

Study Guide for *The Roots of Things* by Alan Grometstein

(chapter summaries with some added material as appropriate)
by Dr. Colton, Honors 225 (last updated: Fall 2019)

“Preface”; Chapter 1: Themes; The Scientific Method

The preface deals in part with the scientific method, which is to say “How do we learn truth about the universe”. Learning truth through science involves *observing* the world around us, developing *hypotheses* to explain those observations—usually expressed in the rigorous language of *mathematics*—and performing *experiments* to test out the hypotheses. Basically: if you think you know something, you need to test it out and make sure that your thoughts about the situation are accurate. The experiments need to be *reproducible* not only by yourself, but by everyone else who performs them as well. If you reach the stage where you can accurately make mathematical predictions, then you can hope to wrap everything up into a *theory* which involves statements of *physical laws* that apply not only to the situations that you have studied but which can be *extrapolated to other situations* as well.

The words which I have italicized comprise the core of the scientific method, to me, namely:

- Observations
- Hypotheses
- Mathematics
- Experiments
- Reproducibility
- Theories
- Physical laws
- Extrapolations

Chapter 1 continues some description of science, but more to the point identifies the approach the author of *Roots*, Alan Grometstein, plans to take as he discusses the science, namely:

- Chronology
- Reasonableness
- Light
- Predictability
- Uncertainty
- Reality
- Beauty
- Simplicity

Math Review; Units; “Conventions, Abbreviations, Symbols”

It has been said that “**mathematics is the language of science.**” This class, and the *Roots* textbook, will expect you to be able to do algebra and basic trigonometry. You need to know what the mathematical constants e and π refer to. You should have a basic scientific calculator that can do mathematical functions such as sin, cos, tan, and exp. More advanced mathematical concepts will arise occasionally, but we will handle those mainly qualitatively.

Powers of 10. Standard metric prefixes indicate powers of ten. In order of increasing size here are the most common ones; I expect you to know these by heart:

- Nano (n) – 10^{-9}
- Micro (μ) – 10^{-6}
- Milli (m) – 10^{-3}
- Centi (c) – 10^{-2} (not really used except in the context of centimeter, cm)
- Kilo (k) – 10^3
- Mega (M) – 10^6
- Giga (G) – 10^9

If/when more obscure metric prefixes show up (e.g. “pico” or “femto”) I will give you their meaning.

As an example, 3 GJ means 3×10^9 joules, which would be 3,000,000,000 joules.

Another example: 4 cm means $4 \times 10^{-2} \text{ m} = 0.04 \text{ m}$.

Short-cut “e” notation can also be used to refer to powers of 10; for example 3.58×10^{13} is 3.58e13.

Units. The standard metric units used in the course are as follows:

- Length – meter (m)
- Mass – kilogram (kg)
- Time – second (s)
- Energy – joule (J)
- Force – newton (N)
- Electric “potential” (voltage) – volt (V)
- Temperature – kelvin (K)
- Power – watt (W), which is energy per time (J/s)

Unit conversions. When non-standard units are given or desired, one can convert between units using *conversion factors*. Here is a list of some of the most common ones; I will give you a similar list on the exams.

- 1 inch = 2.54 cm
- 1 foot = 0.3048 m
- 1 mile = 1.609 km
- 1 mi/hr = 1 mph = 0.44704 m/s
- 1 kg = 2.205 lbs
- 1 cal = 4.186 J
- 1 eV (electron volt) = $1.602 \times 10^{-19} \text{ J}$

Conversion factors represent ratios which equal 1. For example, if 1 inch = 2.54 cm, then that means $\frac{1 \text{ inch}}{2.54 \text{ cm}} = 1$, and for that matter, also $\frac{2.54 \text{ cm}}{1 \text{ inch}} = 1$.

Converting between units can be done by multiplying your starting quantity by a series of ratios which all equal 1 (and therefore don’t change the value of the quantity), with the ratios expressed with the proper numerator and denominator so that the units you do not want in your answer cancel out and leave only the units you do want. For example, to convert 123 inches into meters I would do the following:

$$123 \text{ inch} \times \frac{2.54 \text{ cm}}{1 \text{ inch}} \times \frac{1 \text{ m}}{100 \text{ cm}} = 3.1242 \text{ m}$$

Notice how the inch and cm units each cancel out, leaving just meters.

Temperature conversions. One exception to the above procedure is converting between different temperature units. Different formulas must be used because the temperature scales do not go to zero at the same point (as compared to, for example, distance units where 0 inches = 0 cm). Here are the formulas for temperature conversion; I will give these to you on the exams:

- $T_F = \frac{9}{5} T_C + 32$ (Celsius to Fahrenheit)
- $T_K = T_C + 273.15$ (Celsius to Kelvin)

Fundamental constants. Here is a list of fundamental constants which show up in equations governing physical laws. I will give you a similar list on the exams.

- c = speed of light, $2.998 \times 10^8 \text{ m/s}$
- e = charge of an electron (magnitude), $1.602 \times 10^{-19} \text{ C}$
- h = Planck’s constant, $6.626 \times 10^{-34} \text{ J}\cdot\text{s}$
- k_B = Boltzmann’s constant, $1.381 \times 10^{-23} \text{ J/K}$. In *Roots* this is only labelled k , but to avoid confusion with other quantities which also use the symbol k , I will add the B as a subscript.
- k_C = the Coulomb force constant, sometimes called “Coulomb’s constant”, and has a value of $8.988 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2$. It is not directly used in *Roots*, but is indirectly used in some footnotes to Chapter 10, and I will use it in class.

- m = mass of an electron, 9.109×10^{-31} kg
- σ = Stefan-Boltzmann constant, 5.670×10^{-8} W/(m²·K⁴)
- Wien's law constant (not given a symbol in the book) = 2.898×10^{-3} meters·K

Chapter 2: O Light Divine; Appendix FRE Frequency

Basic wave properties. Although Newton pictured light as a stream of particles, other scientists who were his successors (and in some cases even his contemporaries) soon discovered that light demonstrated features of waves. Wave properties are characterized by:

- *wavelength*, λ (lambda). This is the *distance* over which a wave repeats itself. If, for example, you were to take a picture of a wave so that it's frozen in time, the wavelength is the number of meters between wave crests.
- *period*, T . This is the *time* over which a wave repeats itself. If, for example, you were to watch a wave move past you as you are frozen at a point in space, the period is the number of seconds between wave crests.
- *frequency*, f (Roots uses the symbol ν , nu). This is the inverse of the period: $= \frac{1}{T}$. It describes how many wave crests pass through a particular point in space each second.
- *radians*. We use the terms "oscillations" aka "cycles", as well as angles measured in radians or degrees to quantify how much of a wave has passed by a point. There are 2π radians, which is 360° , in a complete oscillation of the wave.
- *angular frequency*, ω (omega). This is 2π times the frequency: $\omega = 2\pi f$. It represents how many radians of the wave pass by a location per second. It's also related to the period through $\omega = \frac{2\pi}{T}$.
- *wavenumber*, k . This is 2π divided by the wavelength: $= \frac{2\pi}{\lambda}$. It represents how many radians of the wave exist per meter in a wave that is frozen in time.
- *amplitude*, A – This describes how large the wave is, at its largest, measured from the middle of the wave.

The speed of a wave is related to the wavelength and frequency through this equation:

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

Wave speed equation

For light waves, the wave speed is written as c , and is equal to 2.998×10^8 m/s.

Sine waves. We often use sinusoidal functions (i.e. sines and cosines) to represent waves, like this:

$y = A \sin(kx)$ – a wave that has no time dependence (like in a photograph)

$y = A \sin(\omega t)$ – a wave that has no spatial dependence (like if you're restricted to a single point in space)

$y = A \sin(kx - \omega t)$ – a *traveling wave* with both time and space dependence

Here A is the amplitude of the oscillation and y represents the strength of the wave at a particular location in space and/or time (measured in the same units as A).

Phase. Since sine equals 0 when its argument is 0, all of those forms implicitly require a wave which starts at zero. If a particular wave starts at a maximum instead of at zero, a cosine function should be used instead. If the wave starts somewhere between zero and the maximum, it requires a "phase shift" relative to a regular sine or cosine.

Interference. When waves come together, they combine. This is called *interference*. If two waves with the same frequency reach their peaks at the same times/places, they are said to be “in phase” and their interference is *constructive*. The waves add together. If, on the other hand, a wave hits a positive peak at the same point where another wave hits a negative peak (aka a valley or trough), they are said to be “out of phase” and their interference is *destructive*. The waves cancel out.

Other properties of waves. More properties of light include the following; you should review them in the book if needed:

- *Reflection*
- *Refraction*
- *Dispersion*
- *Diffraction*
- *Transverse vs. longitudinal oscillations*
- *Polarization*
- *Coherence*

Chapter 3 - The Young Experiment (1801)

Young’s “two slit experiment” showed unambiguously that light waves interfere with each other, and hence light must be a wave (or so people thought until Einstein). Here’s the setup:

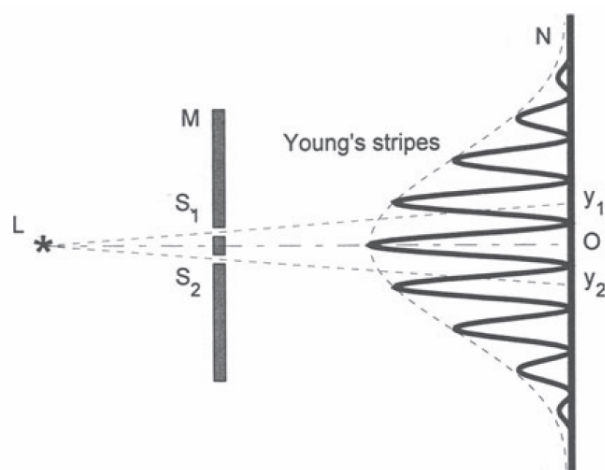


Figure 1. Experimental setup for Young’s two-slit experiment. From *Roots* Fig. 3.3b.

In the figure, L is the light source. M stands for a “mask”, which contains two slits, S_1 and S_2 . The light passes through the mask and gets projected onto a screen labeled N. The stripes on the figure indicate spots of varying intensity or brightness as shown below. The letter y indicates the position on the screen relative to the middle of the pattern (origin O). y_1 and y_2 are just two possible positions.



Figure 2. Examples of bright and dark fringes visible in the TSE. From *Roots* Fig. 3.4.

Using the symbols d and D to refer to the distance between the two slits and the distance between mask and screen, respectively, and assuming that d is much smaller than D , the path length difference (PD) for light rays traveling through S_2 vs. S_1 is:

$$PD = \frac{yd}{D}$$

The bright fringes occur when the path difference is zero, one wavelength, two wavelengths, etc., because that will result in the waves from the two slits being in phase and interfering constructively. The condition for points (y -values) of constructive interference is therefore this:

$$\frac{y_{constr}d}{D} = n\lambda$$

where n is an integer and as mentioned d is the slit separation and D is the distance from the slits to the screen. When rearranged, that gives us this important equation:

$$y_{constr} = \frac{n\lambda D}{d} \quad \text{2-slit constructive interference}$$

The fringe separation distance (which I will call Δy) is constant, and is given by the distance from the origin to the y -value for $n = 1$:

$$\Delta y = \lambda D / d \quad \text{fringe separation distance}$$

Chapter 4 – A Whiff of Ether (1887)

Definitions

- Velocity, v – how fast an object changes position (in m/s): $v = \frac{\text{change in position}}{\text{time it takes for the change}}$
- Acceleration, a – how fast an object's velocity changes (in m/s²): $a = \frac{\text{change in velocity}}{\text{time it takes for the change}}$
- Mass, m – a measure of how much matter is present in an object (in kg)
- Force, F – the strength of a push or a pull on an object (in newtons, N, which is a kg· m/s²)

Newton's three laws of motion

- 1) An object at rest will remain at rest, and an object in motion will remain in constant velocity straight line motion, unless an outside force acts upon it. In other words:

$$\text{if } F_{net} = 0, \text{ then } a = 0$$

The term “net force” must be used in this law and the next one, because (for example) two forces applied opposite to each other can cancel each other out.

- 2) If there is a net force on an object, then there will be an acceleration. The amount of acceleration is given by $a = F/m$. Or, as it's usually written:

$$F_{net} = ma$$

- 3) If there's a force between two objects, then the force of object 1 on object 2 will always be equal and opposite to the force of object 2 on object 1:

$$F_{\text{object 1 on object 2}} = -F_{\text{object 2 on object 1}}$$

A negative force means a force in the opposite direction. Note that this law means that forces always comes in pairs with the “partner forces” acting on two different objects. If (for example) I hit someone’s face with my hand, their face will always apply a force on my hand at the same time in the opposite direction (with the same amount of force).

Example of forces

- Gravity, $F \sim 1/r^2$
- Static electricity, $F \sim 1/r^2$
- Strong and weak nuclear (they are very short ranged, and very complicated)

Gravity and static electric forces are inverse square laws: increase the distance between the objects by $2\times$ and the force gets reduced to $1/4$ of the original value.

Inertial vs. non-inertial observers. Observers must be *inertial* for them to agree that Newton’s laws are valid. Inertial means “at rest or moving at constant velocity” (like in the cargo hold example in *Roots*). Under those conditions everything will be normal. *Non-inertial* observers are ones in a frame of reference that is accelerating or rotating; under those conditions there will be strange effects which objects don’t seem to obey Newton’s laws.

The speed of light. As mentioned above, the speed of light has the symbol c and a measured value of 2.998×10^8 m/s. This value exactly matches the speed of electromagnetic waves predicted by Maxwell’s laws of electricity and magnetism which were published in 1864, and the agreement between theory and experiment was seen both as a great triumph of electromagnetic wave theory and also as incontrovertible evidence of light being a wave. Incidentally, the speed of light was measured with a fair degree of accuracy as early as 1849 (not discussed in book).

Waves generally require a medium in which to travel, so the *Michaelson-Morley experiment* was done to try to measure properties of the *luminiferous ether* which was thought to be the medium for light waves, but which we now know does not exist. The experiment measured interference between light waves starting at the same spot but going back and forth from there along two different paths. The setup of the experiment is given here, assuming equal paths and an ether current or “wind” coming from the right due to the Earth’s motion through the ether:¹

¹ If the ether wind were coming from a different direction the equations would need to be modified but similar principles would hold.

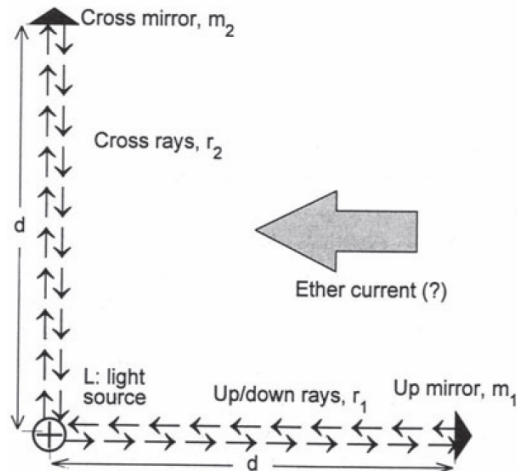


Figure 3. Experimental setup of Michaelson-Morley's experiment, with assumed ether wind from the right. From *Roots* Fig. 4.5a.

Using v to indicate the speed of the ether wind and d for the length of each path, the equations for the time for light to travel along the round-trip horizontal and vertical paths work out to be these:

$$t_{horiz} = \frac{2dc}{c^2 - v^2}$$

$$t_{vert} = \frac{2d}{\sqrt{c^2 - v^2}}$$

The two equations can be combined in a ratio which shows that the horizontal time is larger (because the denominator of the ratio is smaller than the numerator):

$$\frac{t_{horiz}}{t_{vert}} = \frac{c}{\sqrt{c^2 - v^2}}$$

By measuring the differences in the two times, which Michaelson and Morley proposed to do via phase shifts and interference fringes, one can then calculate v . The result of the experiment: they saw no shifting of interference fringes. The conclusion is that there is no difference in times, and hence $v = 0$, and hence there is no ether at all! Since there is no ether, a necessary and confusing consequence is that light emitted by sources will be measured by all observers as traveling at c , regardless of any motion by the sources or observers. This was eventually resolved by Einstein through his theory of relativity in 1905, but (sadly?) is not a topic of this course.

Chapter 5 – Professor Plank is Desperate (1901); Appendix PLA Max Plank

The random motion of atoms and molecules is a form of energy. The hotter the object, the more vigorous the random motion. Like other forms of energy, this is measured in the metric systems in joules. Heat is the transfer of this random *thermal energy*. The three main ways thermal energy can be transferred are conduction, convection, and radiation. We'll focus on radiation: heat transfer through light energy.

Blackbody radiation is the “glow” of hot objects, such as incandescent lights, electric burners, lava, stars, etc. The glow carries away energy from the hot object, which then can be given to cooler objects surrounding it. The three important laws governing blackbody radiation are:

Stefan-Boltzmann law: $P = \epsilon \sigma A T^4$

- P is the radiant energy per time emitted by a glowing object (in watts, which are joules per second). *Roots* gives the equation in the form, $RE = \epsilon \sigma T^4$, where RE = power per surface area of the source, and stands for “radiant emission”.
- Epsilon, ϵ , is called the *emissivity* of the object, and describes how easily the object emits (and absorbs) radiant energy. It is always between 0 and 1 but has a different value for each object.
- Sigma, σ , is the “Stefan-Boltzmann constant”, a fundamental constant with value of $5.670 \times 10^{-8} \text{ W}/(\text{m}^2 \cdot \text{K}^4)$

Wien’s law: $\lambda_{peak} = \frac{2.898 \times 10^{-3} \text{ meters} \cdot \text{K}}{T}$

- λ_{peak} is the peak wavelength of the radiant emission (in meters).
- For some reason the constant in the numerator isn’t given a special name/symbol in *Roots*, although Wikipedia uses the letter b . I’ll just call it the “Wien’s law constant”.

Plank’s law: $SRE = \frac{2\pi h c^2}{\lambda^5 \left(\exp\left(\frac{hc}{\lambda k_B T}\right) - 1 \right)}$

- SRE stands for “spectral radiant emission”, which is the power per surface area, per wavelength, that is emitted by a glowing object at a given wavelength λ .
- h is a new fundamental constant, called “Plank’s constant”; value of $6.626 \times 10^{-34} \text{ J} \cdot \text{s}$.
- k_B is an older constant, called “Boltzmann’s constant”, value of $1.381 \times 10^{-23} \text{ J/K}$. It had previously been discovered in the thermodynamic studies of gases.²
- λ is wavelength (in meters) and c is the speed of light.

Caution: In all three equations the temperatures must be given in kelvin.

Plank’s law actually mathematically *predicts* the Stefan-Boltzmann law and Wien’s law (although that’s not done in *Roots*), and perfectly explains the measured shape of blackbody radiation spectra. Plank derived this equation theoretically via *quantized oscillators*, which means that although light can come in different frequencies, for a given frequency the light can only come in bundles of energy which are multiples of hf .

This is reminiscent of *standing waves*, such as waves which you can make from rubber tubing with *nodes* (unmoving spots) at the edges. Increasing the frequency of oscillation results in special situations where you get successive “humps” (called *antinodes*) in the middle. Here’s a table with the first few modes along with a general “ n -hump” mode:

	Number of antinodes	Wavelength, in terms of length of tubing, L	Frequency, in terms of L and wave speed v (using $f = v/\lambda$)	Frequency, in terms of the first mode, f_1
1 st mode	1	$2L$	$v/(2L)$	f_1
2 nd mode	2	L	v/L	$2 \times f_1$
3 rd mode	3	$(2/3)L$	$v/((2/3)L) = 3v/(2L)$	$3 \times f_1$
n^{th} mode	n	$2L/n$	$v/(2L/n) = nv/(2L)$	$n \times f_1$

² If you care: it’s the ideal gas law constant R in metric units, divided by Avogadro’s number.

You can see that standing waves come in quantized frequencies where only multiples of f_1 are allowed.

Chapters 6 and 7 – The Photoelectric Effect (1902) and Dr. Einstein’s Light Arrows (1905)

In some materials, electrons can be made to jump off of a surface and be attracted to a nearby positive electrode (electrons are negatively charged, and opposites attract). Light helps this process. Whether or not an electron can be made to escape the surface depends on the *spectrum* of the light, but not on the *intensity* of the light. This was studied in the *photoelectric effect* experiments done by Lenard and others:

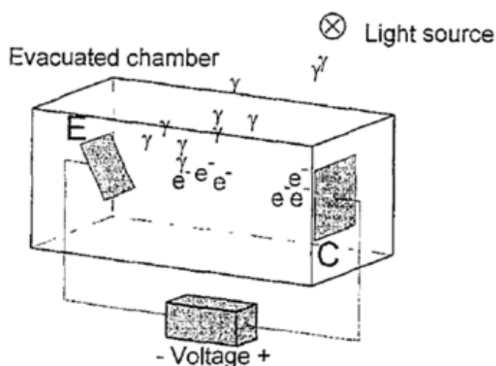


Figure 4. Typical experimental setup to study the photoelectric effect of a material. From *Roots* Fig. 6.2.

The symbol E refers to the “emitter”, which is the material being studied. The γ symbols refer to the “quantized oscillations of electromagnetic energy”, which we (thankfully) now call *photons*, coming from the light source. The e^- symbols refer to electrons which have departed the electrode on the left (material E) and are attracted to the electrode on the right (the “collector”, symbol C). The electrons can come off of the left electrode with varying amounts of energy.

The most important results of a photoelectric effect experiment for a given material are summarized by making a plot of the maximum energy of the electrons vs. the frequency of the applied light. Such a plot is different for each material. Plots for several materials are given in the next figure; the plots show that no electrons are produced until the frequency reaches some threshold (which depends on the material), and thereafter the maximum energy of the electrons increases linearly with frequency with a slope that is equal to Planck’s constant:

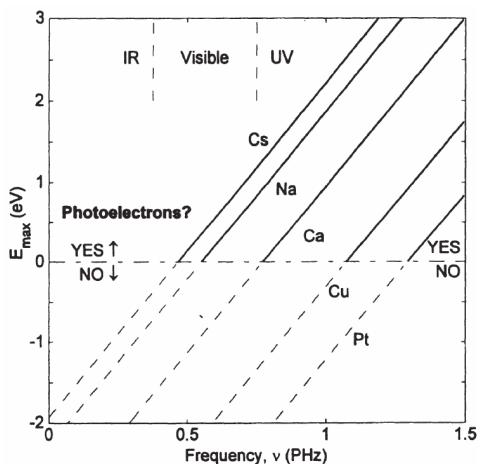


Figure 5. Photoelectric effect results for five materials. From *Roots* Fig. 7.2.

The equation describing the linear result for a given material is this:

$$E_{max} = hf - Wf$$

Wf stands for the *work function* of the material. It is a measure of how deep the “well” is that the electrons must escape from in order to leave the material. The y-intercept of the plot for a given material is the negative of the work function for that material.

Caution: When using that equation, E_{max} and Wf both need to be in joules, so to make a plot like the one above that has eV on the y-axis you would need to convert both E_{max} and Wf to eV.

The results of these experiments were explained by Einstein thusly in 1905: the energy of a light wave comes in bundles, called photons. Einstein’s genius was to give the explanation with only preliminary results, i.e. prior to the full results shown in the figure (those results were from 1916). In J.J. Thomson’s words, the photons in a wave front are like “bright sparkles on a dark background”. Each individual photon has energy

$$E = hf$$

Energy of a photon
(in terms of frequency)

or in terms of the wavelength,

$$E = hc/\lambda$$

Energy of a photon
(in terms of wavelength)

Einstein’s photon idea explains both the photoelectric effect and also the “quantized oscillators” of Planck’s radiation law.

Interlude: Fourier Analysis (replacement for Chapter 2 Appendix FOU)

Waves which vary in time, such as sound, light, voltage, etc., can be visualized via their *frequency spectrum*, which is a plot of how much of each frequency is present in the wave. Finding the frequency representation of a wave is called *Fourier analysis*.

If the wave is periodic, then all of the frequencies present will be multiples of the fundamental frequency (which is the inverse of the period). This is reminiscent of standing waves, which also come only in multiples of the lowest harmonic. In a regular frequency spectrum, the amplitudes of the frequency components are plotted, but often a *power spectrum* is used instead which depicts the amount of power present in each frequency component. A frequency component’s power is proportional to the amplitude of the component, squared.

An important example is a wave which gets shuttered on and off, to form a series of pulses like this:

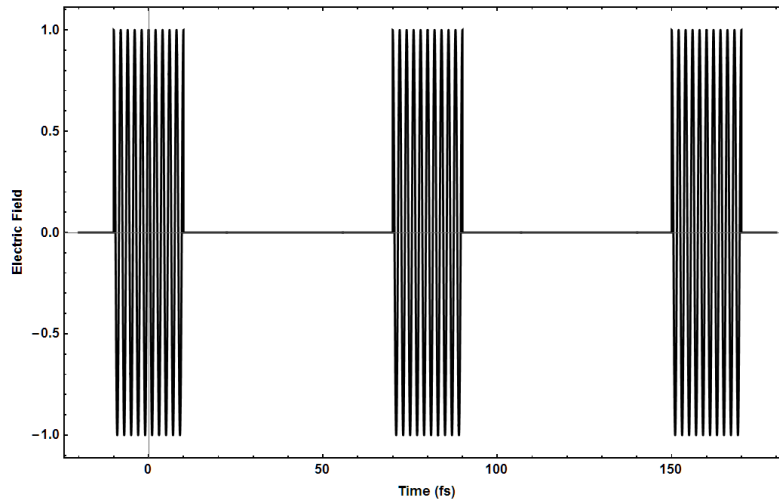


Figure 6. A shuttered wave example. From Fourier Analysis handout Fig. 5.

Here is the power spectrum of that wave:

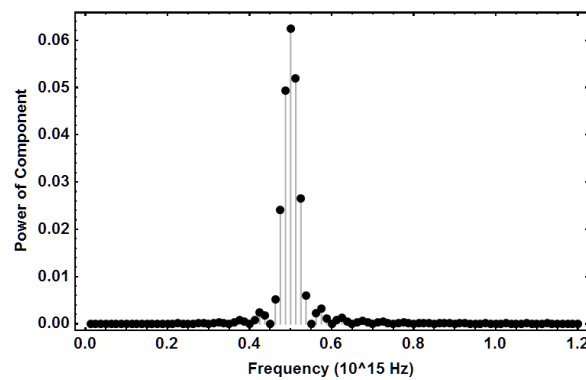


Figure 7. Power spectrum of the shuttered wave example. From Fourier Analysis handout Fig. 6.

As we vary the shuttered wave's overall repetition period and/or its pulse width, we can make several important observations and deductions:

- The power spectrum has a peak at the main oscillation frequency (e.g. the laser's electric field frequency).
- The spacing between frequency points is inversely proportional to the overall repetition period, T . Doubling the period results in the power spectrum points being spaced together $2\times$ more closely.
 - To be even more specific, the spacing between points is *equal* to $1/T$.
 - If the period is lengthened to infinity, the frequency points become spaced infinitely close together to form a continuum of frequencies rather than discrete points.
- The width of the power spectrum is inversely proportional to the pulse width. Doubling the pulse width results in the power spectrum being half as wide.
 - To be even more specific, for this particular shape (a square pulse envelope), the pulse width Δt and the width of the power spectrum Δf are related by $\Delta t \Delta f = 2$.

For a different pulse shape the product of the two widths will be different than that, but the Fourier Uncertainty Principle sets a lower bound on their product:

$$\Delta t \Delta f \geq \frac{1}{4\pi}$$

If you want a wave with a well-defined frequency (a narrow frequency spectrum), it will need to be long in duration. If you want a wave that is localized in time (a short pulse), it will need to have a wide frequency spectrum. You cannot have a wave with a well-defined frequency that is also localized in time.

These ideas, combined with the *wave-particle duality* of quantum mechanics, gives rise to the two famous Heisenberg Uncertainty Principles of quantum mechanics.

$$\Delta E \Delta t \geq \frac{\hbar}{2}$$

(Heisenberg Uncertainty Principle #1)

You cannot have a particle with a well-defined energy that is also localized in time.

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

Heisenberg Uncertainty Principle #2

You cannot have a particle with a well-defined momentum that is also localized in space.

Chapter 8 – Young, Revisited

When very sensitive detectors are used to detect smaller and smaller amounts of light, there comes a point when they start registering individual photons. That can be seen in *Roots* Fig. 8.3, the photos of a woman with differing amounts of exposures.

Much more surprising was this: when Young's double-slit experiment was repeated by G.I. Taylor in 1909, he was able to detect individual photons... but the individual photons still bunched up in interference patterns even though there were no other photons present to interfere with! In some sense the individual photons are passing through both slits at the same time and interfering with themselves! This is crazy!! But apparently it's the way the universe works.

That is depicted in this figure, and there's an actual video of the effect posted to the class website.

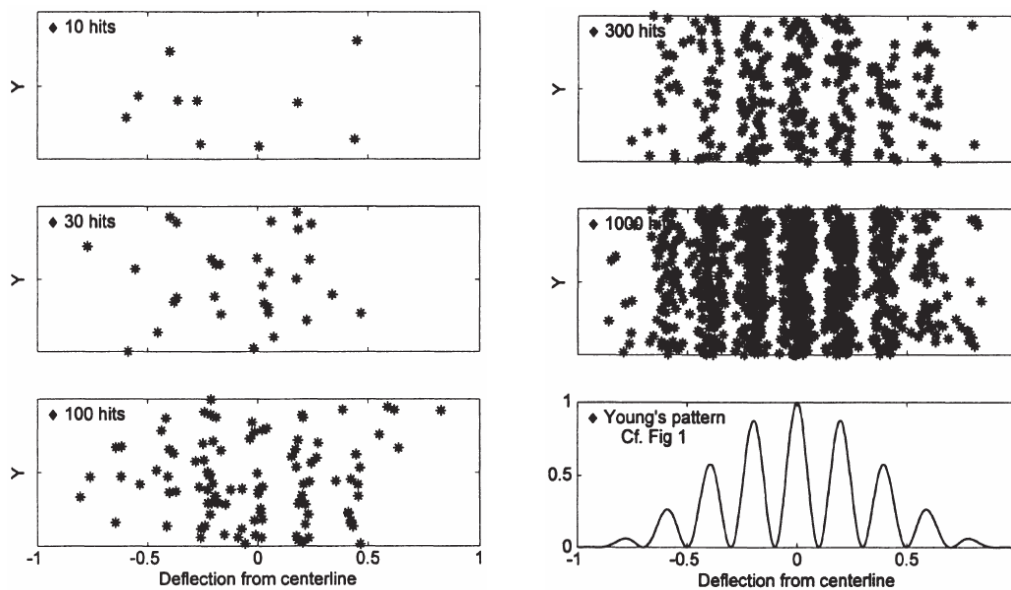


Figure 8. How interference fringes develop from individual photons. From *Roots* Fig. 8.2.

In fact, *every particle* behaves this same way, not just photons. Here's the depiction for electrons in the 1989 Hitachi experiment.

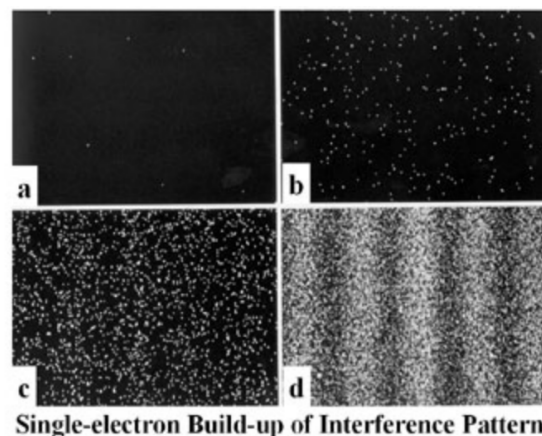


Figure 9. How interference fringes develop from individual electrons.
From <https://physicsworld.com/a/the-double-slit-experiment/>.

To make matters even stranger, if detectors are used to measure/observe which slit the photons or electrons went through, the narrow interference fringes go away and you just get a superposition of two wider bands.

Moral of the story:

Things on a very small scale behave like nothing that you have any direct experience about. They do not behave like waves, they do not behave like particles, they do not behave like clouds, or billiard balls, or weights on springs, or like anything that you have ever seen. —Richard Feynmann, 1963

Chapter 9 – The Nuclear Atom (1904-1912)

The main topic of this chapter is Ernest Rutherford's 1910 experiment where he used alpha particles to bombard nuclei, and from the results was able to deduce that the size of a nucleus is around 10 fm, meaning 10×10^{-15} m. Here's how that went.

Alpha particles are made up of two protons and two neutrons. They have a mass of 6.64424×10^{-27} kg and a charge of $+2e$. They are what's left if you take a helium atom and strip away the electrons, and are often also emitted as independent particles in various radioactive decays. As an alpha particle approaches a nucleus where the positive charge of an atom is concentrated, it experiences a repulsion because both the alpha particle and the nucleus are positively charged. Like charges repel according to Coulomb's law which governs the electric force between any two charged objects:

$$F = \frac{k_C q_1 q_2}{r^2}$$

- F is the amount of the repulsive force (in N)
- q_1 and q_2 are the values of the two charges (in C)
- r is the separation distance between the two charges (in m)
- k_C is the Coulomb force constant, sometimes called "Coulomb's constant", and has a value of $8.988 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2$.

This repulsive force creates what can be called an *energy barrier* for the alpha particles, in much the same way that a ramp creates an energy barrier for a skateboarder. Energy comes in two forms, namely kinetic and potential. Kinetic is the energy inherent in an object's motion; potential is energy inherent in an object's position. The two can be traded back and forth in many circumstances.

The equation for the kinetic energy of an object with mass m and velocity v is:

$$\text{KE} = \frac{1}{2}mv^2$$

There are also equations for potential energy, but the potential energy equation for a given situation depends on the specific force from which the potential energy arises (in the case, the Coulomb force), so we won't go into details on that right now.

Back to the skateboarder analogy. In that example, potential energy arises from gravitational force, and the higher up the ramp a skateboarder is the more potential energy she has. As she approaches the ramp, she has kinetic energy. As she rises up the ramp, the kinetic energy is turned into potential energy. The kinetic energy can be reclaimed when she goes back down the ramp, assuming no energy is lost to effects like friction.

In the Rutherford experiment, the potential energy from the nucleus causes the alpha particle to behave like a skateboarder going up a ramp shaped like this (the small blue ring indicates the size of the nucleus):

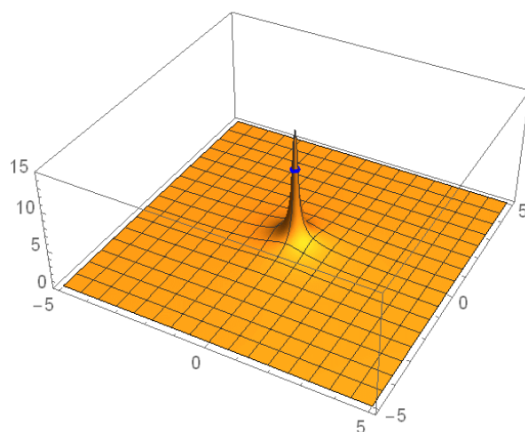


Figure 10. Potential energy seen by an alpha particle approaching a nucleus.

The *impact parameter* describes the closest distance the alpha particle would get to the nucleus if it didn't experience any repulsion at all. (The *distance of closest approach*, or DCA, by contrast, is how close the alpha particle *actually* gets to the nucleus.) For large impact parameters the alpha particles won't be deflected much. But with small impact parameters a substantial deflection can occur. For an impact parameter close to zero, the alpha particle can even “go up the ramp” a ways, then turn around and come straight back! With even more energy, the alpha particle can go all the way up the ramp and collide with the nucleus.

Here are some of Rutherford's key results for alpha particles:

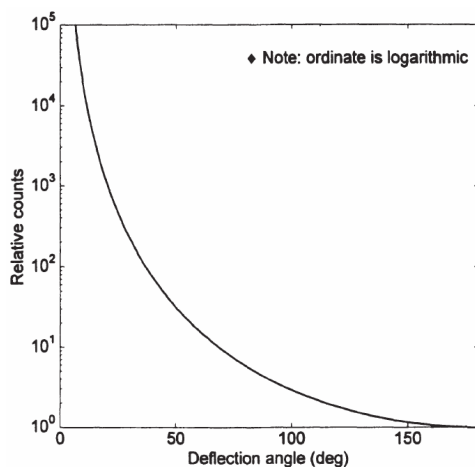


Figure 11. Number of alpha particles deflected at various angles. From *Roots* Fig. 9.5.

These experiments disproved J.J. Thomson's “plum pudding” model of the atom which thought the positive charge was spread out rather than gathered into a nucleus. *Roots* estimates (without showing details) that the deflection of an alpha particle by a single plum pudding-type nucleus would have been about 0.02° in a random direction. That means at each nucleus the alpha particles have a 50-50 chance of going left or right by about 0.02° (up-down deflections have a similar story). This is analogous to flipping a coin, where at each flip the coin has a 50-50 chance of being heads or tails. *Roots* cites a well-known statistical fact that if you flip a coin 10,000 times, then the coin will probably not go heads *exactly* 5,000 times... but one might reasonably expect the number of heads to be within $2\sqrt{10,000} = 200$ heads of

5,000. In other words, the number of heads will likely be $5,000 \pm 200$. The square root dependence comes from analyzing random numbers and is common to many statistical events, such as sampling error in election polls. The factor of 2 comes from defining “very likely” to have a precise statistical meaning (for this specific case) of “within two standard deviations”. To summarize the situation: for N flips the deviation from 50% will very likely be within $2\sqrt{N}$ flips.

This means that if Thomson’s model had been correct, for alpha particles going through a gold foil that is 10,000 atoms thick and with the particles making a 0.02° random deflection at each atom, Rutherford should have seen nearly all of the alpha particles to have deflections less than $0.02^\circ \times 2\sqrt{10,000} = 4^\circ$. But that is not at all what Rutherford observed. He saw many deflections at angles larger than that, some at very large angles. Moreover, by increasing the energy of the alpha particles he was able to see the point at which the alpha particles had enough energy to actually penetrate the nucleus, and in that case the number of alpha particles deflected at very large angles (close to 180° , i.e. straight back) was reduced. All of this data when combined with mathematical models beyond the scope of this class resulted in the conclusion that the positive charge in an atom is concentrated into a nucleus of about 10 fm in size, which was surrounded by a cloud of (presumably) orbiting electrons.

Chapter 10 – Bohr’s Atom (1913-1925)

Rutherford’s “solar system” model of the atom, as it came to be known, suffered from two huge flaws. One, it didn’t explain the observed atomic spectra. Why are only discrete wavelengths observed? And why those particular wavelengths? Secondly, the laws of electricity and magnetism (“Maxwell’s equations”) indicated that due to the loss of energy from electromagnetic waves being given off, if an electron were really orbiting a nucleus it would spiral in and collide within 16 picoseconds or so. Bohr proposed an alternate model of the atom to explain these things. Bohr’s model was a huge break from classical physics and explained some (but not all) features that had been observed. To understand Bohr’s model, we must first review the concepts of momentum, angular momentum, and centripetal acceleration.

Momentum: The momentum of an object, given the symbol p , is defined as its mass times its velocity: $p = mv$. Its usage in physics is pretty close to its usage in English—the more momentum something has, the harder it is to stop.

Angular momentum: The angular momentum of an object, given the symbol L , is defined as the object’s momentum multiplied by a “lever arm” about a central point. For the case of circular motion, the lever arm is just the radius of the circle in which the object is traveling: $L = mvr$. Loosely speaking the angular momentum is a measure of how hard it is to stop something traveling around in a circle or rotating on its own axis.

Centripetal acceleration: As per Newton’s First Law, objects tend to travel in constant velocity *straight-line motion* unless acted upon by a force. In addition to causing speeds to increase or decrease, forces can also cause velocities to *change direction*. A change in a velocity’s direction as opposed to its magnitude is also a type of acceleration. This type of acceleration is called *centripetal*, meaning “center seeking”, because the force is constantly deflecting the object towards the center of its turning circle. It can be shown (beyond the scope of this class) that the amount of acceleration an object must have to move in a circle (radius R) at constant speed v is:

$$a = \frac{v^2}{R}$$

Formula for centripetal acceleration
of object going in a circle

The amount of *force* required to move the object in a circle is that acceleration times the object's mass.

Back to Bohr's model. He proposed three postulates that had no foundation in classical physics... but which seemed to explain the experimental observations related to electrons orbiting their nuclei. They are:

- (1) Angular momentum is quantized. It only exists as multiples of Plank's constant divided by 2π , which as has been mentioned is given the symbol \hbar ("h bar"). In equation form:

$$L = n\hbar$$

where n represents any integer. This means that only certain orbits are possible.

- (2) Electrons in atoms don't follow Maxwell's equations, namely they do not give off radiation as long as they don't switch orbits.
- (3) When an electron does switch orbits, it emits or absorbs a photon to account for the energy difference between the two orbits.

If you combine Bohr's quantized angular momentum idea with formulas for the Coulomb force and centripetal acceleration applied to a hydrogen atom (a single proton attracting a single electron), then two extremely important equations result from the Bohr model, regarding the orbital radius and orbital energy of electrons in the n^{th} orbit.

$$\begin{aligned} r_n &= \frac{\hbar^2}{k_C m e^2} n^2 \\ &= (5.29 \cdot 10^{-11} \text{ meters}) n^2 \end{aligned}$$

Radius of n^{th} orbit
(Bohr model)

$$\begin{aligned} E_n &= -\frac{k_C^2 m e^4}{2\hbar^2} \frac{1}{n^2} \\ &= (-13.61 \text{ eV}) \frac{1}{n^2} \end{aligned}$$

Energy of n^{th} orbit
(Bohr model)

Success! These two equations are still used today. For example, the energy equation combined with the third postulate exactly predicts the known absorption and emission spectral lines of hydrogen atoms. Modified versions of those equations also work well for other atoms which contain only one electron, such as singly ionized helium atoms, doubly ionized lithium atoms, and so forth.

But not total success—Bohr and others could not figure out how to modify the theory for atoms with more than one electron, and experimental findings of atoms in electric and magnetic fields necessitated the creation of other “quantum numbers” besides n without much understanding or justification of them. The additional quantum numbers are labeled in *Roots* as ℓ , m_ℓ , and s . (Disclaimer: what *Roots* calls s is usually called m_s .)

We are skipping the Frank-Hertz experiment of 1914, which also helped verify the discrete nature of electron energy levels in atoms.

Chapter 11 – Compton’s Shift (1923)

We will be skipping this chapter. It describes additional experiments which confirmed the particle nature of photons, specifically collisions between photons and electrons which shifted the wavelength of the photons.

Perhaps the most important equation which came out of those experiments, which we will need to know, is the following which describes the momentum of a photon:

$$p = \frac{h}{\lambda}$$

Momentum of a photon
(in terms of wavelength)

Chapter 12 – A Princely Postulate (1924)

It had been established through experiments such as Young’s double-slit experiment, G.I. Taylor’s follow up with single photons, the photoelectric effect, and the Compton shift, that light displayed both particle and wave aspects. De Broglie took this one step farther and said that *everything* displays both particle and wave aspects. Specifically, the wavelength of a “particle” (or any object with mass) is given by the same equation used for photons:

$$\lambda = \frac{h}{p}$$

Wavelength of a particle
(in terms of momentum)

For macroscopic objects this equation predicts tiny wavelengths, and so their wave nature is never observed. But for particles such as electrons, this predicts wavelengths that are long enough to have real effects on the electrons’ behavior... for example, being able to see interference fringes in the double-slit experiment done with electrons as mentioned in Chapter 8.

Another case where the wave nature of electrons was unambiguously observed was in diffraction by crystals, where a beam of x-rays (or later, electrons) is sent towards a crystal and scattering from the planes of atoms in the crystal causes bright diffraction spots to occur at specific angles which can be determined from Bragg’s law:

$$n\lambda = 2d \cos \frac{\theta}{2}$$

Bragg’s Law

Here n is the diffraction “order”, typically a small integer, λ is the wavelength of the x-rays (or electrons), d is the spacing between planes of atoms, and θ is the diffraction angle as defined in this figure:

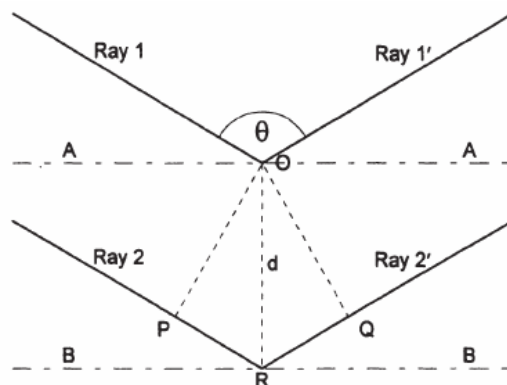


Figure 12. The geometry of x-ray diffraction experiments. From *Roots* Fig. 12.2.

Although the equation was developed by Bragg in 1912 for use with x-rays, in 1927 Davisson and Germer found that electrons diffracted in exactly the same way, with the appropriate wavelength given exactly as predicted by de Broglie's equation.

Since the definitions of kinetic energy (KE) as $\frac{1}{2}mv^2$ and momentum as mv mean that KE and momentum are related by:

$$p = \sqrt{2m \text{ KE}}$$

(test it out by plugging both definitions in, you'll see that it's true), you can equate that to the momentum from de Broglie's equation to find a useful equation that relates wavelength to KE:

$$\boxed{\frac{h}{\lambda} = \sqrt{2m \text{ KE}}}$$

You can use that to determine the wavelength of an electron with a specified energy, then you can use Bragg's law to predict diffraction angles which match up with experimental values.

De Broglie's equation, though powerful and borne out by experiment, is incomplete in that it doesn't explain *how* or *why* electrons act as waves, nor how to reconcile the concept of electrons as waves with other experiments in which electrons seem to act as particles. De Broglie's "matter waves" as we can call them, aka the particle "wave functions", are given the symbol ψ (Greek letter psi).

Chapter 13 – Realism and Other Isms

We are skipping this chapter. It has more of a philosophical bent than I myself feel comfortable teaching. You are welcome to read it on your own if that type of thing interests you. For what it's worth I think that Grometstein's description of himself as a "nondogmatic realist" fits my own view nicely, as well.

Chapter 14 – Inventing the Wave Equation (1926) – part 1

Now we have reached the heart of modern quantum mechanics! Schrödinger's equation gives more insight into what the wave function of a particle is, and allows one to calculate ψ for a given environment

which is characterized by a particle's potential energy (which I will call PE but which *Roots* calls U). We will study this chapter in four parts.

Schrödinger's equation involves derivatives. The derivative of a function is a function in its own right that describes the rate of change of the original function. The canonical example of this is how velocity relates to position: if an object's position is changing very quickly in time, it will have a large velocity at that moment. If the object's position is stationary, it will have zero velocity. If the object's position is going backwards, it will have a negative velocity. A graph of the velocity vs. time will, at each point, correspond to the rate of change (slope) of the graph of the position vs. time.

In *Roots* the derivative of the function f is labeled $D(f)$. A region of f that is sloping upwards will have a positive derivative. A region that is sloping downwards will have a negative derivative. A region which is flat will have a derivative equal to zero. The derivative of the derivative is called the “second derivative” of a function, and given the symbol $D^2(f)$. The 2 doesn't mean squared, it means two sequential derivatives. For some additional assistance on the concept if needed, see this video and this simulation:

- <https://www.youtube.com/watch?v=rAof9Ld5sOg>
- <https://phet.colorado.edu/en/simulation/legacy/calculus-grapher>

Since sines and cosines are so important when discussing waves, we need to know what their derivatives and second derivatives are. The results are as follows (notation note: $\sin kx$ is shorthand for $\sin(kx)$ to avoid using so many parentheses; similarly for $\cos kx$):

$$\begin{aligned} D(\sin kx) &= k \cos kx \\ D(\cos kx) &= -k \sin kx \\ D^2(\sin kx) &= -k^2 \sin kx \\ D^2(\cos kx) &= -k^2 \cos kx \end{aligned}$$

If the sine or cosine function has an amplitude different than 1, then just multiply the derivative by the amplitude as well, e.g.: $D(5 \sin kx) = 5k \cos kx$, etc.

We therefore have an equation for both sine and cosine waves that looks like this:

$$D^2(f) = -k^2 f$$

That's called a “differential equation”, meaning an equation that involves derivatives (which are also sometimes called “differentials”). Unlike regular equations which you solve for *values* of x , y , etc., differential equations are typically used to solve for *functions*. What's more, sine and cosine functions (along with sines and cosines added together) are the *only* functions that will solve that particular differential equation. So, for example, if through some math you arrive at an equation that looks like $D^2(f) = -49f$, the only functions that are solutions to the equation will be: $\sin 7x$, $\cos 7x$, with any amplitudes you like, or any of those functions added together.

In the last chapter we saw that the wavelength of a particle with mass m was related to its kinetic energy according to: $\frac{h}{\lambda} = \sqrt{2m \text{ KE}}$. Since the wavenumber k is related to wavelength via $k = 2\pi/\lambda$, one can write that equation in terms of k instead of λ :

$$k = \frac{\sqrt{2m \text{ KE}}}{\hbar}$$

Keep in mind that the kinetic energy of a particle could be changing in different regions of space, as potential energy gets changed into kinetic and vice versa. To create his equation, Schrödinger said that the differential equation for sines and cosines given above, namely $D^2(f) = -k^2 f$, should hold for matter waves, using the k value for the given region of space as given by the preceding equation. Combined, the two equations become this:

$$D^2(\psi) = -\frac{2m}{\hbar^2} \text{KE } \psi$$

Schrödinger then wrote the equation in terms of the total energy E_{tot} (the sum of kinetic and potential, which in general is conserved), and the potential energy PE, which is a property of the particle's position in space so we write it as $\text{PE}(x)$ as a reminder, meaning potential energy as a function of position x . Schrödinger's equation then becomes:

$$\boxed{D^2(\psi) = -\frac{2m}{\hbar^2} (E_{tot} - \text{PE}(x)) \psi} \quad \text{Schrödinger's equation}$$

This, when extended to three dimensions, is **perhaps the most important equation in science today**. It is the main equation of quantum mechanics, a differential equation that allows one to solve for the wave function ψ and total energy E_{tot} of a particle when given the potential energy landscape it finds itself in (which relates closely to the forces acting on the particle). Although Schrödinger motivated the derivation through ideas and principles such as are given above, it's actually not something that CAN be derived. It's a fundamentally new equation, just like Newton's law of gravity was, or Coulomb's law of electric force, etc., and must be judged based on how closely it predicts reality. And it is indeed amazingly accurate.

Actually, for a given potential energy function, Schrödinger's equation often predicts *multiple* possible wave functions, which we call different possible "states" or "energy states" of the particle. Each possible wave function corresponds to a different energy (unless two states coincidentally have the same energy). And, somewhat strangely, it also predicts that particles can be in *superposition* states that are sort of one state and sort of another. We will learn more about that in the next parts of this chapter.

Chapter 14 – Inventing the Wave Equation (1926) – part 2

Now that we have Schrödinger's equation, or "the Schrödinger equation" as it's also called, we need to focus on what it tells us. If we want to predict the possible states of an electron in an atom, we would need to solve the Schrödinger equation for the potential energy of an electron in the environment of a nucleus, with the wave function and the potential energy being functions of x , y , and z rather than x only. The potential energy is essentially just the opposite of what we previously discussed for an alpha particle near a nucleus, since electrons have a negative charge and alpha particles have a positive charge. In this case when plotted vs. x and y for simplicity, as we did in Fig. 10 above, the potential energy creates a well instead of a skateboard ramp:

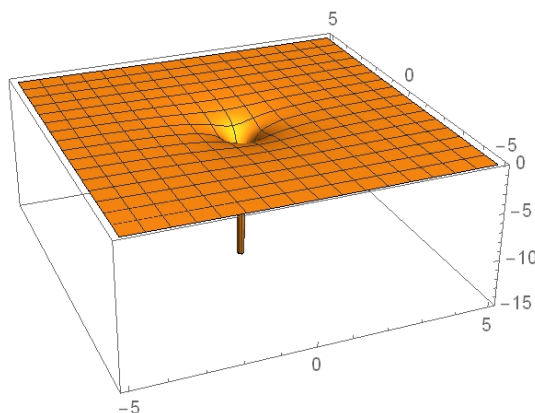


Figure 13. Potential energy seen by electron near a nucleus.

However, that problem is too hard for us to solve (it's solved in a senior-level quantum mechanics class for physics majors). So, we will create a model called an *infinite square well* which is similar in many ways, but much easier to solve. An infinite square well looks like the following but with the sides going up to infinity instead of only to 10; also the sides of the well should really be at arbitrary distances $+d$ and $-d$ instead of at $+1$ and -1 as the figure shows.

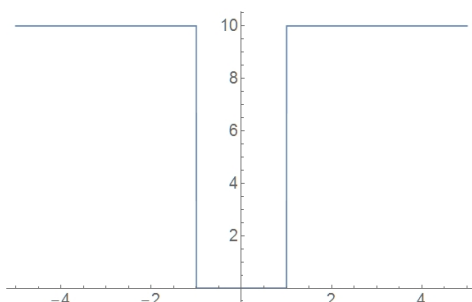


Figure 14. Square well model of potential energy seen by electrons near a nucleus, called “infinite square well” if the sides of well go up to infinity.

Given this potential energy, ψ for the electron is zero outside the well because the sides of the well form a barrier so large that it cannot be overcome. Inside the well the potential energy $V(x)$ is zero, so Schrödinger's equation becomes $D^2(\psi) = -\frac{2m}{\hbar^2} E_{tot} \psi \dots$ and we already know what the solutions are to this differential equation! They are just sine and cosine functions like $\sin kx$, $\cos kx$, or combinations. The wave function inside the well must go to zero at the boundaries in order to match up with the wave function outside the well, so that puts a condition on what values of k are allowed. This results in a set of allowed wave functions that are extremely similar to the standing waves we saw earlier in the semester. Specifically, the first four look like this:

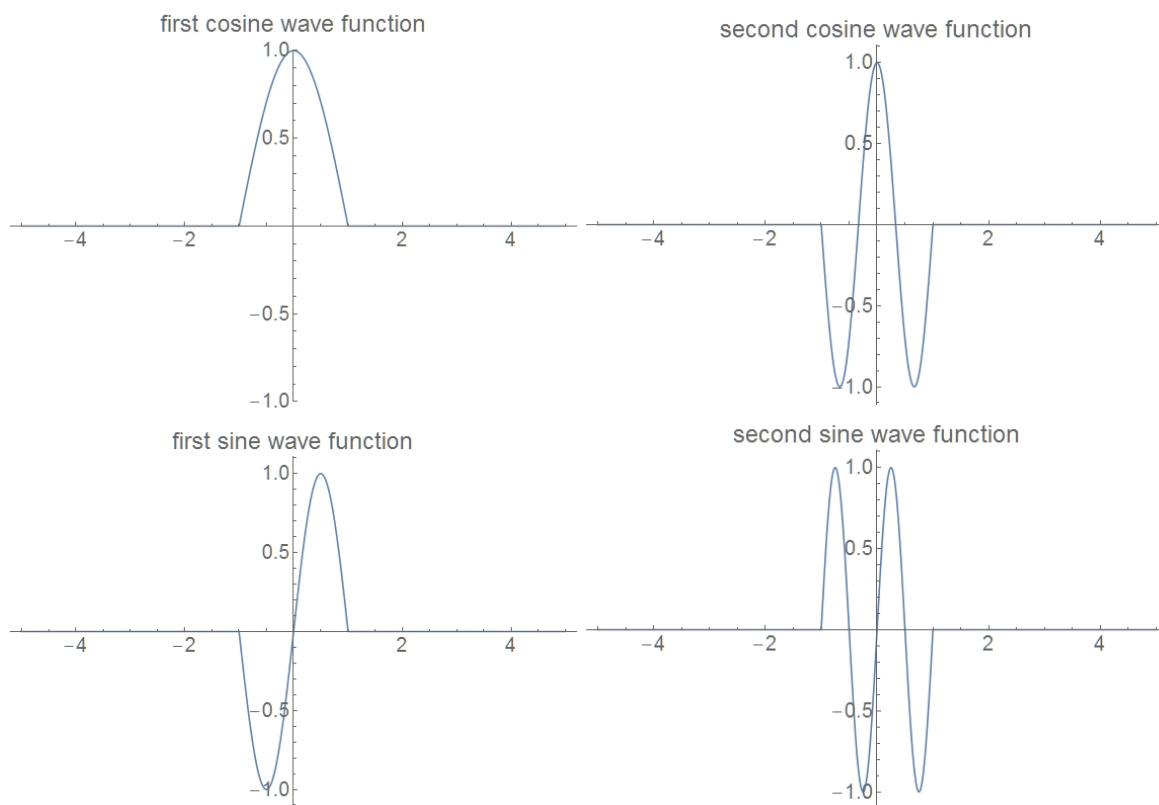


Figure 15. First four allowed wave functions of electron in an infinite square well.

If we shift the origin of coordinates to be at the left hand side of the well instead of the middle (which doesn't affect the physical situation at all), the sines and cosines together can be written as a set of sine functions only, like this, where n is the number of the state which in this case matches the number of antinodes in the corresponding ψ :

$$\psi_n = \sin \frac{n\pi x}{2d}$$

In other words, the first allowed wave function is $\psi_1 = \sin \frac{\pi x}{2d}$ (the top left wave function in Fig. 15), the second one is $\psi_2 = \sin \frac{2\pi x}{2d}$ (the bottom left wave function in Fig. 15), and so forth. These allowed wave functions are called the *eigenfunctions*.

The energy of each state can be found by plugging the equation for ψ back into the Schrödinger equation $D^2(\psi) = -\frac{2m}{\hbar^2} E_{tot} \psi$ and using what we know about second derivatives of sine functions:

$$\begin{aligned} \text{Schrödinger equation with } \psi \text{ plugged in: } D^2 \left(\sin \frac{n\pi x}{2d} \right) &= -\frac{2m}{\hbar^2} E_{tot} \left(\sin \frac{n\pi x}{2d} \right) \\ \text{Evaluate the second derivative: } -\left(\frac{n\pi}{2d} \right)^2 \left(\sin \frac{n\pi x}{2d} \right) &= -\frac{2m}{\hbar^2} E_{tot} \left(\sin \frac{n\pi x}{2d} \right) \end{aligned}$$

Some things cancel out, and we can solve what remains for E_{tot} :

$$E_{tot,n} = \frac{n^2 \pi^2 \hbar^2}{8md^2}$$

Energy of n^{th} wave function in infinite square well model

(This also happens to be the kinetic energy of the n^{th} state, since for this particular model the potential energy inside the well is defined to be zero.)

For an electron in an infinite square well which goes from -0.5 nm to $+0.5$ nm, the energies of the lowest four states are:

$$E_1 = 0.376 \text{ eV}$$

$$E_2 = 1.504 \text{ eV}$$

$$E_3 = 3.384 \text{ eV}$$

$$E_4 = 6.016 \text{ eV}$$

These energies are called the *eigenenergies*. The specific values in this model don't match the energies of electrons in any real atoms, although they are the right order of magnitude... **HOWEVER if one uses the actual potential energy instead of our simplistic model one, and the full 3-dimensional version of the Schrödinger equation instead of our 1-d version, the resulting energies for an electron in a hydrogen atom do exactly match the observed ones.** The wave functions that result from the Schrödinger equation are the *atomic orbitals* that many students learn in chemistry classes. (Or rather, to be technical, the orbitals are usually given as the wave functions squared, rather than the wave functions.)

The Schrödinger equation as given here doesn't have any time dependence. Schrödinger modified the equation to predict what the wave functions will do as time passes—in fact it's possible he did the version that includes time first, and then this version second—and he found that the wave functions oscillate at a frequency f that is related to the total energy of their states, according to:

$$f_n = \frac{E_{tot,n}}{h}$$

Frequency of oscillation of n^{th} wave function

This looks very similar to Einstein's equation for how energy and frequency relate in photons, but the frequency in the equation means something different here. For photons, it's the frequency of the electromagnetic wave that the photons are part of, i.e. the oscillation of the electric and magnetic fields that produce the wave, which when quantized give you photons. For electrons like we're talking about here, this equation is about the frequency of oscillation of the wave function itself, which as we will see in the next two sections involves complex numbers.

Chapter 14 – Inventing the Wave Equation (1926) – part 3

The wave function $\psi(x)$ is an indication of where the particle is located. In 1926 Max Born gave the more specific interpretation that $|\psi(x)|^2$, when properly normalized, indicates the *probability* of finding a particle at the given point x . This is one of the key principles of quantum mechanics. So $|\psi(x)|^2$ is really the measurable quantity, not $\psi(x)$ itself. This is good, because Schrödinger found the oscillating parts of the wave functions involve complex numbers which can't correspond to "real" locations in space (pun intended). But when you take the magnitude of the wave function squared, as indicated by absolute value signs in $|\psi(x)|^2$, you end up with a purely real function.

If the wave function squared represents the probability of finding the particle at various locations, then the proper *normalization* involves making sure that the total probability indicated by $|\psi(x)|^2$ is 100% (since

the particle must be somewhere). In general, that normalization normally requires calculus, so we won't do that; we will just assume that if an eigenfunction has been given to us, it has already been normalized. As a side note, the infinite square well wave functions given previously have *not* been normalized, but we don't really need them for anything except their general shapes so it's OK.

The Schrödinger equation predicts that particles can be in a *quantum superposition* of eigenfunctions. A superposition state is when two or more eigenfunctions are added together in some ratio such as

$$\psi(x) = C_1\psi_1(x) + C_2\psi_2(x) + C_3\psi_3(x) + \dots$$

A superposition wave function

The superposition state must also be normalized; specifically, assuming that the individual ψ_1, ψ_2 , etc., wave functions which make up the superposition have already been individually normalized, then the superposition state can be normalized by just making sure that the squares of the coefficients add up to 1 (since the wave function squared gives the probability):

$$C_1^2 + C_2^2 + C_3^2 + \dots = 1$$

Normalization condition for a superposition wave function

If they don't add up to 1 in that manner, then you must normalize them by dividing all coefficients by the square root of the sum of the squares—and then the squares of the coefficients will add up to 1.

The squares of the normalized coefficients in some sense indicate the percentage that the superposition is in each of the states 1, 2, 3, etc. The expected energy of the superposition state is given by a weighted average of the energies of the eigenstates:

$$E_{tot} = C_1^2 E_1 + C_2^2 E_2 + C_3^2 E_3 + \dots$$

Energy of a superposition state

In other words, each state that is part of the superposition contributes its coefficient squared towards the total energy of the state.

Back to complex numbers. More background on complex numbers is available at these links which were provided in the reading assignment:

- <https://www.youtube.com/watch?v=T647CGsuOVU> - Imaginary Numbers Are Real Part 1
- <https://www.youtube.com/watch?v=2HrSG0fdxLY> - Imaginary Numbers Are Real Part 2
- <https://www.youtube.com/watch?v=N9QOLrfcKNc> - Imaginary Numbers Are Real Part 3
- <https://www.youtube.com/watch?v=DThAoT3q2V4> - Imaginary Numbers Are Real Part 4
- <https://www.youtube.com/watch?v=65wYmy8Pf-Y> - Imaginary Numbers Are Real Part 5
- https://www.youtube.com/watch?v=z5IG_6_zPDo - Imaginary Numbers Are Real Part 6
- <https://www.youtube.com/watch?v=YHvR8siLiD0> - Imaginary Numbers Are Real Part 7
- <https://www.youtube.com/watch?v=sKtloBAuP74> - Euler's Identity

The main points of these videos can be summarized as follows:

1. Imaginary numbers are a useful mathematical construct which can help solve problems that involve real situations.
2. Imaginary numbers lie on an axis that is in some sense perpendicular to the real number line, like the normal y-axis is perpendicular to the normal x-axis, and hence complex numbers (numbers which have both a real and an imaginary part) can be thought of as points in a plane.

3. One can represent those points in polar form, with a magnitude and an angle. For a given complex number z which has real part x and imaginary part iy , the magnitude is given by (via the Pythagorean theorem) $|z| = \sqrt{x^2 + y^2}$, and the angle can be found through some trigonometry.
4. Multiplying together two complex numbers has a geometrical interpretation, namely the angles of the two numbers add together and the magnitudes multiply.
5. Euler's equation is $e^{ix} = \cos x + i \sin x$. By thinking about where the point $(\cos x, \sin x)$ lies in the complex plane, one can deduce that the complex number e^{ix} has a magnitude of 1 and an angle of x . Therefore multiplying a general complex number by e^{ix} causes a rotation of x radians or degrees (depending on what angular units you're using) without changing the magnitude of the complex number. Multiplying by e^{-ix} causes a rotation of $-x$ radians or degrees.

How do complex numbers relate to wave functions? When I previously said that a wave function oscillates at frequency $f = \frac{E_{tot}}{h}$, I didn't mean the wave function just oscillates up and down in amplitude at that frequency. If it did that, there will be many times when the wave function has a very small amplitude and during those times the normalization condition would be violated—there wouldn't be a 100% chance of finding the particle somewhere.

What really happens, is that each eigenfunction oscillates according to $e^{-i\omega t}$, where the angular frequency of oscillation $\omega = 2\pi f$:

$$\psi_n(x, t) = \psi_n(x)e^{-i\omega_n t}$$

How eigenfunctions
oscillate in time

ω_n relates to the total energy of the n^{th} state as follows:

$$\omega_n = 2\pi f_n = 2\pi \left(\frac{E_{tot,n}}{h} \right) = \frac{E_{tot,n}}{\hbar}$$

By Euler's formula we can see that $e^{-i\omega t}$ is $\cos \omega t - i \sin \omega t$. Multiplying by this factor therefore gives rise to sinusoidal oscillations in time, in both the real and the imaginary part of a wave function, but because they are out of phase with each other the normalization condition of $|\psi|^2$ is maintained. This is because $e^{-i\omega t}$ has a magnitude of 1 regardless of what time it is (so when you multiply a complex number by that factor, the magnitude of the resulting complex number is unchanged).

For superposition states, each eigenfunction component oscillates at its own particular frequency given by its own particular energy, like this:

$$\psi(x, t) = C_1\psi_1(x)e^{-i\omega_1 t} + C_2\psi_2(x)e^{-i\omega_2 t} + C_3\psi_3(x)e^{-i\omega_3 t} + \dots$$

How superposition
states oscillate in
time

where $\omega_1 = E_1/\hbar$, $\omega_2 = E_2/\hbar$, $\omega_3 = E_3/\hbar$, etc.

By the way, the complex number nature of the wave function and how eigenfunctions oscillate in time is only referred to in passing in *Roots* and depicted in one figure but without the (hopefully helpful) added detail that I have provided here.

To sum up some of the major points of this and the previous section:

- Potential wells have special solutions called eigenfunctions.

- Each eigenfunction represents a state with a different energy.
- Eigenfunctions have both real and imaginary components.
- Those real and imaginary components oscillate at frequencies that depend on the energies of the given state.
- The magnitude squared of the wave function is a probability function describing where the particle is likely to be.
- Particles can be in superposition states that contain different amounts of two or more eigenfunctions; in that case the state is written as a sum of eigenfunctions with different coefficients (normalized so the squares of the coefficients add up to 1).
 - The squares of the coefficients indicate the probability of being in each of the eigenstates.
 - Each eigenfunction oscillates at its normal frequency but the state as a whole does not oscillate at any one frequency.

All of those points can be demonstrated with the wonderful PhET simulation called “Quantum Bound States”: <https://phet.colorado.edu/en/simulation/legacy/bound-states>

Chapter 14 – Inventing the Wave Equation (1926) – part 4

The final section of Chapter 14 involves a scenario where electrons are shot at a potential “barrier”, which is the opposite of a potential well. The scenario can be demonstrated with the wonderful PhET simulation called “Quantum Tunneling and Wave Packets”, if you set the potential to a few different values as specified in the homework instructions: <https://phet.colorado.edu/en/simulation/legacy/quantum-tunneling>

Some key observations:

- A wave function is made up of a real and an imaginary component, each of which can have spatial oscillations (see *Roots* Fig 14.4).
- A wave function traveling in a region with a flat potential energy will gradually spread out (see *Roots* Fig. 14.5).
- A wave function which impacts a barrier will split into two parts: a part that travels through the barrier and a part that gets reflected off the barrier. The two parts represent probabilities that the wave will transmit or reflect. (See *Roots* Fig. 14.6.)
- If the barrier is sufficiently large, the part that transmits will be very small, essentially zero (see *Roots* Fig. 14.7).
- The reflected portion can interfere with the incoming wave, resulting in spatial oscillations in not only ψ but also $|\psi|^2$ (see *Roots* Figs. 14.7 and 14.8).

Chapter 15 – Collapsing the Wave

If a particle is in a quantum superposition state—that is, a superposition of eigenfunctions, each with its own coefficient—when a measurement is done the wave function will abruptly change, or “collapse”, to one of the eigenfunctions. Which eigenfunction gets chosen is randomly determined with a probability given by its normalized coefficient squared. One example of this is an electron wave packet that hits a barrier with some chance of reflection and some chance of transmission. As was discussed in chapter 14, the wave function splits into two parts: one that reflects and one that transmits. When a measurement is made, the particle is found to be either reflected or transmitted and the wave function collapses to one of those two options. Another example of this is an electron in an atom that is in a superposition of energy states. When it emits a photon the electron collapses to an eigenstate, so the photons are only emitted at energies corresponding to differences between eigenstates.

What does it mean to say that a “measurement is done”? My view is that it means “the quantum system has interacted with a macroscopic system”. I also believe (as I suspect many physicists do, but perhaps is not provable), that if the macroscopic system could also be described in quantum terms then there would be no abrupt wave function collapse but rather a smooth change of the combined wave function. Thus my view is the collapse is an artifact of not being able to describe the macroscopic system in quantum terms because there are too many interacting particles and the combined wave function is too complicated to possibly be worked out. Other physicists may have other views.

This chapter also mentions Heisenberg’s matrix mechanics which was developed in 1925, the year before Schrödinger’s wave equation was formulated. Matrix mechanics is an alternate way of doing quantum mechanics compared to the wave equation; the mathematics behind Heisenberg’s method are different and in some ways more complicated than Schrödinger’s equation so we won’t go into it. Sometime between 1926 and 1939 (I couldn’t locate the exact date) Schrödinger and Dirac proved that the two methods gave rise to mathematically equivalent predictions, even though the details of the calculations were very different. As a side note, a third method for doing quantum mechanics was developed by Richard Feynman in 1948, and this method was *also* proven to be equivalent to Schrödinger’s and Heisenberg’s methods; each method is easier for some problems and harder for others, so all three methods are still used by physicists today.

Chapter 16 – Copenhagen Takes Over (1925-?)

The discussions of Niels Bohr and those who visited the Institute for Theoretical Physics of the University of Copenhagen where he worked (renamed the Niels Bohr Institute in 1965 in his honor) led to the following prevailing view of what quantum mechanics is, how to use it, and what it means. It is called the “Copenhagen Interpretation” and includes the following points:

- 1) The quantum states of particles are described by wave functions, labeled ψ (Schrödinger).
 - a. The eigenfunctions are special solutions to Schrödinger’s equation that oscillate at well-defined frequencies.
 - b. If they are interacting, or *entangled* particles, then ψ represents the state of the combined system rather than an individual particle.
- 2) Quantum mechanical particles exhibit both particle-like and wave-like behavior, which is called the principle of *complementarity* (Bohr).
- 3) $|\psi|^2$ represents a probability function. It should be normalized to 100% (Max Born).
- 4) ψ can be a weighted sum of multiple eigenfunctions; if so, it’s called a *superposition* state.
- 5) During an *observation*, ψ collapses to one of the eigenfunctions.
 - a. One cannot directly measure ψ because the act of measuring it affects the state.
- 6) ψ represents all that can be known about the system prior to the observation.
- 7) Not all properties of the system can be completely known simultaneously; for example position and momentum are *conjugate pairs* and can’t both simultaneously be known to high precision (Heisenberg’s Uncertainty Principle).
- 8) Quantum mechanics will reproduce classical physics in the limit of large quantum numbers (i.e. for large orbits/energies); this is called the *correspondence principle* (Bohr and Heisenberg).

Uncertainty Principles, reviewed: since particles are described by wave functions, Fourier analysis is relevant to these waves. As discussed in the Fourier Analysis section above (and in the handout), a wave which is localized in time must contain many frequency components. Since frequency relates to energy, we can say a wave which is localized in time must contain a range of energies. A wave which is localized in space must contain many spatial frequency components. Since spatial frequency is related to

wavelength according to $k = 2\pi/\lambda$, and wavelength relates to momentum according to $p = h/\lambda$, we can say a wave which is localized in space must contain a range of momentums. Those concepts turn into the two Heisenberg Uncertainty Principles of quantum mechanics,

$$\Delta t \Delta E \geq \frac{\hbar}{2}$$

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

Heisenberg Uncertainty
Principles, again

Roots writes the second of those as Chapter 16, Eq. 1, but does so a little less precisely and is missing the factor of 2 in the denominator. (The factor of 2 comes from a precise definition of Δ as meaning the standard deviation.)

Heisenberg's microscope. Heisenberg demonstrated in a thought experiment with a microscope that uses only a single photon of light to observe an electron, that measuring the position of an electron will invariably affect its momentum. A more and more precise measurement of position will give rise to more and more uncertainty of momentum, in just the same way that the uncertainty principle predicts. Here is an excellent video on the topic: https://www.youtube.com/watch?v=dgoA_jmGlcA

Einstein and Bohr Solvay debates (1927 and 1930). In these two famous conferences in Brussels, Belgium, Einstein made several attempts to prove that quantum mechanics and the Copenhagen Interpretation were wrong, or at least incomplete. Bohr thwarted all his attempts. See this excellent video on the topic: <https://www.youtube.com/watch?v=8PUi0I02Mdc>, although note that with regards to Einstein's 1930 challenge, the video refers to the "mass" and the "weight" of a photon. That's incorrect. Photons have no mass nor weight. The way it's described in *Roots* on page 432 is better, which is that the *energy* of the departing photon creates a loss of measured mass of the box through Einstein's $E = mc^2$ equation.

Chapter 17 – EPR (1935) and EPR-B

In 1935 Einstein, Podolsky, and Rosen published the famous "EPR" paper objecting to quantum mechanics, specifically to entangled states. Their objection can be summarized by the following: if two particles are entangled, then collapsing the wave function of one will necessarily instantaneously collapse the wave function of the other one, *no matter how far apart the two particles may be*. On its surface this appears to violate Einstein's theory of relativity which holds that nothing can travel faster than the speed of light. Their solution was to propose that some properties (or "variables") exist that govern the situation to create what *looks* to be a wave function collapse (but isn't really), but that these variables are hidden so we don't notice them. The variables always existed before they interacted with the macroscopic system, though, so nothing really collapsed faster than the speed of light (in their view). This view is antithetical to quantum mechanics, and for a time people didn't know which was correct, the EPR view or the quantum mechanical one.

In 1951 David Bohm expanded on this a bit, elucidating the EPR position while at the same time defending the quantum mechanical one, in what *Roots* calls the EPR-B paper. And in 1964 John Bell wrote a seminal paper proposing an experiment to test whether the EPR view was correct or whether the quantum mechanical view was correct. Since most tests of Bell's theorem involve polarization of photons, that's what we'll discuss.

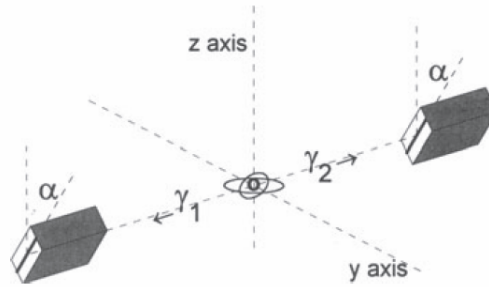


Figure 16. Bohm configuration for entangled photon experiments. Photons are produced between two analyzers. For this configuration, each polarization analyzer is tilted at an angle α to the vertical. From *Roots* Fig. 17.3.

Figure 17 depicts two photons which are produced from a single source in such a way that their quantum states are entangled; specifically, **they are emitted with the same polarization**. This can be done experimentally by exciting calcium atoms with a particular wavelength of light that causes two photons to be produced by the calcium atoms in response, in opposite directions. The photons hit polarizers which the experimenter has positioned to measure the individual polarization of the two photons. In this configuration the polarizers are set for the same angle, α .

Roots has a whole slew of polarization facts, or “P-facts”, in this chapter and the next, but here are some of the most important:

- Polarizations rotated by 180° are equivalent. There is no difference between a polarization of, say, 12.4° and a polarization of 192.4° , or between a polarization of 60° and one of -120° .
- A polarizer acting on a polarized beam of photons will pass $\cos^2 \theta$ of them (expressed as a fraction or percentage), where θ is the angle between the polarizer and the polarization of the beam. Similarly, one can say:
 - (for the beam) $I = I_0 \cos^2 \theta$. This is called Malus’s law. I_0 is the intensity of the beam before the polarizer and I is the intensity after.
 - (for individual photons) Each individual photon’s chance of passing is given by $\cos^2 \theta$.
- The photons that make it through a polarizer are forced to the polarization angle of the polarizer.

In the examples given in this chapter and the next, for the sake of simplicity let’s assume the left photon hits its polarizer first. (Nothing substantial changes if the right photon hits its polarizer first, the situation just get reversed.) Quantum mechanics says that no matter what the state of the left photon before it hits polarizer 1, hitting that polarizer (doing a measurement on its polarization state) will cause its wave function to collapse into either a polarization state of α , in which case it passes, or a polarization state 90° away from that, in which case it is blocked (fails to pass).

Since the two photons are entangled, collapsing photon 1 to α (or $90^\circ + \alpha$) will immediately cause photon 2 to collapse to α (or $90^\circ + \alpha$). It will then either pass or fail its polarizer in exactly the same way that photon 1 did with *its* polarizer. In other words, the two photons will always have the same fate. In the language of *Roots*, 100% of these **SameTilt** runs should result in **SameFate** and 0% of them in **DiffFate**.

What happens when the experiment is done? It’s exactly as quantum mechanics predicts:

100% of the **SameTilt** runs are **SameFate**

Can this be explained within the EPR framework of hidden variables? Yes! However, there is only one possible explanation. Or at least, only one that anyone has come up with so far. As Grometstein writes in the next chapter on page 497, “if you know of a different explanation for [that result], please shout it from

the rooftops.” The EPR explanation is that photons carry a “code”, which can be visualized as a table or a disk, that tells the photon what to do when it meets a polarizer... and that both of the entangled photons in the experiment are created with the exact same code. It might look like this, for example:

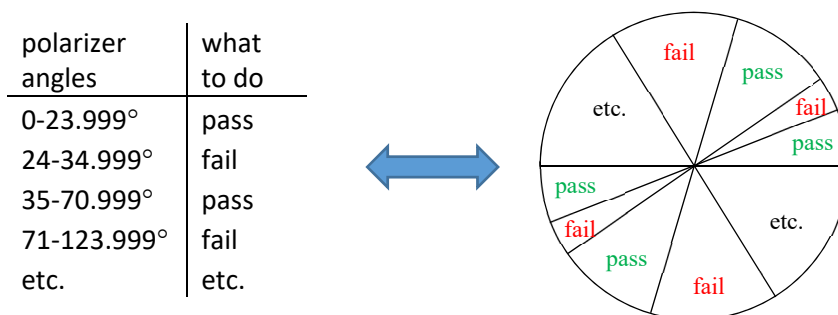


Figure 17. A depiction of a hidden polarization code.

If both photons get produced with the same code, then that naturally leads to them both responding the identical way to polarizers at the same angle.

Chapter 18 – Bell’s Thunderbolt (1964)

Bell’s paper in 1964 asked the question, “What happens if the two polarizers are at *different* angles?” He showed that when the polarizers are set for two different angles, the EPR view and the quantum mechanics view can lead to different predictions. This then gives an experimental measurement that can be done to see whose view was correct, Einstein (EPR) or Bohr (quantum mechanics). The result? Einstein was wrong and Bohr was right.

This chapter examines two different configurations of polarizer 1 vs. polarizer 2.

First configuration

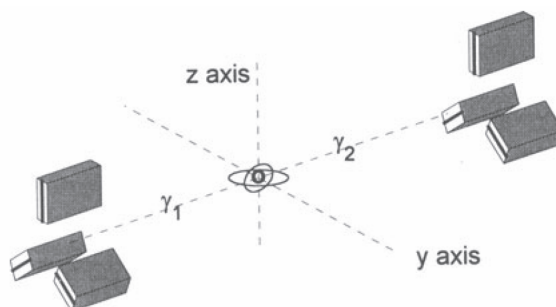


Figure 18. Another configuration for entangled photon experiments. Each side has three possible analyzer angles, which are 0°, 120°, and –120°. From *Roots* Fig. 18.3.

The figure makes it look like there are three different polarizers on each side, but what it really means is there are three different *options* for polarizers: 0°, 120°, and –120° (which is the same as 60°). There are thus 9 different combinations for the two polarizer locations that can be set up; the experimentalist chooses between them randomly during the experiment. We will analyze this situation using both the EPR and the quantum mechanical views, and see which matches the actual experiments.

First, the EPR point of view. This uses the idea of hidden variables described in code tables or disks. Since there are only three different polarizer settings for a given photon, the infinite number of code table options can be broken down into 8 different categories, based on whether a code says “pass” or “fail” for each of the three angles. We label those FFF, FFP, FPF, FPP, PFF, PFP, PPF, and PPP. For example, FFF means a photon will fail at 0°, 120°, and also –120°. “PFF” means a photon will pass at 0° but fail at 120° and –120°. Etc. Those eight possibilities will all have equal chances to be realized. All together then, there are $9 \times 8 = 72$ different possibilities of polarizer configurations and photon codes. To analyze this we can make a large table that describes what will happen for each of the 72 cases. Here ✓ means pass and ✗ means fail, and S and D refer to **SameFate** and **DiffFate** results, respectively.

		PPP code			PPF			PFP			PFF			FPP			FPF			FFP			FFF		
polarizer 1	polarizer 2	phot on 1	phot on 2	same /diff	phot on 1	phot on 2	same /diff	phot on 1	phot on 2	same /diff	phot on 1	phot on 2	same /diff	phot on 1	phot on 2	same /diff	phot on 1	phot on 2	same /diff	phot on 1	phot on 2	same /diff	phot on 1	phot on 2	same /diff
0	0	✓	✓	S	✓	✓	S	✓	✓	S	✓	✓	S	✗	✗	S	✗	✗	S	✗	✗	S	✗	✗	S
0	120	✓	✓	S	✓	✓	S	✓	✗	D	✓	✗	D	✗	✓	D	✗	✓	D	✗	✗	S	✗	✗	S
0	-120	✓	✓	S	✓	✗	D	✓	✓	S	✓	✗	D	✗	✓	D	✗	✗	S	✗	✓	D	✗	✗	S
120	0	✓	✓	S	✓	✓	S	✗	✓	D	✗	✓	D	✓	✗	D	✓	✗	D	✗	✗	S	✗	✗	S
120	120	✓	✓	S	✓	✓	S	✗	✗	S	✗	✗	S	✓	✓	S	✓	✓	S	✗	✗	S	✗	✗	S
120	-120	✓	✓	S	✓	✗	D	✗	✓	D	✗	✗	S	✓	✓	S	✓	✗	D	✗	✓	D	✗	✗	S
-120	0	✓	✓	S	✗	✓	D	✓	✓	S	✗	✓	D	✓	✗	D	✗	✗	S	✓	✗	D	✗	✗	S
-120	120	✓	✓	S	✗	✗	D	✓	✗	D	✗	✗	S	✓	✓	S	✗	✓	D	✓	✗	D	✗	✗	S
-120	-120	✓	✓	S	✗	✗	S	✓	✓	S	✗	✗	S	✓	✓	S	✗	✗	S	✓	✓	S	✗	✗	S

There are 48 **SameFate** cases out of the 72 possibilities in the table ($48/72 = 66.7\%$), so EPR predicts:

SameFate = 66.7%

(EPR prediction for the Fig. 19 configuration)

Or, the way *Roots* describes it, all of the middle cases have five out of nine **SameFate** results ($5/9 = 55.6\%$), and the first and last two columns will only increase that number, so **SameFate $\geq 55.6\%$** .

Now the quantum mechanical point of view. Suppose the left polarizer is in the 0° position. The left photon will either pass or fail.

If the left photon passes, then that’s because it has collapsed to 0°, and due to entanglement the right photon will also collapse to 0°. If the right polarizer is set to 0° (1/3 of the time), the right photon will pass and we’ll get **SameFate** for all of those configurations. If the right polarizer is set to either 120° or –120° (2/3 of the time), the right photon will pass only $\cos^2 120^\circ = 25\%$ of the time for those configurations. So overall for the “left photon passes” case we’ll get **SameFate** $(1/3) \times 100\% + (2/3) \times 25\% = 50\%$ of the time.

If the left photon fails, then that’s because it has collapsed to 90°, and due to entanglement the right photon will also collapse to 90°. If the right polarizer is set to 0° (1/3 of the time), the right photon will NOT pass at all, i.e. will fail 100% of the time, and we’ll get **SameFate** for all of those configurations. If the right polarizer is set to either 120° or –120° (2/3 of the time), both of which are 30° away from 90° and/or –90°, the right photon will pass $\cos^2 30^\circ = 75\%$ of the time. That means the right photon will fail 25% of the time, and we’ll get **SameFate** 25% of the time for those configurations. So just like before overall we’ll get **SameFate** $(1/3) \times 100\% + (2/3) \times 25\% = 50\%$ of the time.

If the left polarizer is in the 120° or -120° positions, the math works out exactly the same. All of the various cases work out to give exactly a 50% chance for **SameFate** results, so overall we have:

$$\text{SameFate} = 50\%$$

(QM prediction for the Fig. 19 configuration)

What do the experiments show? They are simply stated:

$$\text{SameFate} = 50\%$$

(Experimental results for the Fig. 19 configuration)

The experimental results match the prediction of quantum mechanics and not the EPR view.

Second configuration

The second configuration of polarizer 1 vs. polarizer 2 is this; each polarizer has only two basic options:

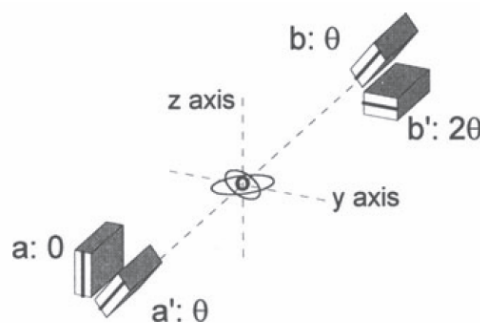


Figure 19. Another configuration for entangled photon experiments. Each side has two possible analyzers, which are 0° and θ for one side, and θ and 2θ for the other side. (The angle θ is chosen arbitrarily.) From *Roots* Fig. 18.4.

Polarizer 1 can be either 0° or some arbitrary angle θ ; polarizer 2 can be either θ or 2θ , where θ is the same angle used for polarizer 1. Those four settings are referred to as a , a' , b , and b' , respectively. The four possible combinations within this configuration are ab , ab' , $a'b$, and $a'b'$.

We are going to completely skip the mathematical analysis now, except to say that it's presented in *Roots* in terms of a *correlation function*: the correlation between polarizers 1 and 2 for a given run is defined as $+1$ if **SameFate** and -1 if **DiffFate**. If we imagine a large number of runs done in a row, we can talk about the average correlation value, C_{ave} . The average correlation for the four combinations can be calculated using both the EPR and the quantum mechanical views, and then we can see which matches the actual experiments.

For reasons that I won't get into, the correlation functions for the four cases are combined in a specific way to construct this function which Grometstein calls B in honor of John Bell:

$$B = |C_{ave}(ab) - C_{ave}(ab')| + |C_{ave}(a'b') + C_{ave}(a'b)|$$

Quantum mechanical point of view: Here's a plot of that B function analyzed from the QM point of view, plotted as a function of θ (in degrees):

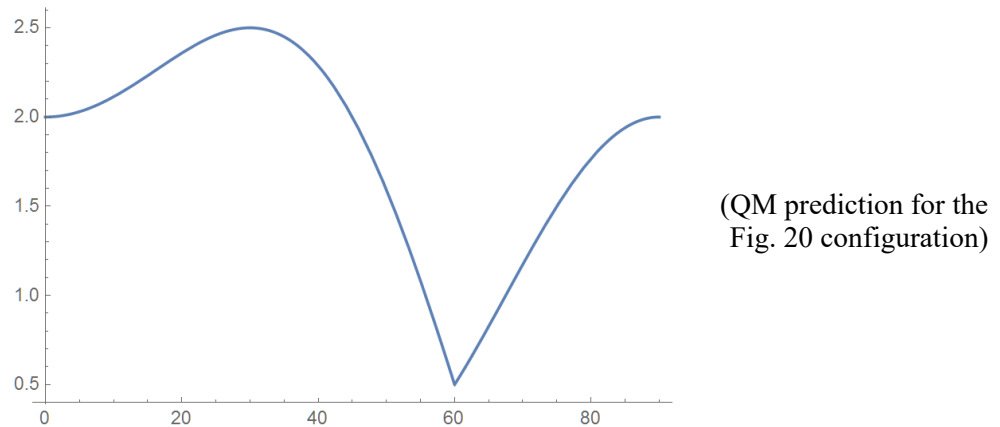


Figure 20. The function B plotted vs. arbitrary polarizer angle θ . Similar to *Roots* Fig. 18.5, but I think there was a small error in his calculation which I have fixed here.

Now the EPR point of view... When in the EPR view *locality* is assumed, namely that the setting of polarizer 1 cannot influence the results of polarizer 2, and vice versa, the result is this:

$$B \leq 2 \quad \text{(EPR prediction for the Fig. 20 configuration, aka "Bell's inequality")}$$

Clearly this disagrees with the quantum mechanical prediction because as you can see in the graph above, the quantum mechanical plot says that B should be greater than 2 for some of the angles.

What do the experiments show? Again we have a testable difference between quantum mechanics and the EPR view. So, what do the experimental results tell us? **An overwhelming number of experiments show Bell's inequality to be violated and the quantum mechanical prediction to be fulfilled.** *Roots* says this is the case for 5 experiments, but the numbers since *Roots* was published are far more conclusive. Dozens and dozens of additional experiments have verified the violation of Bell's inequality, i.e. validating quantum mechanics and disproving the EPR prediction. This Wikipedia article lists some of the notable ones: https://en.wikipedia.org/wiki/Bell_test_experiments. There may still be a little debate as to whether any of the experiments allow for tiny loopholes that still permit an EPR-type of explanation, but generally speaking the sum total of the experiments is now considered to be definitive.

Chapter 19 – The Toll of Bell

So, what does this all mean? Well, it means that the EPR view of hidden variables is wrong. The polarization state of the photons is not predetermined in some sort of code at the moment of creation. The equations of quantum mechanics, on the other hand, give the correct predictions.

The validation of quantum mechanics does not, however, mean that the *Copenhagen Interpretation* itself is necessarily correct. The mathematical theory of quantum mechanics has been proven to be correct, or at least as much as any scientific theory is ever "proven correct"—meaning, future experiments could always potentially find special situations where it doesn't apply, or could show tiny deviations from predictions that aren't visible with today's experimental capabilities. But as best we know for now, the mathematical theory is correct. However, even though the equations yield correct predictions, there may be other ways of *interpreting* the theory that, for example, don't involve the instantaneous collapse of a wave function as described in the Copenhagen Interpretation. Some of these alternate interpretations are

explored in Chapter 19, and many more are listed in this Wikipedia article, https://en.wikipedia.org/wiki/Interpretations_of_quantum_mechanics.

The alternate interpretations are often phrased in terms of their views on *reality* and *locality*. Consider an electron tunneling through a barrier, as we discussed in Chapter 14 part 4. The electron can either go through the barrier or be reflected by it. When a measurement is done, the electron is found to be either in the reflected or transmitted state.

Reality – What is the state of the electron after it encounters the barrier but before the measurement is made? Does a well-defined state even exist? Is the electron always real? The CI says “No; the electron is in a superposition state. Its properties are latent and hence it does not become a real object in the usual sense until the measurement is made.”

Locality – Can remote events influence each other? Can a measurement of the potentially transmitted electron far to the right of the barrier impact the state of the potentially reflected electron far to the left of the barrier? The CI says, “Yes; the electron is in an inherently nonlocal state, and a measurement of one part of the state will immediately impact the other part of the state even if it is far away.

CI says that reality and locality do not both exist. EPR says that reality and locality do both exist. EPR was proven false. But that doesn’t mean the CI is necessarily correct. Perhaps we can give up reality and keep locality? Or give up locality but keep reality? Or discuss shades of both rather than viewing them each as black and white? Those are the types of questions alternate interpretations of quantum mechanics often try to address.

The two main alternate interpretations featured in *Roots* are called the “Many Worlds” theory and the “Pilot Wave” theory. (*Roots* calls the latter the “Theory of Motion” but that’s not a standard name.)

In Many Worlds theory developed by Hugh Everett, when a measurement is made, for example, the electron transmitting through the barrier, the wave function does not collapse but rather the universe itself splits: in one universe the electron transmits and in another universe the electron reflects. The fact that we believe the electron to have transmitted is a statement about which universe we are now in, rather than a statement about anything that happened to the electron. This is basically the source of the science fiction notion of parallel universes.

In the Pilot Wave theory developed by David Bohm, the electron is a real particle. The quantum mechanical wave function does not describe the electron itself but rather can be used to find a guiding “quantum potential” field, *Q*. The *Q* field has a number of strange and unique properties, including substantial nonlocality: a change to the environment of the electron in one spot will instantly impact the *Q* field for the electron everywhere. Each particle in the universe has its own *Q* field, and the respective *Q* fields can be used to (for example) calculate trajectories of particles in a mostly deterministic way. It’s “mostly” deterministic because the uncertainty lies in the inability to precisely know a particle’s initial position and velocity to a high degree of precision, along with the inability to precisely know the *Q* fields. But these quantities are *unknown*, not *unknowable* as they are in the CI.

This then leads us to the end of the book. Is CI correct? Many Worlds theory? Pilot Wave theory? Something else entirely? It’s uncertain. As yet there aren’t experiments which can distinguish between these alternate interpretations. But perhaps at some point in the future the differences between them may be able to be tested by experimental measurements in the same way that the Bell’s inequality experiments have tested and disproved the EPR assumptions.