Part A: Physics and Sound

Chapter 1. Introduction

1a. Why This Book?

There are a number of textbooks available for the non-specialist student about the intersection between sound and science. What sets this particular book apart? Why did I bother to write it?

Most importantly, this book is explicitly aimed at introducing you, the reader, *both* to sound *and* to the science of physics. The target audience is a college level general education course, in which one major goal is to introduce students to the field of physics. The subject of sound provides an interesting and entertaining theme, while the physical approach to understanding the world provides the underpinnings.

To support this, the text provides a careful and thorough introduction to the physical science concepts that underlie the subject. After all, students in such a general education course are likely to have less physics background than average. So, few assumptions are made about what you already understand. For example, this text assumes an initial mathematical skill of only high school algebra. The text doesn't shy away from using math to understand the physical world, but it gives explicit support to readers who may not have a great deal of practice with applying the math. A few math concepts may be completely new to you; this book aims to provide everything needed to introduce those concepts.

Another novel feature of this text is that it has broken down the topics into small nuggets, with a short chapter devoted to each. One chapter in this book is similar in scope to a sub-chapter section in a more typical physical science textbook. However, the chapters here are written to be more independent than such sections. This serves a few purposes.

For one, I hope that it will assist you in contextualizing the content. As you read a chapter, you will know that you are focusing on a specific fact or relationship, and you can at first put aside the distraction of other topics. Once that chapter is understood on its own, you should then seek to understand how that chapter relates to others; it is in those relationships that much of the power of physics lies. But I suggest that you first focus on each concept, and second focus on the connections between them.

Another purpose of the short chapters is to accommodate different approaches. There are many choices for how a reader, or a course instructor, might choose to navigate the physics of sound. I have put the chapters in an order that I find appealing, but they are quite modular. If you wish to learn about only some of the topics, or to take them in a different order, that can be achieved by reading only some of the chapters or taking a different sequence. Of course, some topics require first understanding other topics, so some careful planning is involved. More details about those dependencies can be found in the next section of this chapter.

A third, very practical purpose for the short chapters is to allow an instructor to easily communicate with students about which parts of the text are being included in their particular course. With traditional length chapters, it can be very awkward to include only some of a chapter's elements in a course. Since each chapter or section in this book covers a single concept, the decision to study that section should be all-or-nothing.

Finally, I add a comment on what this textbook is not. There are very few example questions that are worked through, and few numerical calculations. This is not because such things aren't important! Being able to do such mathematical manipulations is the bedrock of physics. The mathematical relationships (that is, equations) that you read here might "make sense," but until you can *use* them to answer questions, you do not truly understand them. When this book is used as a textbook for a course, it is presumed that the course instructor will provide examples, both worked examples showing how to apply the ideas, and questions with which the students can practice doing it themselves.

1b. How to Use This Book

When this book first introduces a new technical word or phrase, it will be shown in **boldface.** When the word is used after that, it will be in normal font. Any such word can be found in the Glossary of Terms near the end of the book with a very short definition.

Each chapter or section covers a relatively narrow topic. Some chapters can be skipped without reducing clarity, and some later parts of the book can be understood without having read some earlier parts. Nevertheless, physics is a cumulative subject (like most fields of inquiry), and most of the chapters rely on your having understood some preceding chapters. Thus, there is a rather complicated network (a **concept map**) describing which chapters are prerequisites for which other chapters. Figure 1.1 can help to navigate those relationships.

Figure 1.1 is especially provided for the benefit of course instructors. who are selecting which subjects to include in a course. This is particularly useful as a way to drop subjects from a course, allowing the determination of which supporting chapters can be dropped along with it. Another way to use Figure 1.1 is to pick a topic that you wish to reach, for instance, Chapter 163 Continuous Refraction. You could then trace back through all the prerequisite chapters needed to reach that destination.

Figure 1.1 shows four basic relationships between the chapters in this book.

- A *solid arrow* from Chapter 10 to Chapter 11 indicates that 10 is a *prerequisite* to 11. That is, Chapter 11 is written with the assumption that you have first read and understood Chapter 10. Of course, if Chapter 10 has any prerequisites, then those ideas may be used in Chapter 11 as well.
- A *dashed arrow* from Chapter 20 to Chapter 23 indicates that 20 is *helpful* for 23. A concept from 20 will aid in understanding 23 and may be mentioned in the text. However, Chapter 23 is written so that 20 is not strictly necessary.
- A *dashed double-ended arrow* indicates that the chapters are *mutually supportive.* There are only a few of these, e.g., between Chapter 44 and Chapter 178. There are similar or related concepts in both chapters. Having studied either one will help in understanding the other, but the order is not important.
- A *double-line dashed arrow* from Chapter 130 back to Chapter 28, which is earlier in the book, indicates that Chapter 130 *explains* something in Chapter 28 more deeply. Chapter 28 is written with the assumption that you will read it first, but it may state some facts without explaining the why or how. Chapter 130 is more advanced or requires more background than Chapter 28. Later study of Chapter 130 will deepen your understanding of Chapter 28.

To assist in tracing connections, Figure 1.1 arrows that are internal to that Part are thicker than arrows that connect to other Parts. Arrows from earlier parts, indicating prerequisites, are red, while arrows to later Parts are black.

There are some connections between chapters that are interesting but not necessary. Those are not indicated in Figure 1.1. For example, Chapter 16 mentions Chapter 15, but there is no corresponding connection in Figure 1.1. This means that you can safely ignore the reference and still understand Chapter 16 just fine.

Some chapters are divided into sections, labeled a, b, c… The sections serve to divide up concepts, but their chapter works as a unit. If a chapter is shown as a prerequisite in Figure 1.1, that means that all such sections are needed. For example, both Sections 10a and 10b are needed before moving on to Chapter 11. However, there are exceptions to this rule. Any sections titled with the word "Extra:" are what you might call stubs. They contain interesting information, but they are not needed for any other chapter in the book, and they can be safely skipped. Sections 7b and 7c are the first examples.

The 10 Parts of the book, labeled A, B, C…, serve to group the chapters into general subject areas. If you are interested in the subject of a Part, you will probably choose to read most of it. If you are not interested in that general subject, you can choose to skip most of that Part. However, you may need to read some of that Part's chapters in support of later Parts. So, for example, you might wish to skip Part C, but you might still need to read certain chapters in Part C so that you can dive into Part D.

Part A: Physics and Sound provides the conceptual basics, giving an overview of this book and its subject matter. Chapter 4 is a refresher on some of the algebra skills. The whole rest of the book assumes that you have read all of Part A.

Part B: Movement of Sound explores situations that can be understood simply on the basis that sound moves, usually at a constant speed and usually in a straight line. It introduces proportions, the physical meaning of speed, and makes some explicit suggestions about strategies for approaching an algebraic physics question.

Part C: Vibration & Oscillation looks at how some objects and physical systems can vibrate on their own, with no outside influence except an initial act to get them started. This involves the basic physics concepts of position, velocity, acceleration, and force, and also introduces the concept of energy. A central focus is on understanding simple harmonic motion, which is a fundamental concept in acoustics.

Part D: Characterizing Sound aims to describe sounds in a quantitative way, identifying the physical numerical quantities most closely associated with perceptions of pitch, loudness, and timbre. The physics meaning of power and intensity are covered, along with Fourier spectra and the decibel unit.

Part E: Hearing & Psychoacoustics describes how humans perceive sounds, including the fundamental theories for how the ear works. A few auditory illusions are included, where they help illustrate those functions. The development of musical scales is also considered.

Part F: Shaping Sound is a general treatment of how sounds, and signals representing sounds, can be affected as they are transmitted through various mechanisms. Applications to speakers, microphones, and human voice are covered, with a very brief discussion of distortion. Vibration damping is included, but for most of the Part this subject is not required.

Part G: Traveling Waves looks at the details of how sound is transmitted through air and other materials, and also the effects that occur as it does so. It starts with generic elements of waves (as opposed to vibrations) using string waves as an example. Mass density and pressure are included, but only some of the Part depends on them.

Part H: Multidimensional Waves details effects which happen when waves, especially sound, can travel in more than one direction. This includes refraction, diffraction, and interference. The Part ends with some notes on directional hearing and with consequences for loudspeakers.

Part I: Standing Waves looks at the waves that can develop in a closed container or finite medium. It is written to rely as little as possible on chapters from the preceding Parts. Since this is the physics underlying much of musical instruments, some readers may want to start by focusing attention here.

Part J: Musical Instruments explores the production of sound by instruments based on strings and tubes, stressing the similarities between instruments rather than their distinctions. It focuses on the mechanisms by which the sound is created and controlled to access specific pitches. Vibrations of surfaces is briefly included.

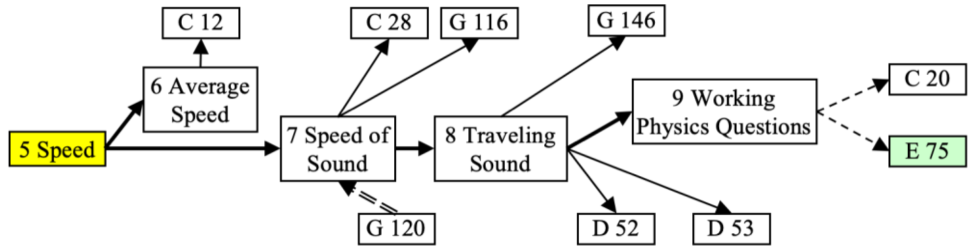

Figure 1.1B

Concept map for Part B. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

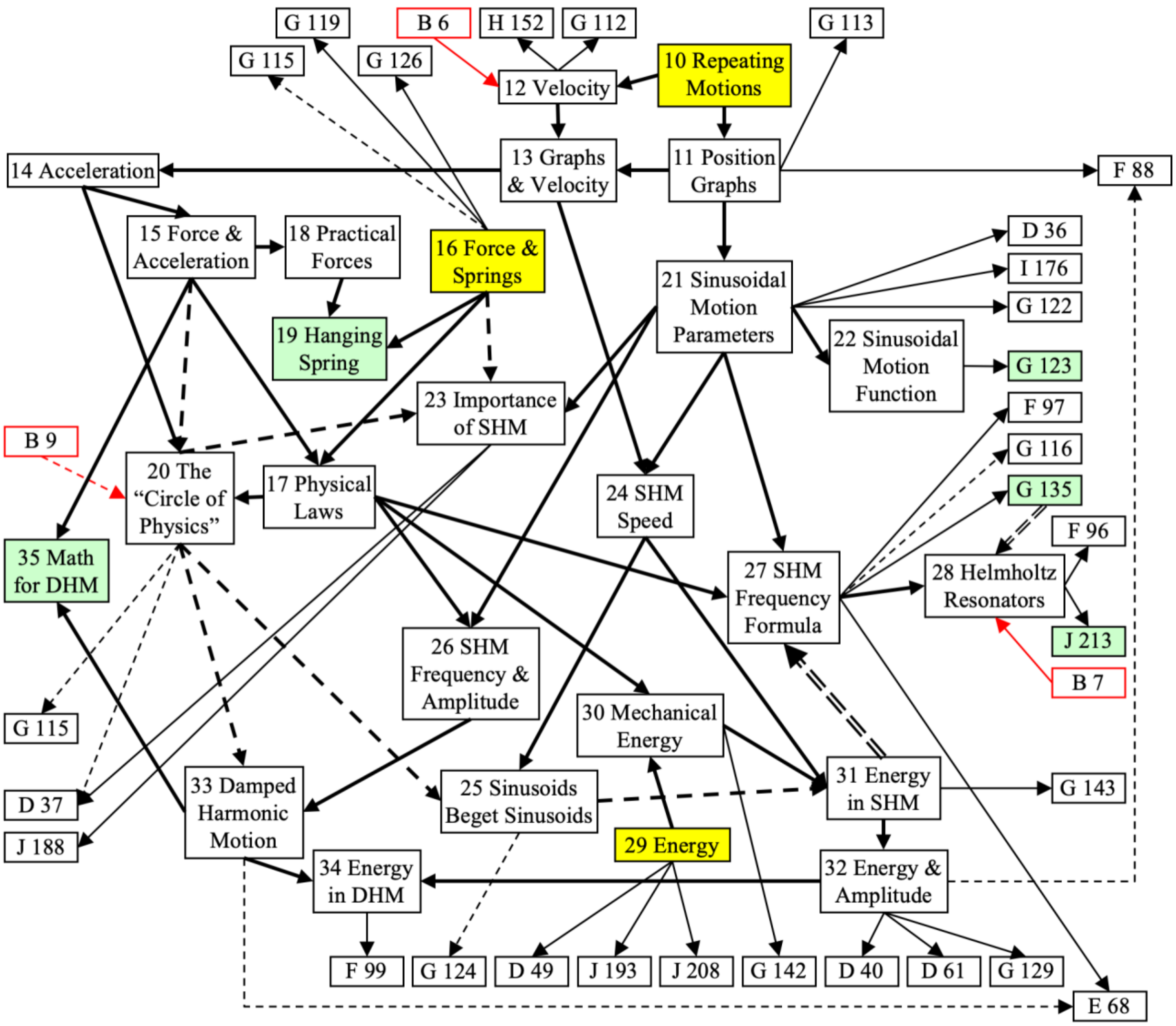

Figure 1.1C

Concept map for Part C. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

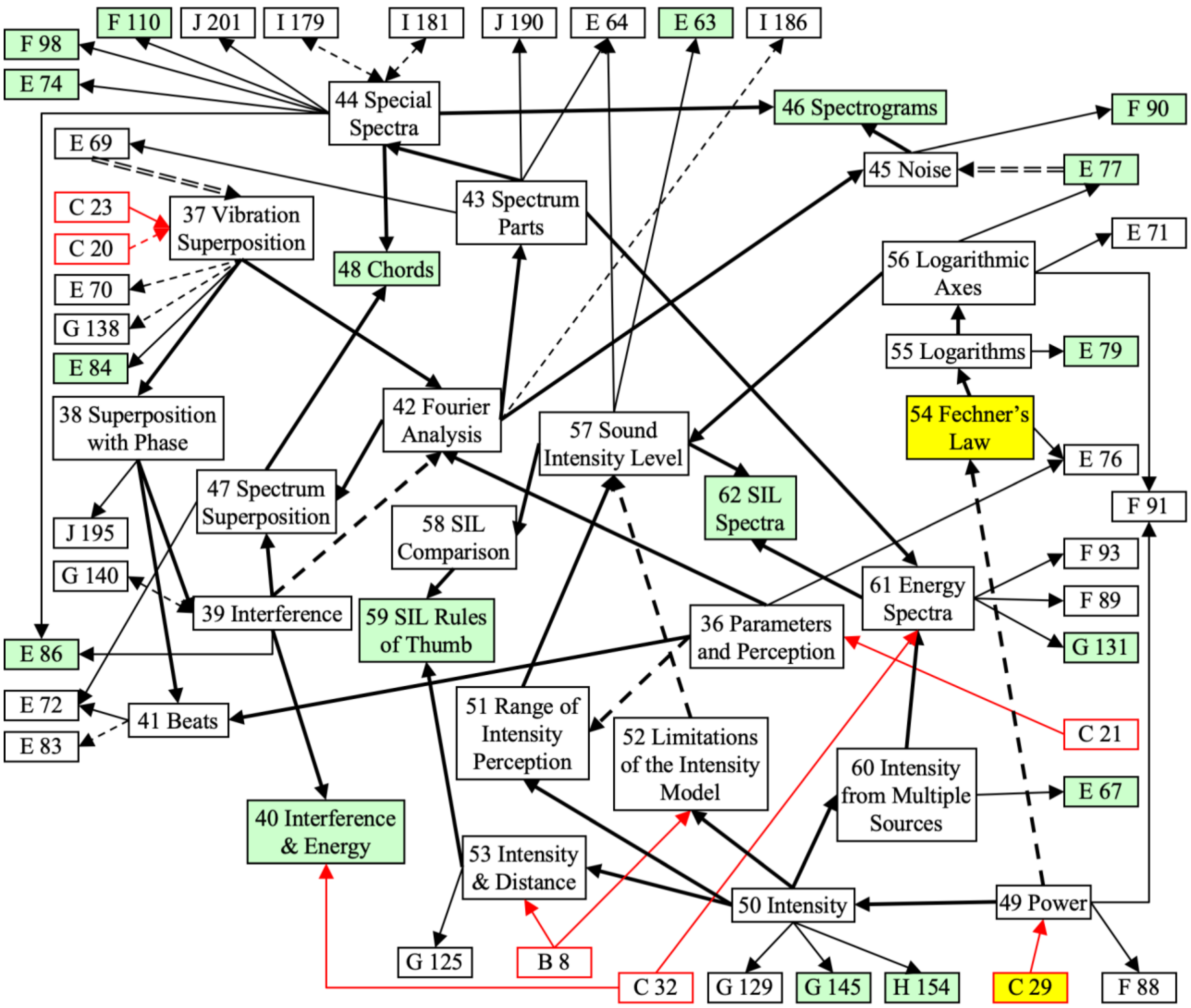

Figure 1.1D

Concept map for Part D. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

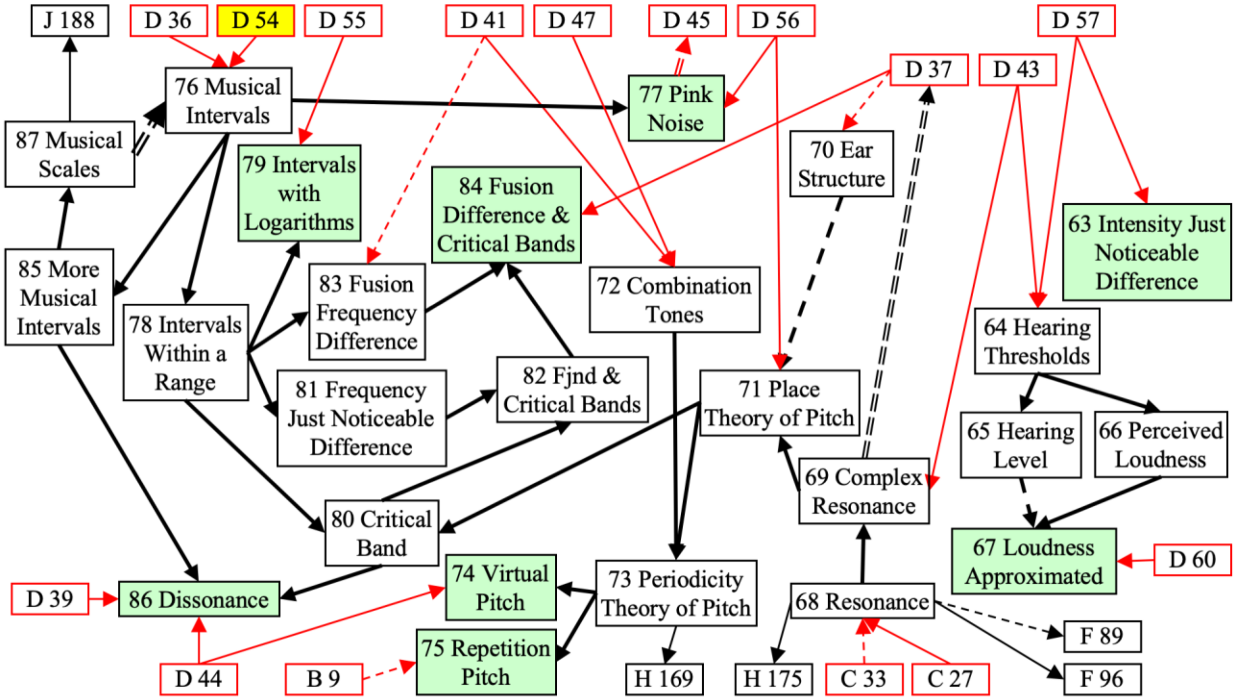

Figure 1.1E

Concept map for Part E. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

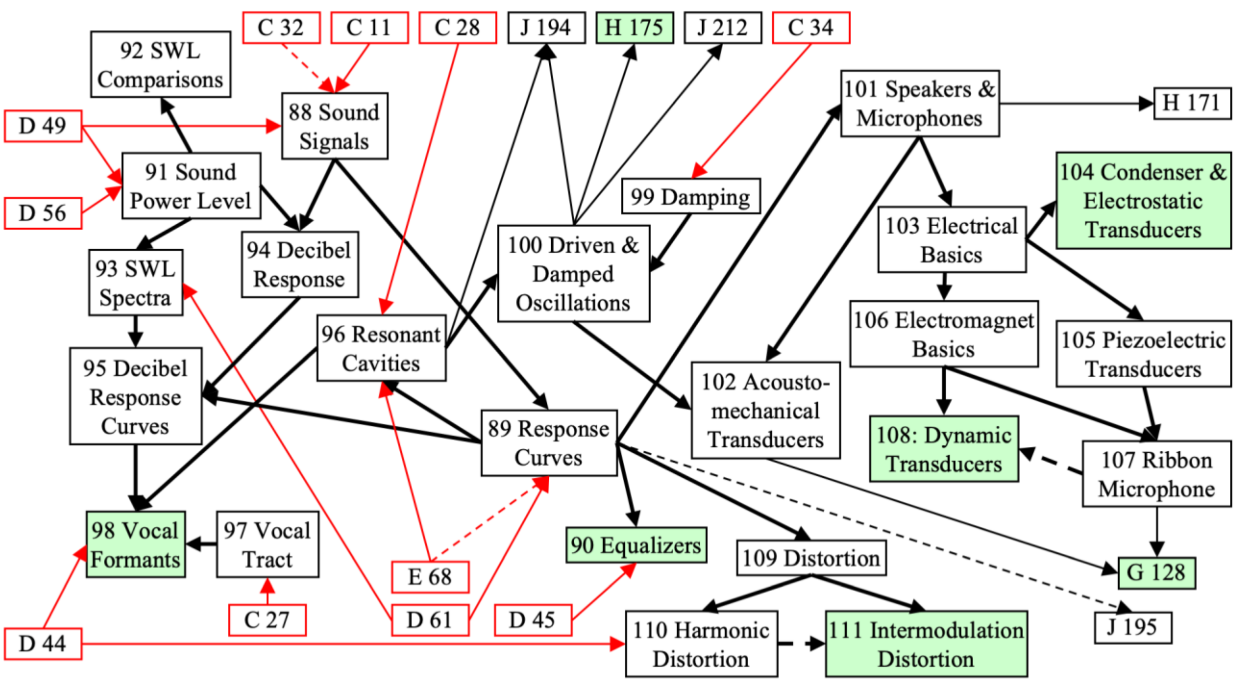

Figure 1.1F

Concept map for Part F. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

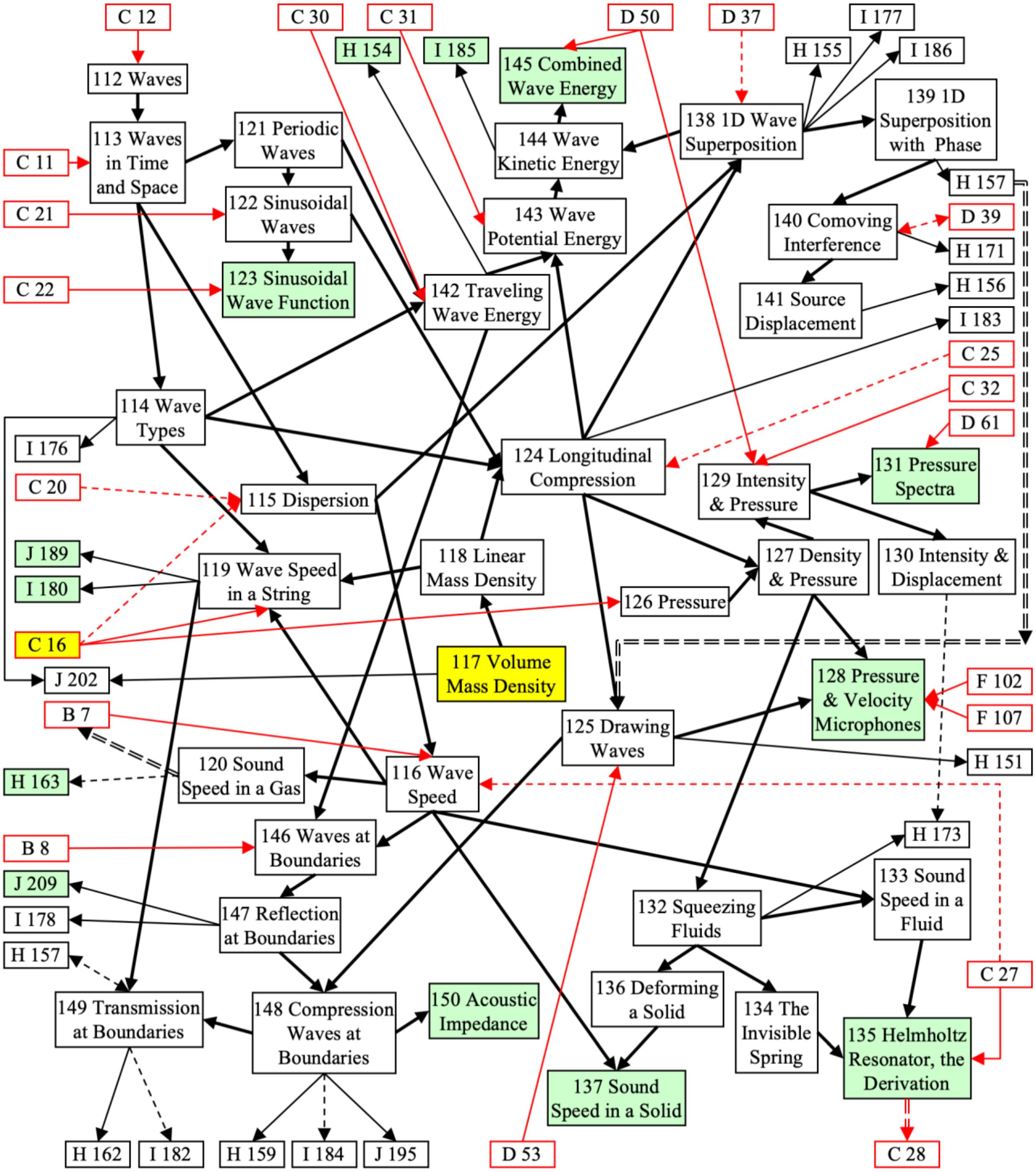

Figure 1.1G

Concept map for Part G. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

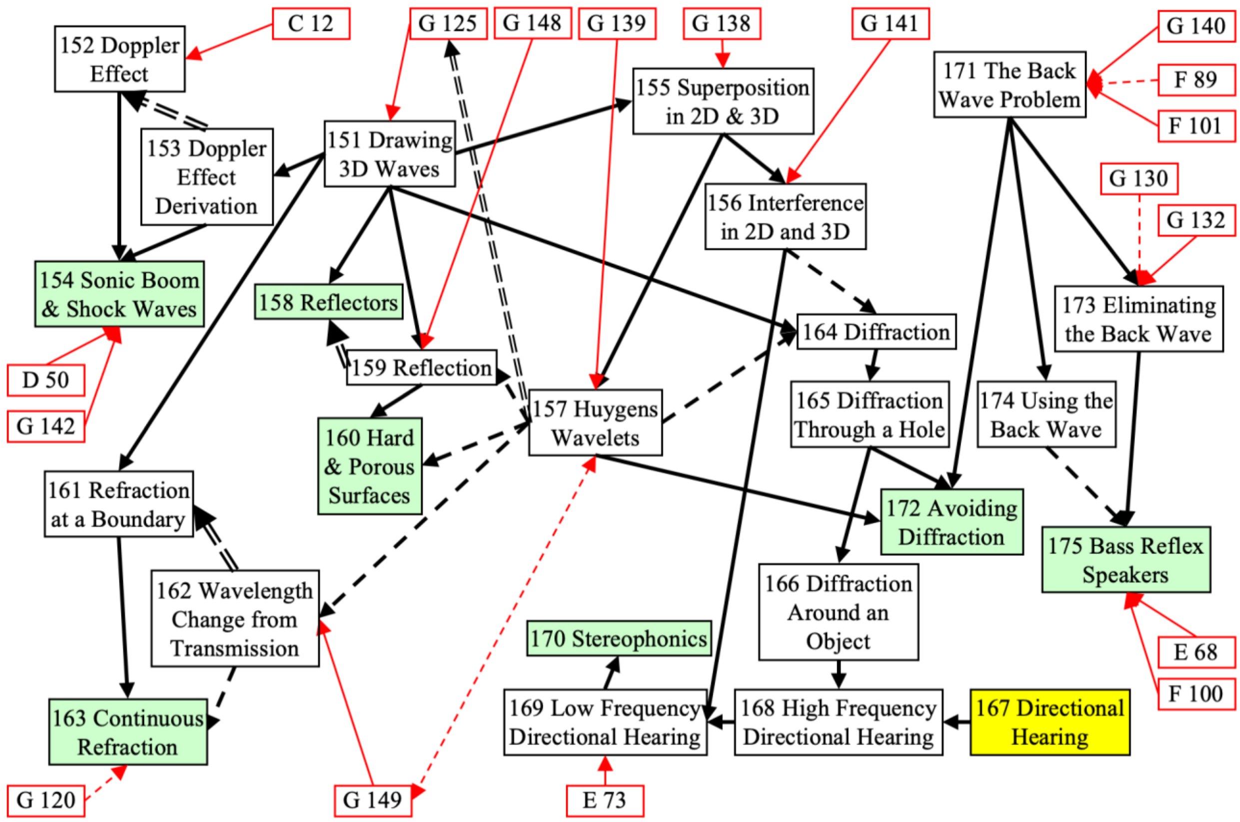

Figure 1.1H

Concept map for Part H. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

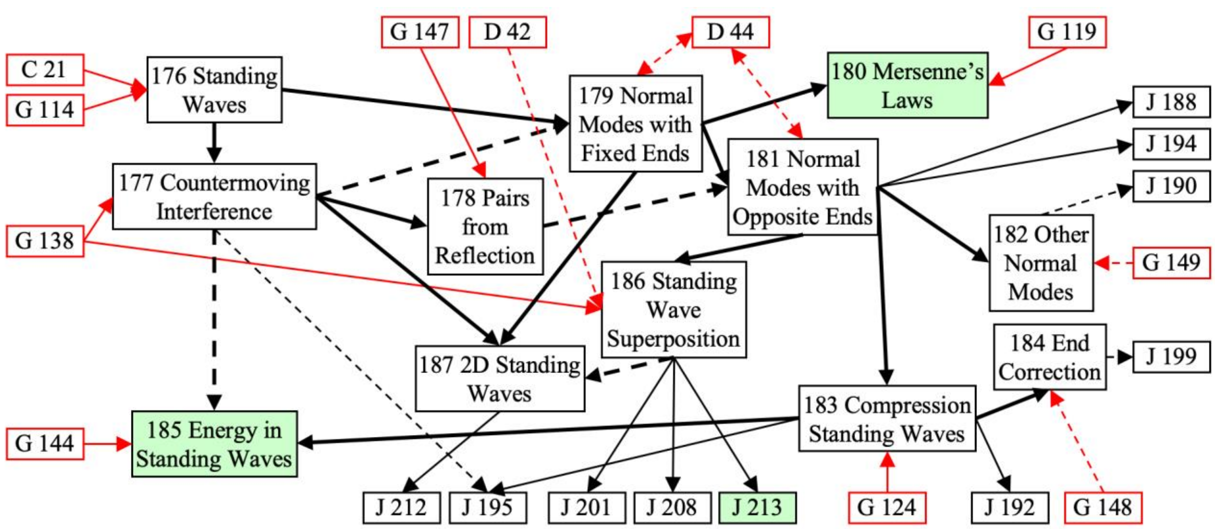

Figure 1.1I

Concept map for Part I. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

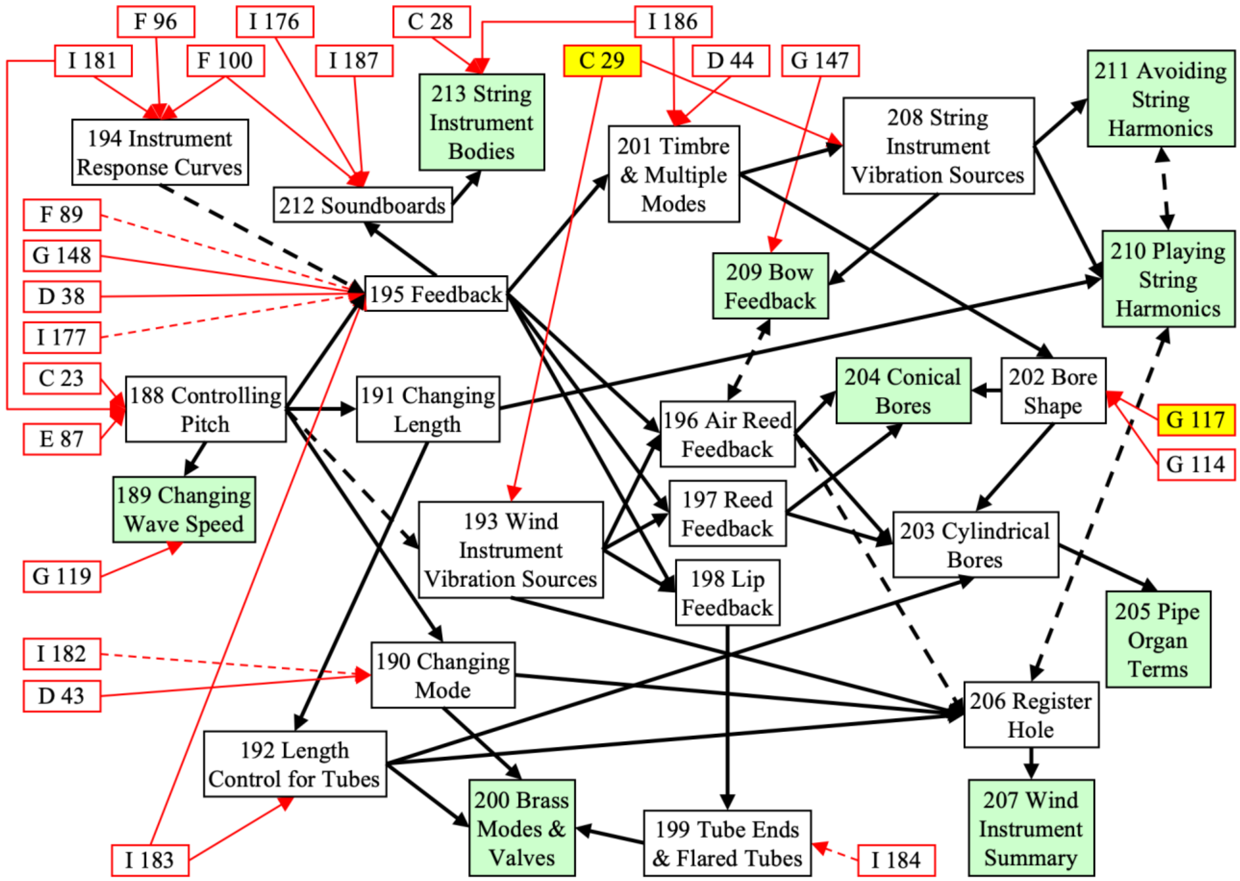

Figure 1.1J

Concept map for Part J. Numbers without titles are chapters in other Parts.
Chapters in yellow have no prerequisites, and chapters in green have no dependents.

Chapter 2. What Is Physics?

2a. Physics Simplifies to Commonalities

Physics is the study of how our natural environment works, in the most general sense. Other sciences study the natural environment, but they differ from physics because they focus on particular parts of the environment. For example, biology focuses on living organisms, chemistry on molecular reactions, and geology on the planet earth. Physics, on the other hand, looks for truths about our environment that cut across all these special areas.

Because of this goal, the physics approach to a question is to simplify it as much as possible. Other sciences often take a special interest in the details of our environment, finding organizing principles for them. In contrast, the physics approach is to strip away as many of the details as we can, arriving at a comparatively simple core.

When the details have been removed, we often find that situations that appeared to be very different are actually very similar. For this reason, the ideas found in physics are useful to many of the other sciences. In fact, physics is sometimes called the most basic, or fundamental, of the sciences. But this should not be taken as a statement that physics is somehow superior to other sciences. It merely reflects a different approach to understanding the universe.

When we ignore details, the results are often more abstract. For example, this book describes features of musical instruments with strings, and musical instruments with narrow tubes of air. It turns out that those two types share many features, so this book lumps them together as "1D instruments," which ignores the detail of whether the vibration is in a string or the air. You may get a feeling that such abstractions make the discussion less "real;" to get comfortable with the abstraction, you have to spend some time reviewing how it applies to the various specific, more "real" cases.

Once it is understood how the core concepts describe a particular situation broadly, details can be added back in for a more complete understanding of that situation. But sometimes that isn't necessary. Depending on what question you are trying to answer or what problem you are trying to solve, understanding the situation broadly may be all you really need.

2b. Physics Finds Models for the Natural World

Once physics has stripped the details from a particular situation, what we are left with is a **model** of that situation. It can be almost guaranteed that the model does not describe that situation exactly. In other words, it is almost surely "wrong." And that's OK. What we require of our models is not that they be exactly right, but rather that they be close enough to be useful.

Throughout this book, many models will be presented. Sometimes, a newer and more refined model will be found to replace an older model. But do not think that makes the older model obsolete. On the contrary, the physics perspective is to use the simplest model that works well enough. So, keep on using that older, simpler model. Only pull out the newer, fancier, more complicated one when the old one isn't sufficient.

A good example of this is the speed of sound through air. You can measure a number for that speed, and then use it quite successfully to answer certain questions. But then, if you study it closely, it turns out that the speed of sound in air is not a single number, but instead depends on the temperature of the air. Section 7b gives a fairly simple equation that relates the temperature and speed, providing a model that improves on the single number. But then Chapter 120 provides another, more complex equation is required if the temperature is very far from typical room temperature. But still, even with that third more accurate model, if temperatures are moderate and high accuracy is not needed, it is perfectly acceptable to use that first, single-speed model.

This means that when you learn about a model, one of the things you should learn is its range of applicability. For the example above, rather than learning "the speed of sound is so-and-so," it is better to learn that "the speed of sound is so-and-so *if* the temperature is moderate." It is not always easy to think of all the ways that a model is limited. An even better description of the first model is "the speed of sound is so-and-so *if* the temperature is moderate *and* the sound is traveling through air." But until you encounter sound traveling through any other substance, it might not occur to you to add the second caveat. Nevertheless, try to be as aware as possible of the limits of applicability for a model.

2c. Physics Is Quantitative

Measurement is fundamental to physics. The physics approach is to find ways to describe the natural world in terms of numerical quantities. This means that the models that we create will most often be mathematical.

Mathematical models are **predictive**. That is, they do not only summarize observations we have made in the past. They can also provide a good guess as to what would happen in situations that we have not yet encountered. Depending on how far we are straying from preceding observations, there is no guarantee that the prediction will be correct. But part of the excitement of physics is that such predictions often do work well. And even when they don't, having a numerical prediction helps to clarify whether a model is working well or not.

Sometimes predictions are requested for specific situations. This is what most physics homework questions ask you to do. You are given a situation and asked to predict a result. We don't often think of homework questions this way, because their purpose is not really to get predictions. Their purpose is to give you practice at selecting pertinent models, and accurately calculating the results of those models. Nevertheless, making predictions is what you must do.

Other times, the predictions create whole new models. If we know that models *A*, *B*, and *C* are all applicable, then we might realize through the math that another relationship *D* must also be true. The process of showing that *D* follows from models *A*, *B*, and *C* is called **deriving** *D*. This is another part of the excitement of physics. Sometimes *A*, *B*, and *C* all seem sensible, and yet *D* is surprising. Other times, *A*, *B*, and *C* are all fairly simple, and yet *D* is quite complex. Physicists love that kind of stuff.

2d. Physical Relationships Can be Described Qualitatively

Very often, we may describe models in non-mathematical terms. Allegory and metaphor might be used, and we may even anthropomorphize inanimate objects. The purpose of these descriptions is to connect the models to experiences and concepts with which we are already familiar. Making connections with things we already know, even if the connection is fanciful, is an excellent way to learn new facts. But in the end, the purpose of these metaphors is to help us think about the mathematical equations.

One way the metaphors help is by telling us what results from a prediction might be reasonable. For instance, we know from first-hand experience that sound travels at a high speed. That is not a numerically accurate statement, but it still has meaning. Suppose we do a calculation that predicts that when I speak to you from across a room, it will take one minute before you hear the sound. This is clearly unreasonable, so we can conclude that either we made a mistake in the calculation, or we are using a model outside its range of applicability. This is a powerful way for you to check your work: ask yourself if the answer that you get makes sense.

2e. Physics Distinguishes Between Similar Yet Different Concepts

Very often, in order to make the physical models match the real world, we find that we must distinguish between concepts that previously were muddled in our minds. For instance, in our everyday life it is often unnecessary to distinguish between the ideas which physicists have named speed, velocity, and acceleration. The same is often true for force, energy, power, and intensity. However, in order to get our mathematical models to describe the world correctly, we need to carefully understand the differences between these, and to know when to use each one.

One example of this that is especially pertinent to the topic of sound is the difference between pitch and loudness. These might seem to be an odd pair to confuse, for we all know that we can hear, and even make with our own mouths, loud and soft low pitches, and loud and soft high pitches. Changing pitch is clearly different from changing loudness. And yet, when these concepts arise in complex or unfamiliar situations, I often have seen students confuse them.

As you work through your course, take the time to learn the distinction between these subtly different concepts, and where to use them. Developing that awareness is more powerful than trying to learn which equation to use for which type of question.

2f. Is Physics Hard?

Physics has a reputation among many people as being difficult to understand. Here are two suggestions for why that might be, and ideas for working through the difficulties.

If you've had a look at Figure 1.1, you might wonder how Section 2a could possibly describe physics with the word "simplify"—that tangle of arrows sure isn't simple! The simplicity is in the individual models, which often correspond to the individual chapters in this book. But then, as you combine models together to answer a question, it is amazing how rapidly complexity can emerge. In fact, this is one aspect of physics that particularly intrigues some physicists. Taken to extremes this is one basis for the study of chaos—meaning not complete disarray, but a field of physics and mathematics.

The material in this book shouldn't lead to extreme complexity, but you may need tools to handle some complications. Your main tool will be algebra, which is discussed at more length in Chapter 4. Here are some general approaches which you may find helpful.

- It is not necessary for you to clearly see *how* you will arrive at an answer to a question before you start working towards the answer. Sometimes, you have to write down some of the (simple) relationships that you know, and then explore combining them with algebra a bit to see if you can get closer to an answer.
- Be especially careful to distinguish between similar but different things. Similar but different *concepts* were described above. But it is even more important to distinguish between different quantities of the same type. For instance, a question might involve several distances, or several time intervals. It is crucial that you not confuse them while working towards a solution.

Another possible reason that some people see physics as difficult is that, once you strip away the details of a situation, you are left with something that is rather abstract. The relationships and models in physics can be very general and widely applicable, which of course is part of their power. But that can also make it more difficult to connect a specific, concrete situation with the appropriate abstract model.

The solution here is the same as the answer to that famous question, "How do you get to Carnegie Hall?"—"Practice, practice, practice." When learning a new model or relationship, connect it to as many concrete examples as you can, and use it in those varied contexts. This means extra work; for instance, it is much easier to simply memorize an equation as a collection of letters, than to have to think about what those letters represent. However, you will be able to *use* that physics equation most effectively if you can connect it to real-world situations. To make an analogy, you could memorize guitar chord fingerings as a set of fret numbers, but that doesn't mean much until you actually try them on a real guitar, and to use them effectively you have to practice them in a variety of tunes.

Chapter 3. What Is Sound?

If one looks in most any dictionary, there are several definitions for the word sound. An important distinction between some of those definitions is emphasized by the old puzzler, "If a tree falls in the forest, and no one is there to hear it, does it make a sound?" In today's society, in which the philosophy of an objective reality is firmly rooted, most people would answer this question with a firm, "Yes!" They are perhaps thinking of sound as a thing in the air, which travels away from the crashing tree. But there are some who would answer, "No." They are likely thinking of sound as an experience or perception. If you look closely, you'll find that most dictionaries have definitions of sound that fit both ideas.

Both of these perspectives on sound can be studied scientifically. The study of the perception of sound is a field named **psychoacoustics**. Psychoacoustics lies at the boundary between the physical and mental worlds. At various points in this book, elements of psychoacoustics will be mentioned.

However, the primary focus of this book is on objective sound, the study of which is named **acoustics**. Acoustics is a branch of physics, but one that is important enough to have its own national societies, separate academic degrees, and people who consider themselves to be acousticians rather than physicists.

One definition of sound in the objective sense is

> Mechanical radiant energy that is transmitted by longitudinal pressure waves in a material medium (as air) and is the objective cause of hearing.[1]

At various points in this book, different elements of this definition will be defined, explained, and expanded upon.

[1] *The Merriam-Webster Dictionary,* Rev. ed. (July 2004), s.v. "sound"

One fundamental part of objective sound is that it travels through some substance, which had to be there before the sound occurred. That substance is called the **medium** for the sound. Perhaps a dolphin would have a different perspective, but for humans sound most frequently implies air as the medium. A very common demonstration is to place a sound source inside a jar from which the air is then evacuated. It becomes impossible for the sound to travel from the source to the outside environment.

Some elements of this definition may broaden your perspective beyond what you first think of as sound. For instance, sound need not necessarily travel through air. Also, although sound is often the cause of hearing, we might wonder heard by who or what? Whales, elephants, dogs, and insects can all hear sounds that are inaudible to humans. From a physics perspective, it is very natural to extend the meaning of the word sound to refer to effects even beyond what any living being can perceive.

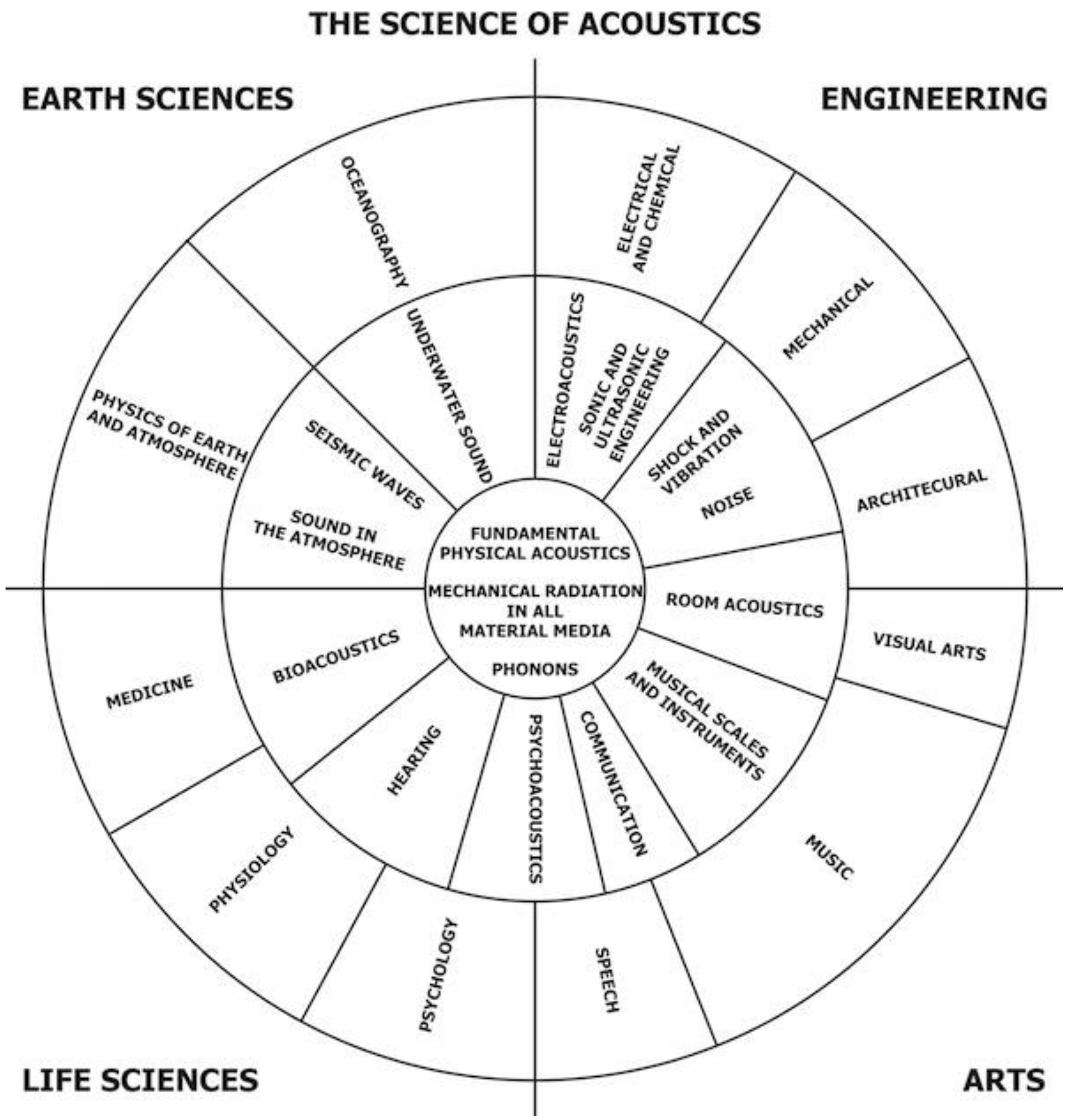

Figure 3.1

Many applications and fields impacted by the science of acoustics.
Source: R. Bruce Lindsay, *J. Acoust. Soc. Am.* 36 (1964): 2241-2243.

A nice overview of the field of acoustics is given in Figure 3.1. Many of the special areas shown in this figure will not be mentioned in this book. But this book will cover fundamental topics that are important in nearly all of them.

Chapter 4. Mathematical Tools for Physics

4a. Numbers and Units

Since in physics we wish to measure quantities, commonly accepted conventions are needed on how to express those measurements. In general, a measurement, such as "3.2 seconds" has two parts: a number and a unit. We'll consider those parts in turn.

Significant Digits

This refers to the digits in a number that serve to make that number more accurate. The concept of significant digits is often used to express, in written form, how accurately a number is known.

All digits in a number are significant *except* for the zeros that only serve to show where the decimal point is. An important skill is counting how many significant digits are in a number. Here are several examples.

number	# sig digits
365.4	4
00365.4	4
0.0034703	5
0.0880	3
250	2 or 3

Points to notice:

- Leading zeros are never significant, regardless of where the decimal point is.
- Zeros in between other digits are always significant.
- Trailing zeros are significant if the number includes a decimal point.

Trailing zeros in a number without a decimal point (as the last example) are problematic. Do those zeros indicate knowledge about those digits, or are they just locating the implied decimal point? Based on what has been written, there is really no way to tell. In the absence of any other information, the less generous interpretation is usually taken.

And now, a short side note about your approach to learning. You have just read an abstract rule (when zeros are significant digits, depending on their placement), as well as a reason behind that rule (relating to the number's accuracy). This is precisely the sort of situation described in Section 2f. You could choose to simply memorize the abstract rule. But a learning approach that is more robust against error is to memorize the rule *and* its background. The way to follow that approach is to practice: for the next week or so, when you see a number (anywhere!) spend a second or two thinking about how many significant digits it has and (most importantly) how any zeros are or are not influencing the number's accuracy.

Scientific Notation

When numbers get very big or very small, it is convenient to write numbers in a different way. Scientific Notation is based on the idea that any number can be broken into a product of two numbers, one that is relatively small (called the **mantissa**), and one that is a power of ten. For example,

$$386.5 = \overbrace{3.865}^{\text{mantissa}} \times 10^2 \tag{4.1}$$
$$0.004437 = 0.4437 \times 10^{-2} = 4.437 \times 10^{-3}$$

The exponent on the 10 will always be an integer, although it may be negative. In the most standardized version of scientific notation, the exponent is chosen so that the mantissa is between 1 and 10. In practice, however, it is not necessary to observe that nicety.

Scientific notation is nothing more than multiplication and raising 10 to a power. But an alternate, and sometimes useful, perspective is that the exponent specifies how many places the decimal point should move to the right. Negative exponents give how far to move the decimal point to the left. Students often get the direction to move the decimal point mixed up. To avoid that mistake, whenever you use this method I recommend pausing afterward to consider whether the result is supposed to be big or small.

Many calculators cannot show exponents and/or don't bother to show the "$\times\ 10$" part of scientific notation. A common abbreviation is to replace the "$\times\ 10$" part with "E", which makes the examples above look like 3.865E2 and 4.437E-3.

Another common shortcut that can lead to confusion is to leave off the first number when it happens to equal one, as in $1.00 \times 10^5 = 10^5$. A common error is to see 10^5 and enter it on a calculator as 10E5, which really means $10.0 \times 10^5 = 10^6$. To avoid problems, imagine that the "$1.00\ \times$" part is present, prompting you to enter $10^5 = 1.00 \times 10^5$ correctly as 1E5.

The number of significant digits in a scientific notation number is equal to the number of significant digits in the mantissa, because the rest of it only indicates where the decimal point goes.

Units and Metric Prefixes

Units connect numbers to the physical world. A distance of 2 inches is vastly different from a distance of 2 miles, and the only difference in how they are expressed is their units. The great majority of physical measurements require a unit, as well as a number, to be complete.

This book will almost always use metric units. Although they may be less familiar to you than American-English units, they make mathematical manipulation easier. In particular, this book will use the *Système International d'Unités* (or SI) version of metric units, which is nearly universal in modern science.

In order to ease the writing of equations, units are frequently expressed using abbreviations, usually of one or two letters. The case of the letters (UPPER CASE or lower case) is very important! If you use the wrong case, you will fail to convey your intended meaning.

In the metric system many types of measurement (length, force, pressure, etc.) have a **root unit**, with a name and abbreviation. From these, we can make new units of different magnitudes by adding prefixes, which act in a way very similar to the multiplier in scientific notation. For example, distance has a root unit of the meter, which is a little longer than a yard. From this, we can make the kilometer, which is 1000 meters, or about $5/8$ of a mile.

Table 4.1 shows important metric prefixes to know, along with examples using the root unit meter. Notice that, except for the last one, *all* the exponents in the table are multiples of three. The last one, centi-, is special because it does not correspond to a multiple-of-three exponent. Centi- will *only* be used for the unit centimeters. There are other prefixes, but they are much less used and will not appear in this book.

As we will see, some measurements have units that are combinations of simpler units. You are probably already familiar with measuring speed in miles per hour, written mathematically as the fraction $\mathrm{miles/hour}$. These are called **derived units**. In fact, it turns out that SI only has seven non-derived units, called **base units**, of which this book will only need four. The metric prefixes are applied to the parts of a derived unit, not the whole. The speed unit $\mathrm{m/ms}$ means "meters per millisecond."

Some derived units, like $\mathrm{m/s}$, are fairly easy to interpret intuitively. But others, for instance $\mathrm{kg/m \cdot s^2}$, certainly are not. Many of those derived units are given their own name, which forms a new root unit that helps remind us of what the quantity means. Each new unit in this book will be described as it becomes useful. Be aware that when performing calculations with units, it is sometimes necessary to break down

Table 4.1

Metric prefixes for units.

Name	Symbol	Meaning (words)	Meaning (math)	example
mega-	M	million	$\times 10^6$	$1\text{ Mm} = 1 \times 10^6\text{ m}$
kilo-	k	thousand	$\times 10^3$	$1\text{ km} = 1000\text{ m}$
(no prefix)			$\times 10^0 = \times 1$	m (meter)
milli-	m	thousandth	$\times 10^{-3}$	$1\text{ mm} = 1 \times 10^{-3}\text{ m}$
micro-	µ	millionth	$\times 10^{-6}$	$1\text{ µm} = 1 \times 10^{-6}\text{ m}$
centi-	c	hundredth	$\times 10^{-2}$	$1\text{ cm} = 1 \times 10^{-2}\text{ m}$

these derived root units into their more basic elements, in order to combine them with other units in the expression.

Units are crucial for conveying meaning to another person, but they also serve an extremely important purpose while doing calculations. Including units in a calculation is a powerful way to find and correct mathematical mistakes. When measurements are put into a physics equation, both the number and the unit should be included. The units then combine with each other in the same way that algebraic symbols do, and must produce a sensible unit for the result. While this idea had been around for several centuries earlier, the suggestion that all physics equations must follow this rule is attributed to Joseph Fourier,[2] who has a very important role to play later in this book.

One example of how you might detect a mistake (yes, I've seen this sort of thing happen!): if you are expecting a distance as the result of a calculation, but the units manipulation yields a result with a unit of seconds, then you know immediately that an error was made in the course of the calculation.

Beginners working with physics equations often have a strong tendency to leave out the units and manipulate only the numbers. It is true that this saves a bit of time, as it means less to keep track of. But it means that you rapidly lose track of what the numbers represent, so that it is far more difficult to notice errors, and you understand less about what is going on. It's like trying to learn to drive a car at night but deciding that it's not really necessary to turn on the headlights. You can make progress this way, but not without many unnecessary bumps and bruises.

Converting Between Units

Since the same measurement can be made using several different units, you may at times need to convert a quantity from one unit to another. With metric system prefixes, this can be a simple exercise in scientific notation and powers of ten. However, even the metric system uses hours, minutes, and seconds. We need a more generally applicable way to convert units.

The following step-by-step method (called the **factor-label method**) will work for any unit conversion, including metric prefixes. We will multiply the quantity by one, but in a form that cancels away the old unit and replaces it with the new one. An example of converting minutes to seconds is shown step-by-step to the right.

1. Write the original quantity, including units.	4.5 min
2. Draw a fraction bar next to it.	4.5 min —

[2] Stephen F. Mason, *A History of the Sciences* (New York: Collier Books, 1962), 169.

3. The unit that you want to get rid of will be either on the top or the bottom of a fraction. Enter the old unit (without a number) on the opposite side of your fraction bar.	$4.5\ \text{min}\ \frac{\quad}{\text{min}}$
4. Enter the new unit you want (without a number) on the other side of the fraction bar.	$4.5\ \text{min}\ \frac{\text{s}}{\text{min}}$
5. Write numbers in front of the units so that the top of your fraction is equal to the bottom. That is, equal in the sense that the top and bottom as a whole are the same measure, even though they have different units and different numbers.	$4.5\ \text{min}\ \frac{60\ \text{s}}{1\ \text{min}}$
6. Do the math.	$4.5\ \cancel{\text{min}}\ \frac{60\ \text{s}}{1\ \cancel{\text{min}}} = 270\ \text{s}$

This method also works when you want to convert just a part of a derived unit. Here is an example with the derived unit m/s^2 (which we will see is a unit of acceleration).

$$3\,\frac{\text{m}}{\text{s}^2} = 3\,\frac{\text{m}}{\text{s}\cdot\text{s}}\,\frac{60\ \text{s}}{1\ \text{min}}\,\frac{60\ \text{s}}{1\ \text{min}} = 10800\,\frac{\text{m}}{\text{min}^2} \tag{4.2}$$

Notice that when a unit is squared, each instance of that unit must be converted separately.

4b. Algebra in Physics

Algebra is a key to solving all but the simplest quantitative physics problems. This book assumes that you have a working knowledge of the mathematical manipulation of algebraic equations, although the next section provides you with some reminders. In this section are some tips for how to best use those skills in the context of physics.

The central concept of algebra is to use **variables** to represent numerical quantities. Variables are normally a single letter; there are only a few specific cases where multiple letters are used together like a word. Using variables is extraordinarily useful. It allows us to think about, and manipulate mathematically, a quantity, even while we don't know the quantity's value.

In physics, algebraic variables represent physical quantities. That is, both the number *and* the unit are included as part of the variable. Some simple examples of defining a variable are "the distance traveled by the thunderclap $= d$," or "$x =$ the length of a table." We will never define a variable that does not contain its units, as in "the distance traveled by the thunderclap $= d$ meters," because doing so limits the power of algebra to represent concepts. Nevertheless, when you define a variable you may find it useful to consider what units might be appropriate, as in

the distance traveled by the thunderclap $= d$ (which could be expressed using meters or km)

Expressing the value of an algebraic variable is a subtly different thing, as in "$x = 1.6$ m." Notice that this does not tell you what the variable means (maybe it is the table length from the previous paragraph, or maybe not). But it does include a unit along with the number.

Variable letters should be chosen to remind you of what they represent. Many students have been trained to use *x* for any unknown quantity. But doing so limits the power of algebra to help you conceptualize a question. As new types of quantities are introduced in this book, the letter that is conventionally used for that type of quantity will be given. As with units and metric prefixes, for algebraic variables it is often important to distinguish between upper-case and lower-case letters.

Often, a physical situation will involve several quantities of the same type. This can sometimes be handled well by using two different letters as variables, but another excellent way to distinguish between them is to add subscripts, as in "time for lightning flash to reach me $= t_L$" or, "time for thunderclap to reach me $=$

t_T". This allows the variable name to represent both the type of the quantity (which helps identify which physics equations are pertinent) and the specific quantity. Very often, being able to answer a question hinges on being able to distinguish between two different quantities of the same type. Sometimes the subscripts chosen are numbers (such as t_1 and t_2), but the numerical subscripts do not imply any mathematical operation.

Occasionally, the same letter is the conventional symbol for two very different types of quantities. For example, we'll see that T is used for both temperature and period. If you need to refer to both in a single question, it is again best to distinguish between them using subscripts. Trying to use an unconventional letter for one of them is more likely to cause confusion.

An **expression** is a collection of algebraic symbols that shows a way to combine quantities mathematically, using mathematical operations such as addition and multiplication. Algebraic relationships are most often in the form of **equations**, that is, an equals sign with a symbol or expression on each side. Notice that an equation tells us that the two sides are numerically the same, not just closely related to each other. For instance, if you find that Bob can run the 50 meter dash in 10 seconds, some students are tempted to write $50\text{ m} = 10\text{ s}$. However, that wouldn't be correct: the two quantities are closely related to each other, but not equal to each other. A **formula** is an equation with a particular purpose to calculate some quantity, so it has a single symbol on one side and an expression on the other side.

4c. Refresher on Algebra Techniques

Although this book assumes that you have a working knowledge of algebra, you may not have exercised that knowledge recently. This section provides a few reminders, to get your mathematical juices flowing.

Expressions

This book will make use of only eight mathematical operations: the usual basic four (addition, subtraction, multiplication, and division), plus exponents (or powers), square root, and the functions cosine and common logarithm (or base-10 logarithm). In particular, the sine function will not appear, although if you feel comfortable with it, you might find it useful. Expressions in this book will extensively use **implied multiplication**, where putting two things next to each other implies multiplication, as in $2d = 2 \times d = 2 \cdot d$. (Enabling this is the main reason to only use single letters for variables.) Your calculator may or may not understand implied multiplication.

One way to manipulate expressions is to use the distributive property

$$k(x_2 - x_1) = kx_2 - kx_1 \quad . \tag{4.3}$$

In this equation, think of the equal sign as meaning "can be converted into." Equation 4.3 can be useful going both ways, either left to right or right to left. Common algebra errors result from trying to use the distributive property where it doesn't apply, as in

$$(d + L)^2 \neq d^2 + L^2 \quad , \tag{4.4}$$

$$\frac{1}{d + L} \neq \frac{1}{d} + \frac{1}{L} \quad . \tag{4.5}$$

Various operations, especially substitution, may result in having a fraction inside a fraction, a so-called **complex fraction**. To simplify those, we have the relations

$$\frac{\left(\frac{A}{B}\right)}{x} = \left(\frac{A}{B}\right)\frac{1}{x} = \frac{A}{Bx} \quad , \tag{4.6}$$

$$\frac{x}{\left(\frac{A}{B}\right)} = x\left(\frac{B}{A}\right) = \frac{xB}{A} \quad . \tag{4.7}$$

I strongly recommend that you use this type of simplification *immediately* when a complex fraction arises. This isn't technically required, but complex fractions can be very confusing, and they are never necessary.

When writing fractions, especially inside a complex fraction, you may be tempted to make your writing less tall by using a slash instead of a horizontal bar. However, my recommendation is that you avoid that whenever possible because of the following confusion:

$$W/A_1\ A_2 = \cdots \qquad = \left(\frac{W}{A_1}\right)A_2 \qquad \text{OR} \qquad = \frac{W}{A_1 A_2} \quad ? \tag{4.8}$$

It is better to just always use a horizontal fraction bar and allow your equations to be tall sometimes. In complex fractions, use a really large fraction bar and/or parentheses to keep it clear which fraction is inside which.

Equations

The most basic rule of working with algebraic equations is that you may apply any operation that you like, as long as you apply it to both sides of the equation. This, of course, means that you need to have an equation to start with. An expression isn't very useful unless paired with another one via an equals sign.

When you have an equation with a fraction on both sides, a shortcut that is often handy is cross-multiplying, as in

$$\frac{F_1}{m_1 + m_2} = \frac{F_3}{m_3} \quad \Rightarrow \quad F_1 \cdot m_3 = F_3 \cdot (m_1 + m_2) \quad . \tag{4.9}$$

In other situations, a useful operation is to flip both fractions over, as in

$$\frac{F_1}{m_1 + m_2} = \frac{F_3}{m_3} \quad \Rightarrow \quad \frac{m_1 + m_2}{F_1} = \frac{m_3}{F_3} \quad . \tag{4.10}$$

In a situation where you are working with multiple equations, the most fundamental way to combine those equations is to use **substitution**. This can be done whenever the same algebraic variable appears in both equations. Equations 4.11 illustrate the process. In step (a), both equations contain k. In step (b), the first equation (either one could have been chosen) is **solved for** k, meaning that it is algebraically rearranged so that k is by itself on one side of the equation. In step (c), the k in the second equation is substituted by the matching expression in the first equation. Finally, some simplification is done.

a) $$F = -k\,x \quad , \qquad f = \sqrt{\frac{k}{m}}$$

b) $$\frac{-F}{x} = k \quad , \qquad f = \sqrt{\frac{k}{m}} \tag{4.11}$$

c) $$f = \sqrt{\frac{\frac{-F}{x}}{m}} = \sqrt{\frac{-F}{x\,m}}$$

Geometry

When working with spatial problems especially, you may need the following algeraic relationships describing geometrical shapes.

When there is a triangle with two perpendicular sides, called a **right triangle**, the lengths of the three sides are related by the famous **Pythagorean theorem**. Naming the lengths of the two shorter sides (the ones

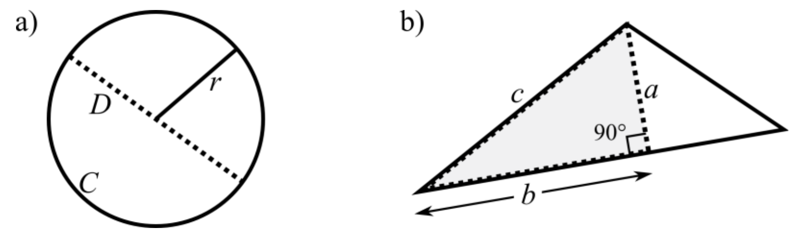

Figure 4.1

Circle and triangle with lengths labeled, as used in the geometry equations.

next to the 90° angle) a and b, and naming the length of the longer side (opposite the 90° angle, called the **hypotenuse**) c, as for the dotted triangle in Figure 4.1(b), the theorem is

$$a^2 + b^2 = c^2 \quad . \tag{4.12}$$

Many triangles are not right triangles! Make sure that a triangle has a 90° angle before applying Eq. 4.12. If you are working with a situation involving a non-right triangle, it may be helpful to divide it into two right triangles as illustrated by the solid-line triangle in Figure 4.1(b).

For a circle (see Figure 4.1(a)), the relationship between the diameter D, radius r, and circumference C is

$$C = \pi D = 2\pi r \quad . \tag{4.13}$$

Sometimes people get confused between this circumference formula and the formula for a circle's area,

$$A = \pi r^2 \quad . \tag{4.14}$$

The key to distinguishing these is that both circumference and radius are lengths, measured in something like meters. Equation 4.14, on the other hand, results in a quantity with units of length squared. This is a perfect example of how relying on memorization is more difficult than relying on an understanding of the equation.

This book also assumes that you have a familiarity with the concept of volume, with traditional variable capital V and SI unit cubic meters, symbolized by m^3. The concept will come up for several different shapes, but the only shape used in a calculation will be a rectangular box, for which the formula is

$$V = lwh \quad , \tag{4.15}$$

where l, w, and h are the length, width, and height of the box respectively.

4d. Proportions

Many physical models express an observation that two different quantities that we can measure are related to each other. Sometimes it then extends to more than two, but we'll stick with just two for now. The simplest sort of relationship occurs when a change of one quantity corresponds to the same sort of change in the other quantity. For example, doubling one means that the other doubles as well. Such a relationship is called a **direct proportion**, often referred to simply as a **proportion**.

Let's illustrate that idea with sound. One element in the definition of sound is that sound is *transmitted*, i.e., it travels. We could quantify how fast it travels with a simple experiment. Start with a sound source from which you can easily see the moment when the sound is made. Suppose that you move 100 m away from the source, and then use a good stopwatch to measure the time that elapses between when you see the sound being created and when you hear the sound. You might measure that time to be 0.294 s. You now have a way to describe how fast sound moves: Sound covers a distance of 100 m in 0.294 s.

Of course, that is not a very general description. In other situations, you might be interested in other distances or times. In these cases, your own physical intuition would probably serve you well in making a prediction. To travel $50\text{ m} = \frac{1}{2} \times 100\text{ m}$ sound would require $\frac{1}{2} \times 0.294\text{ s} = 0.147\text{ s}$. In a time of $0.588\text{ s} = 2 \times 0.294\text{ s}$, sound can be expected to travel $2 \times 100\text{ m} = 200\text{ m}$. These predictions reflect an underlying prediction that the distance traveled by a sound is proportional to the time required for that travel.

A general definition of proportionality is

> Two quantities are proportional if, when comparing different situations, **ratios** of corresponding quantities are always equal.

"Corresponding" can mean "being the same type of measurement." For instance, the ratio of distances traveled is equal to the ratio of times required. You may have learned to represent such ratios with notation like this

$$200\text{ m} : 100\text{ m} :: 2 : 1 \quad , \quad 2 : 1 :: 0.588\text{ s} : 0.294\text{ s} \quad . \tag{4.16}$$

But ratio is really just another word for fraction, so a better way to represent those ratios is

$$\frac{100\text{ m}}{200\text{ m}} = \frac{1}{2} = \frac{0.294\text{ s}}{0.588\text{ s}} \quad . \tag{4.17}$$

This is a better representation because it will allow us to use the rules of algebraic manipulation to rearrange the numbers. Suppose that you hear thunder 4.2 s after a lightning flash, and you wish to know how far away the lightning strike was. Unlike the previous examples, 4.2 s is not a simple multiple of 0.294 s. But we can find the answer by using the same sort of ratio equation,

$$\begin{gathered} \frac{?}{100\text{ m}} = \frac{4.2\text{ s}}{0.294\text{ s}} \quad , \\ ? = \frac{4.2\text{ s}}{0.294\text{ s}} 100\text{ m} = 1429\text{ m} = 1.429\text{ km} \quad . \end{gathered} \tag{4.18}$$

"Corresponding" can also mean, "coming from the same situation." With that interpretation, we would write the ratios or fractions in the form

$$\frac{100\text{ m}}{0.294\text{ s}} = \frac{200\text{ m}}{0.588\text{ s}} \quad , \tag{4.19}$$

which is just an algebraic rearrangement of Eq. 4.17. In either case, there are actually two correspondences that must be lined up properly: same type of quantity (horizontally lined up in Eq. 4.19), and from the same situation (vertically lined up in Eq. 4.19).

If we rearrange Eq. 4.18 in this new configuration

$$\frac{?}{4.2\text{ s}} = \frac{100\text{ m}}{0.294\text{ s}} \quad , \tag{4.20}$$

then we might notice that every time we want to make such a comparison, we will have to calculate the division on the right side. It would be convenient to do the division right now, once and for all. This is the motivation for defining the numerical quantity **speed**, with the result

$$\text{speed of sound} = \frac{100\text{ m}}{0.294\text{ s}} = 340\,\frac{\text{m}}{\text{s}} \quad . \tag{4.21}$$

Notice how the units of speed, such as m/s, are predetermined by the mathematical relationship. Speed is sometimes defined as "the distance traveled in one unit of time," where the phrase "one unit of time" refers, in this example, to one second. However, the speed can be calculated by dividing *any* distance by its *corresponding* time.

There is a symbol to represent the relationship of proportionality. If we choose to represent the distance traveled by a sound with d and the time required for that travel with t, then their proportionality is expressed by

$$d \propto t \quad . \tag{4.22}$$

This looks something like an equation, but it is not! For instance, you *cannot* plug numerical quantities into the right side of a proportion and arrive at a numerical value for the left side. What you *can* do with a proportion is to create an equation, such as

$$\frac{d_1}{d_2} = \frac{t_1}{t_2} \quad , \tag{4.23}$$

where the subscripts refer to two different situations, such as the two situations in Eq. 4.17.

You might wonder why this book has now spent two pages dissecting the idea of speed, which probably could have been defined in two sentences. There are a few reasons.

First, this discussion shows how the relationships that we find in physics are not simply invented out of the brains of scientists. There are objective observations that motivate the relationships, and mathematical patterns which lead to definitions even as simple as speed. In fact, there was just one arbitrary decision made along the way. In Eq. 4.19–4.21, there was no particular reason to put the distance on top. It would have been equally reasonable to write Eq. 4.21 as

$$\frac{0.294\ \mathrm{s}}{100\ \mathrm{m}} = 0.00294 \frac{\mathrm{s}}{\mathrm{m}} \quad . \tag{4.24}$$

In fact, this ratio is sometimes (although rarely) used, and is named the **slowness** of sound.

Second, this illustrates a theme that we will see repeatedly throughout the book. There will be many pairs of quantities that are found to be proportional to each other. And in nearly every case, it will be advantageous to use their ratio to define a new quantity, as we did for speed here. These are all small examples of the search for quantitative commonalities, as described in Sections 2a and 2c. The speed of sound is commonly shared by all the different situations in which sound is traveling through air.

Combining Eqs. 4.21 and 4.23 in another way gives us a pattern to look for,

$$d_1 = \left(340 \frac{\mathrm{m}}{\mathrm{s}}\right) t_1 \quad , \tag{4.25}$$

$$\text{variable 1} = \text{constant} \times \text{variable 2} \quad . \tag{4.26}$$

If an equation like this shows up, it tells us that variable 1 is proportional to variable 2. Because of the similarity of Eq. 4.25 to proportion 4.22, the constant in Eq. 4.26 is called the **constant of proportionality** for the two variables.

4e. Axes, Scales, and Graphs

In order to visually compare numbers that represent the same type of quantity, they are often marked off on an axis, which is a line with numbers marked along it in a specific way. If the numbers are physical quantities, then they most likely have associated units. Standard practice is to use the same unit for all the numbers, and to write the unit at the end of the line in parentheses, which saves having to copy it for every number label on the axis. For example, see the two axes in Figure 4.2

Most commonly, the numbers are arranged so that equal steps along the line correspond to equal increments in the numbers. That is, position on the line (as measured from the position of zero) is *proportional to* the number represented. This is a called a **linear axis.** As happens with all proportions, the ratio of (represented number) to (distance on the paper) is useful, and it's called the **scale factor.** In fact, axes like this are so common that they are sometimes called scales.

When two physical quantities are related to each other, the relation can be powerfully illustrated by a **graph**, with an axis for each quantity. One sort of graph is a map, for instance of a road or a trail, as in Figure 4.2. In order to refer to a specific place, such as the house, linear axes are overlaid on the map, commonly labeled *x* and *y* because they specify positions. The position can then be described with **coordinates,** a set of numbers specifying where the position is along each axis. It is common for science graphs (although not so much for maps) to place the axes so that they cross at the zero point of both axes. That zero point is called the **origin** of the axis, and the position with zero for all coordinates is the **origin** of the graph. For any graph in this book, the axis origins will be assumed to be at the axis crossing point unless explicitly indicated otherwise.

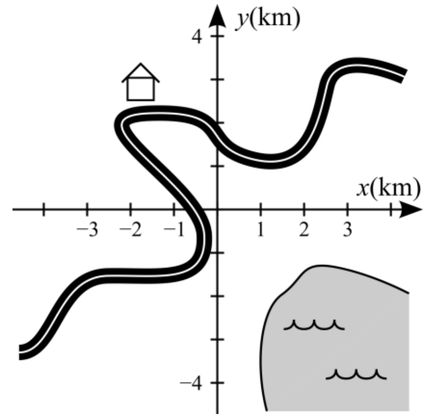

Figure 4.2
A map, an example of a graph.

The main difference between the map and the real thing is the scale factor. Distance d_p measured on the paper parallel to the *x*-axis, is proportional to the real-life distance d_r between the corresponding places. The scale factor is given by

$$S = \frac{d_r}{d_p} \quad , \tag{4.27}$$

which might, for instance, be something like 5 miles/inch. Graphs of other physical quantities will have axes with different scales, including different units.

In fact, for a map, the scale factor is usually chosen to be the same for both axes. This has the result that the same scale factor applies for a distance along *any* diagonal direction as well. However, for nearly every graph in this book the scales for the two axes will be different. In that case, there is *no useful meaning* to diagonally measured distances.

When the algebraic relationship between two physical quantities is not known, then a graph can illustrate the relationship through data points (as measured in some experiment), or perhaps a curve (drawn by someone who has seen enough data points to know what the general trend is). This can be useful, but it has limited power to help understanding the relationship. It can be quite difficult to find the corresponding algebraic relationship. Curved lines that look similar can result from very different mathematical functions. Even if an equation is found that matches the graph, that equation is based only on observation, rather than reasoning about physical relationships. The equation is therefore called **phenomenological** (based on observation of the phenomenon).

When the quantities are related by a known equation, a curve on a graph can reveal a much more intuitive sense of the relationship than an algebraic equation does. The easiest graph shape to be confident about is a straight line. That describes a **linear** relationship, for which the algebraic relationship is the equation $Y = MX + B$. Here X and Y represent the **variable** quantities on the two graph axes, even though they are not necessarily positions x and y. M and B are **constant** quantities that describe the graph as a whole, called **parameters**. Physicists who are trying to interpret observational data often look for ways to make the resulting graphs be straight lines.