Steven S. Andrews

Light and Waves

A Conceptual Exploration of Physics

Springer

Light and Waves

Steven S. Andrews

Light and Waves

A Conceptual Exploration of Physics

Springer

Steven S. Andrews
Department of Bioengineering
University of Washington
Seattle, WA, USA

ISBN 978-3-031-24096-6 ISBN 978-3-031-24097-3 (eBook)
https://doi.org/10.1007/978-3-031-24097-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Light is simultaneously familiar and mysterious. It is something that nearly every one of us sees every waking day of our lives, generally without giving it much thought. But each one of those commonplace light rays has the remarkable property of being composed of both waves and particles at the same time. Furthermore, those rays likely arose from the strange quantum mechanical transitions of individual atoms within a light bulb, or the sun, or that took place thousands of years ago in unimaginably distant stars. This combination of familiarity and mystery has led scholars to study the nature of light from antiquity to the current day, even now revealing surprising new details.

This book explores all types of waves. These include light waves in particular but also string waves, sound waves, water waves, seismic waves, the bizarre matter waves of quantum mechanics, and the gravitational waves that ripple through spacetime. It also focuses on particles of matter and light, which naturally leads to explorations of velocity, forces, momentum, and other properties of physical objects. Waves carry energy, which leads to the deep concepts of energy conservation and energy transfer. Through these explorations, this book tours a substantial fraction of the field of physics.

The emphasis here is on building a strong conceptual understanding of how light and waves work. Topics are illustrated with examples that are drawn from familiar experiences or that are just simply fascinating, such as the science behind bird coloration, how musical instruments work, and the causes of global warming. The language of mathematics is often clearer than the equivalent paragraphs of prose, so this book helps teach the necessary mathematical skills and then uses them to build a deeper understanding of physics. These mathematics, which never extend beyond introductory algebra, are covered in the main text and several appendices.

This book was written with several audiences in mind. The first is undergraduates who are not science majors but who are required to take a science course. In my experience, these students are highly heterogeneous, with some who enjoy science but have chosen to pursue other interests, and others who prefer to avoid anything quantitative. A class on light and waves works well in this case because it is widely accessible, while retaining the interest of all students by covering topics that most have not seen before. The second audience is high school students taking a physics elective. Again, studying light and waves cuts across traditional physics topics, enabling the course to engage students with different backgrounds. This book is aligned with the goals of the Next Generation Science Standards, a set of high school science education standards that have been adopted by most US states, which emphasize building a conceptual understanding, presenting scientific practice frequently, and focusing on the key underlying ideas. The third audience is adults who would like to learn about light and other waves on their own, whether for personal interest or building skills. They will appreciate the fact that the text is self-contained, without requiring supplementation from lectures, homework, or online resources.

Each chapter opens with a question and closes with a list of exercises. The opening questions are intended to inspire curiosity, encouraging the student to seek out answers while reading the text. The exercises at the end are divided into three categories. “Questions” test the student’s understanding of the material while also addressing common misconceptions. “Problems” generally require numerical calculation, which help develop numerical fluency, provide a sense of scale for the relevant phenomena, and build a quantitative understanding of the topic. And “Puzzles” are challenging but interesting problems that explore the topic in depth. Many of these are best approached during small-group instruction, where they serve as tasks that groups of students can discuss, puzzle over, and solve together. The back of the book includes solutions to the odd-numbered problems.

The book is divided into four parts: Waves, Rays, Light, and Modern Physics. Waves introduces widely-applicable physical concepts, like resonance and superposition, which arise repeatedly throughout the rest of the book. Rays explores geometric optics, including shadows, reflection, and refraction; these topics are largely empirical but important for understanding light behaviors. Light investigates the perception and physics of light, going into topics such as color, polarization, and thermal radiation. Finally, Modern Physics applies the same physics that students have learned throughout the book to make sense of photons, basic quantum mechanics, and gravitational waves. This section highlights the many connections between quantum and classical mechanics showing, for example, how quantized matter waves are analogous to standing waves on guitar strings.

There is more material here than can be reasonably covered in one quarter or even one semester. One approach is to cover all chapters, but to skip some sections within each chapter (optional sections are labeled with asterisks). Alternatively, some of the chapters could be skipped entirely. While it’s probably best to start with the Waves section, the rest of the chapters are sufficiently independent that they could be covered in pretty much any order.

I thank David Boness, the physics department chair at Seattle University, who gave me an opportunity to teach a course titled “The World of Light”, from which this book evolved, and who ordered frequent encouragement. I also benefited from numerous discussions with the Seattle University physics department faculty and majors. Discussions with Bernhard Mecking and Danny Desurra were helpful as well. In addition, this book benefited substantially from feedback given to me by the students in my “World of Light” classes. Sam Harrison and others at Springer provided invaluable assistance in the final publishing stages. Although their

contributions were less well defined, I also thank my professors and colleagues who taught me math, science, and scientific writing. These include John Finn, Jane Lipson, Ollie Zafiriou, Steve Boxer, Steve Chu, Dennis Bray, Adam Arkin, Jay Groves, Roger Brent, Herbert Sauro, and many others. Finally, my family has been a constant source of support and joy throughout this project.

Seattle, USA
2023

Steven S. Andrews

Contents

Part IV Modern Physics

*** Optional section**

Appendices

1 Theories of Light

Figure 1.1 Beams of sunlight (called crepuscular rays) over Cardigan Bay, UK.

Opening question
What is a ray of light made of?
(a) a stream of particles
(b) a series of waves
(c) energy
(d) electric and magnetic fields
(e) all of the above

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-3_1

Light is the energy that warms you when you sit in the sun, the energy that grows the world's plants, and the energy that powers the Earth's wind. Light is also the stuff of rainbows, the medium by which we view the world around us, and essentially our only connection to the rest of the universe.

But what actually is light? It would be nice if one could put a beam of light under a microscope and look at it, to see what it really is. However, this isn't possible. Simply put, if light enters our eyes, then we see brightness, and if it doesn't, then we don't. There's no more to see directly than that. Thus, instead of looking at light with a microscope, we have to figure out what light is by investigating its behaviors and then inferring what it must be from them.

This chapter takes a historical approach in answering what light is, starting with a variety of ancient ideas and ending with the modern understanding of particle-wave duality. This history began with numerous philosophical ideas in ancient Greece, India, and China, became more advanced in medieval Egypt, and then progressed to a scientific understanding that was developed primarily in Europe and then North America during the 16th to 19th centuries. Further scientific advances are now pursued worldwide.

1.1 Ancient Ideas About Light

1.1.1 Extramission Theory

The ancient Greeks came up with perhaps the earliest theory of how light worked. They claimed that vision worked by people's eyes emitting beams of light in what is now called the ***extramission theory***. They believed that sight worked in somewhat the same way as the sense of touch. To determine the texture of something, for example, you reach out your hand and touch it. Your hand then feels the surface, through complicated processes, and sends signals back to your brain. Your brain interprets those signals to determine whether the surface is smooth, rough, soft, slippery, or whatever. Likewise, the ancient Greeks imagined that the eyes sent out rays that would sense the colors and shapes of objects and then report back about what they found.

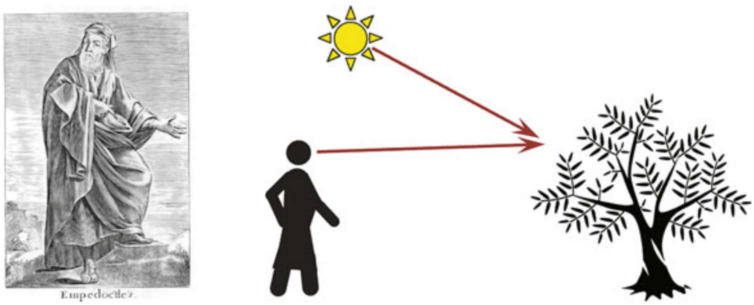


Figure 1.2 (Left) Empedocles. (Right) Extramission theory, in which eye rays interact with sun rays.

To make this more precise, Empedocles, a 5th century BCE scholar (Figure 1.2), postulated that everything was composed of fire, air, earth, and water. He claimed that Aphrodite, the Greek goddess of love, made people's eyes out of all four elements and then lit the fire in their eyes, which shone out and constituted sight.

It was soon recognized that this extramission theory would imply that people would be able to see equally well at night as during the day, which obviously is not the case. Empedocles addressed this by modifying his theory to claim that vision arose from an interaction between the rays from the eyes and rays from light sources, such as the sun. This was better, but then people realized that one can go outside and see the stars immediately without having to wait for beams of light emitted from the eyes to get there and back, showing another problem with the theory. Empedocles ignored this issue, but the subsequent Greek philosophers Plato (429–347 BCE) and his student Aristotle (384–322 BCE) addressed it by claiming that light traveled infinitely fast. This version of the extramission theory was not improved upon further.

However, it was not believed universally. In fact, Aristotle himself didn't fully believe this theory which he had helped create, but stated that "In general, it is unreasonable to suppose that seeing occurs by something issuing from the eye"[1]. He thus supported the correct ***intromission theory***, in which peoples' eyes do not emit light, but only observe the light that shines into them. Nevertheless, Plato's supposed authority led to the extramission theory remaining as the dominant theory of vision up until the Middle Ages.

Remarkably, several important optics discoveries were made during ancient times, despite widespread belief in the extramission theory. Euclid (c. 300 BCE) described rays of light as traveling in straight lines and also wrote a treatise about vision that was largely correct in its treatment of perspective, Hero of Alexandria (c. 10–70 CE) figured out the mathematics of reflection, and Ptolemy (c. 100–170 CE) quantitatively described the bending of light as it shines from air into water, called refraction. All of these philosophers explained their results within the context of the extramission theory. Meanwhile, the Chinese philosopher Mo Zi (c. 470–391 BCE), or perhaps one of his followers, made many of these same discoveries but with the correct intromission theory. Those ideas were not pursued by later Chinese philosophers and were not known in the West until much later.

The extramission theory, despite being unphysical, continued to have a surprising number of adherents long after it was disproved in the Middle Ages. Leonardo da Vinci, the quintessential Renaissance man, expressed extramission views in the 1490s. Even today, children often believe in extramission until they are taught otherwise. Furthermore, recent research showed that about half of modern college students believe that eyes emit rays during vision[2].

The extramission theory also survives in modern popular culture and modern fiction. For example, giving someone the "evil eye" could be interpreted as some sort

[1] Lindberg, David C. (1976) *Theories of Vision from Al-Kindi to Keplar*, University of Chicago Press: Chicago.

[2] Winer, G.A., J.E. Cottrell, V. Gregg, J.S. Fournier, and L.A. Bica (2002) *American Psychologist* 67:417.

of beam shining from the eye. Also, various superheroes and cartoon characters, from Superman to X-Men, can emit beams of light from their eyes (Figure 1.3). Finally, it should be pointed out that "ray-tracing" computer graphics, which are widely used in movie special effects, are based upon what is essentially an extramission concept: rays are drawn outward from the eye or camera lens to determine what is seen in each particular direction.

Figure 1.3 Superboy using his X-ray vision to see a pie in a closed oven. From May 1950, issue #8 ©DC Comics.

1.1.2 Particle Theory

A separate ***particle theory*** of light started at about the same time as the extramission theory with a group of people called the ***atomists*** in both ancient Greece and ancient India, around the 4th or 5th centuries BCE. They claimed that everything in the universe, including light, was composed of tiny particles that were indivisible and indestructible. These particles generally differed according to shape, such as by being pointy or smooth, and could also hook onto other particles to form clusters. This theory is often traced to Democritus, a Greek philosopher born about 460 BCE, and, separately, to Kanada, an Indian philosopher who lived at roughly the some point, between the 2nd and 6th centuries BCE (Figure 1.4). Both atomistic theories had many different versions. For example, within Indian atomism, some believed that there were 4 types of atoms, others claimed 24 types, and yet others held that there were infinitely many types.

Atomism supposedly explained the observation that "pure" materials, such as air, metal, and water, could be broken down, and then re-formed again. Regarding light, the atomists believed that the sun and other bright objects emitted minute atoms of light that traveled very fast, or perhaps infinitely fast, and which people

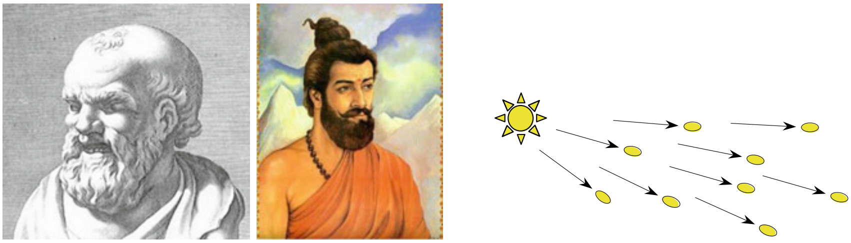


Figure 1.4 Democritus (left) and Kanada (middle), who where two early philosophers who advanced the particle view of light. (Right) Light particles.

perceived when the particles hit their eyes (this intromission explanation contradicted the extramission theory of vision that was widely believed at the time).

In retrospect, the atomists' ideas about the particle nature of light were prescient, as explained below, although this has to assigned more to chance than to scientific deduction. The particle description of light largely held through antiquity and the Middle Ages.

1.2 Islamic Golden Age

The first important discoveries about the nature of light that were based on careful analysis were made by Muslim scientists about a thousand years later. They were made during the Islamic Golden Age, which was an especially prolific period for Islamic culture, science, and mathematics and extended from about the 8th to 14th centuries. Our use of Arabic numerals and algebra date to this period.

Ibn al-Haytham was an Arab scientist born around the year 965 in what is now Iraq, who later moved to Egypt (Figure 1.5). He combined the intromission theory that Aristotle introduced with Euclid's explanation of perspective and Ptolemy's description of refraction. He further advanced these theories, finally yielding a moderately correct explanation of how vision works. In particular, he was the first philosopher in the Western history who convincingly showed that vision occurs through the intromission method.

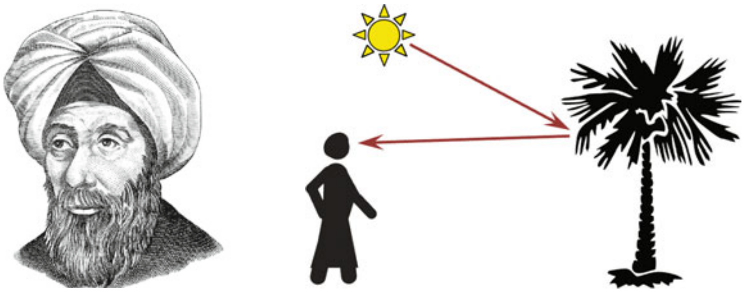


Figure 1.5 (Left) Ibn al-Haytham. (Right) Intromission theory, in which sunlight reflects off an object and goes into a person's eye.

Al-Haytham argued that light extramission didn't make sense because eyes that used extramission would lose an enormous amount of vision substance when looking out into space. Also, light clearly enters our eyes, by the fact that we feel pain when looking at very bright lights. He claimed, correctly, that every point on every object sends light beams out in all directions, of which a very few of them enter our eyes. His final major argument was one of simplicity: given that light is known to reflect off objects and then go to our eyes, then there's no need for any assumption about rays being emitted from our eyes, so that assumption might as well be dropped. Thus, he concluded that the intromission theory is clearly the correct one. Al-Haytham's work greatly influenced European science through the 17th century.

1.3 The Particle-Wave Debate

1.3.1 Wave Theory

A major flaw with the particle theory was that it did not explain refraction very well (Figure 1.6). In 1637, the French scientist René Descartes addressed this problem by hypothesizing that light was composed of waves, thus introducing the ***wave theory*** of light. It was already known that light travelled at a fast but finite speed, so he claimed that light waves bent as they moved from a faster medium to a slower medium, just as water waves and sound waves do, which were understood at the time[3]. This made sense, so the wave theory started to gain acceptance. Subsequent scientists, including Christiaan Huygens and Robert Hooke, built on these ideas to formulate a mathematical theory of light waves.

Figure 1.6 (Left) Three rays of light being refracted by a plastic block. Early wave theory proponents: (Middle) René Descartes and (Right) Christiaan Huygens.

The great English physicist Isaac Newton (1643–1727), who also developed large parts of physics and much of calculus, did not agree with these new wave ideas but preferred the older particle description that had been introduced by the ancient

[3] Descartes was correct that refraction arises from different wave speeds, but incorrectly believed that light travels faster in water than in air.

atomists (Figure 1.7). Furthermore, he conducted experiments in which he looked for the spreading of narrow light beams, with the logic that they should spread if light was made of waves. He observed no spreading, so he concluded that light was made of particles. While referring to these views in a letter to Robert Hooke, a proponent of the wave theories, he famously wrote "if I have seen further, it is by standing on the shoulders of giants." This is typically interpreted as showing Newton's humility, showing that even such a great scientist as himself has built upon others' work. However, the context of this phrase suggests that it was actually a polite dismissal of the wave proponents, claiming that his particle ideas were better because they were based upon the work of more important scientists[4]. Newton published his particle view of light in 1704 (not coincidentally, just after Hooke died), arguing in favor of a particle theory for light. Due to his scientific prowess across all areas of physics, this largely shifted the scientific consensus back to particles.

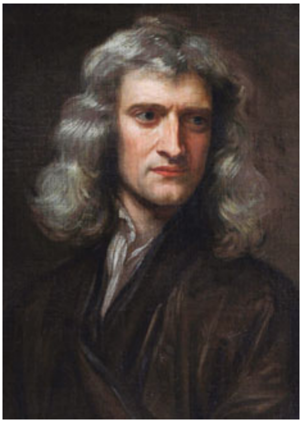


Figure 1.7 Isaac Newton, a proponent of the particle theory of light.

Nevertheless, this did not stop the accumulation of evidence in favor of the wave theory. In 1803, Thomas Young projected light through two parallel slits and found that it bent around behind the slits and then created a pattern of light and dark stripes on a screen (Figure 1.8). This could be explained with a wave interpretation of light but not a particle interpretation. The bending of waves around the slit edges is called ***diffraction*** and the combination of the light from the two slits to create the light and dark stripes is called ***interference***.

[4] The letter reads, in part, "What Des-Cartes did was a good step. You have added much several ways, and especially in taking the colours of thin plates into philosophical consideration. If I have seen further it is by standing on the shoulders of Giants. But I make no question but you have divers very considerable experiments besides those you have published, ..." (Letter from Isaac Newton to Robert Hooke, 1675, available at https://digitallibrary.hsp.org/index.php/Detail/objects/9792)

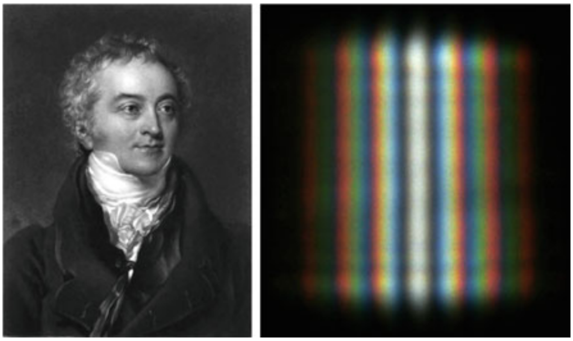


Figure 1.8 (Left) Thomas Young. (Right) Pattern on light and dark lines produced by shining light through two narrow slits, arising from diffraction and interference.

While performing research on electricity and magnetism, which seemed at the time to be completely separate from research on light, the Scottish scientist James Clerk Maxwell made the remarkable discovery that electric and magnetic fields could propagate as waves (Figure 1.9). Furthermore, he calculated that these waves would propagate at same speed as that of light which suggested, correctly as it turned out, that light was an ***electromagnetic wave***. It also led to the concept of other electromagnetic waves, including what we now call radio waves.

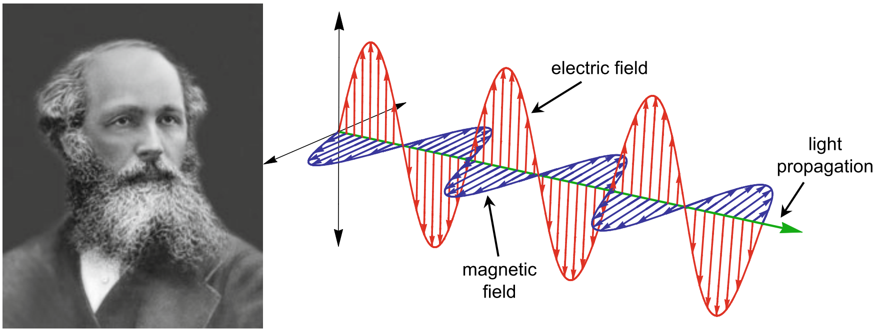


Figure 1.9 (Left) James Clerk Maxwell. (Right) Maxwell's description of an electromagnetic wave.

Heinrich Hertz experimentally verified these ideas a few years later. He produced radio waves by making an electric spark with one piece of equipment, those waves propagated across his laboratory, and they then induced a spark in a completely separate piece of equipment.

By the end of the 1880s, when Hertz had completed these experiments, there was no remaining doubt that light could be correctly described as being a wave. Like other types of waves, light refracted when it changed speeds, it diffracted when it went through narrow slits, and it interfered to create light and dark stripes. Also, it had been mathematically described as a wave with the theory of electricity and magnetism, and those explanations had been experimentally confirmed. At long last, the debate between particles and waves was over, with the wave theory being the decisive winner. Or so it seemed.

1.4 Particle-Wave Duality

In the 1890s, just a few years after Hertz had confirmed the wave nature of light, Max Planck was investigating the colors of light that are emitted by hot objects, which is called ***thermal radiation*** (Figure 1.10). After extensive effort, he came up with an equation that agreed with the experimentally observed emission essentially perfectly, where the close agreement suggested that his equation was probably correct. However, almost everyone, including Planck, thought that his equation had to be wrong because it was based on the assumption that light could only be emitted in integer amounts, which didn't make sense.

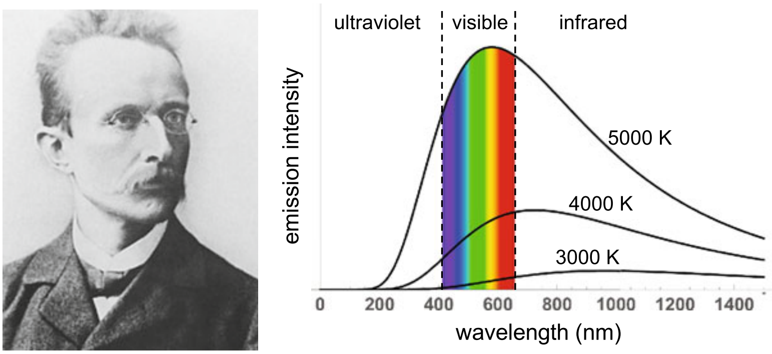


Figure 1.10 (Left) Max Planck. (Right) Emission spectra of hot objects, such as light bulb filaments.

Albert Einstein (Figure 1.11), an unknown Swiss physicist who was unsuccessfully trying to find a faculty position at the time, didn't worry about what made sense, but followed where the equations led. In this case, he made the bold proposal that perhaps Planck's integer amounts of light actually represented physical particles of light. Planck disagreed, writing "The theory of light would be thrown back not by decades, but by centuries, into the age when Christiaan Huygens dared to fight against the mighty emission theory of Isaac Newton..."[5]. However, Einstein wasn't claiming that light was *only* particles, but instead that light was both a particle and a wave at the same time, which is now called ***particle-wave duality***. This made even less sense, but there was strong evidence that it was correct because it immediately solved several other problems at the time. Most importantly, it solved the ***photoelectric effect***,

[5] Quoted from F. Todd Baker, "Atoms and Photons and Quanta, Oh My!: Ask the Physicist about Atomic Nuclear, and Quantum Physics", Morgan & Claypool Publishers, 2015, p. 2-4.

in which ultraviolet light can knock electrons off metal but visible light cannot. Similarly, it explains why we get sunburns from ultraviolet light but not from visible light.

Figure 1.11 Albert Einstein.

In another astonishing development, the French physicist Louis de Broglie argued that if light can be both a wave and a particle at the same time, then perhaps all "normal" particles are also waves. In particular, perhaps electrons, which were already known at the time to be tiny negatively charged particles, were both particles and waves. This far-fetched proposal was experimentally verified three years later when it was shown that electrons could exhibit diffraction and interference, just like other waves.

The wave nature of particles was further developed over the next several decades to create the field of quantum mechanics. It enabled physicists to finally answer a long list of questions that had eluded scientists for centuries, such as why atoms emit specific wavelengths of light (e.g. why neon signs are red), how chemical bonds work, and why metals conduct electricity but non-metals don't.

1.4.1 Today

Now, a century later, a vast amount of experimental evidence has accumulated that supports the notion of particle-wave duality for both light and matter. These results show that light and matter are non-intuitive in many ways but that it is nevertheless possible to build an understanding of how they behave. Furthermore, this understanding helps us make sense of many phenomena in the natural world, ranging from the shimmering iridescence of hummingbird feathers to the average temperature of the entire Earth. In addition, engineers have used this knowledge to build a tremendous array of modern technology, including lasers, computer chips, fiber optic communication networks, and solar panels.

1.5 Looking Ahead

Subsequent chapters delve further into these topics, exploring the nature of what light and waves really are. They start with an exploration of waves, including not only light but also string, sound, water, and other waves. We'll consider waves by themselves, such as how to measure their sizes and how fast they propagate. We'll then consider interactions between waves, including the diffraction and interference topics that were introduced above and which formed critical evidence in favor of the wave nature of light. We'll then move on to the interactions between waves and matter, leading to the absorption and emission of waves, followed by a chapter on sound, water, and other mechanical waves.

Next, the book takes a break from making sense of waves, and simply investigates what light does, by exploring shadows, reflection, and refraction. These chapters explain how mirrors and lenses work, along with rainbows and many other optical phenomena.

We'll then return to investigating waves in depth, this time focusing primarily on light. This naturally leads into quantum mechanics, a topic that can be quite advanced but is addressed on a more intuitive level here. Finally, we'll consider the most elusive waves of all, which are waves in the gravitational field.

1.6 Summary

Light has fascinated philosophers and scientists for thousands of years (Figure 1.12). Several theories about light began in roughly the 6th century BCE. The ancient Greeks came up with the extramission theory, in which vision works through emission of light rays from the eyes. Other ancient Greeks, including Aristotle, believed instead in the correct intromission theory, in which people's eyes only see the light that shines into them. Yet other philosophers believed in a particle theory of light, in which light (and everything else) is composed of tiny indivisible particles.

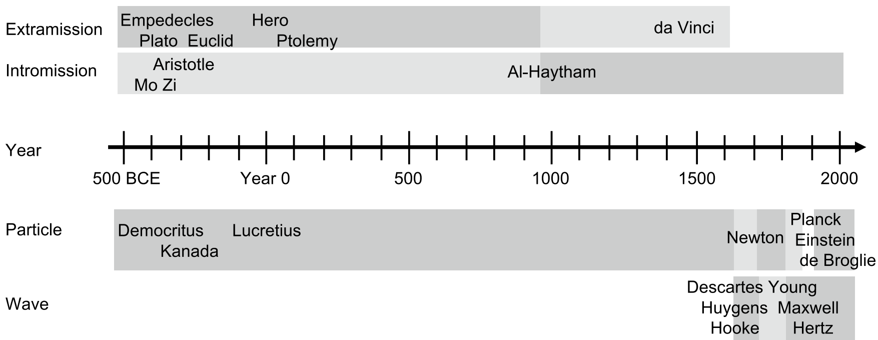


Figure 1.12 Timeline of different theories of light. Although necessarily imprecise, dark grey regions represent the dominant scientific view and light grey regions represent alternative views.

In the late 900s, Ibn al-Haytham provided strong evidence for the intromission theory, improved several of the optics theories that the ancient Greeks had developed, and developed the first reasonably correct explanation of human vision.

In the 1600s, Descartes showed that a wave theory of light was able to explain refraction. However, Newton denounced it based on his observations that light rays travel in perfectly straight lines, which returned the general consensus back to particles. The wave theory returned to the forefront when Young and others showed that light exhibits diffraction and interference. It was then thoroughly confirmed in the late 1800s by a theoretical explanation of light waves by Maxwell and its validation by radio wave experiments.

This conclusion was modified by scientific breakthroughs from the 1890s to 1920s that overturned the understandings of both light and matter. Einstein proposed that light exhibits particle and wave properties simultaneously, called particle-wave duality, which explained Planck's thermal radiation results as well as the photoelectric effect. Also, de Broglie and others showed that normal particles, such as electrons, are also waves, meaning that they also exhibit particle-wave duality. Further development of these ideas led to quantum mechanics, which explains atomic spectra, chemical bonds, metallic behavior, and other phenomena.

1.7 Exercises

Questions

1.1. What did Ibn Al-Haytham do?
   (a) correctly explained how vision works
   (b) promoted the particle explanation of light
   (c) promoted the wave explanation of light
   (d) developed the extramission theory
   (e) explained particle-wave duality

1.2. What did Isaac Newton do?
   (a) developed key ideas in calculus
   (b) developed key ideas in physics
   (c) performed experiments on light
   (d) promoted the particle theory of light
   (e) all of the above.

1.3. When do people's eyes emit light beams?
   (a) when they are angry
   (b) when they are feeling romantic
   (c) whenever their eyes are open
   (d) at night
   (e) never

1.4. What is light?
   (a) particles, not waves
   (b) waves, not particles
   (c) either waves or particles, but never both at once
   (d) both waves and particles at the same time
   (e) still unknown

1.5. What are electrons?
   (a) particles, not waves
   (b) waves, not particles
   (c) either waves or particles, but never both at once
   (d) both waves and particles at the same time
   (e) still unknown

1.6. Matching. For each word on the left, give its definition from those on the right.

(a) diffraction	(1) waves bend when changing speed (e.g. light shining into water)
(b) interference	(2) waves combine, sometimes forming light and dark stripes
(c) refraction	(3) waves bounce off surfaces
(d) reflection	(4) waves bend when going past sharp edges

1.7. What were the dominant beliefs about light during the early Middle Ages (5th to 10th centuries CE)? Choose one for each part. (a) Extramission or intromission theories, (b) finite or infinite light speed, (c) particles or waves?

1.8. Would human hearing be best described as using extramission or intromission methods?

1.9. Give at least two reasons for how we know that the intromission theory is correct and the extramission theory is incorrect.

1.10. List two experimental results that support the wave explanation for light.

1.11. List two experimental results that support the modern particle explanation for light (photons).

Puzzles

1.12 Ostriches supposedly stick their heads in the sand because they think that they can’t be seen then. Would this work if the extramission theory were true? Discuss. (This isn’t actually true; ostriches don’t actually stick their heads in the sand.)

Part I

Waves

Properties of Waves

2

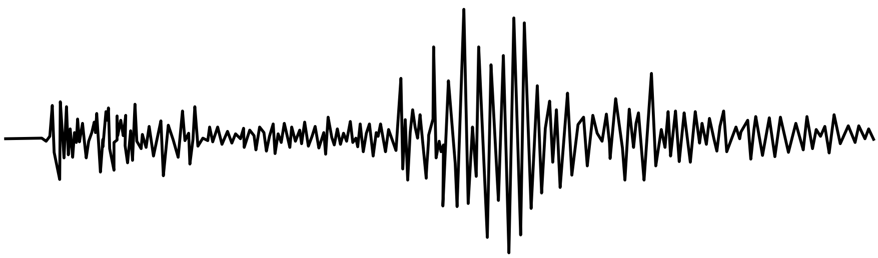


Figure 2.1 Seismogram of the first 30 seconds of California's 1989 Loma Prieta earthquake.

Opening question
A lightning strike creates a flash and "clap" of thunder. Which happen?
(a) Both sound and light transport energy; the sound also transports air
(b) Both sound and light transport energy; no air is transported
(c) The light transports energy but the sound does not
(d) Neither the sound nor light transport energy, although the air moves some

To better understand what light is, it helps to broaden our scope to investigate other types of waves as well. These include waves along taut strings, waves on the surface of water, and the waves of other bands of electromagnetic radiation, such as X-rays and radio waves. It also includes the wave-like properties of electrons, and the gravitational waves that ripple across the universe.

All types of waves involve displacements away from a "normal" state, and all can be characterized by their wavelengths, frequencies, and speeds. However, they also vary. They have different displacement types and different displacement directions. In some cases, they have billion-fold differences between short and long waves, or slow and fast waves. Some waves can propagate through empty space, whereas

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-3_2

others require a medium. This chapter explores these similarities and differences. In the process, it introduces the terminology and essential mathematics for waves.

2.1 Introduction to Waves

2.1.1 What Is a Wave?

Waves are disturbances that propagate. Consider, for example, shaking one end of a rope when the other end is being held fixed by a friend, a doorknob, or something else, as in Figure 2.2. As you move your hand up and down, you create a disturbance in the rope by displacing it away from the reasonably straight shape that it had initially. These disturbances then propagate away from your hand to become waves that travel along the rope.

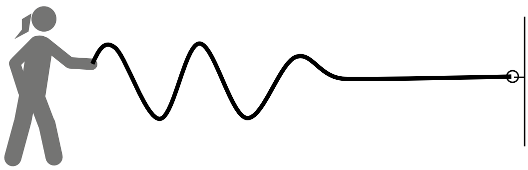


Figure 2.2 A wave on a rope.

The wave propagates because each tiny section of the rope is attached to the sections that are on each side of it. When a section gets pulled up by the section that's behind it, it moves upward. This pulls the next section up, and that pulls the next section up, and so on. A little while later, this section gets pulled down. It responds by moving down, which pulls down on the next section, and then the next section, and so on. During this process, each section of the rope only goes up and down a small amount, but the waves propagate along the length of the rope.

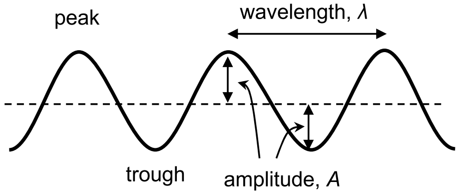
If you keep shaking the rope, you might notice your arm getting tired. This reflects the fact that your arm is constantly putting energy into the waves, which then carry the energy away from you and along the rope. Thus, another important property of waves is that they transport energy. Eventually, the doorknob, or whatever else the other end of the rope is tied to, might shake loose, again showing the transfer of energy from one end to the other.

2.1.2 Amplitude and Wavelength

Figure 2.3 shows a picture of a wave. This particular wave is a smooth up-and-down oscillation, called a ***sine wave*** or ***sinusoidal wave***; we'll consider more complicated wave shapes later on, in Chapter 3. The tops of the individual waves are called either ***peaks*** or ***crests*** and the bottoms are called ***troughs***.

The height of the peaks above the mid-level, which is the same as the depths of the troughs below the mid-level, is called the ***amplitude***. It is typically represented

Figure 2.3 Wave peaks, troughs, wavelength, and amplitude.



in math equations as A. The amplitude corresponds to the amount of energy in the wave. It is the difference between a gentle ocean swell and a storm-tossed sea; it determines the brightness of light, the loudness of sound, or the amount of shaking in an earthquake.

The distance between one peak and the next peak is called the ***wavelength***, represented with the Greek letter λ (lambda). The wavelength could also be measured from one trough to the next trough, or any other two corresponding points on the waves. It is the primary measure for the size of a wave, dictating its behavior, what it interacts with, and how we perceive it. It is the difference between long ocean swells, shorter wind-driven waves, and very short ripples; it is the pitch of a musical note, the color of light, or the type of electromagnetic radiation.

The sizes of the things that produce or interact with a wave are often similar to the wave's wavelength. For example, musical wind instruments, including everything from tin whistles to trombones, produce sound waves with wavelengths that are similar to the instrument's length; short instruments play high notes with short wavelengths (a few centimeters), and long instruments play low notes with long wavelengths (a few meters). As we will see many times later, we will define the sizes of things, such as mirror roughness, water depth, or sizes of obstructions, based on the dimensions of the object relative to the wave's wavelength. For example, waist-deep water is shallow when compared to 10 m waves, but deep relative to 10 cm waves.

2.2 Characterizing Waves

2.2.1 Types of Waves

There are many different types of waves.

First, there is a large class of waves called ***mechanical waves***, all of which arise from displacements within a physical object that is called the ***medium***. ***String waves*** are a type of mechanical wave; these are waves on strings, ropes, wires, and similar media, including the rope waves described above. A plucked guitar string is another example of a string wave. These are particularly simple because they propagate in a single dimension and are easy to create and visualize. ***Sound waves*** are mechanical waves in which the air is the medium. When a stereo speaker plays music, it alternately pushes and pulls on the air to displace it away from its normal uniform pressure, leading to pressure waves that propagate away from the speaker, which are sound waves. ***Seismic waves***, which create earthquakes, are like sound waves but have rock as their medium rather than air. They can travel through the upper levels of Earth, where the rock is solid, or near the middle of the Earth, where the rock is molten. They can also travel across the surface of the Earth, rather like water waves. Finally, ***water waves*** are mechanical waves that have water as their medium. These are the familiar waves that we see on puddles, lakes, the ocean, and even a plain glass of water.

All of these mechanical waves are transmitted through a medium, where the medium is displaced back and forth slightly as each wave passes through it, but the medium does not move as a whole. For example, sound is different from wind; sound is air moving back and forth, while wind is air moving in a constant direction. Likewise, ocean waves that crash on a beach might have been produced by a storm a thousand miles away, but the water molecules that wash against the shore were never in that storm. Instead, the storm's energy got transmitted to those molecules through water waves.

Mechanical waves clearly cannot exist outside of their media. A string wave cannot propagate past the end of the string and a seismic wave cannot propagate outside of the Earth. Explosions in deep space are always silent.

Electromagnetic waves, including light, are not mechanical waves and do not require a medium. Instead, they are oscillating electric and magnetic fields, each one of which creates the other one. In their case, the "normal" state of some region of space is to have no electric or magnetic fields (or, more precisely, electric and magnetic fields that don't change over time). Suddenly perturbing one of these fields, such as by moving a bunch of electrons from one place to another, creates a disturbance that then propagates away from that region as an electromagnetic wave. Based on analogies with mechanical waves, scientists assumed for many years that electromagnetic waves would require a medium as well, so they came up with the concept that all of space was filled with a light-transmitting medium called the ***luminiferous aether***. Maxwell expressed this assumption by writing "The undulatory theory of light also assumes the existence of a medium"[1]. However, it was troubling that this supposed aether didn't appear to exert any friction on the planets' orbits and then, more importantly, multiple attempts to detect the effect of the aether on the motion of light failed. Eventually, scientific consensus shifted to the conclusion that electromagnetic waves do not require a medium. We now know that outer space is truly empty, and that light waves travel through it perfectly well.

Electromagnetic waves include radio waves, microwaves, infrared light, visible light, ultraviolet light, X-rays, and gamma rays, shown in Figure 2.4. For convenience, we will sometimes call all of these light, where ***visible light*** is the electromagnetic radiation that we can see and ***invisible light*** is the electromagnetic radiation that we cannot see.

Matter waves are the wave natures of electrons, atoms, and other physical particles. These are not waves *in* matter, which are just familiar mechanical waves, but the quantum mechanical wave nature *of* the matter itself. In other words, an electron is both a particle and a wave at the same time, and its wave description is called a matter wave. As with light waves, matter waves are non-mechanical waves and do not require

[1] James Clerk Maxwell: "A Treatise on Electricity and Magnetism/Part IV/Chapter XX", 1873.

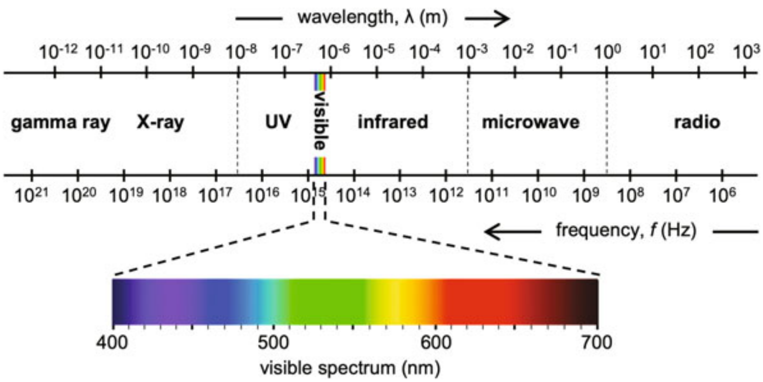


Figure 2.4 Electromagnetic spectrum, showing bands and wavelengths.

a medium. Matter waves are typically important for extremely small particles, like electrons, protons, and individual atoms but irrelevant for larger objects. Even a single molecule is generally large enough that its matter waves can be ignored. Although the mathematical descriptions of matter waves have been thoroughly worked out and those predictions invariably agree with experiments, the exact nature of what matter waves are still remains mysterious. Nevertheless, if we build on our statement that a wave is a disturbance that propagates, then this shows that the "normal" state for a region of space is that there is nothing there, and that the matter wave displaces this result to the possibility something might be there. We will largely postpone discussing matter waves until Chapter 14.

Gravitational waves are yet another type of non-mechanical wave, and again do not require a medium[2]. Gravitational waves are bizarre ripples in space itself or, more accurately, in space-time. Einstein predicted their existence in 1916, showing that they must arise whenever any object that has mass, meaning any object at all, is accelerated. This means that accelerating cars, falling boulders, and orbiting planets all produce gravitational waves. However, they are extraordinarily weak and so only become measurable when masses and accelerations are extremely large. The first experimental evidence for gravitational waves was not found until 1974, when two astronomers discovered that a pair of neutron stars that orbited each other lost energy at a rate that was consistent with it being carried away by gravitational waves. After a long search, gravitational waves were first detected directly in 2015. Again, we will largely ignore this topic for now, but will return to it in Chapter 15.

[2] Note that gravitational waves should not be called "gravity waves." This is because ordinary water waves were called "gravity waves" long before gravitational waves were conceived of, so the term "gravity waves" still refers to water waves.

2.2.2 Transverse and Longitudinal Waves

It can be helpful to categorize waves by considering the direction of the displacement relative to the direction of wave propagation. There are two options: waves are ***transverse*** if the displacement is perpendicular to the propagation direction and ***longitudinal*** if the displacement is parallel to the propagation direction (Figure 2.5). Thinking back to the rope example from before, those were transverse waves because the rope was displaced up and down while the waves travelled forward, which are perpendicular axes.

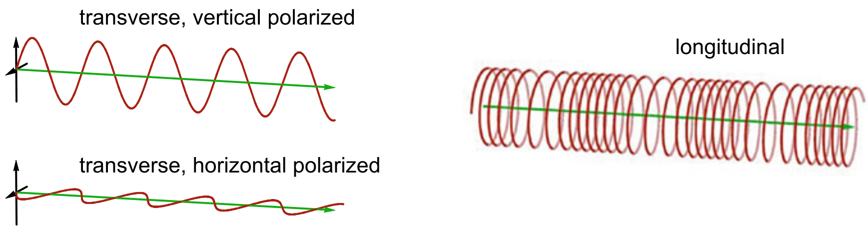


Figure 2.5 Transverse and longitudinal waves.

All transverse waves have two possible ***polarizations***, describing the displacement direction. In the rope example, it was shaken up and down, so the rope was displaced vertically and those waves were vertically polarized. If it had been shaken left and right instead, then the waves would have been horizontally polarized. Figure 2.5 shows both possibilities.

Light waves are another example of transverse waves, in their case because both their electric and magnetic fields, which are their displacements, are perpendicular to the direction of wave propagation (see Figure 1.9). Light waves can also be polarized either vertically or horizontally, which is defined in their case by whether the electric field goes up-and-down or left-and-right (the electric field was chosen because it's typically more important than the magnetic field).

Longitudinal waves occur when there is compression and spreading along the direction of the wave propagation. A Slinky toy that is first stretched out on a table and then given a quick push at one end forms a nice example. The quick push creates a compressed region that travels the length of the Slinky as a longitudinal wave. Longitudinal waves cannot exhibit polarization because there is only one possible axis for the displacement. Sound waves are a particularly important type of longitudinal waves. They can be created by a stereo speaker alternately pushing and pulling on air, much like a person creates Slinky waves by pushing and pulling on a Slinky.

Water waves are interesting because they are both transverse *and* longitudinal. The transverse part is obvious; if you observe water waves, it's clear that the displacement goes up and down while the waves move forward, so the displacement is perpendicular to the direction of propagation and they are transverse waves. Considering the longitudinal part, suppose you watch a cork or some other object that is floating on the water as waves pass by it. You would see that the cork doesn't just go up and down but actually moves in a circle, going up, then forward, then down, and then backward, round and round. The forward and backward motions represent the longitudinal components of water waves.

Seismic waves are even more complicated, exemplifying all of these wave motions. They are generated when rocks scrape past each other deep underground, at the ***earthquake hypocenter*** or ***earthquake focus***[3]. One outcome of this is that it creates sudden pressure changes in the surrounding rock, which then propagate outward through the rock as pressure waves. Except for the fact that the medium is rock rather than air, they are just like sound waves and, likewise, are longitudinal waves. These pressure waves are called ***P-waves***, which is officially short for ***primary-waves*** because they travel faster than other seismic waves and so are usually the first to get detected; however, it can be easier to remember them by thinking of P-waves as standing for "pressure-waves." The rock sliding motion also creates side-to-side displacements in the surrounding rock, which propagate outward as transverse waves. These are called ***S-waves***, which officially stands for ***secondary-waves*** because they are the second wave type to get detected; again though, these can be easier to remember by thinking of the "S" as standing for "shear," "sliding," or "sideways." Like all transverse waves, S-waves have two possible polarizations, vertical and horizontal. Both the P-waves and S-waves change when they get to the surface of the Earth. There, they become yet different types of waves and then travel along the Earth's surface with both transverse and longitudinal motions, rather like water waves.

2.2.3 Waves in 1, 2, and 3 Dimensions

Transverse waves on strings, and longitudinal waves on a Slinky, are ***one dimensional waves***, meaning that they take place along a relatively long and thin object. The fact that waves displace the string away from a straight line doesn't affect the fact that they are still considered to be one dimensional.

Water waves are ***two dimensional waves*** or ***surface waves*** because they take place at the two-dimensional surface of the water (Figure 2.6). Again, the waves displace the surface away from being totally flat, but these waves are still called two dimensional. The membrane head of a drum also supports two dimensional waves,

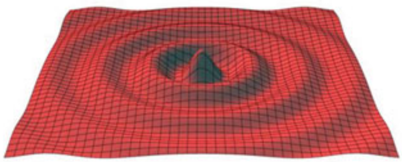


Figure 2.6 Transverse surface wave with vertical polarization.

[3] The earthquake epicenter is the point on the Earth's surface that is directly above the hypocenter.

3 Superposition

Figure 3.1 Ripples produced by rain falling on water.

Opening question

Two stones are dropped into a pond, each producing ripples. When the sets of ripples meet, which happens:
(a) The ripples reflect off each other, each going back toward where it came from.
(b) The ripples add together when they overlap and then pass through each other.
(c) The ripples cancel out wherever they meet, producing flatter water.
(d) The ripples bend around each other, creating complex wave patterns.

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-3_3

When two tennis balls collide, they bounce off each other. However, when two waves collide, they don't. Instead, they smoothly pass through each other, temporarily combining to form new wave shapes in the process. Once they have finished passing through each other, they continue on their separate ways as if nothing had ever happened.

This process is stunningly simple, arising from nothing more than the addition of the two sets of wave displacements. However, it leads to a broad array of interesting behaviors. Two sets of waves can add to create bigger waves, smaller waves, or even no waves at all. They can also add to create waves that oscillate in time but don't move in space. The principles of wave addition explain how sound waves can curve around buildings, why oil drops on wet pavement produce colorful rings, and how holiday laser lights can create elaborate designs on walls. They even explain how light waves can add together to produce bright spots in the centers of reasonably large shadows.

Perhaps the showiest examples arise in the brilliant colors of many animals, including the shimmering iridescence of hummingbirds, the vibrant blue of blue morpho butterflies, and the gaudy colors of peacock tails.

3.1 Superposition of Waves

3.1.1 The Superposition Principle

Suppose a rope is being held by one person at each end. Then, both people put ***pulses*** into the rope, where a pulse is simply a brief disturbance. The pulses propagate toward each other, as in Figure 3.2. What happens when the pulses meet?

Figure 3.2 Two people putting pulses into a rope that travel toward each other.

The answer, in a sense, is very uninteresting. The pulses don't interact in any way at all. They don't collide off each other, annihilate each other, break the rope, or do anything else dramatic. Instead, their displacements simply add together as the pulses pass each other, which is called the ***superposition principle***. It is shown in Figure 3.3 for two pulses with either the same ***polarity*** or opposite polarity, where the polarity is the side of the rope that the pulse is on. Initially, the pulses are separate and moving together, then their displacements add to form either extra large or extra small displacements when the pulses are at the same place, and finally the pulses pass each other. They then continue on separately, with no remaining evidence of their prior interactions.

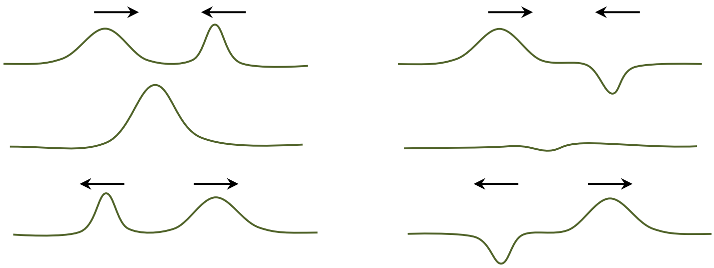
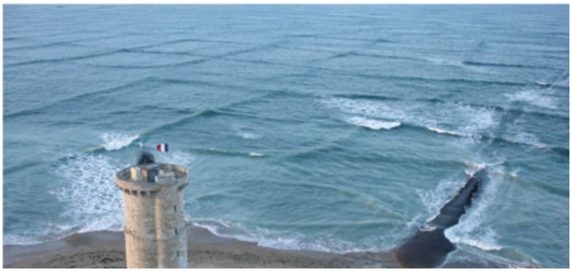


Figure 3.3 Superposition of pulses on a rope. The two pulses have the same polarity on the left side and opposite polarity on the right side. The top row shows the pulses before they meet, the middle row shows them meeting, and the bottom row shows them after they have passed.

The superposition principle holds true regardless of pulse lengths, pulse shapes, pulse amplitudes, and pulse polarity. It also applies to all types of waves, including water waves, sound waves, and light waves. For example, Figure 3.4 shows the superposition of two sets of water waves. Here, the water displacement at any point on the surface is the sum of the displacements from each of the two sets of waves. Likewise, if two people talk at the same time, then their sound waves add together to include both sounds at once. Also, if two laser beams pass through the same point in space, then their fields add together where they cross, and the beams propagate on as if nothing had ever happened. These behaviors are sufficiently important, and different from what happens with solid objects, that the superposition principle is a defining characteristic of wave behavior[1].

Figure 3.4 Superposition of water waves.



[1] Despite the widespread validity of the superposition principle, which is assumed in almost all parts of this book, there are some exceptions. These arise from ***non-linear effects***, which are simply defined as wave behaviors that violate the superposition principle. Nonlinear effects are typically negligible for low amplitude waves, but can become important when waves have large amplitudes. They are studied in the fields of nonlinear optics and nonlinear acoustics. The primary situation in which nonlinear effects are commonly observed is for water waves, where they give rise to phenomena such as wave breaking (note that some high amplitude waves are breaking in Figure 3.4) and wave development over time.

3.1.2 Superposition with Different Frequencies

Wave superposition often gives real waves rough and complicated shapes. For example, the left side of Figure 3.5 shows multiple ***wave components*** adding together on the ocean: there are wind-driven waves, ripples from rain drops, and the boat's ***wake*** (waves made by the boat), all added together. The displacements from these different wave components add to create the resulting water surface shape. It is somewhat possible to identify wave peaks and troughs in this photograph, but they aren't well defined. The waves' wavelengths and amplitudes are even less well defined. The right side of Figure 3.5 helps make sense of this situation by showing that the sum of two sine waves that have different wavelengths, each of which is a separate wave component, also creates a complicated result. However, here, it's possible to see which features arose from which component. As in the photograph, the peaks, troughs, wavelengths, and amplitudes are well-defined for the component waves but not for the total wave.

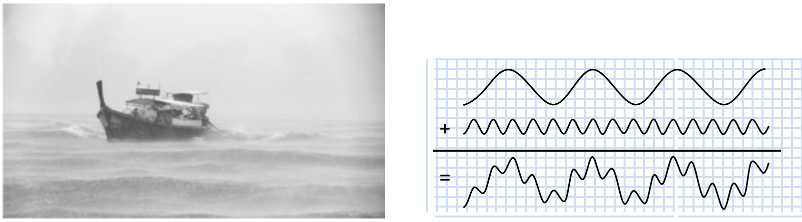


Figure 3.5 Superposition of waves with different wavelengths. (Left) Waves from wind, rain, and boat wake all added together. (Right) Two sine waves with different wavelengths added together.

Remarkably, the summing process can also be reversed, allowing one to decompose a final wave shape into its separate components. This can be performed mathematically using a method called the ***Fourier transform***, named for the French mathematician Jean-Baptiste Fourier, which is central to the quantitative analysis of oscillations and waves but is beyond the scope of this book. However, waves can also be decomposed experimentally in some cases. For example, white light can be separated into the different colors that compose it using a prism, producing a rainbow. Each of the colored rays leaving the prism has a single well-defined wavelength. Also, our ears decompose sound waves into their separate frequencies, detecting the high frequency components just inside the ear drum and the lower frequency components deeper into the spiral shaped tubes of the inner ear.

An important point here is that all real waves, regardless of how complex their shapes might be, are nothing more than sums of simple waves, such as sine waves or smooth pulses. This means that we can focus our attention on learning how simple waves work, knowing that they can always be added back together to give complete wave shapes.

Color

10

Figure 10.1 Painting titled "Last Painter on Earth" by James Doolin (1932–2002), which depicts Last Chance Canyon, California. Painted 1983, 72 × 120 inches, oil on linen.

Opening question

A painter mixes together equal amounts of magenta and cyan paints. What color paint does he create?

(a) Red
(b) Black
(c) Blue
(d) Green
(e) Violet

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-3_10

Think about a bright red strawberry. It's undoubtedly red, but what does it mean to be red? Would it still be red if you closed your eyes? What if it were illuminated with only blue light? Might it be a red strawberry to you, but an orange strawberry to someone else, a green strawberry to your pet dog, and a color that we can't even imagine to a parrot?

Fundamentally, color is about the perception of light by our eyes and brains. It arises from the different frequencies of light waves that enter our eyes, but doesn't truly attain the concept of color until those light waves are absorbed in our eyes and the resulting signals are transformed into mental images in our brains. Perhaps surprisingly, most people seem to transform these signals in nearly the same way, so colors appear about the same for everyone. However, there are exceptions. Color perception is different at night than during the day, it depends to some extent on the culture that people grew up in, and, of course, it is different for people who are color blind. Animals are different yet, often perceiving either more or fewer total colors than people can.

The biology of color perception also leads to interesting effects when different colors are combined. For example, the red strawberry would appear neither red nor blue when illuminated with blue light, but black. Color mixing, which is now highly scientific, creates all of the colors that we see on printed materials, phone displays, and television screens.

10.1 Color Vision

10.1.1 How Vision Works

The left panel of Figure 10.2 shows a ray of light entering an eye. It is focused by the cornea, anterior chamber, and lens to create an image of what we're looking at on the eye's retina, which is at the back of the eye. At the retina, shown in the middle panel, the light shines through several layers of signaling cells, through the majority of several ***photoreceptor cells***, and finally gets absorbed by bent ***retinal*** pigment molecules that are toward the back side of the photoreceptor cells[1]. When a retinal molecule absorbs a photon of light (right panel), the light's energy momentarily breaks a weak chemical bond. This allows the retinal molecule to snap into a lower-energy straight shape. The shape change then distorts nearby proteins, which initiates a chemical signal within the cell. That signal gets passed back through the photoreceptor cell and into the signaling cells, which do some image processing such as to detect motion, object edges, and colors. These processed signals then continue on to the brain where they get combined with many other signals to form a mental image.

[1] Retinal is a form of Vitamin A, which the body synthesizes from carotene, which we ingest when eating carrots and other vegetables. Hence the common advice to eat carrots to improve your vision.

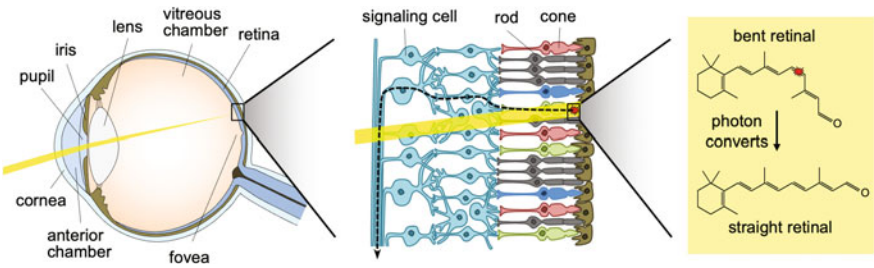


Figure 10.2 (Left) A diagram of a yellow light ray entering a human eye and being absorbed at the retina. (Middle) A small portion of the retina, showing signaling, rod, and cone cells. Cone cells colors show whether a particular cell is sensitive to red, green, or blue light. The dark patch inside each cell is its nucleus. The dashed line shows a signal going from a triggered cell to the brain. (Right) Retinal molecules that are within rod and cone cells. Light absorption temporarily breaks one of the chemical bonds, enabling the molecule to snap into a straight shape.

There are two types of photoreceptor cells, ***rods*** and ***cones***. Rods are extremely sensitive to light, even able to be triggered by a single photon. They are also very sensitive for observing motion, due to the ways that they are connected together by signaling cells. Rods are abundant over almost the entire retina, but *not* in the spot at the center of our vision, which is called the***fovea***. The fact that the fovea has very few rods means that the center of our vision isn't very good in low light conditions, such as at night. On the other hand, the rods over the rest of the retina give us sensitive ***peripheral vision***, meaning vision of things that we are not looking at directly. For example, you may have noticed that a flashing light at night, such as from an airplane that's flying overhead, appears bright when you see it "out of the corner of your eye", meaning with your peripheral vision; however, it doesn't appear as bright when you look at it directly.

Cones are the dominant photoreceptor cells in the fovea. They are much less sensitive to light than rods, but they enable color vision by coming in three separate types; one is primarily sensitive to red light, another to green light, and the third to blue light. Because there are many cones in the fovea, we are good at seeing the colors of things when we look directly at them.

Both rods and cones detect light using the same retinal molecules, but vary what wavelengths these molecules are sensitive to by surrounding them with different proteins[2]. Figure 10.3 shows the resulting spectral sensitivities. It shows that "red" cone cells have at least some sensitivity from red to violet light, but are most sensitive to yellow light, "green" cones are sensitive to slightly shorter wavelengths, and "blue" cones are only sensitive over the range from cyan to violet. These blue cones are sensitive into the ultraviolet too, but we can't see that light because it is blocked by the lenses in our eyes.

[2] The proteins around the retinal molecules, called opsins, have electrically charged atoms in them. These atoms exert electric fields on the retinal molecules, and the different electric fields, arising from slightly different protein structures, create different retinal absorption spectra through a mechanism

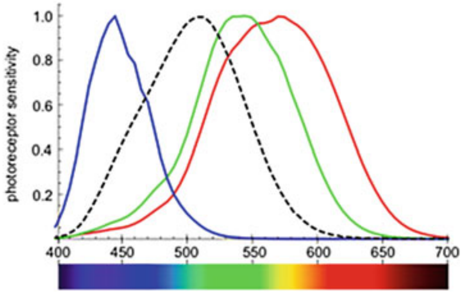


Figure 10.3 Spectral sensitivity of human eye photoreceptors. The black dashed line is for rods and colored lines are for cones. These graphs have been normalized so that each one has a peak value of 1.

The fact that the sensitivity curves overlap implies that most light wavelengths are able to excite multiple photoreceptor cell types. These individual cells can't detect what color of light they receive, but simply report the amount of light that they detect. As a result, it is up to the neighboring signaling cells and then the brain to determine the color of light. They do this by combining the information that they receive from many different cone cells, along with their knowledge of which photoreceptor cell is which[3].

For example, suppose you look at a street light that emits yellow light at 589 nm. From Figure 10.3, your red cone cells would be fully excited, your green cone cells would be about 50% excited, and your blue cone cells would not be excited at all. Your brain has learned that this pattern of excitation corresponds to yellow light, so the light appears yellow to you.

10.1.2 Light and Dark Adaptation

When we walk from bright sunlight into a dark room, we see only blackness initially. Soon though, our vision returns as our eyes transition from being ***light adapted*** to ***dark adapted***. A quick but relatively minor part of this adaptation process is that the irises of our eyes expand to let more light through our pupils (when looking at someone's eye, the pupil is the central black region and the iris is the brown, blue, or green ring that surrounds it). Further dark adaption occurs as the photoreceptor cells gradually convert straight retinal, which was made straight by the large quantities of sunlight that had been entering our eyes before, into bent retinal. Having more bent retinal means that more of the light entering our eyes gets absorbed, thus making it easier to see when there isn't much light.

If the room is dark enough, such as a dark movie theater, it becomes too dark for our cones to be of much use, leading us to see almost exclusively with our rods.

called the Stark effect. See Kochendoerfer, Gerd G., et al. "How color visual pigments are tuned." *Trends in biochemical sciences* 24.8 (1999): 300–305.

[3] While it might seem reasonable that the rod cells would contribute to color perception, using the fact that their sensitivity spectra are different from those of the cone cells, they do not actually appear to do so.

Matter Waves

14

Figure 14.1 Neon dragon, from Museum of Neon Art, Glendale, CA.

Opening question
What causes atoms to have distinct energy levels?
(a) Standing electron waves
(b) Quantization of light into photons
(c) Electrons bumping into each other
(d) Interactions between neighboring atoms
(e) It's still unknown

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-3_14

Look around and observe your surroundings. You probably see furniture, people, buildings, trees, and other familiar objects, all with their respective shapes and colors. These are large objects, in contrast to the truly tiny world of electrons and protons, where quantum mechanics becomes important. Also, these objects behave normally, unlike the bizarre world of quantum mechanics where waves are particles, and particles are waves, and nothing can be known exactly.

And yet, it could also be said that essentially everything about your surroundings is the way it is because of quantum mechanics. The wave-like properties of electrons create most of the colors that you see. They also explain how electrons, protons, and neutrons make atoms, how atoms bind together to form molecules, and how molecules bind together to form liquids and solids. Quantum mechanics explains why air is transparent, blood is red, and security lights are yellow. It is also the theory behind much modern technology, from computers to lasers, neon signs, and even fabric whiteners.

This chapter introduces the basic principles of quantum mechanics. Remarkably, most of them are just the same physics of waves that we've already seen.

14.1 Matter Waves

14.1.1 De Broglie Relations

Louis de Broglie was a French aristocrat who majored in history and intended to pursue a career in humanities. These plans were derailed by World War I, in which he served as a radio engineer; then, his brother inspired him to turn his attention to physics, where he ended up writing one of the world's more famous PhD dissertations. In it, he extended Einstein's discovery that light is made of particles to propose that matter is made of waves.

These matter waves needed to have wavelengths and frequencies, so de Broglie simply rearranged the two most important equations for photons — the photon momentum equation, which gives a photon's momentum as $p = h/\lambda$, and the Planck-Einstein relation, which gives a photon's energy as $E = hf$ (see Chapter 13). This led to the ***de Broglie relations*** for matter waves,

$$\lambda = \frac{h}{p} \qquad \text{and} \qquad f = \frac{E}{h}. \tag{14.1}$$

The former equation is the ***de Broglie wavelength*** and the latter is the ***de Broglie frequency***. As usual, h is Planck's constant, equal to $6.626 \cdot 10^{-34}$ J s.

In contrast to the situation for photons, de Broglie set the momentum and energy values in these equations to their standard values from classical physics. The classical momentum of an object is

$$p = mv_{part.}, \tag{14.2}$$

where m is its mass and $v_{part.}$ is its velocity; the "part." subscript clarifies that this is the velocity of the particle, as opposed to its matter wave, which will turn out to be relevant later on. The classical energy of an object is

$$E = \frac{1}{2} m v_{part.}^2 + U(x). \tag{14.3}$$

The first term is the kinetic energy, meaning the energy of the particle's motion. The second term, $U(x)$, represents the object's potential energy as a function of its position, which we'll return to in the next section. These equations show that heavier and/or faster objects have more of pretty much everything. They have more energy, more momentum, and faster frequencies. On the other hand, they have *shorter* wavelengths (like light waves, where shorter wavelengths correspond to higher energy photons). Quantum effects are most important when objects have *long* wavelengths, so electrons and other lightweight particles tend to exhibit the strongest quantum effects.

These four equations form the foundation of ***non-relativistic quantum mechanics***. This is the physics of systems that are small enough that particle-wave duality is important and that move slowly enough that Einstein's theory of relativity can be safely ignored[1]. Most of the mathematical results in the rest of this chapter follow directly from these equations.

Example. Find de Broglie wavelengths for (a) a helium atom at a temperature of 2 K, which moves at 100 m/s (mass is $6.65 \cdot 10^{-27}$ kg), (b) a 0.059 kg golf ball moving at 60 m/s.

Answer.
(a) Use the de Broglie wavelength equation for helium:

$$\lambda = \frac{h}{p} = \frac{h}{m v_{part.}} = \frac{6.626 \cdot 10^{-34}\ \mathrm{J\,s}}{(6.646 \cdot 10^{-27}\ \mathrm{kg})(100\ \mathrm{m/s})} = 1.0 \cdot 10^{-9}\ \mathrm{m}.$$

(b) For the golf ball:

$$\lambda = \frac{h}{p} = \frac{h}{m v_{part.}} = \frac{6.626 \cdot 10^{-34}\ \mathrm{J\,s}}{(0.059\ \mathrm{kg})(60\ \mathrm{m/s})} = 1.9 \cdot 10^{-34}\ \mathrm{m}.$$

The helium atom has a "large" wavelength of about 1 nm, which is similar to the atom's size, so quantum effects are likely to be important. In contrast, the golf ball's wavelength is so tiny that quantum effects are completely irrelevant.

[1] The de Broglie relations are still valid with relativity, but the momentum and energy equations need to be replaced by their relativistic versions. These are: $p = \gamma m_0 v$ and $E = \gamma m_0 c^2$, where m_0 is the rest mass and $\gamma = 1/\sqrt{1 - \frac{v^2}{c^2}}$. This relativistic energy includes the mass energy along with the kinetic energy, which increases the de Broglie frequency and wave phase velocity, but doesn't affect any observable results.

14.1.2 Wave Functions

The news of de Broglie's work spread quickly and generally received a positive reception. For example, Einstein wrote that de Broglie had "lifted a corner of the great veil"[2]. However, de Broglie's wave theory was incomplete because it didn't explain how the matter waves moved or even what they really were.

Erwin Schrödinger, an Austrian physicist, made the next major advance. While on a Christmas ski vacation with his mistress[3] in the Swiss Alps in 1925, he came up with the idea that each particle was represented by a ***wave function***, given by $\psi(x)$. This function was a wave that propagated through space, much like electromagnetic waves, but was for matter instead of light. He also figured out an equation, now called the ***Schrödinger equation***, that described how these waves propagated[4].

One of Schrödinger's insights was that the mathematics became simpler if the wave function displacements were complex numbers. Complex numbers have two parts to them, called the real and imaginary components; for example, $2 + 3i$ is a complex number in which the 2 part is real and the $3i$ part is imaginary. The i factor represents the value of $\sqrt{-1}$, which has no physical meaning but allows for some clever math such as the facts that $i^2 = -1$ and $i \cdot (-i) = 1$.

The top of Figure 14.2 shows a wave function for a particle as a function of the position along the x axis, with its real component on the vertical axis and its imaginary component on the axis that comes out of the page. As with photons, matter waves are best understood as probability waves, where large displacements imply high probability of finding the particle at that location and small displacements imply low probabilities. More precisely, the actual probability of the particle being at some position along the x-axis is given by the ***absolute square*** of the wave function at that location; the absolute square is essentially the same as the square of a number, but is a purely real number without any imaginary component[5]. The bottom of Figure 14.2 shows this probability; it shows the likelihood of the particle being at each position along the x-axis. The area under this probability function is equal to 1, meaning that the electron has 100% probability of being somewhere.

---

[2] APS News, "This Month in Physics History. October 18, 1933: Louis de Broglie elected to Academy" vol. 19, no. 9, p. 2, October 2010.

[3] In addition to his physics work, Schrödinger is well known for his womanizing, such as that he had three daughters by three different mistresses, and he effectively had two wives for a while. There doesn't seem to be a record of which mistress accompanied him during the Christmas 1925 trip. See "The lone ranger of quantum mechanics" by Dick Teresi, *New York Times* Section 7, page 14, Jan. 7, 1990.

[4] The Schrödinger equation for wave propagation in one dimension is the differential equation $i\hbar\frac{\partial}{\partial t}\psi(x,t) = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U(x,t)\right]\psi(x,t)$, where $\hbar = \frac{h}{2\pi}$ and $U(x,t)$ is the potential energy as a function of position and time.

[5] To take the absolute square, one multiplies a number by its complex conjugate, where the complex conjugate is the same as the original number except that the signs of the imaginary components are reversed. For example, the complex conjugate of $2 + 3i$ is $2 - 3i$ and the absolute square of $2 + 3i$ is $(2 + 3i)(2 - 3i) = 4 + 6i - 6i - 9i^2 = 4 + 9 = 13$. Complex conjugates are denoted by asterisks, so the absolute square of $\psi(x)$ is $\psi(x)\psi^*(x)$.

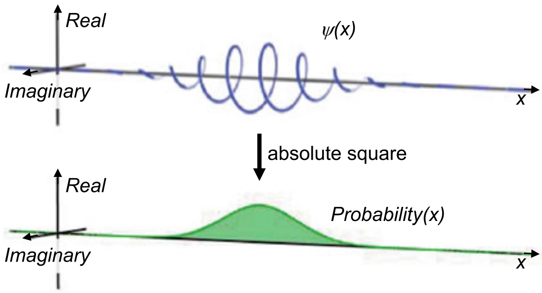


Figure 14.2 (Top) A matter wave function, showing a wave packet that is propagating along the x-axis. (Bottom) The absolute square of the wave function, which is the probability of the particle being at any particular position along the x-axis.

14.1.3 What is Waving?

While de Broglie and Schrödinger showed that matter waves were clearly important, a persistent question was what was actually waving. This was expressed by the German scientist Erich Hückel (and translated into English by the Swiss physicist Felix Bloch) in the poem

> Erwin with his psi can do
> Calculations quite a few.
> But one thing has not been seen:
> Just what does psi really mean.

The answer to this mystery is that there simply is no physical reality for matter waves. Instead, matter waves are only mathematical entities; they are figments of our imagination[6]. That said though, matter waves behave like other waves and lead to physically sensible predictions, such as the result that squaring the wave displacements produces the probability of finding the corresponding particle at any given position. They also enable accurate, elegant, and sometimes even intuitive explanations of the physical world. Thus, as most physicists do, we will treat them here as though they are actual waves.

14.2 Traveling Matter Waves

14.2.1 Free Particles

Consider a particle, such as an electron, proton, or atom, that's just moving along through space, with nothing in its way. This ***free particle*** has some kinetic energy, because it's moving, but it doesn't have any potential energy because it's not inter-

[6] Even Niels Bohr, one of the founders of quantum mechanics, questioned their reality; he wrote "There is no quantum world. There is only an abstract quantum mechanical description. It is wrong to think that the task of physics is to find out how Nature *is*. Physics concerns what we can say about Nature." From A. Peterson, *Bulletin of the Atomic Scientist* 19:12, 1963.

acting with anything. This means that we can set the potential energy to zero,

$$U(x) = 0. \tag{14.4}$$

If we define the particle's mass as m and its velocity as $v_{part.}$, we can then compute its momentum, energy, de Broglie wavelength, and de Broglie frequency:

$$p = mv_{part.} \qquad E = \frac{1}{2}mv_{part.}^2 \qquad \lambda = \frac{h}{mv_{part.}} \qquad f = \frac{mv_{part.}^2}{2h}. \tag{14.5}$$

We want to learn more about the particle's matter waves, so we compute their velocity from the frequency-wavelength relation, $v_{phase} = \lambda f$, which is true for all types of waves. The "phase" subscript reminds us that this is the wave phase velocity, meaning how fast the wave peaks move. Substituting in from above and cleaning up lead to

$$v_{phase} = \lambda f = \frac{h}{mv_{part.}} \cdot \frac{mv_{part.}^2}{2h} = \frac{v_{part.}}{2}. \tag{14.6}$$

This is a surprising result. It says that the wave peaks don't move at the same speed as the particle. Instead, they only move half as fast.

This is reminiscent of the situation with water waves, in which a group of waves travelled with a "group velocity" that was different from the waves' "phase velocity". In that case, the difference arose from dispersion, meaning that the phase velocity depended on the wavelength. To see if the same thing occurs here, we rearrange the de Broglie wavelength to $v_{part.} = h/m\lambda$ and then substitute that into Eq. 14.6 to give

$$v_{phase} = \frac{h}{2m\lambda}. \tag{14.7}$$

Indeed, this shows that the matter wave phase velocity depends on the wavelength, implying that they are dispersive.

The fact that matter waves are dispersive has a couple of important consequences. First, as just discussed, they have different phase and group velocities. We won't derive it here, but it can be shown that the group velocity is the same as the particle's velocity. This makes sense because the group of waves is the particle, so it should move at the same speed as the particle. On the other hand, the phase velocity is largely meaningless; it's half as fast as the group velocity in the example shown, but can be different in other situations (the phase velocity would have been different if we had set the potential energy to some value other than zero in Eq. 14.4).

Second, it means that matter waves spread out as they propagate. Figure 14.3 illustrates this by showing the matter waves at three time points for a single particle that moves from left to right. This spreading is the same as that seen in water waves produced from a rock that's dropped in a pond, where a single initial wave spreads out to a broad group of waves. In terms of the matter waves, it means that the particle starts with a narrow range of possible positions initially but then quickly spreads out to having a wide range of possible positions.

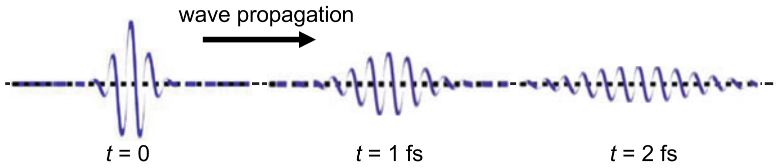


Figure 14.3 The same matter wave at three points in time, showing that it spreads out. The time is measured in femtoseconds (10^{-15} s).

14.2.2 Classical and Quantum Roller Coasters

For particles that aren't free, we need to consider how the potential energy varies over space. The left panel of Figure 14.4 illustrates this by showing a normal roller coaster. The cars start at the top of the hill at the left end, where they have high potential energy, and roll down the hill, speeding up as they go. In the process, they convert potential energy to kinetic energy, which reaches its greatest value when the cars reach the bottom of the first dip (blue cars). The energy conversion process then reverses as they coast up the next hill, where they slow down and convert their kinetic energy back to potential energy, eventually going slowly over the top of the hill (orange cars). Then they go down and speed up again.

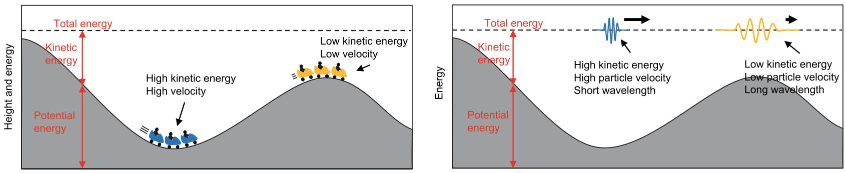


Figure 14.4 (Left) A classical roller coaster. (Right) A quantum roller coaster. Gray shading represents a potential energy surface.

These hills can be seen as a ***potential energy surface***, meaning that they represent the cars' potential energy as a function of their position. As a math expression, it is the $U(x)$ function introduced above. The cars' total energy, shown by the dashed black line at the top of the figure, stays constant over the whole journey due to the conservation of energy (we ignore friction). Subtracting the potential energy from this total energy yields the kinetic energy, which can be visualized in the figure as the distance between the potential energy surface and the dashed line (upper red arrow). For example, we can immediately see that the cars' kinetic energy is low at the hill tops and high at the hill bottoms.

The right panel shows a quantum roller coaster in which two particles move from left to right, each represented by a wave function. The gray region still represents a potential energy surface, now created by electrical forces rather than gravity but that doesn't really change anything, and the dashed line still represents the total energy. The distance between these heights still represents the kinetic energy. The only changes are stylistic: (1) this diagram shows wave functions instead of cars, and

15 Gravitational Waves

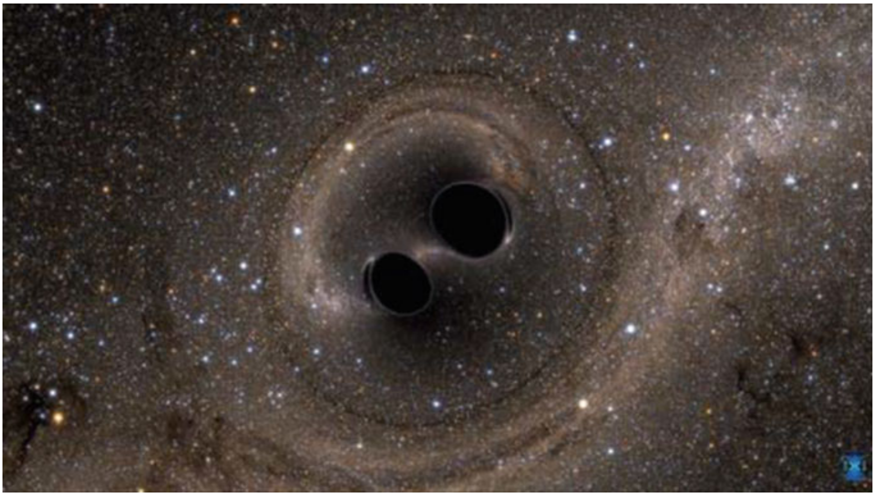


Fig. 15.1 Simulated appearance of two black holes orbiting each other, which produces gravitational waves. The swirl pattern arises from gravitational lensing of the background stars.

Opening question
What does gravity do? Select all that are appropriate.
(a) Keeps planets and moons in their orbits
(b) Can either attract or repel objects
(c) Bends light rays
(d) Warps space and time
(e) Propagates in waves

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-3_15

A new era of astronomy began at 9:50 am UTC on September 14, 2015. Throughout the entirety of human history before then, every direct observation of the universe beyond our solar system had been performed by observing electromagnetic radiation. People only observed visible light up until the late 1800s and then infrared, ultraviolet, and other bands more recently, but it was always some sort of electromagnetic radiation. This changed on that morning in September 2015 with the first direct detection of gravitational waves. Those waves arose over a billion light-years away, from two black holes that merged into one. Other gravitational waves have been observed since then and, undoubtedly, many more will be observed in the future.

Gravitational waves are waves in the gravitational field, much as electromagnetic waves are waves in the electric and magnetic fields. They arise from accelerating masses, such as orbiting stars or black holes, or even just orbiting planets and moons. Gravitational waves are very difficult to detect due to their being naturally weak, and because strong ones tend to arise very far away. Einstein's theory of general relativity showed that gravity bends space and distorts time, so gravitational waves are accurately described as ripples in the fabric of spacetime.

15.1 Gravity

Before discussing gravitational waves, it helps to back up a little and explore gravity first, which is the focus of this section.

15.1.1 Newtonian Gravity

In the early 1600s, Galileo provided conclusive evidence for the ***heliocentric model*** of the solar system, in which the sun is at the center of the solar system and the planets go around it. This then raised the problem of what keeps the planets in their orbits. Half a century later, Isaac Newton famously solved the problem when he saw an apple fall off a nearby tree. The falling apple led Newton to realize that the Earth exerts a force on objects around it, such as apples on trees, and birds in the sky, and even the Moon and the sun. He realized that this force, which he called gravity[1], could attract the Moon to the Earth sufficiently strongly to cause the Moon to orbit the Earth. Likewise, the concept of gravity explained how other planets could orbit the sun, and even how other moons could orbit other planets.

From these insights, Newton came up with his ***law of universal gravitation*** for describing the gravitational force between two objects. In modern notation, it is

$$F = \frac{Gm_1m_2}{r^2} \qquad G = 6.674 \cdot 10^{-11}\ \mathrm{m^3kg^{-1}s^{-2}}. \tag{15.1}$$

[1] Newton did not invent the word "gravity", which comes from the Latin word *gravitas* for the quality of heaviness. However, he shifted its meaning and was the first to use it in the context of a force.

Here, G is a number called the ***gravitational constant*** and has the value shown above, m_1 and m_2 are the masses of the objects, and r is the distance between them. This equation shows that gravitational force increases with larger object masses and decreases with larger separation between the objects. Note that this force is symmetric with respect to the two objects, meaning that each object pulls on the other one with the same amount of force. Also, masses are always positive, so the gravitational force is always positive, and gravity is always attractive. Anti-gravity, in which masses repel each other, is a nice science fiction concept but violates Newton's law of gravitation and has never been observed.

Example. The Earth's mass is $5.97 \cdot 10^{24}$ kg, the Moon's mass is $7.35 \cdot 10^{22}$ kg, and the distance between them is $3.84 \cdot 10^{8}$ m. What is the gravitational force acting on the Moon from the Earth?

Answer. Plugging these numbers into Newton's law of gravitation shows that the force is $1.99 \cdot 10^{20}$ N. This is the force of the Earth on the Moon, and of the Moon on the Earth.

15.1.2 Tides

Newton realized that the distance dependence of his law of gravitation helped explain why the Earth has tides, including why many places (including Britain, where Newton lived) have two high tides per day.

Leaving out the Earth for a moment, imagine three balls that are in a line, all reasonably close to each other and all falling toward the Moon. If the line of balls is perpendicular to the direction the balls are falling, as in the left panel of Figure 15.2, then this means that the balls are all about the same distance from the Moon. In this case, they experience the same gravitational force, so they fall at the same rate and stay the same distance from each other. On the other hand, if the line of balls is parallel to the direction they are falling, as in the middle panel of Figure 15.2, then they are at different distances from the Moon. In this case, the closest ball feels the

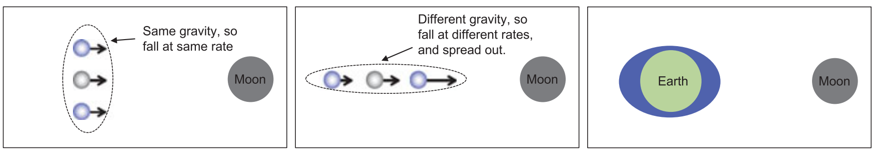


Fig. 15.2 Diagrams of tides. (Left, Middle) Balls falling toward the moon only spread out if they have different distance from the moon. (Right) The Earth and its oceans spreading out.

most gravity and the farthest ball feels the least gravity, so they fall at different rates and spread out.

To explain tides, the Earth is like the middle ball, while the oceans are like the outer balls, shown in the right panel of Figure 15.2. The part of the ocean that is closest to the Moon falls fastest, so it gets pulled away from the Earth. Likewise, the part of the ocean that is farthest from the Moon falls slowest, so the Earth pulls away from it. Together, these cause the ocean to be raised up on the sides of the Earth closest and farthest from the Moon in ***tidal bulges***. The Earth makes one full rotation per day, which then explains why many places have two high and two low tides every day. Of course, the Earth doesn't actually fall into the Moon, or the Moon into the Earth, which is due to to the Moon's fast velocity around the Earth; both Earth and Moon are constantly falling toward each other, but the Moon's velocity converts this falling into a nearly circular orbit.

Real tides are more complicated than this due to the additional influence of the sun's gravity, the inertia of the ocean water, the drag exerted on the ocean's water by the Earth's rotation, resonance effects, and other factors (see Section 6.5.6). However, this explanation still captures the essential aspects of the tide's driving force.

15.1.3 Gravitational Fields

Newton's conception of gravity was an "action at a distance" model, meaning that one object exerts a gravitational force on another object over some intervening distance. Much as the action-at-a-distance concepts of electrical and magnetic forces got reinterpreted as electric and magnetic fields during the 19th century, the same happened with gravitation. Rather than thinking of a gravitational force acting between, say, the Earth and a falling apple, over the space between them, the idea was is to think of the Earth's mass as creating a ***gravitational field*** around itself. Figure 15.3 illustrates this field. In this interpretation, the Earth's gravitational field at the apple's position exerts a force on the apple, and that force causes it to fall toward the Earth.

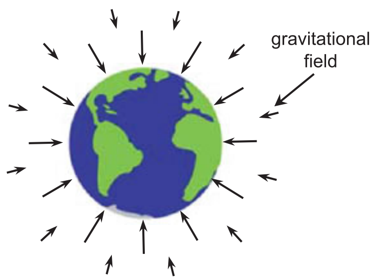


Fig. 15.3 Gravitational field of the Earth. Longer arrows represent stronger gravity.

A benefit of this new interpretation is that gravity becomes a local interaction, in which objects are only influenced by the gravitational field at their position, rather than by masses that are far away. The gravitational field is a vector field and all objects within it experience a gravitational force in the field's direction. In the example with the Earth, the field points toward the center of the Earth.

The strength of the gravitational field comes directly from Newton's law of gravitation, Eq. 15.1. Both sides of the equation are divided by an object's mass, yielding the acceleration that an object experiences due to gravitational attraction,

$$a = \frac{GM}{r^2}. \tag{15.2}$$

Here, M is the mass of the object that's producing the gravitational field. For example, the acceleration of an apple falling toward the Earth uses M as the Earth's mass and r as the Earth's radius (because we're assuming the apple is close to the Earth's surface), which is $6.37 \cdot 10^6$ m. Plugging these numbers into Eq. 15.2 gives the apple's acceleration as 9.80 m/s^2. We've seen this value many times before. It is typically just called the Earth's gravitational acceleration and denoted g. Note that the apple's mass wasn't necessary for this calculation; everything near the Earth's surface falls with this same acceleration, regardless of its mass.

If the gravitational field arises from multiple objects, then the separate fields add together to yield the total field. For example, the total gravitational field halfway between the Earth and Moon is the sum of the fields from the Earth, Moon, and sun, all measured at that point. To be thorough, one could also add in the fields other planets, although those influences would be extremely small.

Example. What is the sun's gravitational field at the Earth's position? The sun's mass is $1.99 \cdot 10^{30}$ kg and it is $1.50 \cdot 10^8$ km from the Earth.

Answer. Use the gravitational field equation with the sun's mass and distance:

$$a = \frac{GM}{r^2} = \frac{(6.674 \cdot 10^{-11}\ \text{m}^3\text{kg}^{-1}\text{s}^{-2})(1.99 \cdot 10^{30}\ \text{kg})}{(1.50 \cdot 10^{11}\ \text{m})^2} = 0.0059\ \text{m}\,\text{s}^{-2}.$$

This means that the Earth is constantly accelerating toward the sun at a rate of about 0.6 cm/s^2. It doesn't fall into the sun because its velocity around the sun keeps it in its orbit.

15.2 Gravitational Waves

15.2.1 First Direct Detection

1.3 billion years ago, when life on Earth consisted solely of single-celled organisms, two black holes in the southern sky were whirling around each other at high speed. The changing positions of the masses, each of which was about 30 times heavier than our sun, caused the gravitational field in the region to oscillate wildly. This produced waves in the gravitational field, called ***gravitational waves***[2], that propagated outward. The waves carried away energy, so the black holes fell closer together in what's called an ***inspiral***. This caused them to orbit even faster, making the waves stronger, and releasing more energy. This built up to create a powerful burst of gravitational waves until, suddenly, the black holes touched and merged together. At this point, the masses stopped orbiting each other, so the gravitational waves stopped, too.

That burst of waves expanded outward at the speed of light and passed by the Earth 1.3 billion years later, on September 14, 2015. By this time, humans had evolved, they had developed gravitational wave observatories, and two observatories had just been upgraded for greater sensitivity. Those observatories detected the passing burst of gravitational waves, which were the first gravitational waves to be detected directly. Astronomers named this event GW150914, for gravitational waves and the date. Their results, shown in Figure 15.4, show the wave displacements on the y-axis as the amount of strain on the detectors (the distance they moved divided by their separation distance), which represents changes in the gravitational field at the Earth.

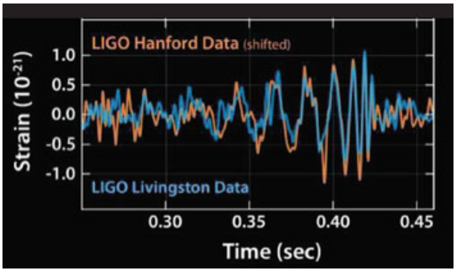


Fig. 15.4 The first gravitational waves directly observed. The two colors show the waves as detected at two different observatories, of which the orange lines were shifted to account for the different observatory orientations and for the time taken for the wave to travel from one to the other.

The individual waves in this figure directly represent the positions of the two black holes as they orbited around each other. Their high frequency is remarkable. From the graph's x-axis, the gravitational waves had a period of only about 0.02 s initially, which then decreased to about 0.004 s at the end, meaning that the frequency sped up from 50 Hz to 250 Hz, before the black holes finally touched and merged. These frequencies would be very low in the electromagnetic spectrum, and near the middle of the audible spectrum, but are extremely fast for such massive objects to orbit each other.

15.2.2 What are Gravitational Waves?

As stated before, gravitational waves are waves in the gravitational field. Also, the gravitational field is the acceleration due to gravity. Gravitational waves can't be observed with a single object but if you have two objects, say two balls, then a

[2] These are called "gravitational waves" because the term "gravity waves" was already taken; those are normal water waves in which gravity provides the restoring force.

Units

B

B.1 Units Are Your Friends

How are you supposed to remember a long list of equations? The simple answer is that you aren't. Memorizing a long list of equations is a quick route to frustration and is not what successful scientists or students do. Thinking about units is an important way to avoid memorizing equations.

Suppose, for example, that you want to compute a velocity, v, from a wavelength, λ, and a frequency, f. You remember that there is an equation that relates these variables, but you can't remember which term goes where. So, you try

$$v\lambda = f \quad \text{(incorrect)}.$$

To check if this is correct, compute the units on the left side: v is in m/s and λ is in m, and these combine to give m^2/s. Meanwhile, the units on the right side are s^{-1}. These are not the same, showing that your guess is incorrect. With more trial and error, you try one of

$$v = \lambda f \quad \text{or} \quad \lambda = \frac{v}{f} \quad \text{or} \quad f = \frac{v}{\lambda} \quad \text{(correct)}.$$

In each of these cases, the units match on the two sides of the equation and, in fact, all of these rearrangements are correct. Thus, the units enabled you to find correct equations and to easily discard incorrect ones.

Similarly, when you come up with your own equations, it's a good idea to check that the units are valid. If they're not, then the equation is certainly incorrect. This is a simple check that can catch a large fraction of mistakes.

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-1

B.2 The Metric System

Everyone used “customary” units of measure, which varied by region, up until near the end of the 18th century. Lengths were measured in British feet (30.48 cm), French feet (32.48 cm), Norwegian feet (31.37 cm), Japanese shaku (about 30 cm), or any of many other commonly used units, depending on where one lived. These units sometimes varied substantially by region within the country and also varied over time. This was inconvenient because there was no way to accurately describe how big something was without bringing the actual object along. It was also a problem for international trade and scientific communication. Furthermore, even within the systems of units of individual countries, it was difficult to convert values between different sizes of units because they weren’t separated by regular intervals. In the modern American system of volumes, for example, there are 3 teaspoons to a tablespoon, 16 tablespoons in a cup, 2 cups in a pint, 2 pints in a quart, 4 quarts in a gallon, and, roughly, 7.5 gallons in a cubic foot. These irregular intervals make conversion difficult.

The metric system was developed to address these problems and is now used universally in science. It is also the primary system of measurement in most of the world, with the primary exception of the United States. It is often called the SI system, from the French *Système International*.

The metric system is built upon a small collection of ***base units*** including the meter (m) for length, the kilogram (kg) for mass, and the second (s) for time[2]. There are also ***derived units*** that are built from base units. For example, energy is often measured in joules (J), a derived unit, where 1 joule is equal to 1 kg m^2 s^{-2}. Table B.1 lists the most important base and derived units. If you express all quantities with these base or derived units, then further unit conversion is generally not required.

Table B.1 Metric base and derived units

quantity	dimension	unit	symbol	in base units
length	L	meter	m	m
mass	M	kilogram	kg	kg
time	T	second	s	s
frequency	T^{-1}	hertz	Hz	s^{-1}
energy	$L^2 M T^{-2}$	joule	J	$kg\,m^2\,s^{-2}$
power	$L^2 M T^{-3}$	watt	W	$kg\,m^2\,s^{-3}$
force	$L M T^{-2}$	newton	N	$kg\,m\,s^{-2}$
pressure	$L^{-1} M T^{-2}$	pascal	Pa	$kg\,m^{-1}\,s^{-2}$

[2] This book uses the modern “mks” convention in which meters, kilograms, and seconds are the base units. This largely supplanted the older “cgs” convention, which is based on centimeters, grams, and seconds.

The “dimension” column in the table shows the dimension of the quantity, where L represents length, M represents mass, and T represents time. It can be useful, although the final column, “in base units” is equivalent and typically more useful.

Larger and smaller versions of the base and derived units can be formed using the prefixes that are listed in Table B.2. These prefixes often make numbers more convenient, but it is generally best to stick with meters, kilograms, and seconds within calculations; if you start mixing in centimeters, milligrams, or other units, then you’ll need to keep track of the factors of 10. Thus, the best approach is typically to convert all numbers to base or derived units at the beginning of a calculation, do the calculation with those units, and then add the appropriate prefix to your answer at the end to make it more meaningful.

Table B.2 Metric prefixes

prefix	abbreviation	value	expanded value
tera	T	10^{12}	1,000,000,000,000
giga	G	10^{9}	1,000,000,000
mega	M	10^{6}	1,000,000
kilo	k	10^{3}	1000
centi	c	10^{-2}	0.01
milli	m	10^{-3}	0.001
micro	μ	10^{-6}	0.000001
nano	n	10^{-9}	0.000000001
pico	p	10^{-12}	0.000000000001

In this table, the “expanded value” column might present numbers in a more familiar way than the “value” column but is actually less useful. For example, suppose you are given a number as 43 km and you want to enter it into your calculator in meters, as you should. You could enter it as 43 and then tack on three zeros, or you could enter it as 43 and then multiply by 1000. However, the best approach is to think of this number as $43 \cdot 10^3$ m, where the 10^3 part is the kilo part; enter this number in a calculator as 43, then the EE key, and then 3; see Appendix A.

B.3 Unit Math

As mentioned above, it is a good idea to check the units in equations to ensure that they agree on both sides. If they don’t agree, the equation is certainly incorrect. A simple approach is to substitute in the units for each variable, while dropping the unit prefixes. The prefixes aren’t necessary because they only change the value, but not the underlying unit. Here is an example using the relationship among velocity, wavelength, and frequency:

$$v = \lambda f \quad \rightarrow \quad \frac{\text{m}}{\text{s}} = (\text{m})(\text{s}^{-1})$$

These units agree, confirming that the equation is reasonable (and in fact is correct).

Here are the rules for unit math.

- Ignore all numbers. This includes all values like 2, 4.7, and -58, all factors of π, and all metric prefixes.
- Change derived units to base units if it will help. For example, change Hz to s^{-1} and W to J s^{-1}. On the other hand, it might not help. For example, changing J to $\text{kg m}^2\text{s}^{-2}$ may or may not lead to a simpler result.
- Countable objects are optional as units. For example, a wavelength can be expressed equivalently as 0.1 m, or as 0.1 m/wave. Likewise, a rotational frequency can be given equivalently as 2000 rotations/s or 2000 s^{-1}.
- Addition and subtraction: units must be the same for both terms, and units stay the same at the end. For example, 2 apples + 3 apples = 5 apples; this is valid because the units are the same in both terms and then become the same in the end. However, 2 apples + 3 oranges is not valid because the terms don't have the same units. 2 s + 3 s^{-1} is equally incorrect.
- Multiplication and division: combine units. For each type of unit, such as "m" for meter, add up all of the exponent values in the numerator and subtract any of the exponent values in the denominator. For example, multiplying m by m gives m^2 and dividing m by s gives m s^{-1}. Note that a unit with a negative exponent is equivalent to dividing by that unit, so $\text{m/s} = \text{m s}^{-1}$ and $\text{m/s}^2 = \text{m s}^{-2}$.

Suppose you have the equation $\lambda = \frac{h}{p}$, where h is measured in J s, and p is measured in kg m/s. You want to know the units for λ. The equation and variables have meanings, of course, but that is irrelevant here.

$$\lambda = \frac{h}{p} \quad \rightarrow \quad \lambda = \frac{\text{J s}}{\text{kg m s}^{-1}} = \frac{\text{kg m}^2\text{s}^{-2}\text{s}}{\text{kg m s}^{-1}} = \text{m}$$

Here, we first replaced the variables with their units, we then found that the joules in the numerator didn't cancel with anything in the denominator so we expanded it in terms of the base units ($\text{J} = \text{kg m}^2\text{s}^{-2}$), and we finally collected together all of the kg, m, and s terms to find the total number of each. As it turned out, all that was left was a meter unit, showing that λ must have units of meters.

Once you are done simplifying, a fraction should only have a single numerator and a single denominator. For example, the unit m/s/s doesn't make sense. Does this mean (m/s)/s, which simplifies to m/s^2, or does this mean m/(s/s), which simplifies to m? It's unclear, so it's best to avoid such constructions. This leads to the question of how to deal with fractions of fractions, which is a very common situation. Suppose, you want to divide m/s by m/wave. A simple approach is to replace each of the two fractions with a single row of units by using negative exponents, and to then count up how many you have of each type of unit,

$$\frac{\frac{\text{m}}{\text{s}}}{\frac{\text{m}}{\text{wave}}} = \frac{\text{m s}^{-1}}{\text{m wave}^{-1}} = \text{wave s}^{-1} = \frac{\text{wave}}{\text{s}}$$

B.4 Unit Conversion

As mentioned above, a typical approach for solving a problem is to express all numbers with their values in meters, kilograms, and seconds (and joules, newtons, pascals, etc., as given in Table B.1), do the math with these base units, and then make the result more tidy by replacing any powers of 10 with the appropriate prefix.

This unit conversion is simple if you work entirely with metric values, often allowing you to account for units on the fly while entering numbers into a calculator. For example, suppose you want to find the frequency of green light, which has a wavelength of 500 nm. To do this, rearrange the equation $v = \lambda f$ to get $f = \frac{v}{\lambda}$. The velocity is $3 \cdot 10^8$ m/s and λ is 500 nm, so enter the following into a calculator

$$f = \text{3e8} \div \text{500e-9}$$

The result is $6 \cdot 10^{14}$. Because everything was entered in meters and seconds, the result is also in metric base units. In this case, the result is in s^{-1} which can be verified using unit math. This answer of $6 \cdot 10^{14}\,\text{s}^{-1}$ is fully sufficient, but could also be rescaled by taking 12 factors of 10 off it and using the tera prefix to give the frequency as 600 THz.

Unit conversion is more complicated if you aren't working entirely in the metric system. Most people still instinctively start by reaching for their calculators, but this is generally a mistake. Instead, it's better to start with paper and pencil, write down the unit conversion equation, and then use a calculator at the very end. The procedure for this unit conversion equation starts by rephrasing the question as a mathematical expression. For example, suppose the question is how many feet are in 1 meter? This is rephrased as

$$?\text{ feet} = \frac{1\text{ m}}{1}$$

Writing the right hand side as a fraction with a 1 in the denominator isn't necessary, but can help avoid confusion later on. Having the left side there is nice because it reminds you of where you're going. Note that this equation is indeed an equality, in that some number of feet is actually exactly the same thing as 1 meter. Next, multiply the right side of this equation by 1 as many times as needed until you get the correct units, where this "1" is a conversion factor. Units can be crossed off, as they cancel with each other. Continuing with this example, we convert meters to centimeters by multiplying by the fraction $\frac{100\text{ cm}}{1\text{ m}} = 1$ and then cross off meters:

$$?\text{ feet} = \frac{1\,\cancel{\text{m}}}{1} \cdot \frac{100\text{ cm}}{1\,\cancel{\text{m}}}$$

Continuing gives

$$? \text{ feet} = \frac{1 \cancel{\text{m}}}{1} \cdot \frac{100 \cancel{\text{cm}}}{1 \cancel{\text{m}}} \cdot \frac{1 \cancel{\text{inch}}}{2.54 \cancel{\text{cm}}} \cdot \frac{1 \text{ foot}}{12 \cancel{\text{inch}}}$$

At the very end, enter all of these numbers into a calculator, which gives the result that 1 meter equals 3.28 feet.

More complicated types of units make the task a little more complicated but not fundamentally different. For example, the density of water is 1 g/cm^3. What is it in pounds per cubic foot? Here, the unit conversion equation looks like:

$$? \frac{\text{lbs}}{\text{ft}^3} = \frac{1 \cancel{\text{g}}}{\cancel{\text{cm}^3}} \cdot \frac{1 \cancel{\text{kg}}}{1000 \cancel{\text{g}}} \cdot \frac{2.20 \text{ lbs}}{1 \cancel{\text{kg}}} \cdot \frac{2.54 \cancel{\text{cm}}}{1 \cancel{\text{inch}}} \cdot \frac{2.54 \cancel{\text{cm}}}{1 \cancel{\text{inch}}} \cdot \frac{2.54 \cancel{\text{cm}}}{1 \cancel{\text{inch}}} \cdot \frac{12 \cancel{\text{inch}}}{1 \text{ foot}} \cdot \frac{12 \cancel{\text{inch}}}{1 \text{ foot}} \cdot \frac{12 \cancel{\text{inch}}}{1 \text{ foot}}$$

Plugging these numbers into a calculator gives the answer as 62.3 lbs/ft^3.

As a final point, note that Google does an excellent job of unit conversion, making this entire task largely unnecessary. Nevertheless, unit conversion is a useful skill to have in case you need a number in certain units and you don't have internet access at that moment. Table B.3 lists some common unit conversions.

Table B.3 Some unit conversions

1 inch	=	2.54 cm	1 hour	=	60 minutes
1 foot	=	12 inches	1 minute	=	60 seconds
1 mile	=	5280 feet	1 day	=	24 hours
1 acre	=	43560 ft^2	1 kg	=	2.205 lbs
1 ft^3	=	7.48 gallons	1 calorie	=	4.184 J
1 gallon	=	3.785 liters	1 kilocalorie	=	4184 J

B.5 Exercises

Problems

B.1. For each part, use the unit conversion method presented here and show your work. (a) How many km is 4000 miles? (b) How many nm is 17 μm (the width of a human hair)? (c) How many cycles per day is 97 MHz?

B.2. For each part, use the unit conversion method presented here and show your work. (a) How many gallons are in 0.25 acre-feet (the annual water use of a typical family)? (b) How many km/liter are equal to 40 miles per gallon? (c) How many US$/gallon is equal to 1.28 euros/liter (the price of gasoline in France), assuming that 1 euro equals $1.14?

H Useful Facts and Figures

Table H.1 Fundamental constants

Constant	Symbol	Approximate value
Speed of light in vacuum	c	$2.998 \cdot 10^{8}$ m/s
Elementary charge	e	$1.602 \cdot 10^{-19}$ C
Planck's constant	h	$6.626 \cdot 10^{-34}$ J s
Gravitational constant	G	$6.674 \cdot 10^{-11}$ $m^3kg^{-1}s^{-2}$
Electric permittivity of space	ϵ_0	$8.854 \cdot 10^{-12}$ $C^2N^{-1}m^2$
Electron mass	m_e	$9.109 \cdot 10^{-31}$ kg
Proton mass	m_p	$1.673 \cdot 10^{-27}$ kg
Rydberg constant	R_∞	$1.097 \cdot 10^{7}$ m^{-1}
Boltmann's constant	k_B	$1.381 \cdot 10^{-23}$ J/K
Stefan-Boltzmann constant	σ	$5.670 \cdot 10^{-8}$ $W\,m^{-2}K^{-4}$
Wien's displacement constant	b	$2.898 \cdot 10^{-3}$ m K
Avagadro's number	N_A	$6.022 \cdot 10^{23}$ mol^{-1}

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-1

Table H.2 More useful numbers

Constant	Symbol	Value	Notes
Speed of sound in air		340 m/s	Typical range is 330-350 m/s
Speed of sound in fresh water		1481 m/s	At 20° C
Speed of sound in seawater		1500 m/s	Typical range is 1480-1540 m/s
Density of water		$1000\ \mathrm{kg\,m^{-3}}$	At 4° C
Gravitational acceleration	g	$9.80\ \mathrm{m\,s^{-2}}$	Average for Earth's surface
Sun: Mass	$M_\odot$	$1.989 \cdot 10^{30}$ kg	
Radius	$R_\odot$	$6.96 \cdot 10^{8}$ m	Radius is larger at equator than poles
Surface temperature	$T_\odot$	5778 K	The interior is much hotter
Earth: Mass	$M_\oplus$	$5.972 \cdot 10^{24}$ kg	
Radius	$R_\oplus$	$6.37 \cdot 10^{6}$ m	Radius is larger at equator than poles
Moon: Mass		$7.348 \cdot 10^{22}$ kg	
Radius		$1.74 \cdot 10^{6}$ m	Radius is larger at equator than poles
Earth-Sun distance		$1.496 \cdot 10^{11}$ m	Between object centers
Earth-Moon distance		$3.844 \cdot 10^{8}$ m	Between object centers

Table H.3 Refractive indices (from Table 2.2)

medium	n
Vacuum	1 (by definition)
Air	1.0003
Ice	1.31
Water	1.33
PMMA (Acrylic)	1.49
Glass (crown & window glass)	1.52
Sapphire	1.77
Cubic zirconia (fake diamond)	2.16
Diamond	2.42

Table H.4 Greek alphabet

Name	U.C.	L.C.	Name	U.C.	L.C.
Alpha	A	α	Nu	N	ν
Beta	B	β	Xi	Ξ	ξ
Gamma	Γ	γ	Omicron	O	o
Delta	Δ	δ	Pi	Π	π
Epsilon	E	ε	Rho	P	ρ
Zeta	Z	ζ	Sigma	Σ	σ
Eta	H	η	Tau	T	τ
Theta	Θ	ϑ	Upsilon	Υ	υ
Iota	I	ι	Phi	Φ	φ
Kappa	K	$\varkappa$	Chi	X	χ
Lambda	Λ	λ	Psi	Ψ	ψ
Mu	M	μ	Omega	Ω	ω

Table H.5 Metric base and derived units (from Table B.1)

Quantity	Dimension	Unit	Symbol	In base units
Length	L	meter	m	m
Mass	M	kilogram	kg	kg
Time	T	second	s	s
Frequency	T^{-1}	hertz	Hz	s^{-1}
Energy	L^2MT^{-2}	joule	J	$\mathrm{kg\,m^2\,s^{-2}}$
Power	L^2MT^{-3}	watt	W	$\mathrm{kg\,m^2\,s^{-3}}$
Force	LMT^{-2}	newton	N	$\mathrm{kg\,m\,s^{-2}}$
Pressure	$L^{-1}MT^{-2}$	pascal	Pa	$\mathrm{kg\,m^{-1}\,s^{-2}}$

Table H.6 Metric prefixes (from Table B.2)

Prefix	Abbreviation	Value
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
centi	c	10^{-2}
milli	m	10^{-3}
micro	µ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}

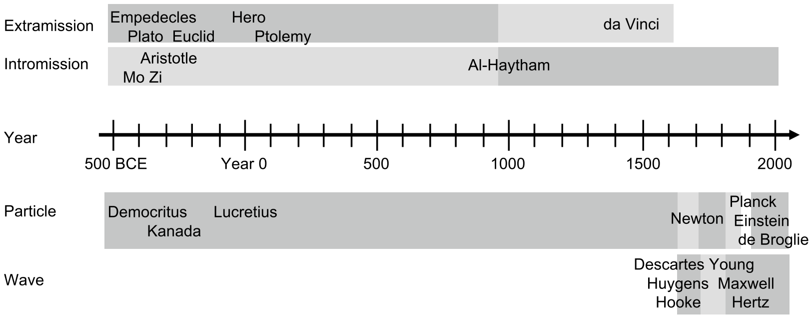


Figure H.1 Optics timeline (from Figure 1.12).

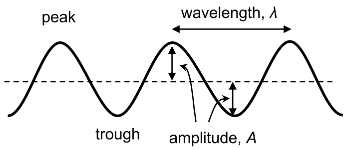


Figure H.2 Wave terminology (from Figure 2.3).

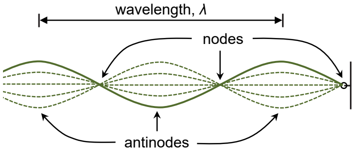


Figure H.3 Standing waves (from Figure 3.11).

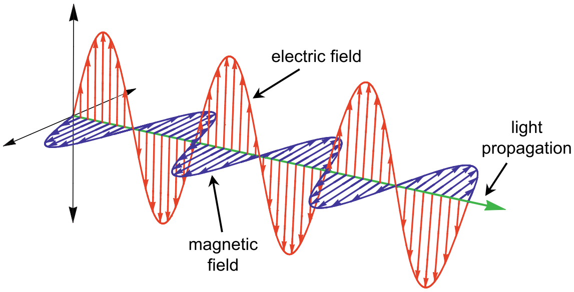


Figure H.4 Electromagnetic waves (from Figure 11.7).

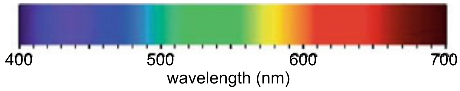


Figure H.5 Visible spectrum (from Figure 2.4).

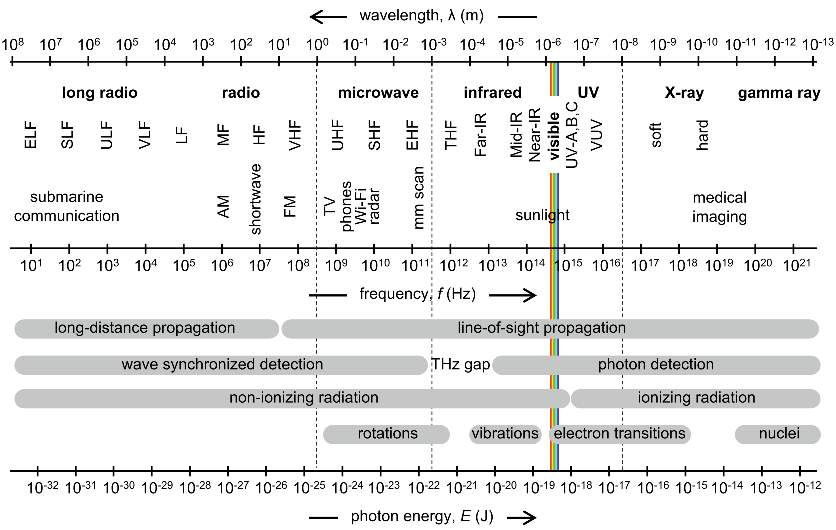


Figure H.6 Electromagnetic spectrum (from Figure 11.10).

Figure H.7 Scientific pitch notation (from Figure 6.8).

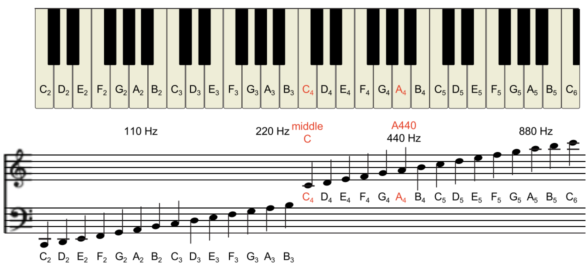


Table H.7 Musical note frequencies (from Figure 6.9)

Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz
C1	32.7	C2	65.4	C3	130.8	C4	261.6	C5	523.3	C6	1046.5	C7	2093.0
C♯1	34.6	C♯2	69.3	C♯3	138.6	C♯4	277.2	C♯5	554.4	C♯6	1108.7	C♯7	2217.5
D1	36.7	D2	73.4	D3	146.8	D4	293.7	D5	587.3	D6	1174.7	D7	2349.3
D♯1	38.9	D♯2	77.8	D♯3	155.6	D♯4	311.1	D♯5	622.3	D♯6	1244.5	D♯7	2489.0
E1	41.2	E2	82.4	E3	164.8	E4	329.6	E5	659.3	E6	1318.5	E7	2637.0
F1	43.7	F2	87.3	F3	174.6	F4	349.2	F5	698.5	F6	1396.9	F7	2793.8
F♯1	46.2	F♯2	92.5	F♯3	185.0	F♯4	370.0	F♯5	740.0	F♯6	1480.0	F♯7	2960.0
G1	49.0	G2	98.0	G3	196.0	G4	392.0	G5	784.0	G6	1568.0	G7	3136.0
G♯1	51.9	G♯2	103.8	G♯3	207.7	G♯4	415.3	G♯5	830.6	G♯6	1661.2	G♯7	3322.4
A1	55.0	A2	110.0	A3	220.0	A4	440.0	A5	880.0	A6	1760.0	A7	3520.0
A♯1	58.3	A♯2	116.5	A♯3	233.1	A♯4	466.2	A♯5	932.3	A♯6	1864.7	A♯7	3729.3
B1	61.7	B2	123.5	B3	246.9	B4	493.9	B5	987.8	B6	1975.5	B7	3951.1

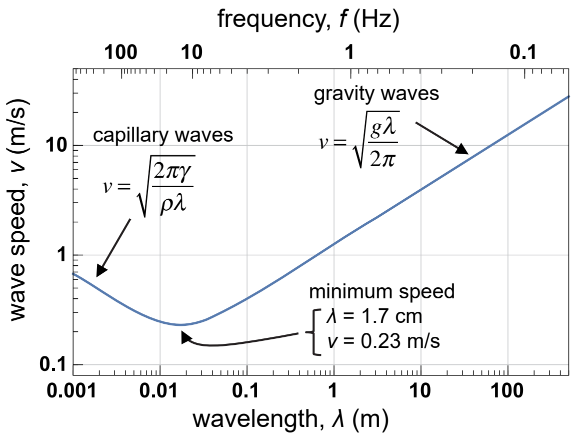


Figure H.8 Deep water wave speed (from Figure 6.14).

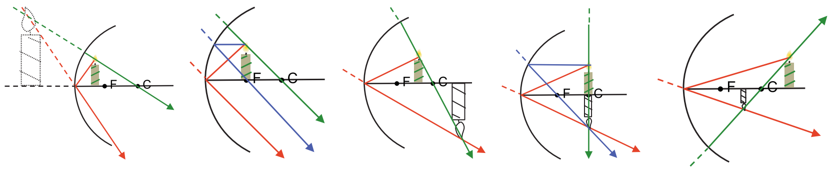


Figure H.9 Concave mirrors (from Figure 8.15).

Figure H.10 Convex mirrors (from Figure 8.17).

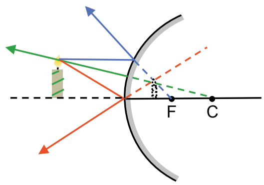
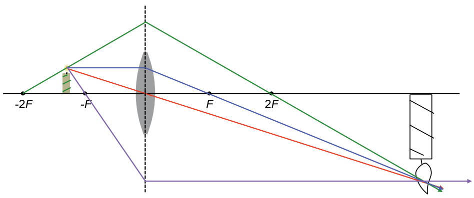


Figure H.11 Lens principal rays (from Figure 9.14).



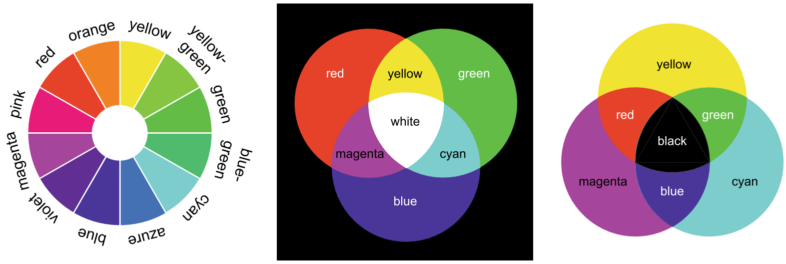


Figure H.12 Color models (from Figures 10.9, 10.10, and 10.12).

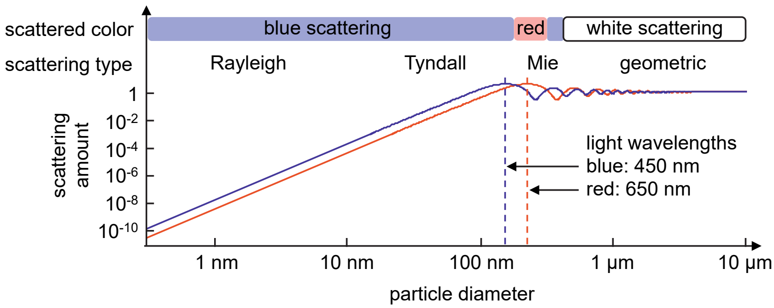


Figure H.13 Scattering (from Figure 11.14).

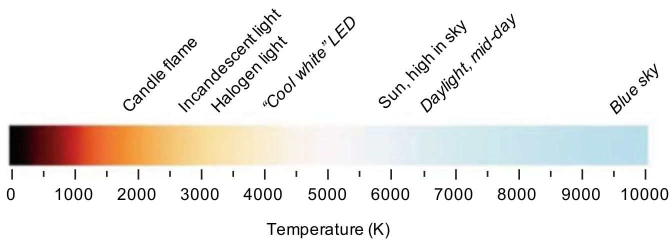


Figure H.14 Color temperature (from Figure 12.4).

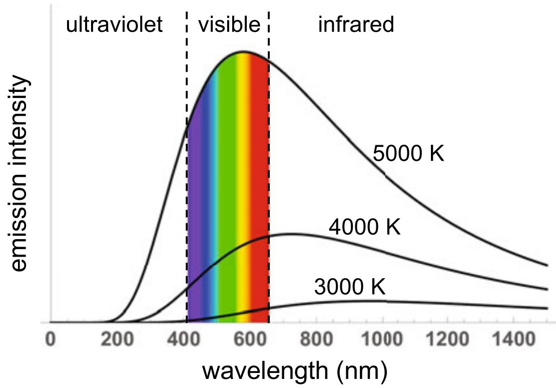


Figure H.15 Blackbody spectra (from Figure 12.3).

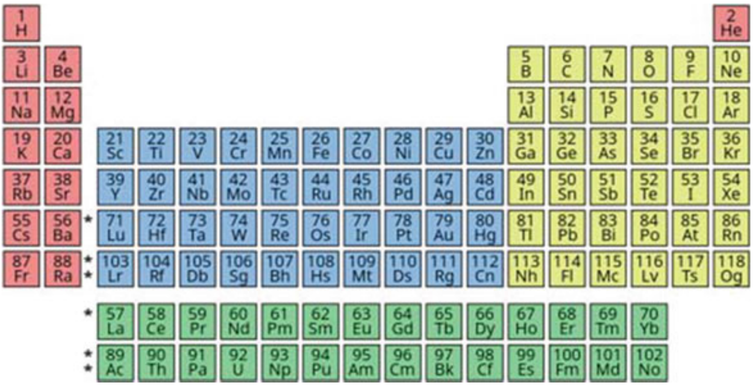


Figure H.16 Periodic table.

Index

S. S. Andrews, *Light and Waves*,
https://doi.org/10.1007/978-3-031-24097-1