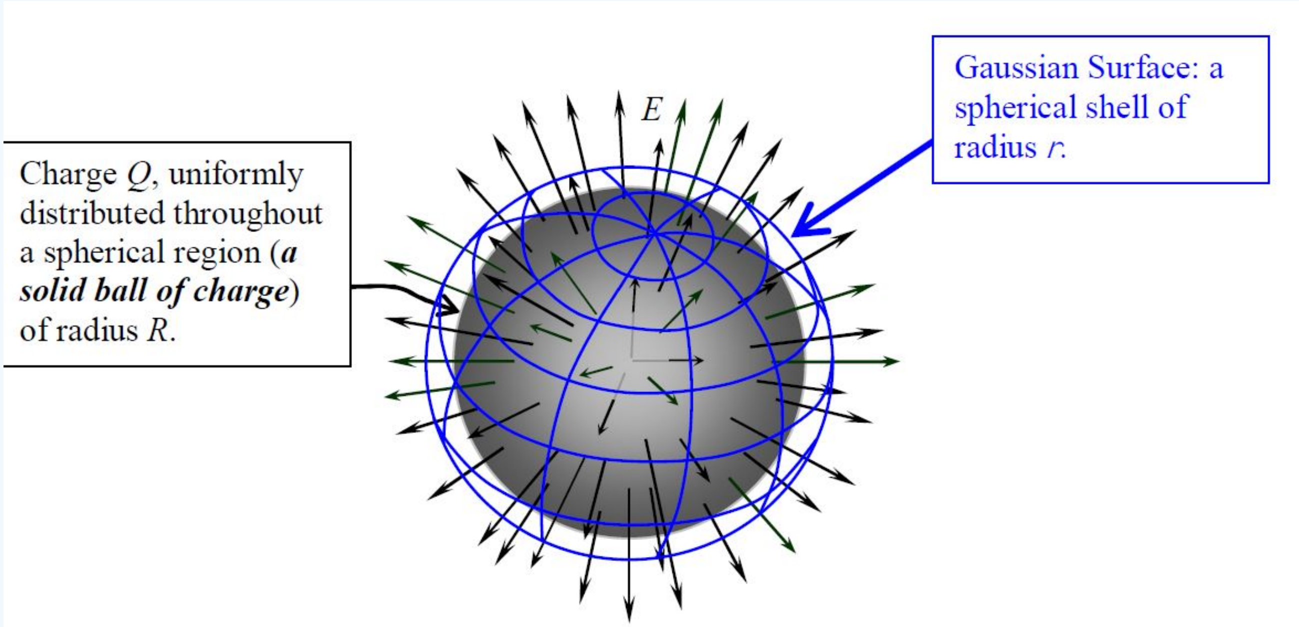
A couple of pages back we used Gauss's Law to arrive at the relation $E 4\pi r^2 = \frac{Q_{\text{enclosed}}}{\epsilon_o}$ and now we have something to plug in for Q_{enclosed}. Doing so yields:

$$E 4\pi r^2 = \frac{\left(\frac{r^3}{R^3}\right) Q}{\epsilon_o}$$

$$E = \frac{Q}{4\pi\epsilon_o R^3} r$$

This is our result for the magnitude of the electric field due to a uniform ball of charge at points inside the ball of charge $(r \le R)$. E is directly proportional to the distance from the center of the charge distribution. E increases with increasing distance because, the farther a point is from the center of the charge distribution, the more charge there is inside the spherical shell that is centered on the charge distribution and upon which the point in question is situated. How about points for which $r \ge R$?

If $r \ge R$,



the analysis is identical to the preceding analysis up to and including the point where we determined that:

$$E 4\pi r^2 = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

But as long as $r \ge R$, no matter by how much r exceeds R, all the charge in the spherical distribution of charge is enclosed by the Gaussian surface. "All the charge" is just Q the total amount of charge in the uniform ball of charge. So,

$$E 4\pi r^2 = \frac{Q}{\epsilon_o}$$

$$E = \frac{1}{4\pi\epsilon_o} \frac{Q}{r^2}$$

The constant $\frac{1}{4\pi\epsilon_o}$ is just the Coulomb constant k so we can write our result as:

$$E = \frac{kQ}{r^2}$$

This result looks just like Coulomb's Law for a point charge. What we've proved here is that, at points outside a spherically-symmetric charge distribution, the electric field is the same as that due to a point charge at the center of the charge distribution.

B35: Gauss's Law for the Magnetic Field and Ampere's Law Revisited

Gauss's Law for the Magnetic Field

Remember Gauss's Law for the electric field? It's the one that, in conceptual terms, states that the number of electric field lines poking outward through a closed surface is proportional to the amount of electric charge inside the closed surface. In equation form, we wrote it as:

$$\oint \vec{E} \cdot \overrightarrow{dA} = \frac{Q_{\text{enclosed}}}{\epsilon_o}$$

We called the quantity on the left the electric flux $\Phi_E = \oint \vec{E} \cdot \overrightarrow{dA}$.

Well, there is a Gauss's Law for the magnetic field as well. In one sense, it is quite similar because it involves a quantity called the magnetic flux which is expressed mathematically as $\Phi_B = \oint \vec{B} \cdot \overrightarrow{dA}$ and represents the number of magnetic field lines poking outward through a closed surface. The big difference stems from the fact that there is no such thing as "magnetic charge." In other words, there is no such thing as a magnetic monopole. In Gauss's Law for the electric field we have electric charge (divided by ϵ_o) on the right. In Gauss's Law for the magnetic field, we have 0 on the right:

$$\oint \vec{B} \cdot \overrightarrow{dA} = 0$$

As far as calculating the magnetic field, this equation is of limited usefulness. But, in conjunction with Ampere's Law in integral form (see below), it can come in handy for calculating the magnetic field in cases involving a lot of symmetry. Also, it can be used as a check for cases in which the magnetic field has been determined by some other means.

Ampere's Law

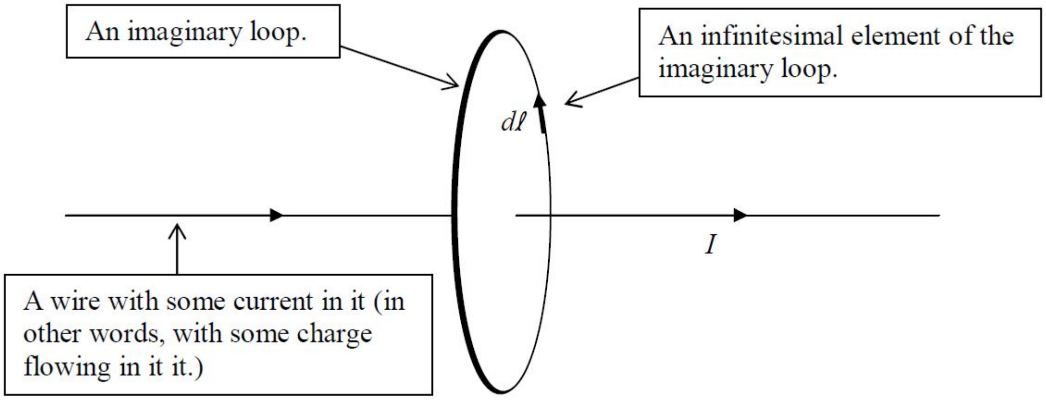
We've talked about Ampere's Law quite a bit already. It's the one that says a current causes a magnetic field. Note that this one says nothing about anything changing. It's just a cause and effect relation. The integral form of Ampere's Law is both broad and specific. It reads:

$$\oint \vec{B} \cdot \overrightarrow{d\ell} = \mu_o I_{THROUGH}$$

where:

- ○ the circle on the integral sign, and, $\overrightarrow{d\ell}$, the differential length, together, tell you that the integral (the infinite sum) is around an imaginary closed loop.
- $\vec{B}$ is the magnetic field,
- $\overrightarrow{d\ell}$ is an infinitesimal path element of the closed loop,
- μ_o is a universal constant called the magnetic permeability of free space, and
- $I_{THROUGH}$ is the current passing through the region enclosed by the loop.

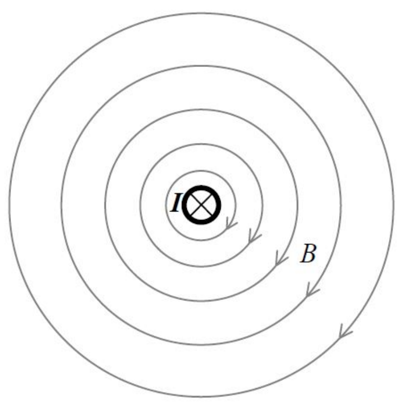
What Ampere's Law in integral form says is that, if you sum up the magnetic-field along-a-pathsegment times the length of the path segment for all the path segments making up an imaginary closed loop, you get the current through the region enclosed by the loop, times a universal constant. The integral $\oint \vec{B} \cdot \overrightarrow{d\ell}$ on whatever closed path upon which it is carried out, is called the circulation of the magnetic field on that closed path. So, another way of stating the integral form of Ampere's Law is to say that the circulation of the magnetic field on any closed path is directly proportional to the current through the region enclosed by the path. Here's the picture:



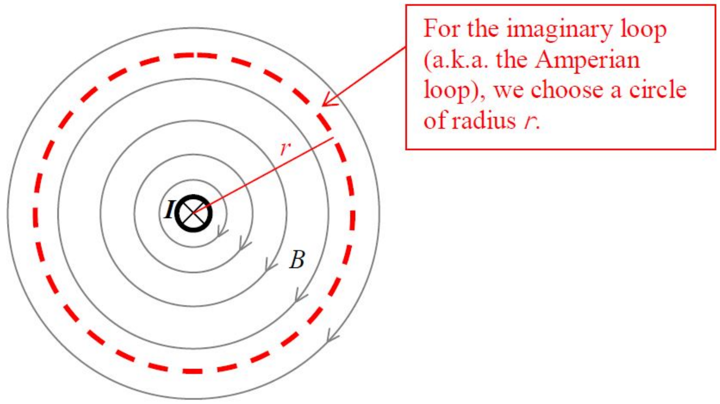
In the picture, I show everything except for the magnetic field. The idea is that, for each infinitesimal segment $\overrightarrow{d\ell}$ of the imaginary loop, you dot the magnetic field $\vec{B}$, at the position of the segment, into $\overrightarrow{d\ell}$. Add up all such dot products. The total is equal to μ_0 times the current I through the loop.

So, what's it good for? Ampere's Law in integral form is of limited use to us. It can be used as a great check for a case in which one has calculated the magnetic field due to some set of current-carrying conductors some other way (e.g. using the Biot-Savart Law, to be introduced in the next chapter). Also, in cases involving a high degree of symmetry, we can use it to calculate the magnetic field due to some current.

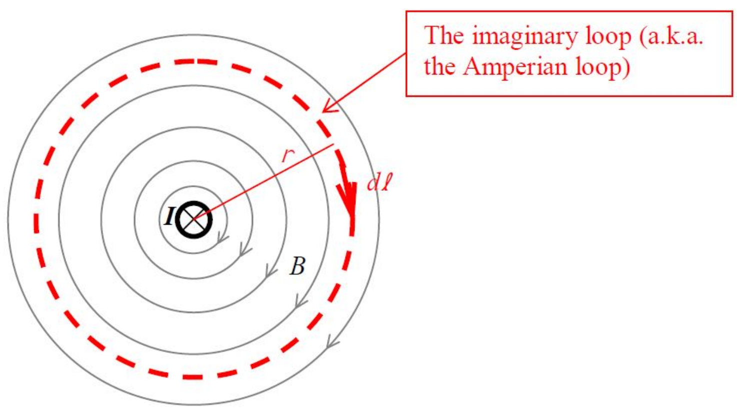
For example, we can use Ampere's Law to get a mathematical expression for the magnitude of the magnetic field due to an infinitely-long straight wire. I'm going to incorporate our understanding that, for a segment of wire with a current in it, the current creates a magnetic field which forms loops about the wire in accord with the right-hand rule for something curly something straight. In other words, we already know that for a long straight wire carrying current directly away from you, the magnetic field extends in loops about the wire, which, from your vantage point, are clockwise.



From symmetry, we can argue that the magnitude of the magnetic field is the same for a given point as it is at any other point that is the same distance from the wire as that given point. In implementing Ampere's Law, it is incumbent upon us to choose an imaginary loop, called an Amperian Loop in this context, that allows us to get some useful information from Ampere's Law. In this case, a circle whose plane is perpendicular to the straight wire and whose center lies on the straight wire is a smart choice.



At this point I want to share with you some directional information about the integral form of Ampere's Law. Regarding the $\overrightarrow{d\ell}$: each $\overrightarrow{d\ell}$ vector can, from a given point of view, be characterized as representing either a clockwise step along the path or a counterclockwise step along the path. And, if one is clockwise, they all have to be clockwise. If one is counterclockwise, they all have to be counterclockwise. Thus, in carrying out the integral around the closed loop, the traversal of the loop is either clockwise or counterclockwise from a specified viewpoint. Now, here's the critical direction information: Current that passes through the loop in that direction which relates to the sense (clockwise or counterclockwise) of loop traversal in accord with the right-hand rule for something curly something straight (with the loop being the something curly and the current being the something straight) is considered positive. So, for the case at hand, if I choose a clockwise loop traversal, as viewed from the vantage point that makes things look like:



then, the current I is considered positive. If you curl the fingers around the loop in the clockwise direction, your thumb points away from you. This means that current, through the loop that is directed away from you is positive. That is just the kind of current we have in the case at hand. So, when we substitute the I for the case at hand into the generic equation (Ampere's Law),

$$\oint \vec{B} \cdot \overrightarrow{d\ell} = \mu_o I_{THROUGH}$$

for the current $\mu_o I_{THROUGH}$ it goes in with a "+" sign.

$$\oint \vec{B} \cdot \overrightarrow{d\ell} = \mu_o I$$

Now, with the loop I chose, every $\overrightarrow{d\ell}$ is exactly parallel to the magnetic field $\vec{B}$ at the location of the $\overrightarrow{d\ell}$, so, $\vec{B} \cdot \overrightarrow{d\ell}$ is simply $B\, d\ell$. That is, with our choice of Amperian loop, Ampere's Law simplifies to:

$$\oint B d\ell = \mu_o I$$

Furthermore, from symmetry, with our choice of Amperian loop, the magnitude of the magnetic field B has one and the same value at every point on the loop. That means we can factor the magnetic field magnitude B out of the integral. This yields:

$$B \oint d\ell = \mu_o I$$

Okay, now we are on easy street. The $\oint d\ell$ is just the sum of all the $d\ell$'s making up our imaginary loop (a circle) of radius r. Hey, that's just the circumference of the circle $2\pi r$. So, Ampere's Law becomes:

$$B(2\pi r) = \mu_o I$$

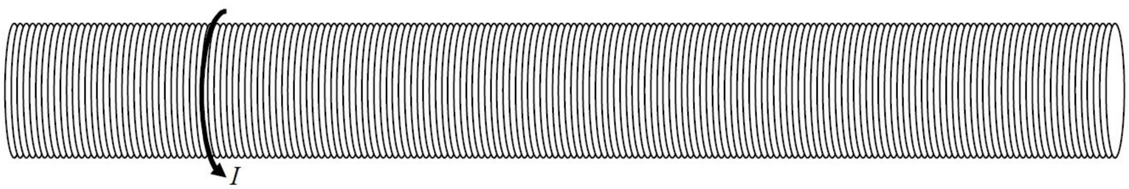
which means

$$B = \frac{\mu_o I}{2\pi r}$$

This is our end result. The magnitude of the magnetic field due to a long straight wire is directly proportional to the current in the wire and inversely proportional to the distance from the wire.

A Long Straight Solenoid

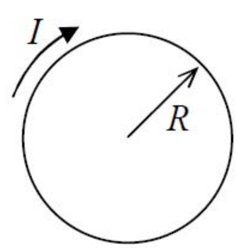
A solenoid is a coil of wire in the form of a cylindrical shell. The idealized solenoid that we consider here is infinitely long but, it has a fixed finite radius R and a constant finite current I.



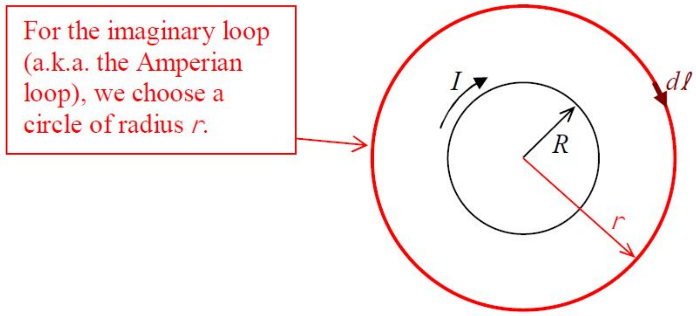
It is also characterized by its number-of-turns-per-length, n, where each "turn" (a.k.a. winding) is one circular current loop. In fact, we further idealize our solenoid by thinking of it as an infinite set of circular current loops. An actual solenoid approaches this idealized solenoid, but, in one turn (in the view above), the end of the turn is displaced left or right from the start of the turn by an amount equal to the diameter of the wire. As a result, in an actual solenoid, we have (in the view above) some left-to-right or right-to-left (depending on which way the wire wraps around) current. We neglect this current and consider the current to just go "round and round."

Our goal here is to find the magnetic field due to an ideal infinitely-long solenoid that has a number-of-turns-per-length n, has a radius R, and carries a current I.

We start by looking at the solenoid in cross section. Relative to the view above, we'll imagine looking at the solenoid from the left end. From that point of view, the cross section is a circle with clockwise current:



Let's try an Amperian loop in the shape of a circle, whose plane is perpendicular to the axis of symmetry of the solenoid, a circle that is centered on the axis of symmetry of the solenoid.

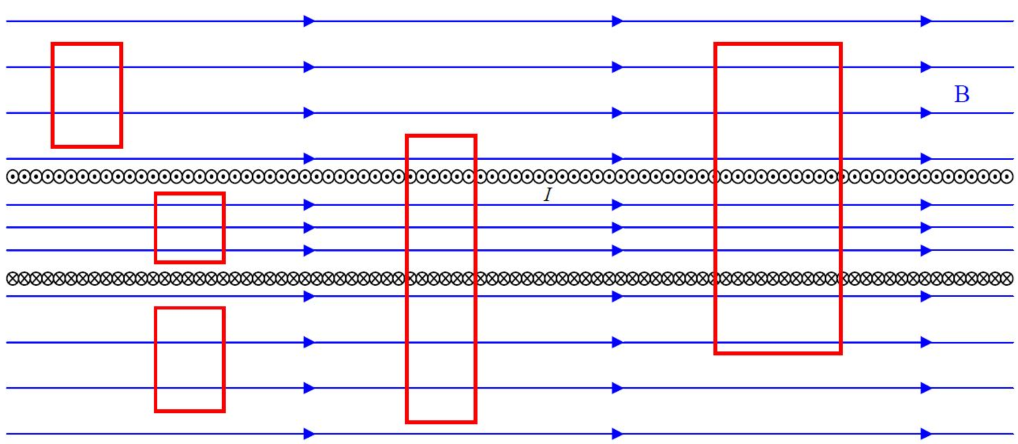


From symmetry, we can argue that if the magnetic field has a component parallel to the depicted $d\ell$, then it must have the exact same component for every $d\ell$ on the closed path. But this would make the circulation $\oint \vec{B} \cdot \overrightarrow{d\ell}$ non-zero in contradiction to the fact that no current passes through the region enclosed by the loop. This is true for any value of r. So, the magnetic field can have no component tangent to the circle whose plane is perpendicular to the axis of symmetry of the solenoid, a circle that is centered on the axis of symmetry of the solenoid.

Now suppose the magnetic field has a radial component. By symmetry it would have to be everywhere directed radially outward from the axis of symmetry of the solenoid, or everywhere radially inward. In either case, we could construct an imaginary cylindrical shell whose axis of symmetry coincides with that of the solenoid. The net magnetic flux through such a Gaussian surface would be non-zero in violation of Gauss's Law for the magnetic field. Hence the solenoid can have no radial magnetic field component.

The only kind of field that we haven't ruled out is one that is everywhere parallel to the axis of symmetry of the solenoid. Let's see if such a field would lead to any contradictions.

Here we view the solenoid in cross-section from the side. At the top of the coil, we see the current directed toward us, and, at the bottom, away. The possible longitudinal (parallel to the axis of symmetry of the solenoid) magnetic field is included in the diagram.



The rectangles in the diagram represent Amperian loops. The net current through any of the loops, in either direction (away from you or toward you) is zero. As such, the circulation $\oint \vec{B} \cdot \overrightarrow{d\ell}$ is zero. Since the magnetic field on the right and left of any one of the loops is perpendicular to the right and left sides of any one of the loops, it makes no contribution to the circulation there. By symmetry, the magnetic field at one position on the top of a loop is the same as it is at any other point on the top of the same loop. Hence, if we traverse any one of the loops counterclockwise (from our viewpoint) the contribution to the circulation is $-B_{TOP}L$ where L is the length of the top and bottom segments of whichever loop you choose to focus your attention on. The "-" comes from the fact that I have (arbitrarily) chosen to traverse the loop counterclockwise, and, in doing so, every $\overrightarrow{d\ell}$ in the top segment is in the opposite direction to the direction of the magnetic field at the top of the loop. The contribution to the circulation by the bottom segment of the same loop is $+B_{BOTTOM}L$. What we have so far is:

$$\oint \vec{B} \cdot \overrightarrow{d\ell} = \mu_o I_{THROUGH}$$

$$\oint \vec{B} \cdot \overrightarrow{d\ell} = 0$$

(where the net current through any one of the loops depicted is zero by inspection.)

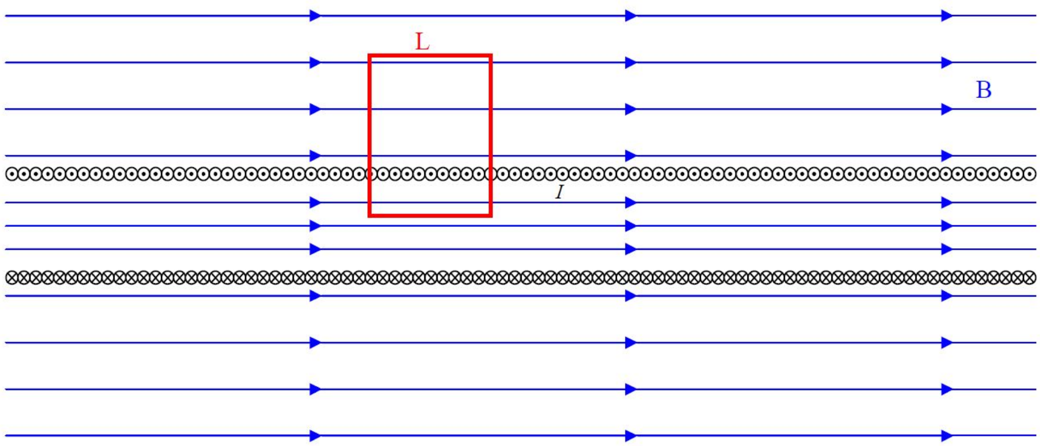
$$0 + -B_{TOP} L + 0 + B_{BOTTOM} L = 0$$

(with the two zeros on the left side of the equation being from the right and left sides of the loop where the magnetic field is perpendicular to the loop.)

Solving for B_{BOTTOM} we find that, for every loop in the diagram (and the infinite number of loops enclosing a net current of zero just like them):

$$B_{BOTTOM} = B_{TOP}$$

What this means is that the magnetic field at all points outside the solenoid has one and the same magnitude. The same can be said about all points inside the solenoid, but, the inside the solenoid value may be different from the outside value. In fact, let's consider a loop through which the net current is not zero:



Again, I choose to go counterclockwise around the loop (from our viewpoint). As such, by the right-hand rule for something curly something straight, current directed toward us through the loop is positive. Recalling that the number-of-turns-per-length-of-the-solenoid is n, we have, for the loop depicted above,

$$\oint \vec{B} \cdot \overrightarrow{d\ell} = \mu_o I_{THROUGH}$$

$$0 + -B_{TOP} L + 0 + B_{BOTTOM} L = \mu_o n L I$$

$$B_{BOTTOM} = B_{TOP} + \mu_o n I$$

The bottom of the loop is inside the solenoid and we have established that the magnitude of the magnetic field inside the solenoid has one and the same magnitude at all points inside the solenoid. I'm going to call that B_{INSIDE}, meaning that $B_{BOTTOM} = B_{INSIDE}$. Similarly, we have found that the magnitude of the magnetic field has one and the same (other) value at all points outside the solenoid. Let's call that $B_{OUTSIDE}$, meaning that $B_{TOP} = B_{OUTSIDE}$. Thus:

$$B_{INSIDE} = B_{OUTSIDE} + \mu_o n I$$

This is as far as I can get with Gauss's Law for the magnetic field, symmetry, and Ampere's Law alone. From there I turn to experimental results with long finite solenoids. Experimentally, we find that the magnetic field outside the solenoid is vanishingly small, and that there is an appreciable magnetic field inside the solenoid. Setting

$$B_{OUTSIDE} = 0$$

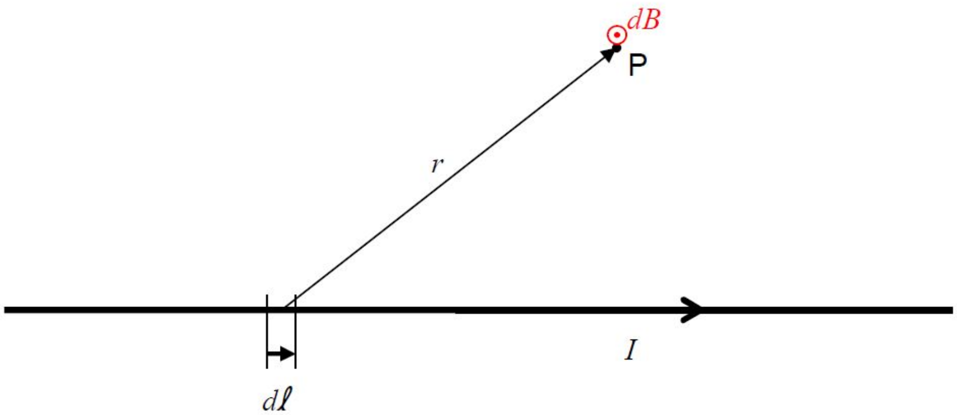
we find that the magnetic field inside a long straight solenoid is:

$$B_{INSIDE} = \mu_o n I$$

B36: The Biot-Savart Law

The Biot-Savart Law provides us with a way to find the magnetic field at an empty point in space, let's call it point P, due to current in wire. The idea behind the Biot-Savart Law is that each infinitesimal element of the current-carrying wire makes an infinitesimal contribution to the magnetic field at the empty point in space. Once you find each contribution, all you have to do is add them all up. Of course, there are an infinite number of contributions to the magnetic field at point P and each one is a vector, so, we are talking about an infinite sum of vectors. This business should seem familiar to you. You did this kind of thing when you were calculating the electric field back in *Chapter 30 The Electric Field Due to a Continuous Distribution of Charge on a Line*. The idea is similar, but here, of course, we are talking about magnetism.

The Biot-Savart Law gives the infinitesimal contribution to the magnetic field at point P due to an infinitesimal element of the current-carrying wire. The following diagram helps to illustrate just what the Biot-Savart Law tells us.



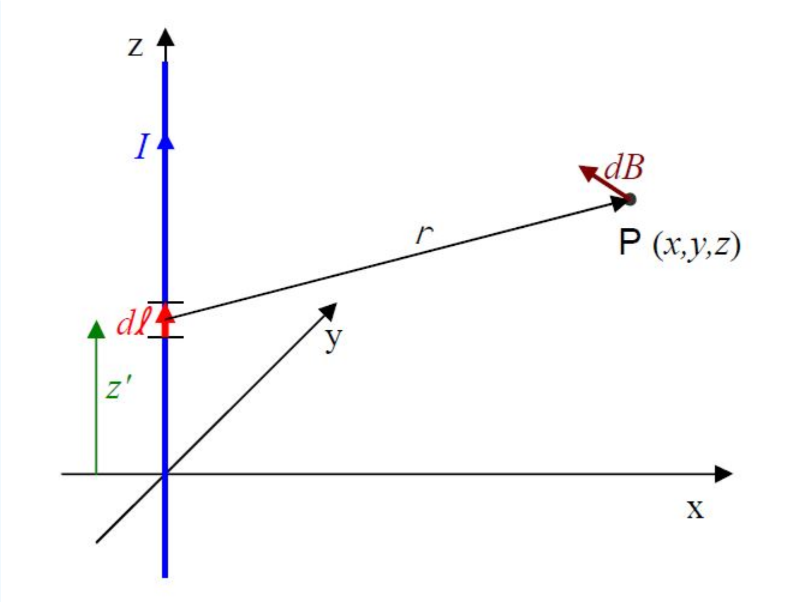
The Biot-Savart Law states that:

$$\overrightarrow{dB} = \frac{\mu_o}{4\pi} \frac{I\overrightarrow{dl} \times \vec{r}}{r^3}$$

The Biot-Savart Law represents a powerful straightforward method of calculating the magnetic field due to a current distribution.

Calculate the magnetic field due to a long straight wire carrying a current I along the z axis in the positive z direction. Treat the wire as extending to infinity in both directions.

Solution



Each infinitesimal element of the current-carrying conductor makes a contribution $\overrightarrow{dB}$ to the total magnetic field at point P.

The $\vec{r}$ vector extends from the infinitesimal element at $(0, 0, z')$ to point P at (x, y, z).

$$\vec{r} = (x\hat{i} + y\hat{j} + z\hat{k}) - z'\hat{k}$$

$$\vec{r} = x\hat{i} + y\hat{j} + (z - z')\hat{k}$$

The magnitude of $\vec{r}$ is thus:

$$r = \sqrt{x^2 + y^2 + (z - z')^2}$$

The $\overrightarrow{dl}$ vector points in the +z direction so it can be expressed as $\overrightarrow{dl} = dz'\hat{k}$

With these expressions for $\vec{r}$, r, and $\overrightarrow{dl}$ substituted into the Biot-Savart Law,

$$\overrightarrow{dB} = \frac{\mu_o}{4\pi} \frac{I\overrightarrow{dl} \times \vec{r}}{r^3}$$

we obtain:

$$\overrightarrow{dB} = \frac{\mu_o I}{4\pi} \frac{dz'\hat{k} \times \left(x\hat{i} + y\hat{j} + (z - z')\hat{k}\right)}{\left[x^2 + y^2 + (z - z')^2\right]^{3/2}}$$

$$\overrightarrow{dB} = \frac{\mu_o I}{4\pi} \frac{\left(x\hat{k} \times \hat{i} + y\hat{k} \times \hat{j} + (z - z')\hat{k} \times \hat{k}\right)}{\left[x^2 + y^2 + (z - z')^2\right]^{3/2}}$$

$$\overrightarrow{dB} = \frac{\mu_o I}{4\pi} \frac{dz'(x\hat{j} - y\hat{i})}{\left[x^2 + y^2 + (z - z')^2\right]^{3/2}}$$

$$\overrightarrow{dB} = -\frac{\mu_o I}{4\pi} y \frac{dz'}{\left[x^2 + y^2 + (z - z')^2\right]^{3/2}} \hat{i} + \frac{\mu_o I}{4\pi} x \frac{dz'}{\left[x^2 + y^2 + (z - z')^2\right]^{3/2}} \hat{j}$$

Let's work on this a component at a time. For the x component, we have:

$$dB_x = -\frac{\mu_o I}{4\pi} y \frac{dz'}{\left[x^2+y^2+(z-z')^2\right]^{3/2}}$$

Integrating over z' from $-\infty$ to ∞ yields:

$$B_x = -\frac{\mu_o I}{4\pi} y \int_{-\infty}^{\infty} \frac{dz'}{\left[x^2+y^2+(z-z')^2\right]^{3/2}}$$

I'm going to go with the following variable substitution:

$$u = z - z'$$

$$du = -dz', \text{ so, } dz' = du$$

Upper Limit: Evaluating $u = z - z'$ at $z' = \infty$ yields $-\infty$ for the upper limit of integration.

Lower Limit: Evaluating $u = z - z'$ at $z' = -\infty$ yields ∞ for the upper limit of integration.

So, our integral becomes:

$$B_x = -\frac{\mu_o I}{4\pi} y \int_{-\infty}^{\infty} \frac{-du}{(x^2+y^2+u^2)^{3/2}}$$

I choose to use one of the minus signs to interchange the limits of integration:

$$B_x = -\frac{\mu_o I}{4\pi} y \int_{-\infty}^{\infty} \frac{du}{(x^2+y^2+u^2)^{3/2}}$$

Using $\int \frac{dx}{(x^2+a^2)^{3/2}} = \frac{1}{a^2}\frac{x}{\sqrt{x^2+a^2}}$ from your formula sheet; and; identifying x^2+y^2 as a^2, and, u as the x appearing on the formula sheet, we obtain:

$$B_x = -\frac{\mu_o I}{4\pi} y \frac{1}{x^2+y^2} \frac{u}{\sqrt{u^2+x^2+y^2}}\Big|_{-\infty}^{\infty}$$

Now, I need to take the limit of that expression as u goes to ∞ and again as u goes to $-\infty$. To facilitate that, I want to factor a u out of the square root in the denominator. But, I have to be careful. The expression $\sqrt{u^2+x^2+y^2}$, which is equivalent to $\sqrt{(z-z')^2+x^2+y^2}$ is a distance. That means it is inherently positive, whether u (or z' for that matter) is positive or negative. So, when I factor u out of the square root, I'm going to have to use absolute value signs. For the denominator: $\sqrt{u^2+x^2+y^2} = \sqrt{u^2\left(1+\frac{x^2}{u^2}+\frac{y^2}{u^2}\right)} = |u|\sqrt{1+\frac{x^2}{u^2}+\frac{y^2}{u^2}}$, so,

$$B_x = -\frac{\mu_o I}{4\pi} y \frac{1}{x^2+y^2} \frac{u}{|u|} \frac{1}{\sqrt{1+\frac{x^2}{u^2}+\frac{y^2}{u^2}}}\Big|_{-\infty}^{\infty}$$

$$B_x = -\frac{\mu_o I}{4\pi} y \frac{1}{x^2+y^2} \left(1\frac{1}{\sqrt{1+0+0}} - -1\frac{1}{\sqrt{1+0+0}}\right)$$

$$B_x = -\frac{\mu_o I}{4\pi} y \frac{1}{x^2+y^2}(2)$$

$$B_x = -\frac{\mu_o I}{2\pi} y \frac{1}{x^2+y^2}$$

Now for the y component. Recall that we had:

$$\overrightarrow{dB} = -\frac{\mu_o I}{4\pi} y \frac{dz'}{\left[x^2+y^2+(z-z')^2\right]^{3/2}}\hat{i} + \frac{\mu_o I}{4\pi} y \frac{dz'}{\left[x^2+y^2+(z-z')^2\right]^{3/2}}\hat{j}$$

so,

$$dB_y = \frac{\mu_o I}{4\pi} x \frac{dz'}{\left[x^2+y^2+(z-z')^2\right]^{3/2}}$$

But, except for the replacement of $-y$ by x, this is the same expression that we had for dB_x. And those, (the $-y$ in the expression for dB_x and the x in the expression for dB_y), are, as far as the integration over z' goes, constants, out front. They don't affect the integration, they just "go along for the ride." So, we can use our B_x result for B_y if we just replace the $-y$, in our expression for B_x, with x. In other words, without having to go through the entire integration process again, we have:

$$B_y = \frac{\mu_o I}{2\pi} x \frac{1}{x^2+y^2}$$

Since we have no z component in our expression

$$\overrightarrow{dB} = -\frac{\mu_o I}{4\pi} y \frac{dz'}{\left[x^2+y^2+(z-z')^2\right]^{3/2}} \hat{i} + \frac{\mu_o I}{4\pi} y \frac{dz'}{\left[x^2+y^2+(z-z')^2\right]^{3/2}} \hat{j}$$

$\vec{B}$ itself must have no z component.

Substituting our results for $B_x, B_y,$ and B_z into $\hat{i}, \hat{j}, \hat{k}$ expression for $\vec{B}$, (Namely, $\vec{B} = B_x\hat{i} + B_y\hat{j} + B_z\hat{k}$), we have:

$$\vec{B} = -\frac{\mu_o I}{2\pi} y \frac{1}{x^2+y^2} \hat{i} + \frac{\mu_o I}{2\pi} x \frac{1}{x^2+y^2} \hat{j} + 0\hat{k}$$

$$\vec{B} = \frac{\mu_o I}{2\pi} \frac{1}{x^2+y^2} (-y\hat{i} + x\hat{j})$$

The quantity x^2+y^2 is just r^2, the square of the distance that point P is from the current carrying wire (recall that we are finding the magnetic field due to a wire, with a current I, that extends along the z axis from $-\infty$ to ∞)

$$\vec{B} = \frac{\mu_o I}{2\pi} \frac{1}{r^2} (-y\hat{i} + x\hat{j})$$

Furthermore, the vector $(-y\hat{i} + x\hat{j})$ has magnitude $\sqrt{(-y)^2+x^2} = \sqrt{x^2+y^2} = r$. Hence, the unit vector $\hat{u}_B$ in the same direction as $(-y\hat{i} + x\hat{j})$ is given by

$$\hat{u}_B = \frac{-y\hat{i} + x\hat{j}}{r} = -\frac{y}{r}\hat{i} + \frac{x}{r}\hat{j}$$

and, expressed as its magnitude times the unit vector in its direction, the vector $(-y\hat{i} + x\hat{j})$ can be written as:

$$(-y\hat{i} + x\hat{j}) = r\hat{u}_B$$

Substituting $(-y\hat{i} + x\hat{j}) = r\hat{u}_B$ into our expression $\vec{B} = \frac{\mu_o I}{2\pi} \frac{1}{r^2} (-y\hat{i} + x\hat{j})$ yields:

$$\vec{B} = \frac{\mu_o I}{2\pi} \frac{1}{r^2} r\hat{u}_B$$

$$\vec{B} = \frac{\mu_o I}{2\pi} \frac{1}{r} \hat{u}_B$$

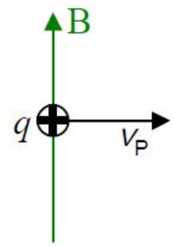
Note that the magnitude of $\vec{B}$ obtained here, namely $B = \frac{\mu_o I}{2\pi} \frac{1}{r}$, is identical to the magnitude obtained using the integral form of Ampere's Law. The direction $\hat{u}_B = -\frac{y}{r}\hat{i} + \frac{x}{r}\hat{j}$ for the magnetic field at any point P having coordinates (x, y, z), is also the same as, "the magnetic field extends in circles about that wire, in that sense of rotation (counterclockwise or clockwise) which is consistent with the right hand rule for something curly something straight with the something straight being the current and the something curly being the magnetic field."

B37: Maxwell's Equations

In this chapter, the plan is to summarize much of what we know about electricity and magnetism in a manner similar to the way in which James Clerk Maxwell summarized what was known about electricity and magnetism near the end of the nineteenth century. Maxwell not only organized and summarized what was known, but he added to the knowledge. From his work, we have a set of equations known as Maxwell's Equations. His work culminated in the discovery that light is electromagnetic waves.

In building up to a presentation of Maxwell's Equations, I first want to revisit ideas we encountered in chapter 20 and I want to start that revisit by introducing an easy way of relating the direction in which light is traveling to the directions of the electric and magnetic fields that are the light.

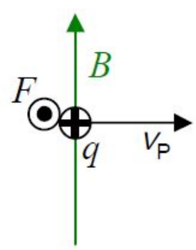
Recall the idea that a charged particle moving in a stationary magnetic field



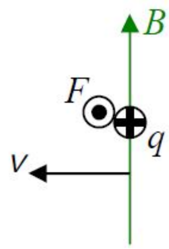
experiences a force given by

$$\vec{F} = q\vec{V}_p \times \vec{B}$$

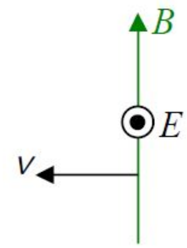
This force, by the way, is called the Lorentz Force. For the case depicted above, by the righthand rule for the cross product of two vectors, this force would be directed out of the page.



Viewing the exact same situation from the reference frame in which the charged particle is at rest we see a magnetic field moving sideways (with velocity $\vec{v} = -\vec{v}_p$) through the particle. Since we have changed nothing but our viewpoint, the particle is experiencing the same force.



We introduce a "middleman" by adopting the attitude that the moving magnetic field doesn't really exert a force on the charged particle, rather it causes an electric field which does that. For the force to be accounted for by this middleman electric field, the latter must be in the direction of the force. The existence of light indicates that the electric field is caused to exist whether or not there is a charged particle for it to exert a force on.



The bottom line is that wherever you have a magnetic field vector moving sideways through space you have an electric field vector, and, the direction of the velocity of the magnetic field vector is consistent with

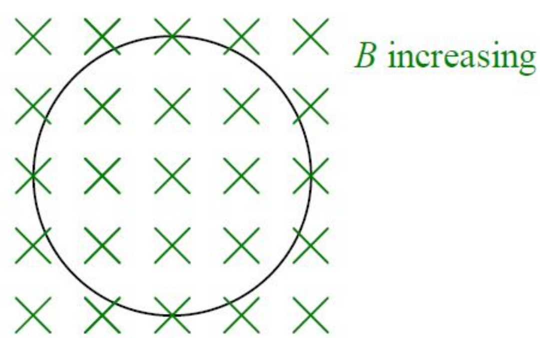
$$\text{direction of}\,\vec{V} = \text{direction of}\,\vec{E}\times\vec{B}.$$

You arrive at the same result for the case of an electric field moving sideways through space. (Recall that in chapter 20, we discussed the fact that an electric field moving sideways through space causes a magnetic field.)

The purpose of this brief review of material from chapter 20 was to arrive at the result $\text{direction of}\,\vec{V} = \text{direction of}\,\vec{E}\times\vec{B}$. This direction relation will come in handy in our discussion of two of the four equations known as Maxwell's Equations.

One of Maxwell's Equations is called Faraday's Law. It brings together a couple of things we have already talked about, namely, the idea that a changing number of magnetic field lines through a loop or a coil induces a current in that loop or coil, and, the idea that a magnetic field vector that is moving sideways through a point in space causes an electric field to exist at that point in space. The former is a manifestation of the latter. For instance, suppose you have an increasing number of downward directed magnetic field lines through a horizontal loop. The idea is that for the number of magnetic field lines through the loop to be increasing, there must be magnetic field lines moving sideways through the conducting material of the loop (to get inside the perimeter of the loop). This causes an electric field in the conducting material of the loop which in turn pushes on the charged particles of the conducting material of the loop and thus results in a current in the loop. We can discuss the production of the electric field at points in space occupied by the conducting loop even if the conducting loop is not there. If we consider an imaginary loop in its place, the magnetic field lines moving through it to the interior of the loop still produce an electric field in the loop; there are simply no charges for that field to push around the loop.

Suppose we have an increasing number of downward-directed magnetic field lines through an imaginary loop. Viewed from above the situation appears as:

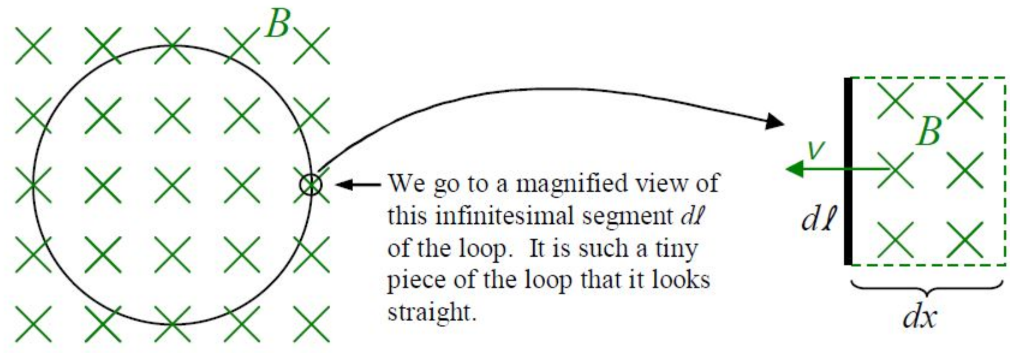


The big idea here is that you can't have an increasing number of downward-directed magnetic field lines through the region encircled by the imaginary loop without having, either, downward directed magnetic field lines moving transversely and *inward* through the loop into the region encircled by the loop, or, upward-directed magnetic field lines moving transversely and *outward* through the loop out of the region encircled by the loop. Either way you have magnetic field lines cutting through the loop and with each magnetic field cutting through the loop there has to be an associated *electric* field with a component tangent to the loop. Our technical expression for the "number of magnetic field lines through the loop" is the magnetic flux, given, in the case of a uniform (but time-varying) magnetic field by

$$\Phi_B = \vec{B}\cdot\vec{A}$$

where A is the area of the region encircled by the loop.

Faraday's Law, as it appears in Maxwell's Equations, is a relation between the rate of change of the magnetic flux through the loop and the electric field (produced by this changing flux) in the loop. To arrive at it, we consider an infinitesimal segment dl of the loop and the infinitesimal contribution to the rate of change of the magnetic flux through the loop resulting from magnetic field lines moving through that segment dl into the region encircled by the loop.



If the magnetic field depicted above is moving sideways toward the interior of the loop with a speed $v = \frac{dx}{dt}$ then all the magnetic field lines in the region of area $A = dl\, dx$, will, in time dt, move leftward a distance dx. That is, they will all move from outside the loop to inside the loop creating a change of flux, in time dt, of

$$d\phi_B = B\, dA$$

$$d\phi_B = B\, dl\, dx$$

Now, if I divide both sides of this equation by the time dt in which the change occurs, we have

$$\frac{d\phi_B}{dt} = B\, dl \frac{dx}{dt}$$

which I can write as

$$\dot{\phi_B} = B\, dl\, v$$

or

$$\dot{\phi_B} = v\ \ B\, dl$$

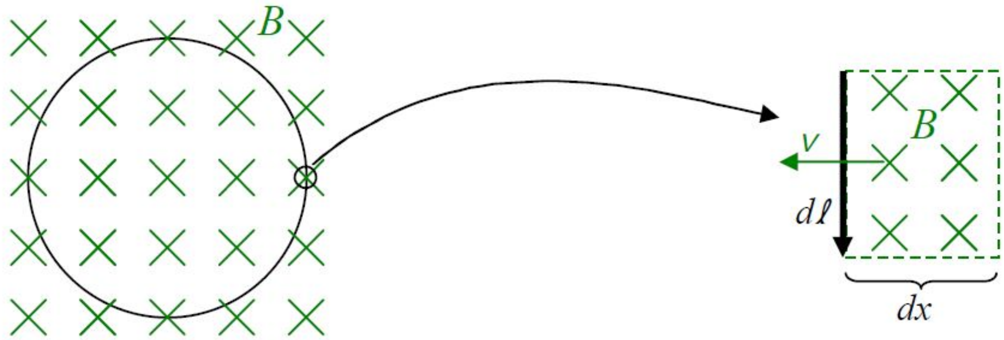
For the case at hand, looking at the diagram, we see that $\vec{B}$ and $\vec{v}$ are at right angles to each other so the magnitude of $\vec{v} \times \vec{B}$ is just vB. In that case, since $\vec{E} = -\vec{v} \times \vec{B}$ (from equation 20-1 with $-\vec{v}$ in place of $\vec{v}_P$), we have $E = vB$. Replacing the product vB appearing on the right side of equation 37-1 $(p\dot{h}i_B = vB\, dl)$ yields $\dot{\phi}_B = Edl$ which I copy at the top of the following page:

$$\dot{\phi}_B = Edl$$

We can generalize this to the case where the velocity vector $\vec{v}$ is not perpendicular to the infinitesimal loop segment in which case $\vec{E}$ is not along $\overrightarrow{dl}$. In that case the component of $\vec{E}$ that is along $\overrightarrow{dl}$, times the length dl itself, is just $\vec{E} \cdot \overrightarrow{dl}$ and our equation becomes

$$\dot{\phi}_B = -\vec{E} \cdot \overrightarrow{dl}$$

In this expression, the direction of $\overrightarrow{dl}$ is determined once one decides on which of the two directions in which a magnetic field line can extend through the region enclosed by the loop is defined to make a positive contribution to the flux through the loop. The direction of $\overrightarrow{dl}$ is then the one which relates the sense in which $\overrightarrow{dl}$ points around the loop, to the positive direction for magnetic field lines through the loop, by the right hand rule for something curly something straight. With this convention the minus sign is needed to make the dot product have the same sign as the sign of the ongoing change in flux. Consider for instance the case depicted in the diagram:



We are looking at a horizontal loop from above. Downward is depicted as into the page. Calling downward the positive direction for flux makes clockwise, as viewed from above, the positive sense for the $\overrightarrow{dl}$'s in the loop, meaning the $\overrightarrow{dl}$ on the right side of the loop is pointing toward the bottom of the page (as depicted). For a downward-directed magnetic field moving leftward into the loop, $\vec{E}$ must be directed toward the top of the page (from $\text{direction of}\, \vec{v} = \text{direction of}\, \vec{E} \times \vec{B}$). Since $\vec{E}$ is in the opposite direction to that of $\overrightarrow{dl}$, $\vec{E} \cdot \overrightarrow{dl}$ must be negative. But movement of downward-directed magnetic field lines into the region encircled by the loop, what with downward being considered the positive direction for flux, means a positive rate of change of flux. The left side of $\dot{\phi}_B = -\vec{E} \cdot \overrightarrow{dl}$ is thus positive. With $\vec{E} \cdot \overrightarrow{dl}$ being negative, we need the minus sign in front of it to make the right side positive too. Now $\dot{\phi}_B$ is the rate of change of magnetic flux through the region encircled by the loop due to the magnetic field lines that are entering that region through the one infinitesimal $\overrightarrow{dl}$ that we have been considering. There is a $\dot{\phi}_B$ for each infinitesimal $\overrightarrow{dl}$ making up the loop. Thus there are an infinite number of them. Call the infinite sum of all the $\dot{\phi}_B$'s $\dot{\Phi}_B$ and our equation becomes:

Index

A

Ampere’s Law
B35: Gauss’s Law for the Magnetic Field and Ampere’s Law Revisited

Archimedes' Principle
33A: Fluids: Pressure, Density, Archimedes' Principle

B

beats
32A: Beats and the Doppler Effect

Bernoulli’s principle
34A: Pascal’s Principle, the Continuity Equation, and Bernoulli’s Principle

C

Capacitance
B8: Capacitors, Dielectrics, and Energy in Capacitors

capacitor
B8: Capacitors, Dielectrics, and Energy in Capacitors

centripetal acceleration
18A: Circular Motion - Centripetal Acceleration

charging the capacitor
B8: Capacitors, Dielectrics, and Energy in Capacitors

circuit
9B: Electric Current, EMF, and Ohm's Law

Circular motion
18A: Circular Motion - Centripetal Acceleration
20A: Torque & Circular Motion

conductor
B4: Conductors and the Electric Field

Conservation of angular momentum
5A: Conservation of Angular Momentum

Continuity equation
34A: Pascal’s Principle, the Continuity Equation, and Bernoulli’s Principle

continuity principle
34A: Pascal’s Principle, the Continuity Equation, and Bernoulli’s Principle

Coulomb's Law
B1: Charge & Coulomb's Law

Cross product
21A: Vectors - The Cross Product & Torque

current
9B: Electric Current, EMF, and Ohm's Law

D

dielectric
B8: Capacitors, Dielectrics, and Energy in Capacitors

doppler effect
32A: Beats and the Doppler Effect

E

Elastic collision
4A: Conservation of Momentum

electric field
B2: The Electric Field - Description and Effect

F

farad (Units)
B8: Capacitors, Dielectrics, and Energy in Capacitors

Faraday’s Law
B18: Faraday's Law and Lenz's Law

first law of thermodynamics
37A: The First Law of Thermodynamics

fluid
33A: Fluids: Pressure, Density, Archimedes' Principle

freefall
13A: Freefall, a.k.a. Projectile Motion

G

Gauss's law
B35: Gauss’s Law for the Magnetic Field and Ampere’s Law Revisited

H

heat
35A: Temperature, Internal Energy, Heat and Specific Heat Capacity

Heat capacity
35A: Temperature, Internal Energy, Heat and Specific Heat Capacity

I

inelastic
4A: Conservation of Momentum

internal energy
35A: Temperature, Internal Energy, Heat and Specific Heat Capacity

K

Kirchhoff’s First Rule
B12: Kirchhoff’s Rules, Terminal Voltage

Kirchhoff’s Second Rule
B12: Kirchhoff’s Rules, Terminal Voltage

L

Lenz's Law
B18: Faraday's Law and Lenz's Law

loop rule
B12: Kirchhoff’s Rules, Terminal Voltage

M

Maxwell's equations
B37: Maxwell’s Equations

Moment of Inertia
5A: Conservation of Angular Momentum

O

Ohm’s law
9B: Electric Current, EMF, and Ohm's Law

P

Pascal’s Principle
34A: Pascal’s Principle, the Continuity Equation, and Bernoulli’s Principle

power
B11: Resistivity and Power

Projectile motion
13A: Freefall, a.k.a. Projectile Motion

R

ray tracing
B28: Thin Lenses - Ray Tracing

RC circuit
B13: RC Circuit

resistivity
B11: Resistivity and Power

rotational kinetic energy
3A: Conservation of Mechanical Energy II: Springs, Rotational Kinetic Energy

S

specific heat
35A: Temperature, Internal Energy, Heat and Specific Heat Capacity

T

Temperature
35A: Temperature, Internal Energy, Heat and Specific Heat Capacity

Torque
5A: Conservation of Angular Momentum
20A: Torque & Circular Motion
21A: Vectors - The Cross Product & Torque

Glossary

absolute pressure | sum of gauge pressure and atmospheric pressure [OpenStax]

acceleration due to gravity | acceleration of an object as a result of gravity [OpenStax]

acceleration vector | instantaneous acceleration found by taking the derivative of the velocity function with respect to time in unit vector notation [OpenStax]

accuracy | the degree to which a measured value agrees with an accepted reference value for that measurement [OpenStax]

Achimedes' principle | buoyant force on an object equals the weight of the fluid it displaces [OpenStax]

action-at-a-distance force | type of force exerted without physical contact [OpenStax]

amplitude (A) | maximum displacement from the equilibrium position of an object oscillating around the equilibrium position [OpenStax]

angular acceleration | time rate of change of angular velocity [OpenStax]

angular frequency | ω, rate of change of an angle with which an object that is moving on a circular path [OpenStax]

angular momentum | rotational analog of linear momentum, found by taking the product of moment of inertia and angular velocity [OpenStax]

angular position | angle a body has rotated through in a fixed coordinate system [OpenStax]

angular velocity | time rate of change of angular position [OpenStax]

anticommutative property | change in the order of operation introduces the minus sign [OpenStax]

antinode | location of maximum amplitude in standing waves [OpenStax]

antiparallel vectors | two vectors with directions that differ by 180° [OpenStax]

aphelion | farthest point from the Sun of an orbiting body; the corresponding term for the Moon's farthest point from Earth is the apogee [OpenStax]

apparent weight | reading of the weight of an object on a scale that does not account for acceleration [OpenStax]

associative | terms can be grouped in any fashion [OpenStax]

average acceleration | the rate of change in velocity; the change in velocity over time [OpenStax]

average power | work done in a time interval divided by the time interval [OpenStax]

average speed | the total distance traveled divided by elapsed time [OpenStax]

average velocity | the displacement divided by the time over which displacement occurs [OpenStax]

banked curve | curve in a road that is sloping in a manner that helps a vehicle negotiate the curve [OpenStax]

base quantity | physical quantity chosen by convention and practical considerations such that all other physical quantities can be expressed as algebraic combinations of them [OpenStax]

base unit | standard for expressing the measurement of a base quantity within a particular system of units; defined by a particular procedure used to measure the corresponding base quantity [OpenStax]

beat frequency | frequency of beats produced by sound waves that differ in frequency [OpenStax]

beats | constructive and destructive interference of two or more frequencies of sound [OpenStax]

Bernoulli's equation | equation resulting from applying conservation of energy to an incompressible frictionless fluid:

$$p + \frac{1}{2}\rho v^2 + \rho g h = constant,$$

throughout the fluid [OpenStax]

Bernoulli's principle | Bernoulli's equation applied at constant depth:

$$p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2$$

[OpenStax]

black hole | mass that becomes so dense, that it collapses in on itself, creating a singularity at the center surround by an event horizon [OpenStax]

bow wake | v-shaped disturbance created when the wave source moves faster than the wave propagation speed [OpenStax]

breaking stress (ultimate stress) | value of stress at the fracture point [OpenStax]

bulk modulus | elastic modulus for the bulk stress [OpenStax]

bulk strain (or volume strain) | strain under the bulk stress, given as fractional change in volume [OpenStax]

bulk stress (or volume stress) | stress caused by compressive forces, in all directions [OpenStax]

buoyant force | net upward force on any object in any fluid due to the pressure difference at different depths [OpenStax]

center of gravity | point where the weight vector is attached [OpenStax]

center of mass | weighted average position of the mass [OpenStax]

centripetal acceleration | component of acceleration of an object moving in a circle that is directed radially inward toward the center of the circle [OpenStax]

centripetal force | any net force causing uniform circular motion [OpenStax]

closed system | system for which the mass is constant and the net external force on the system is zero [OpenStax]

commutative | operations can be performed in any order [OpenStax]

component form of a vector | a vector written as the vector sum of its components in terms of unit vectors [OpenStax]

compressibility | reciprocal of the bulk modulus [OpenStax]

compressive strain | strain that occurs when forces are contracting an object, causing its shortening [OpenStax]

compressive stress | stress caused by compressive forces, only in one direction [OpenStax]

conservative force | force that does work independent of path [OpenStax]

conserved quantity | one that cannot be created or destroyed, but may be transformed between different forms of itself [OpenStax]

constructive interference | when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs [OpenStax]

conversion factor | a ratio that expresses how many of one unit are equal to another unit [OpenStax]

Coriolis force | inertial force causing the apparent deflection of moving objects when viewed in a rotating frame of reference [OpenStax]

corkscrew right-hand rule | a rule used to determine the direction of the vector product [OpenStax]

critically damped | condition in which the damping of an oscillator causes it to return as quickly as possible to its equilibrium position without oscillating back and forth about this position [OpenStax]

cross product | the result of the vector multiplication of vectors is a vector called a cross product; also called a vector product [OpenStax]

density | mass per unit volume of a substance or object [OpenStax]

derived quantity | physical quantity defined using algebraic combinations of base quantities [OpenStax]

derived units | units that can be calculated using algebraic combinations of the fundamental units [OpenStax]

destructive interference | when two identical waves arrive at the same point exactly out of phase; that is, precisely aligned crest to trough [OpenStax]

difference of two vectors | vector sum of the first vector with the vector antiparallel to the second [OpenStax]

dimension | expression of the dependence of a physical quantity on the base quantities as a product of powers of symbols representing the base quantities; in general, the dimension of a quantity has the form $L^a M^b T^c I^d \Theta^e N^f J^g$ for some powers a, b, c, d, e, f, and g [OpenStax]

dimensionally consistent | equation in which every term has the same dimensions and the arguments of any mathematical functions appearing in the equation are dimensionless [OpenStax]

dimensionless | quantity with a dimension of $L^0 M^0 T^0 I^0 \Theta^e N^0 J^0 = 1$; also called quantity of dimension 1 or a pure number [OpenStax]

direction angle | in a plane, an angle between the positive direction of the x-axis and the vector, measured counterclockwise from the axis to the vector [OpenStax]

discrepancy | the difference between the measured value and a given standard or expected value [OpenStax]

displacement | the change in position of an object [OpenStax]

displacement | change in position [OpenStax]

displacement vector | vector from the initial position to a final position on a trajectory of a particle [OpenStax]

distance traveled | the total length of the path traveled between two positions [OpenStax]

distributive | multiplication can be distributed over terms in summation [OpenStax]

Doppler effect | alteration in the observed frequency of a sound due to motion of either the source or the observer [OpenStax]

Doppler shift | actual change in frequency due to relative motion of source and observer [OpenStax]

dot product | the result of the scalar multiplication of two vectors is a scalar called a dot product; also called a scalar product [OpenStax]

drag force | force that always opposes the motion of an object in a fluid; unlike simple friction, the drag force is proportional to some function of the velocity of the object in that fluid [OpenStax]

dynamics | study of how forces affect the motion of objects and systems [OpenStax]

elapsed time | the difference between the ending time and the beginning time [OpenStax]

elastic | object that comes back to its original size and shape when the load is no longer present [OpenStax]

elastic | collision that conserves kinetic energy [OpenStax]

elastic limit | stress value beyond which material no longer behaves elastically and becomes permanently deformed [OpenStax]

elastic modulus | proportionality constant in linear relation between stress and strain, in SI pascals [OpenStax]

elastic potential energy | potential energy stored as a result of deformation of an elastic object, such as the stretching of a spring [OpenStax]

energy conservation | total energy of an isolated system is constant [OpenStax]

English units | system of measurement used in the United States; includes units of measure such as feet, gallons, and pounds [OpenStax]

equal vectors | two vectors are equal if and only if all their corresponding components are equal; alternately, two parallel vectors of equal magnitudes [OpenStax]

equilibrium | body is in equilibrium when its linear and angular accelerations are both zero relative to an inertial frame of reference [OpenStax]

equilibrium point | position where the assumed conservative, net force on a particle, given by the slope of its potential energy curve, is zero [OpenStax]

equilibrium position | position where the spring is neither stretched nor compressed [OpenStax]

escape velocity | initial velocity an object needs to escape the gravitational pull of another; it is more accurately defined as the velocity of an object with zero total mechanical energy [OpenStax]

estimation | using prior experience and sound physical reasoning to arrive at a rough idea of a quantity's value; sometimes called an "order-of-magnitude approximation," a "guesstimate," a "back-of-the-envelope calculation", or a "Fermi calculation" [OpenStax]

event horizon | location of the Schwarzschild radius and is the location near a black hole from within which no object, even light, can escape [OpenStax]

exact differential | is the total differential of a function and requires the use of partial derivatives if the function involves more than one dimension [OpenStax]

explosion | single object breaks up into multiple objects; kinetic energy is not conserved in explosions [OpenStax]

external force | force applied to an extended object that changes the momentum of the extended object as a whole [OpenStax]

external force | force acting on an object or system that originates outside of the object or system [OpenStax]

first equilibrium condition | expresses translational equilibrium; all external forces acting on the body balance out and their vector sum is zero [OpenStax]

fixed boundary condition | when the medium at a boundary is fixed in place so it cannot move [OpenStax]

flow rate | abbreviated Q, it is the volume V that flows past a particular point during a time t, or $Q = \frac{dV}{dt}$ [OpenStax]

fluids | liquids and gases; a fluid is a state of matter that yields to shearing forces [OpenStax]

force | push or pull on an object with a specific magnitude and direction; can be represented by vectors or expressed as a multiple of a standard force [OpenStax]

force constant (k) | characteristic of a spring which is defined as the ratio of the force applied to the spring to the displacement caused by the force [OpenStax]

free boundary condition | exists when the medium at the boundary is free to move [OpenStax]

free fall | situation in which the only force acting on an object is gravity [OpenStax]

free fall | the state of movement that results from gravitational force only [OpenStax]

free-body diagram | sketch showing all external forces acting on an object or system; the system is represented by a single isolated point, and the forces are represented by vectors extending outward from that point [OpenStax]

frequency (f) | number of events per unit of time [OpenStax]

friction | force that opposes relative motion or attempts at motion between systems in contact [OpenStax]

fundamental | the lowest-frequency resonance [OpenStax]

fundamental frequency | lowest frequency that will produce a standing wave [OpenStax]

gauge pressure | pressure relative to atmospheric pressure [OpenStax]

gravitational field | vector field that surrounds the mass creating the field; the field is represented by field lines, in which the direction of the field is tangent to the lines, and the magnitude (or field strength) is inversely proportional to the spacing of the lines; other masses respond to this field [OpenStax]

gravitational torque | torque on the body caused by its weight; it occurs when the center of gravity of the body is not located on the axis of rotation [OpenStax]

gravitationally bound | two object are gravitationally bound if their orbits are closed; gravitationally bound systems have a negative total mechanical energy [OpenStax]

harmonics | the term used to refer collectively to the fundamental and its overtones [OpenStax]

hearing | perception of sound [OpenStax]

Hooke's law | in a spring, a restoring force proportional to and in the opposite direction of the imposed displacement [OpenStax]

hydraulic jack | simple machine that uses cylinders of different diameters to distribute force [OpenStax]

hydrostatic equilibrium | state at which water is not flowing, or is static [OpenStax]

ideal banking | sloping of a curve in a road, where the angle of the slope allows the vehicle to negotiate the curve at a certain speed without the aid of friction between the tires and the road; the net external force on the vehicle equals the horizontal centripetal force in the absence of friction [OpenStax]

ideal fluid | fluid with negligible viscosity [OpenStax]

impulse | effect of applying a force on a system for a time interval; this time interval is usually small, but does not have to be [OpenStax]

impulse-momentum theorem | change of momentum of a system is equal to the impulse applied to the system [OpenStax]

inelastic | collision that does not conserve kinetic energy [OpenStax]

inertia | ability of an object to resist changes in its motion [OpenStax]

inertial force | force that has no physical origin [OpenStax]

inertial reference frame | reference frame moving at constant velocity relative to an inertial frame is also inertial; a reference frame accelerating relative to an inertial frame is not inertial [OpenStax]

instantaneous acceleration | acceleration at a specific point in time [OpenStax]

instantaneous angular acceleration | derivative of angular velocity with respect to time [OpenStax]

instantaneous angular velocity | derivative of angular position with respect to time [OpenStax]

instantaneous speed | the absolute value of the instantaneous velocity [OpenStax]

instantaneous velocity | the velocity at a specific instant or time point [OpenStax]

intensity (I) | power per unit area [OpenStax]

interference | overlap of two or more waves at the same point and time [OpenStax]

internal force | force that the simple particles that make up an extended object exert on each other. Internal forces can be attractive or repulsive [OpenStax]

Kepler's first law | law stating that every planet moves along an ellipse, with the Sun located at a focus of the ellipse [OpenStax]

Kepler's second law | law stating that a planet sweeps out equal areas in equal times, meaning it has a constant areal velocity [OpenStax]

Kepler's third law | law stating that the square of the period is proportional to the cube of the semi-major axis of the orbit [OpenStax]

kilogram | SI unit for mass, abbreviated kg [OpenStax]

kinematics | the description of motion through properties such as position, time, velocity, and acceleration [OpenStax]

kinematics of rotational motion | describes the relationships among rotation angle, angular velocity, angular acceleration, and time [OpenStax]

kinetic energy | energy of motion, one-half an object's mass times the square of its speed [OpenStax]

kinetic friction | force that opposes the motion of two systems that are in contact and moving relative to each other [OpenStax]

laminar flow | type of fluid flow in which layers do not mix [OpenStax]

law | description, using concise language or a mathematical formula, of a generalized pattern in nature supported by scientific evidence and repeated experiments [OpenStax]

law of conservation of angular momentum | angular momentum is conserved, that is, the initial angular momentum is equal to the final angular momentum when no external torque is applied to the system [OpenStax]

Law of Conservation of Momentum | total momentum of a closed system cannot change [OpenStax]

law of inertia | see Newton's first law of motion [OpenStax]

lever arm | perpendicular distance from the line that the force vector lies on to a given axis [OpenStax]

linear mass density | the mass per unit length λ of a one dimensional object [OpenStax]

linear mass density | λ, expressed as the number of kilograms of material per meter [OpenStax]

linear wave equation | equation describing waves that result from a linear restoring force of the medium; any function that is a solution to the wave equation describes a wave moving in the positive x-direction or the negative x-direction with a constant wave speed v [OpenStax]

linearity limit (proportionality limit) | largest stress value beyond which stress is no longer proportional to strain [OpenStax]

longitudinal wave | wave in which the disturbance is parallel to the direction of propagation [OpenStax]

loudness | perception of sound intensity [OpenStax]

magnitude | length of a vector [OpenStax]

mechanical energy | sum of the kinetic and potential energies [OpenStax]

mechanical wave | wave that is governed by Newton's laws and requires a medium [OpenStax]

meter | SI unit for length, abbreviated m [OpenStax]

method of adding percents | the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation [OpenStax]

metric system | system in which values can be calculated in factors of 10 [OpenStax]

model | representation of something often too difficult (or impossible) to display directly [OpenStax]

moment of inertia | rotational mass of rigid bodies that relates to how easy or hard it will be to change the angular velocity of the rotating rigid body [OpenStax]

momentum | measure of the quantity of motion that an object has; it takes into account both how fast the object is moving, and its mass; specifically, it is the product of mass and velocity; it is a vector quantity [OpenStax]

natural angular frequency | angular frequency of a system oscillating in SHM [OpenStax]

neap tide | low tide created when the Moon and the Sun form a right triangle with Earth [OpenStax]

net external force | vector sum of all external forces acting on an object or system; causes a mass to accelerate [OpenStax]

net work | work done by all the forces acting on an object [OpenStax]

neutron star | most compact object known—outside of a black hole itself [OpenStax]

newton | SI unit of force; 1 N is the force needed to accelerate an object with a mass of 1 kg at a rate of 1 m/s2 [OpenStax]

Newton's first law of motion | body at rest remains at rest or, if in motion, remains in motion at constant velocity unless acted on by a net external force; also known as the law of inertia [OpenStax]

Newton's law of gravitation | every mass attracts every other mass with a force proportional to the product of their masses, inversely proportional to the square of the distance between them, and with direction along the line connecting the center of mass of each [OpenStax]

Newton's second law for rotation | sum of the torques on a rotating system equals its moment of inertia times its angular acceleration [OpenStax]

Newton's second law of motion | acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system and is inversely proportional to its mass [OpenStax]

Newton's third law of motion | whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts [OpenStax]

node | point where the string does not move; more generally, nodes are where the wave disturbance is zero in a standing wave [OpenStax]

non-conservative force | force that does work that depends on path [OpenStax]

non-Euclidean geometry | geometry of curved space, describing the relationships among angles and lines on the surface of a sphere, hyperboloid, etc. [OpenStax]

non-renewable | energy source that is not renewable, but is depleted by human consumption [OpenStax]

noninertial frame of reference | accelerated frame of reference [OpenStax]

normal force | force supporting the weight of an object, or a load, that is perpendicular to the surface of contact between the load and its support; the surface applies this force to an object to support the weight of the object [OpenStax]

normal mode | possible standing wave pattern for a standing wave on a string [OpenStax]

normal pressure | pressure of one atmosphere, serves as a reference level for pressure [OpenStax]

notes | basic unit of music with specific names, combined to generate tunes [OpenStax]

null vector | a vector with all its components equal to zero [OpenStax]

orbital period | time required for a satellite to complete one orbit [OpenStax]

orbital speed | speed of a satellite in a circular orbit; it can be also be used for the instantaneous speed for noncircular orbits in which the speed is not constant [OpenStax]

order of magnitude | the size of a quantity as it relates to a power of 10 [OpenStax]

orthogonal vectors | two vectors with directions that differ by exactly 90°, synonymous with perpendicular vectors [OpenStax]

oscillation | single fluctuation of a quantity, or repeated and regular fluctuations of a quantity, between two extreme values around an equilibrium or average value [OpenStax]

overdamped | condition in which damping of an oscillator causes it to return to equilibrium without oscillating; oscillator moves more slowly toward equilibrium than in the critically damped system [OpenStax]

overtone | frequency that produces standing waves and is higher than the fundamental frequency [OpenStax]

overtones | all resonant frequencies higher than the fundamental [OpenStax]

parallel axis | axis of rotation that is parallel to an axis about which the moment of inertia of an object is known [OpenStax]

parallel vectors | two vectors with exactly the same direction angles [OpenStax]

parallel-axis theorem | if the moment of inertia is known for a given axis, it can be found for any axis parallel to it [OpenStax]

parallelogram rule | geometric construction of the vector sum in a plane [OpenStax]

pascal (Pa) | SI unit of stress, SI unit of pressure [OpenStax]

Pascal's principle | change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container [OpenStax]

percent uncertainty | the ratio of the uncertainty of a measurement to the measured value, expressed as a percentage [OpenStax]

perfectly inelastic | collision after which all objects are motionless, the final kinetic energy is zero, and the loss of kinetic energy is a maximum [OpenStax]

perihelion | point of closest approach to the Sun of an orbiting body; the corresponding term for the Moon's closest approach to Earth is the perigee [OpenStax]

period (T) | time taken to complete one oscillation [OpenStax]

periodic motion | motion that repeats itself at regular time intervals [OpenStax]

phase shift | angle, in radians, that is used in a cosine or sine function to shift the function left or right, used to match up the function with the initial conditions of data [OpenStax]

phon | numerical unit of loudness [OpenStax]

physical pendulum | any extended object that swings like a pendulum [OpenStax]

physical quantity | characteristic or property of an object that can be measured or calculated from other measurements [OpenStax]

physics | science concerned with describing the interactions of energy, matter, space, and time; especially interested in what fundamental mechanisms underlie every phenomenon [OpenStax]

pitch | perception of the frequency of a sound [OpenStax]

plastic behavior | material deforms irreversibly, does not go back to its original shape and size when load is removed and stress vanishes [OpenStax]

Poiseuille's law | rate of laminar flow of an incompressible fluid in a tube:

$$Q = \frac{(p_2 - p_1)\pi r^4}{8\eta l}.$$

[OpenStax]

Poiseuille's law for resistance | resistance to laminar flow of an incompressible fluid in a tube:

$$R = \frac{8\eta l}{\pi r^4}$$

[OpenStax]

polar coordinate system | an orthogonal coordinate system where location in a plane is given by polar coordinates [OpenStax]

polar coordinates | a radial coordinate and an angle [OpenStax]

position | the location of an object at a particular time [OpenStax]

position vector | vector from the origin of a chosen coordinate system to the position of a particle in two- or threedimensional space [OpenStax]

potential energy | function of position, energy possessed by an object relative to the system considered [OpenStax]

potential energy diagram | graph of a particle's potential energy as a function of position [OpenStax]

potential energy difference | negative of the work done acting between two points in space [OpenStax]

power | (or instantaneous power) rate of doing work [OpenStax]

precession | circular motion of the pole of the axis of a spinning object around another axis due to a torque [OpenStax]

precision | the degree to which repeated measurements agree with each other [OpenStax]

pressure | force per unit area exerted perpendicular to the area over which the force acts [OpenStax]

pressure | force pressing in normal direction on a surface per the surface area, the bulk stress in fluids [OpenStax]

principle of equivalence | part of the general theory of relativity, it states that there no difference between free fall and being weightless, or a uniform gravitational field and uniform acceleration [OpenStax]

projectile motion | motion of an object subject only to the acceleration of gravity [OpenStax]

pulse | single disturbance that moves through a medium, transferring energy but not mass [OpenStax]

radical coordinate | distance to the origin in a polar coordinate system [OpenStax]

range | maximum horizontal distance a projectile travels [OpenStax]

reference frame | coordinate system in which the position, velocity, and acceleration of an object at rest or moving is measured [OpenStax]

relative velocity | velocity of an object as observed from a particular reference frame, or the velocity of one reference frame with respect to another reference frame [OpenStax]

renewable | energy source that is replenished by natural processes, over human time scales [OpenStax]

resonance | large amplitude oscillations in a system produced by a small amplitude driving force, which has a frequency equal to the natural frequency [OpenStax]

restoring force | force acting in opposition to the force caused by a deformation [OpenStax]

resultant vector | vector sum of two (or more) vectors [OpenStax]

Reynolds number | dimensionless parameter that can reveal whether a particular flow is laminar or turbulent [OpenStax]

rocket equation | derived by the Soviet physicist Konstantin Tsiolkovsky in 1897, it gives us the change of velocity that the rocket obtains from burning a mass of fuel that decreases the total rocket mass from mi down to m [OpenStax]

rolling motion | combination of rotational and translational motion with or without slipping [OpenStax]

rotational dynamics | analysis of rotational motion using the net torque and moment of inertia to find the angular acceleration [OpenStax]

rotational kinetic energy | kinetic energy due to the rotation of an object; this is part of its total kinetic energy [OpenStax]

rotational work | work done on a rigid body due to the sum of the torques integrated over the angle through with the body rotates [OpenStax]

scalar | a number, synonymous with a scalar quantity in physics [OpenStax]

scalar component | a number that multiplies a unit vector in a vector component of a vector [OpenStax]

scalar equation | equation in which the left-hand and right-hand sides are numbers [OpenStax]

scalar product | the result of the scalar multiplication of two vectors is a scalar called a scalar product; also called a dot product [OpenStax]

scalar quantity | quantity that can be specified completely by a single number with an appropriate physical unit [OpenStax]

Schwarzschild radius | critical radius (RS) such that if a mass were compressed to the extent that its radius becomes less than the Schwarzschild radius, then the mass will collapse to a singularity, and anything that passes inside that radius cannot escape [OpenStax]

second | the SI unit for time, abbreviated s [OpenStax]

second equilibrium condition | expresses rotational equilibrium; all torques due to external forces acting on the body balance out and their vector sum is zero [OpenStax]

shear modulus | elastic modulus for shear stress [OpenStax]

shear strain | strain caused by shear stress [OpenStax]

shear stress | stress caused by shearing forces [OpenStax]

shock wave | wave front that is produced when a sound source moves faster than the speed of sound [OpenStax]

SI units | the international system of units that scientists in most countries have agreed to use; includes units such as meters, liters, and grams [OpenStax]

significant figures | used to express the precision of a measuring tool used to measure a value [OpenStax]

simple harmonic motion (SHM) | oscillatory motion in a system where the restoring force is proportional to the displacement, which acts in the direction opposite to the displacement [OpenStax]

simple harmonic oscillator | a device that oscillates in SHM where the restoring force is proportional to the displacement and acts in the direction opposite to the displacement [OpenStax]

simple pendulum | point mass, called a pendulum bob, attached to a near massless string [OpenStax]

sonic boom | loud noise that occurs as a shock wave as it sweeps along the ground [OpenStax]

sound | traveling pressure wave that may be periodic; the wave can be modeled as a pressure wave or as an oscillation of molecules [OpenStax]

sound intensity level | unitless quantity telling you the level of the sound relative to a fixed standard [OpenStax]

sound pressure level | ratio of the pressure amplitude to a reference pressure [OpenStax]

space-time | concept of space-time is that time is essentially another coordinate that is treated the same way as any individual spatial coordinate; in the equations that represent both special and general relativity, time appears in the same context as do the spatial coordinates [OpenStax]

specific gravity | ratio of the density of an object to a fluid (usually water) [OpenStax]

spring tide | high tide created when the Moon, the Sun, and Earth are along one line [OpenStax]

stable equilibrium point | point where the net force on a system is zero, but a small displacement of the mass will cause a restoring force that points toward the equilibrium point [OpenStax]

standing wave | wave that can bounce back and forth through a particular region, effectively becoming stationary [OpenStax]

static equilibrium | body is in static equilibrium when it is at rest in our selected inertial frame of reference [OpenStax]

static friction | force that opposes the motion of two systems that are in contact and are not moving relative to each other [OpenStax]

strain | dimensionless quantity that gives the amount of deformation of an object or medium under stress [OpenStax]

stress | quantity that contains information about the magnitude of force causing deformation, defined as force per unit area [OpenStax]

stress-strain diagram | graph showing the relationship between stress and strain, characteristic of a material [OpenStax]

superposition | phenomenon that occurs when two or more waves arrive at the same point [OpenStax]

surface mass density | mass per unit area σ of a two dimensional object [OpenStax]

system | object or collection of objects whose motion is currently under investigation; however, your system is defined at the start of the problem, you must keep that definition for the entire problem [OpenStax]

tail-to-head geometric construction | geometric construction for drawing the resultant vector of many vectors [OpenStax]

tangential accleration | magnitude of which is the time rate of change of speed. Its direction is tangent to the circle. [OpenStax]

tensile strain | strain under tensile stress, given as fractional change in length, which occurs when forces are stretching an object, causing its elongation [OpenStax]

tensile stress | stress caused by tensile forces, only in one direction, which occurs when forces are stretching an object, causing its elongation [OpenStax]

tension | pulling force that acts along a stretched flexible connector, such as a rope or cable [OpenStax]

terminal velocity | constant velocity achieved by a falling object, which occurs when the weight of the object is balanced by the upward drag force [OpenStax]

theory | testable explanation for patterns in nature supported by scientific evidence and verified multiple times by various groups of researchers [OpenStax]

theory of general relativity | Einstein's theory for gravitation and accelerated reference frames; in this theory, gravitation is the result of mass and energy distorting the space-time around it; it is also often referred to as Einstein's theory of gravity [OpenStax]

thrust | reaction force that pushes a body forward in response to a backward force [OpenStax]

tidal force | difference between the gravitational force at the center of a body and that at any other location on the body; the tidal force stretches the body [OpenStax]

timbre | number and relative intensity of multiple sound frequencies [OpenStax]

time of flight | elapsed time a projectile is in the air [OpenStax]

torque | cross product of a force and a lever arm to a given axis [OpenStax]

torsional pendulum | any suspended object that oscillates by twisting its suspension [OpenStax]

total accleration | vector sum of centripetal and tangential accelerations [OpenStax]

total displacement | the sum of individual displacements over a given time period [OpenStax]

total linear acceleration | vector sum of the centripetal acceleration vector and the tangential acceleration vector [OpenStax]

trajectory | path of a projectile through the air [OpenStax]

transducer | device that converts energy of a signal into measurable energy form, for example, a microphone converts sound waves into an electrical signal [OpenStax]

transverse wave | wave in which the disturbance is perpendicular to the direction of propagation [OpenStax]

turbulence | fluid flow in which layers mix together via eddies and swirls [OpenStax]

turbulent flow | type of fluid flow in which layers mix together via eddies and swirls [OpenStax]

turning point | position where the velocity of a particle, in one-dimensional motion, changes sign [OpenStax]

two-body pursuit problem | a kinematics problem in which the unknowns are calculated by solving the kinematic equations simultaneously for two moving objects [OpenStax]

uncertainty | a quantitative measure of how much measured values deviate from one another [OpenStax]

underdamped | condition in which damping of an oscillator causes the amplitude of oscillations of a damped harmonic oscillator to decrease over time, eventually approaching zero [OpenStax]

unit vector | vector of a unit magnitude that specifies direction; has no physical unit [OpenStax]

unit vectors of the axes | unit vectors that define orthogonal directions in a plane or in space [OpenStax]

units | standards used for expressing and comparing measurements [OpenStax]

universal gravitational constant | constant representing the strength of the gravitational force, that is believed to be the same throughout the universe [OpenStax]

vector | mathematical object with magnitude and direction [OpenStax]

vector components | orthogonal components of a vector; a vector is the vector sum of its vector components [OpenStax]

vector equation | equation in which the left-hand and right-hand sides are vectors [OpenStax]

vector product | the result of the vector multiplication of vectors is a vector called a vector product; also called a cross product [OpenStax]

vector quantity | physical quantity described by a mathematical vector—that is, by specifying both its magnitude and its direction; synonymous with a vector in physics [OpenStax]

vector sum | resultant of the combination of two (or more) vectors [OpenStax]

velocity vector | vector that gives the instantaneous speed and direction of a particle; tangent to the trajectory [OpenStax]

viscosity | measure of the internal friction in a fluid [OpenStax]

wave | disturbance that moves from its source and carries energy [OpenStax]

wave function | mathematical model of the position of particles of the medium [OpenStax]

wave number | $\frac{2\pi}{\lambda}$ [OpenStax]

wave speed | magnitude of the wave velocity [OpenStax]

wave velocity | velocity at which the disturbance moves; also called the propagation velocity [OpenStax]

wavelength | distance between adjacent identical parts of a wave [OpenStax]

weight | force $\vec{w}$ due to gravity acting on an object of mass m [OpenStax]

work | done when a force acts on something that undergoes a displacement from one position to another [OpenStax]

work done by a force | integral, from the initial position to the final position, of the dot product of the force and the infinitesimal displacement along the path over which the force acts [OpenStax]

work-energy theorem | net work done on a particle is equal to the change in its kinetic energy [OpenStax]

work-energy theorem for rotation | the total rotational work done on a rigid body is equal to the change in rotational kinetic energy of the body [OpenStax]

Young's modulus | elastic modulus for tensile or compressive stress [OpenStax]

Detailed Licensing

Overview

Title: Calculus-Based Physics (Schnick)

Webpages: 87

All licenses found:

- CC BY-SA 2.5: 89.7% (78 pages)
- Undeclared: 10.3% (9 pages)

By Page

- Calculus-Based Physics (Schnick) - *CC BY-SA 2.5*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
    - Why Write the Book, and, Why Release it for Free? - *CC BY-SA 2.5*
  - Volume A: Kinetics, Statics, and Thermodynamics - *CC BY-SA 2.5*
    - 1A: Mathematical Prelude - *CC BY-SA 2.5*
    - 2A: Conservation of Mechanical Energy I: Kinetic Energy & Gravitational Potential Energy - *CC BY-SA 2.5*
    - 3A: Conservation of Mechanical Energy II: Springs, Rotational Kinetic Energy - *CC BY-SA 2.5*
    - 4A: Conservation of Momentum - *CC BY-SA 2.5*
    - 5A: Conservation of Angular Momentum - *CC BY-SA 2.5*
    - 6A: One-Dimensional Motion (Motion Along a Line): Definitions and Mathematics - *CC BY-SA 2.5*
    - 7A: One-Dimensional Motion: The Constant Acceleration Equations - *CC BY-SA 2.5*
    - 8A: One-Dimensional Motion: Collision Type II - *CC BY-SA 2.5*
    - 9A: One-Dimensional Motion Graphs - *CC BY-SA 2.5*
    - 10A: Constant Acceleration Problems in Two Dimensions - *CC BY-SA 2.5*
    - 11A: Relative Velocity - *CC BY-SA 2.5*
    - 12A: Gravitational Force Near the Surface of the Earth, First Brush with Newton's 2nd Law - *CC BY-SA 2.5*
    - 13A: Freefall, a.k.a. Projectile Motion - *CC BY-SA 2.5*
    - 14A: Newton's Laws #1: Using Free Body Diagrams - *CC BY-SA 2.5*
    - 15A: Newton's Laws #2: Kinds of Forces, Creating Free Body Diagrams - *CC BY-SA 2.5*
    - 16A: Newton's Laws #3: Components, Friction, Ramps, Pulleys, and Strings - *CC BY-SA 2.5*
    - 17A: The Universal Law of Gravitation - *CC BY-SA 2.5*
    - 18A: Circular Motion - Centripetal Acceleration - *CC BY-SA 2.5*
    - 19A: Rotational Motion Variables, Tangential Acceleration, Constant Angular Acceleration - *CC BY-SA 2.5*
    - 20A: Torque & Circular Motion - *CC BY-SA 2.5*
    - 21A: Vectors - The Cross Product & Torque - *CC BY-SA 2.5*
    - 22A: Center of Mass, Moment of Inertia - *CC BY-SA 2.5*
    - 23A: Statics - *CC BY-SA 2.5*
    - 24A: Work and Energy - *CC BY-SA 2.5*
    - 25A: Potential Energy, Conservation of Energy, Power - *CC BY-SA 2.5*
    - 26A: Impulse and Momentum - *CC BY-SA 2.5*
    - 27A: Oscillations: Introduction, Mass on a Spring - *CC BY-SA 2.5*
    - 28A: Oscillations: The Simple Pendulum, Energy in Simple Harmonic Motion - *CC BY-SA 2.5*
    - 29A: Waves: Characteristics, Types, Energy - *CC BY-SA 2.5*
    - 30A: Wave Function, Interference, Standing Waves - *CC BY-SA 2.5*
    - 31A: Strings, Air Columns - *CC BY-SA 2.5*
    - 32A: Beats and the Doppler Effect - *CC BY-SA 2.5*
    - 33A: Fluids: Pressure, Density, Archimedes' Principle - *CC BY-SA 2.5*
    - 34A: Pascal's Principle, the Continuity Equation, and Bernoulli's Principle - *CC BY-SA 2.5*
    - 35A: Temperature, Internal Energy, Heat and Specific Heat Capacity - *CC BY-SA 2.5*
    - 36A: Heat: Phase Changes - *CC BY-SA 2.5*
    - 37A: The First Law of Thermodynamics - *CC BY-SA 2.5*
  - Volume B: Electricity, Magnetism, and Optics - *CC BY-SA 2.5*

- B1: Charge & Coulomb's Law - *CC BY-SA 2.5*
- B2: The Electric Field - Description and Effect - *CC BY-SA 2.5*
- B3: The Electric Field Due to one or more Point Charges - *CC BY-SA 2.5*
- B4: Conductors and the Electric Field - *CC BY-SA 2.5*
- B5: Work Done by the Electric Field and the Electric Potential - *CC BY-SA 2.5*
- B6: The Electric Potential Due to One or More Point Charges - *CC BY-SA 2.5*
- B7: Equipotential Surfaces, Conductors, and Voltage - *CC BY-SA 2.5*
- B8: Capacitors, Dielectrics, and Energy in Capacitors - *CC BY-SA 2.5*
- 9B: Electric Current, EMF, and Ohm's Law - *CC BY-SA 2.5*
- B10: Resistors in Series and Parallel; Measuring I & V - *CC BY-SA 2.5*
- B11: Resistivity and Power - *CC BY-SA 2.5*
- B12: Kirchhoff’s Rules, Terminal Voltage - *CC BY-SA 2.5*
- B13: RC Circuit - *CC BY-SA 2.5*
- B14: Capacitors in Series & Parallel - *CC BY-SA 2.5*
- B15: Magnetic Field Introduction - Effects - *CC BY-SA 2.5*
- B16: Magnetic Field - More Effects - *CC BY-SA 2.5*
- B17: Magnetic Field: Causes - *CC BY-SA 2.5*
- B18: Faraday's Law and Lenz's Law - *CC BY-SA 2.5*
- B19: Induction, Transformers, and Generators - *CC BY-SA 2.5*
- B20: Faraday’s Law and Maxwell’s Extension to Ampere’s Law - *CC BY-SA 2.5*
- B21: The Nature of Electromagnetic Waves - *CC BY-SA 2.5*
- B22: Huygens’s Principle and 2-Slit Interference - *CC BY-SA 2.5*
- B23: Single-Slit Diffraction - *CC BY-SA 2.5*
- B24: Thin Film Interference - *CC BY-SA 2.5*
- B25: Polarization - *CC BY-SA 2.5*
- B26: Geometric Optics, Reflection - *CC BY-SA 2.5*
- B27: Refraction, Dispersion, Internal Reflection - *CC BY-SA 2.5*
- B28: Thin Lenses - Ray Tracing - *CC BY-SA 2.5*
- B29: Thin Lenses - Lens Equation, Optical Power - *CC BY-SA 2.5*
- B30: The Electric Field Due to a Continuous Distribution of Charge on a Line - *CC BY-SA 2.5*
- B31: The Electric Potential due to a Continuous Charge Distribution - *CC BY-SA 2.5*
- B32: Calculating the Electric Field from the Electric Potential - *CC BY-SA 2.5*
- B33: Gauss’s Law - *CC BY-SA 2.5*
- B34: Gauss’s Law Example - *CC BY-SA 2.5*
- B35: Gauss’s Law for the Magnetic Field and Ampere’s Law Revisited - *CC BY-SA 2.5*
- B36: The Biot-Savart Law - *CC BY-SA 2.5*
- B37: Maxwell’s Equations - *CC BY-SA 2.5*

- Back Matter - *Undeclared*
  - Index - *Undeclared*
  - Glossary - *Undeclared*
  - Detailed Licensing - *Undeclared*