Quantum Mechanics and Astrophysics **Summer 2018**

Lecture 1: Introduction to Astrophysics and Quantum Mechanics

Lecturers: Kavish Gandhi and Yuan Lee

1.1 Introduction

In this course, we will be covering a wide array of topics in astrophysics and quantum mechanics, making sometimes direct and sometimes tangential connections between them. In this first lecture, we will make no attempt at this, but rather provide a background in each subject. We hope you enjoy!

1.2 Astrophysics: Distance Ladder

A typical starting point in astrophysics is the distance to objects we see. To understand how large, how bright, how massive the twinkling point sources we see in the sky every night truly are, we first need a way of figuring out how far away they are. In astrophysics, we typically can only calculate these distances in terms of smaller distance. For instance, to know the distance to nearby stars, it is impractical or impossible to actually send a probe to measure the distance. Rather, instead, we use indirect methods, which rely on us knowing precisely the distance between the Earth and the sun, for example. This necessity of measuring smaller distances to compute larger distances is the basis for what is called the "distance ladder" of astronomy.

1.2.1 What is the radius of the Earth?

We know, from a multitude of evidence (e.g. the Earth produces a curved shadow on the moon, and ships travelling on the ocean disappear "over" the horizon), that Earth is approximately a sphere. Eratosthenes, a Greek polymath known for inventing the field of geography (!) and being the chief librarian at the Library of Alexandria, was the first to seek to measure the circumference of the Earth as a sphere, and did so to remarkable accuracy! In particular, legend goes that one day, he heard that, at Aswan, the sun's rays shine directly overhead at the summer solstice. Doing the same measurement at the same time in Alexandria, he measured a shadow of approximately 7.12 degrees.

From here, we can use this surprising difference between the angle of each shadow to determine the radius of the Earth. In particular, using that the distance to the sun is much greater than the radius of the Earth, we can assume that the sunlight at Aswan and Alexandria are approximately parallel to one another. This situation is illustrated in Figure 1.1.

Using this, we have similar triangles, so the relative angle should be equal to the angle of the sector formed by these two cities on a great circle on the sphere of the Earth (see if you can justify this for yourself). Since the distance between the two cities is approximately 500 miles, having been measured directly, Eratosthenes then estimated the circumference of the Earth at $500 \cdot \frac{360}{7.12} \approx 25280.9$ miles. Converting to kilometers and solving for the radius, we get an estimate of the Earth's radius of 6475 kilometers. This is less than 1.6% off the true radius, 6371 kilometers, measured by near-Earth satellites, which is quite a feat!

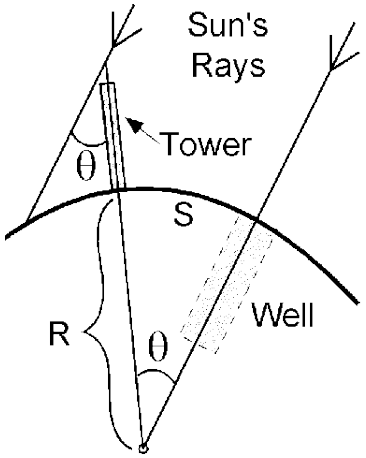

Figure 1.1: Diagram illustrating Eratosthenes's geometric setup for calculating the radius of the Earth, to first order approximation `http://www.tikalon.com/blog/blog.php?article=2011/Eratosthenes`.

1.2.2 Distance to the moon and sun

Next, and possibly most importantly, we calculate the distance from the Earth to the sun. This distance forms the basis for all other calculations done in this lecture, and is so important that it names its own distance unit: in particular, the (mean) distance from the Earth to the sun is known as the **astronomical unit**. To approximate exactly how long this is, an ancient Greek astronomer, Aristarchus, used the moon as a stepping stone. In particular, first, during a lunar eclipse, he first measured that Earth's shadow is approximately two Earth radii is diameter, that the eclipse lasts approximately 3 hours at maximum, and the moon's orbit around the Earth is 1 month. From there, notice that we can express the moon's velocity in two ways: first, using the size of the shadow, as $v = \frac{2R_{earth}}{3 \text{ hours}}$. Second, assuming an approximately circular orbit, we get that

$$v = \frac{2\pi d_{earth-moon}}{730 \text{ hours}}.$$

Solving for $d_{earth-moon}$, we get that it is approximately $77R_{earth}$.

From here, Aristarchus noted that, at a half-moon, the angle between the sun and the moon was approximately $\theta = 87$ degrees (using modern day techniques, this is wildly inaccurate). From there, he noted that, by basic trigonometry, we should get that $d_{earth-sun} = d_{earth-moon} \cdot \sec(\theta)$. Using his value of θ, we get that the distance from the earth to the sun is $9.37 \cdot 10^9$ meters, which is quite a bit off from the now accepted value of $1.496 \cdot 10^{11}$ meters. However, in subsequent measurements, θ has been corrected to a much more accurate 89 degrees, 51 minutes, which gives an estimate of $1.87 \cdot 10^{11}$ meters, much closer to the actual result.

1.2.3 Parallax

Now that we know the distance to (relevant) bodies in our solar system (we can in fact use this earth-sun distance, or more accurate radar techniques, to find the distance to all planets), we can now find the distance to stars in the neighborhood of the sun, using a technique called *parallax*. The spirit of this technique is as follows: consider a nearby star S, and observe it from Earth with respect to "background" stars, which are stars that are very far away and thus remain fixed over the orbit of the Earth (such stars can be empirically picked). As shown in Figure 1.2, by measuring the parallax angle p, in radians, between two observations of

this star at opposite points in the Earth's orbit, and using that the distance between the Earth and the sun is 1 AU, we get that the distance to a nearby star is

$$d = \frac{1 \text{ AU}}{\tan(p)}.$$

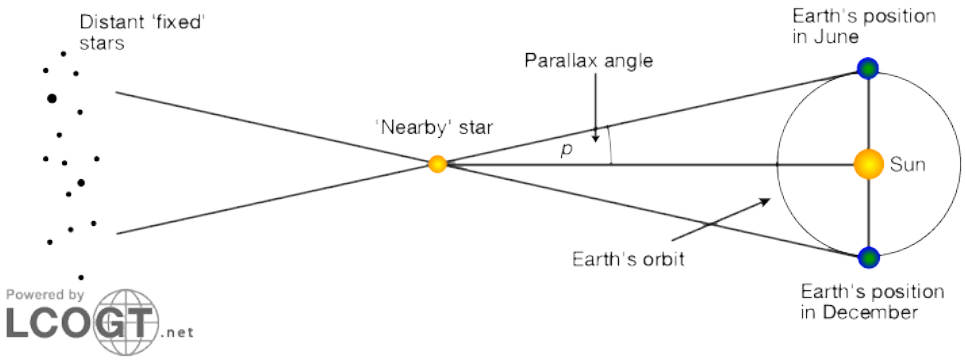

Figure 1.2: Diagram illustrating the parallax technique, taken from `https://lco.global/spacebook/parallax-and-distance-measurement/`.

Now, normally, a "nearby" star is in fact quite far, in which case p is very small. In this case, we can use the **small-angle approximation**, which gives that $\tan(\theta) \approx \theta$ when θ is close to 0, to get that $d \approx \frac{1 \text{ AU}}{p}$.

Often, we will measure p in arcseconds, rather than radians. In this case, using that 1 radian is approximately $206,265$ arseconds and letting p_{as} be the parallax angle in arcseconds, we get that $d \approx \frac{206{,}265 \text{ AU}}{p_{as}}$.

This naturally motivates a new unit of distance, the **parsec** (abbreviated **pc**), which is exactly equal to $206,265$ AU, which gives us, finally, the normal form of the parallax formula,

$$d = \frac{1 \text{ pc}}{p_{as}}.$$

By the nature of the technique, more and more distant stars require a more and more sensitive measure of the parallax angle. To this end, the European Space Agency (ESA) launched the Hipparcos Space Astrometry Mission from 1989-1993, with the goal of accurately measuring parallax angles for a small number of stars up to one-thousandth of an arcsecond, and for millions of stars up to about one-hundredth of an arcsecond. However, for all of the information that this mission provided, it only surveyed the neighborhood of the sun; a quick calculation gives that the farthest star measured was at 1000 pc, whereas the distance from the sun to Sagittarius A, the supermassive black hole at the center of the Milky Way, is approximately 7860 pc (and this is just in our galaxy!). Thus, though this technique is powerful, and the ongoing Gaia mission promises to extend our parallax radius even further, further methods will be needed to measure distances to other galaxies, other galaxy clusters, and to the edge of our universe.

1.2.4 Standard Candles

In this section, we will discuss **standard candles**, one such technique that employs special stellar objects whose intrinsic brightness is known.

1.2.4.1 Absolute Magnitude, Apparent Magnitude, and the Distance Modulus

Before we discuss examples of standard candles, we first need to understand *why* they are useful. To explain this, we first define the **luminosity** of a star (or any astronomical object, for that matter) as the amount of energy emitted per unit time. This is an *intrinsic* property of the star; that is, it does not depend on the (stationary) reference frame from which we observe it, and in particular does not depend on the *distance* of the observer. We next define the **flux** of a star at a given point to be the energy per unit area per unit time. Assuming, as we usually do, that stars emit energy isotropically (there are a few complexities here, but nothing that concerns us too much at this moment), the "luminosity" of a star at a given distance d is spread equally over the entire surface area of a sphere of radius d, and thus we get that

$$F = \frac{L}{4\pi d^2}.$$

At this point, we invoke a logarithmic change in notation to the form of **astronomical magnitude**. The history of this scale is interesting but irrelevant at this moment, similar to the decibel scale, so suffice to say that it was chosen such that a five unit change in magnitude corresponded to a hundredfold change in brightness. In particular, the **absolute magnitude** of a star is defined as

$$M = -\sqrt[5]{100}\log_{10}(L/L_0),$$

where L_0 is a reference luminosity, and the **apparent magnitude** of a star is defined as

$$m = -\sqrt[5]{100}\log_{10}(F/F_0),$$

where F_0 is the reference flux of a star at luminosity L_0 and distance $d_0 = 10$ parsecs. From here, note that we can relate these two metrics using our relationship between flux and luminosity, getting that

$$m = -\sqrt[5]{100}\log_{10}(F/F_0) = -\sqrt[5]{100}\log_{10}(L/L_0) + 2\sqrt[5]{100}\log_{10}(d/d_0) = M + 2\sqrt[5]{100}\log_{10}(d/d_0).$$

The final term in this expression is known as the **distance modulus**; we see immediately that, given the apparent and absolute magnitudes, we can solve for the distance to the object in question. Given a telescope, estimating the (mean) apparent magnitude is not a difficult task. However, to use this technique, we need some stellar objects for which the absolute magnitude is easily found...

1.2.4.2 Cepheid variables, Henrietta Leavitt, and Andromeda

Enter Cepheid variable! These are "variable stars," so named because they vary in brightness over time, where the variation in brightness is periodic and can be directly traced to expansion and contraction of the star (caused by the so-called κ-mechanism, which I can provide further detail about to anyone interested). The first such star, Delta Cephei, was discovered in 1784, but these stars were not well understood until much later, in the early 1900's. In particular, in 1908, Henrietta Leavitt discovered that the pulsation period and intrinsic luminosity of Cepheid variables (note, importantly, that she was able to measure this intrinsic luminosity, in large part, using the distance modulus equation for Cepheids with a known distance, found from parallax: the distance ladder strikes again!) had a very strong (polynomial) empirical relation between them, encapsulated in the below equation for the absolute magnitude of a cepheid variable:

$$M = -2.76\log(P-1) - 4.16,$$

where P is the pulsation period in days. This was a huge discovery, and seemed to hold for all Cepheid variables discovered. Using this, then, astronomers were able to compute the distance to a number of astronomical objects by identifying Cepheid variables in them, observing their pulsation period, and extracting

the absolute magnitude from that. From that and the mean apparent magnitude measured, the distance modulus equation directly yields an estimate of the distance to these objects!

This empirical relation was one of the main pieces evidence that ended one the "Great Debate" in astrophysics at the time; whether there existed other galaxies outside of the Milky Way. The main candidate at the time, the Andromeda Nebula, was argued by many to be inside the Milky Way, as calculations had yielded that, for it to be a separate galaxy, its distance would have to be on the order of millions of light years, which most astronomers at the time could not accept as possible. However, by using Cepheid variables that could provably be shown to exist in the nebula, Edwin Hubble, who later proved one of the most important relations in extragalactic astronomy, Hubble's Law, showed that this distance was fact valid for Andromeda, thereby giving us our first evidence of another galaxy!

In subsequent years, estimates using Cepheid variables have been refined using various techniques. In particular, a few systematic errors in initial measurements of variables were corrected, and it was also discovered by Walter Baade that there exists two distinct populations of Cepheids; type I (or classical) Cepheids, which follow the period-luminosity relation given above, and type II Cepheids, which are fainter and follow a slightly translated and rotated period-luminosity relation. The diagram in Figure **??** illustrates the empirical data behind this relation, as well as the data for yet another standard candle, RR Lyrae, which are older, fainter stellar objects of shorter average period.

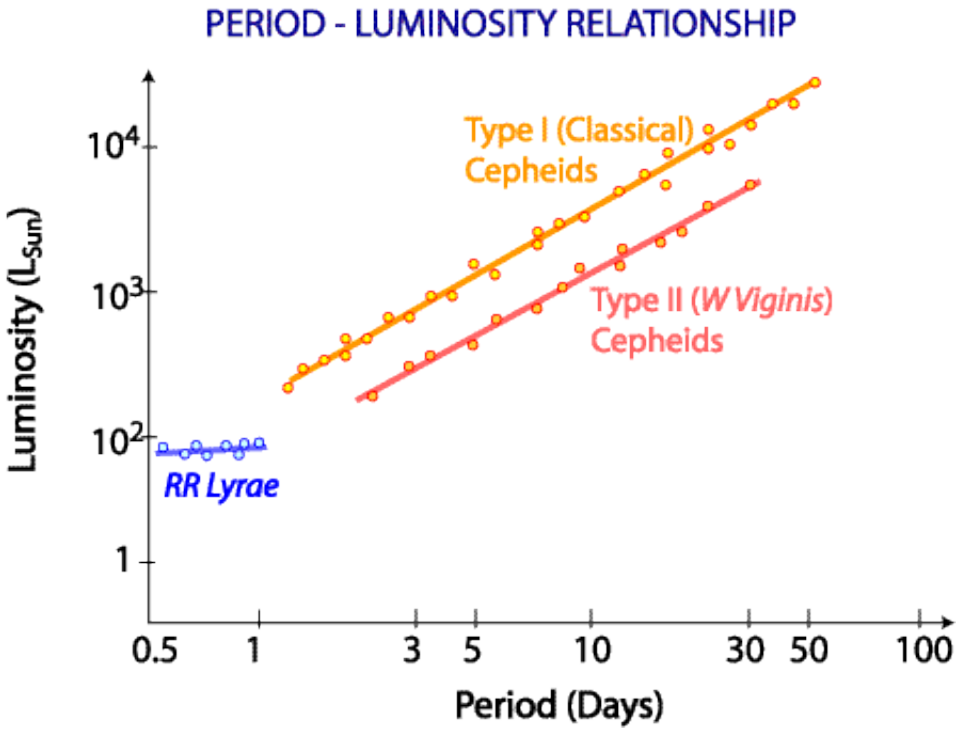

Figure 1.3: Diagram illustrating the empirical linear relation for Cepheid variables, taken from `http://www.atnf.csiro.au/outreach/education/senior/astrophysics/variable_cepheids.html`.

1.2.4.3 Type 1a supernovae

Another stellar object of a known brightness is the supernova, which occurs when a star explodes and can temporarily outshine even a galaxy! In particular, the supernova in question involves a white dwarf, the

bright core of a former star of initial mass $M \lesssim 6M_\odot$. We will discuss stellar evolution in more detail later, but in essence, this remnant was formed after the original star exhausted all of its hydrogen fuel, expelled its outer layers to become a red giant, which eventually faded and left just the hot, carbon-oxygen core behind. White dwarves, in particular, are incredibly dense stars, with just one teaspoon of matter weighing almost five tons; because of this extreme density, the internal forces are not balanced by the normal Coulomb and gravitational forces, but rather by what is known as "electron degeneracy," which we will discuss at great length in a further lecture. Suffice to say that this pressure enforces a strict upper limit on the mass of a white dwarf, the **Chandrasekhar limit**, which is approximately $1.4M_\odot$. When a white dwarf is in a binary system (i.e. it is orbiting around another star), its incredibly high density gives it a strong gravitational pull, and thus, in the right conditions, it accretes matter from the other star; when it exceeds the Chandrasekhar limit, it explodes, thus creating a type 1a supernova.

Because all type 1a supernova are produced by an explosion at approximately the same mass, they have a astonishingly fixed peak absolute magnitude, which is approximately -19.5. From here, then, we have a standard candle; by using the distance-modulus relation and the apparent brightness of a type 1a supernova, we can back out the distance to an object!

Note, as usual, that there is more complexity to the situation. As illustrated in Figure 1.4, there are many different shapes for light curves for type 1a supernova over time, and in fact each of these correspond to different peak brightnesses. In particular, the decay of the light curve is governed by the radioactive of a particular isotope of cobalt, and it turns out that, by studying this, astronomers have derived the **luminosity decay rate relation**, which is an empirical formula relating the rate at which a type 1a supernova's light curve decays over the first 15 day period after the explosion, and the peak brightness of the supernova. Using this correction, then, we can use type 1a supernova as a standard candle, and thereby explore our galaxy even further!

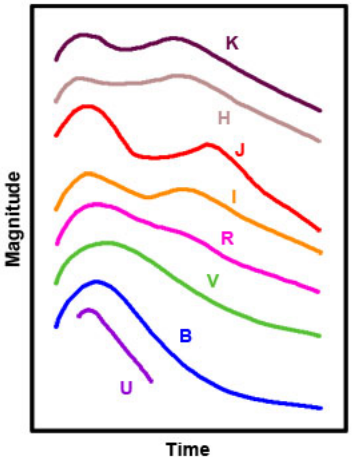

Figure 1.4: Diagram illustrating different light curves for type 1a supernova `http://astronomy.swin.edu.au/cosmos/T/Type+Ia+Supernova+Light+Curves`.

1.2.5 Assorted Other Techniques

In this section, we briefly discuss additional techniques that allow astronomers to measure distances on the scale of galaxies, galaxy clusters, and even the universe. These most prominently include the Tully-Fischer relation, the Sonyaev-Zeldovich effect, and Hubble's Law Because we did not cover these in much detail in class, we discuss the Tully-Fischer relation as an example, and leave the rest to your interest; if you want

more detail about them, feel free to reach out to us for literature recommendations.

1.2.5.1 Tully Fischer Relation

First, the Tully-Fischer relation is an empirical relation between the rotational velocity of a spiral galaxy and its intrinsic luminosity, first discovered by Richard Tully and James Fischer. Given such an empirical relationship, note that we can measure the (approximate) rotational velocity by measuring the width of spectral line. It is outside the scope of this lecture to explain the direct relationship here, but here is an intuitive explanation. In a spiral galaxy, since it is rotating, some of the stars are moving away from us and some towards us, so the spectral lines from the former are redshifted and the latter blueshifted, thus causing a widening of the corresponding spectral line. The magnitude of this widening is linearly related to the rotational velocity, and thus can be empirically measured, most commonly using the 21 cm hydrogen line. From here, it is clear that the distance modulus relation directly yields an estimate of the distance, so it suffices to describe this relationship.

First, we give what has been empirically found. The data in Figure 1.5, to first approximation, shows a linear relationship between the magnitude of a spiral galaxy and its rotational velocity, and we can see that, given two clusters at different distances, the relationship appears to be identical, except translated upwards (thus corresponding to the distance difference). This hints strongly at an (approximate) relationship of the form

$$L \propto V_{rot}^{\alpha},$$

where α is the slope of the relationship in Figure 1.5.

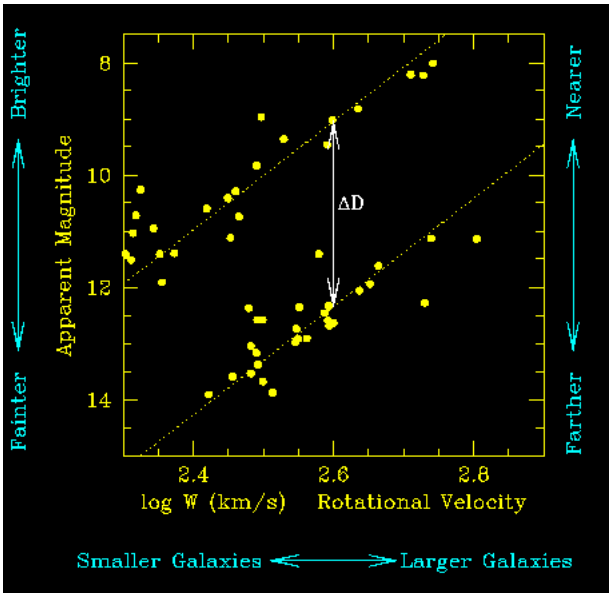

Figure 1.5: Empirical data relating rotational velocity to absolute magnitude for two star clusters, Abell 1367 and the Fornax cluster, taken from `https://www.noao.edu/staff/shoko/tf.html`.

Aggregating data from many studies, it has been shown that α is consistently somewhere between 3.5 and 4, depending on the specific characteristics of the galaxy, as well as its neighborhood. This is a rather astonishing result, and the mechanics behind it are not very well understood. Nonetheless, we can do a "back-of-the-envelope" approximation to justify why $\alpha = 4$ might be a reasonable result.

In particular, for a spiral galaxy, the acceleration of a test mass at distance R is identically described by

$$\frac{GM(R)}{R^2} = \frac{V_{rot}^2}{R}.$$

If we make the approximation that $\frac{L}{M} = C$ is constant for spiral galaxies, which has some empirical basis, we can replace M with L and get that

$$L = C_1 V_{rot}^2 R,$$

where $C_1 = \frac{C}{G}$. Now, making another reasonable assumption that spiral galaxies have a relatively constant surface brightness over their radius (this may not be exactly accurate near the edges, but since most mass is concentrated on the interior, it is still reasonable), we have that $L = C_2 R^2$ for some constant C_2. Equating these two, we get that $C_1 V_{rot}^2 R = C_2 R^2 \Rightarrow R = C_3 V_{rot}^2$, for $C_3 = \frac{C_1}{C_2}$. Plugging this into our expression above, we get that

$$L = C_1 C_3 V_{rot}^4 \Rightarrow L \propto V_{rot}^4,$$

as desired.

1.3 Conclusion

There are myriad techniques for computing distance in astronomy, each on a firm and fascinating physical basis. What is important is not for you to know all of them, but to know the concept: in astrophysics, distance is not direct, but indirect, each calculation resting firmly on the distance "rung" below it.

1.4 Quantum Mechanics: Experiments

Quantum physics was borne out of experimental results that could not be explained by classical theory. As it turns out, the world not only behaves very differently on extremely large scales – it also behaves weirdly on extremely small scales. The experiments described here laid the foundations for quantum theory, which is used to describe phenomena on the small scale.

Note: Some key experimental results are not covered here, because the calculations for those experiments are more complicated. Nonetheless, these experiments played a big part in establishing quantum mechanics as a theory of nature. These experiments include Planck's blackbody radiation, the hydrogen spectrum, the Franck-Hertz experiment, the Stern-Gerlach experiment, and the Davisson-Germer experiment. Later, further experiments uncovered other effects that could only be explained with quantum theory, such as the quantum Hall effect.

1.4.1 Photoelectric Effect

People thought for a long time that if you shine light on a clean sheet of metal for a sufficiently long period of time, the electrons in the metal will absorb enough energy to escape from the metal surface. This phenomenon is known as the photoelectric effect, and the escaped electrons are known as "photoelectrons". The least tightly-bound of these electrons were originally bound with an energy of ϕ, which is the minimum amount of work electrons in the metal must do before they can escape from the metal surface. As a result, ϕ is also known as the work function, and it is fixed for a given type of metal.

However, calculations showed that it would take years for an electron to absorb enough energy to escape from the metal. In practice, photoelectrons were ejected almost instantaneously.

In the late 1800s and the early 1900s, experiments uncovered more inconsistencies with the classical prediction. For example, the energy of emitted electrons increased with the frequency of the incident light. In comparison, classical theory predicts that the energy of the emitted photoelectrons will be proportional to the intensity of the radiation, and independent of the frequency. Moreover, no photoelectrons were emitted if the incident light were below a certain threshold frequency, even though classical theory predicts no such threshold.

In 1905, Albert Einstein came up with a heuristic description of light that could explain the photoelectric effect. He postulated that light came in packets of energy $h\nu$, where $h \approx 6.63 \times 10^{-34}\,\mathrm{J\,s}$ is a proportionality constant known as Planck's constant and ν is the frequency of the incident light. An electron absorbs the full energy of an incident photon, or else it absorbs no energy at all. If an electron absorbs enough energy from a photon, it can escape from the metal surface; otherwise, it dissipates or re-emits that energy.

Now, we can explain each of these observations in turn. Photoelectrons were ejected almost instantaneously because it takes a very short time for electrons to absorb the energy of incident photons. Electrons absorb at most $h\nu$ in energy, so photoelectrons are produced only if $h\nu > \phi$. (Recall that ϕ is the minimum energy needed for electrons to leave the metal surface.) This gives a threshold value for the frequency: if $\nu < \phi/h$, then no photoelectrons will be produced. Moreover, if $\nu > \phi/h$, an electron that absorbs $h\nu$ in energy from a photon loses at least ϕ in energy when it escapes from the metal surface, so emitted photoelectrons have at most $h\nu - \phi$ in energy. This clearly increases with frequency.

The packets of light energy were then known as "light-quanta", and each quantum of light has energy $h\nu$. However, well into the 1920s, physicists still felt uncomfortable with the idea that light could be described as particles. At the time, light had been described as electromagnetic waves for the better part of a century, and the wave description (using Maxwell's equations) had been hugely successful. It would take more definitive evidence to convince physicists that light could be described as particles too.

1.4.2 Compton Effect

In the early 1900s, physicists were also interested in how electromagnetic waves (i.e. light waves) interact with matter. The common understanding then was that charged particles in atoms (e.g. electrons) would absorb the incident radiation and re-emit electromagnetic waves with the same frequency. This is because the incident electromagnetic wave generates an oscillating electric field at the position of the charged particle, causing the particle to oscillate at the same frequency as the electromagnetic wave. Accelerating particles then emit radiation with the same frequency as the frequency of oscillation.

In 1923, however, Arthur Compton observed that some of the X-rays scattered off a graphite sheet had a larger wavelength than the incident X-rays. From the theory of electromagnetic waves, we know that the frequency ν and wavelength λ of light are related by: $\nu\lambda = c$ where $c \approx 3 \times 10^8\,\mathrm{m\,s^{-1}}$ is the speed of light. This means that the frequency of emitted radiation is smaller than the frequency of incident radiation. (The above phenomenon is known as the Compton effect, see Figure 1.6.)

Compton then went on to explain this phenomenon by assuming that light came in discrete quanta, and that each quantum has a particle-like momentum. Compton's original derivation included the effects of relativity, but for simplicity we will mainly use ideas from classical mechanics to derive the wavelength shift of scattered X-rays. The one key fact from relativity that we require is the expression for the energy E of a particle:

$$E^2 = m^2c^4 + p^2c^2$$

where m is the mass of the particle, p is the momentum of the particle and c is the speed of light.

The energy of the X-rays is much larger than the ionization energy of the electrons in the material, so it is reasonable to assume that that the material contains free electrons. When X-rays are scattered off the material, particle-like quanta of light collide and exchange momentum with the electrons. We know today that free electrons in a material have a very low average velocity, so we can assume that the electrons were originally stationary. The collision causes the light quanta to impart some of its momentum to the electron.

So far, we have not yet discussed the origins of the light quantum's momentum. We know from the energy relation above that massless particles ($m = 0$) like light quanta obey the relation $E = pc$. From the photoelectric effect, we also know that the energy of a quantum of light is $E = h\nu = hc/\lambda$. Therefore, we get $p = h/\lambda$, which is an expression for the light quantum's momentum.

Momentum must be conserved in any closed system, so we know that the momentum of the incident quantum of light must equal the sum of the momenta of the scattered quantum and the electron. (Recall that the electron is initially stationary.) Energy must also be conserved in this non-dissipative system. Let the scattering angle of the light be θ, the incident wavelength be λ_0, the scattered wavelength be λ_θ, and the final speed of the electron be v. Also let the mass of the electron be m, the momentum of the electron be p, and the scattering angle of the electron (i.e. the angle of the electron relative to the direction of travel of the incident light) be φ. Then, conservation of momentum (in the parallel and perpendicular directions) and conservation of energy give

$$\begin{aligned}
\frac{h}{\lambda_0} &= \frac{h}{\lambda_\theta}\cos\theta + p\cos\varphi \\
0 &= \frac{h}{\lambda_\theta}\sin\theta - p\sin\varphi \\
mc^2 + \frac{hc}{\lambda_0} &= \frac{hc}{\lambda_\theta} + \sqrt{m^2c^4 + p^2c^2}.
\end{aligned}$$

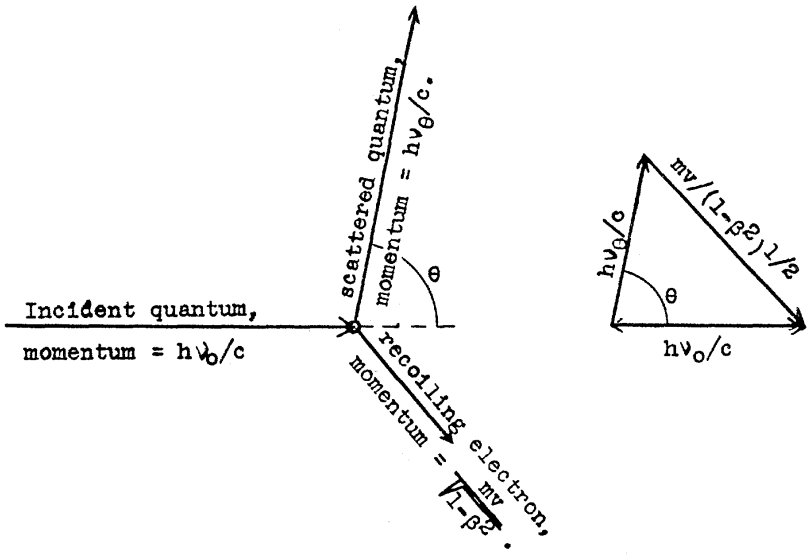

Figure 1.6: Diagram of the collision between a quantum of light energy and an electron. Here, the momentum of the electron is written as $mv/\sqrt{1-\beta^2}$ due to relativistic effects. The incident light has frequency ν_0 and the scattered light has frequency ν_θ. From the main text, the momentum of a light quantum is $p = h/\lambda = h\nu/c$, where we use the relationship $\nu\lambda = c$. Image taken from Compton's original 1923 paper in *Physical Review*.

Rearranging the momentum equations and using $\cos^2\varphi + \sin^2\varphi = 1$,

$$(p\cos\varphi)^2 + (p\sin\varphi)^2 = \left(\frac{h}{\lambda_0} - \frac{h}{\lambda_\theta}\cos\theta\right)^2 + \left(\frac{h}{\lambda_\theta}\sin\theta\right)^2$$
$$\implies p^2 = \left(\frac{h}{\lambda_0}\right)^2 + \left(\frac{h}{\lambda_\theta}\right)^2 - 2\left(\frac{h}{\lambda_0}\right)\left(\frac{h}{\lambda_\theta}\right)\cos\theta.$$

Substituting the energy equation,

$$m^2c^2 + 2mhc\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_\theta}\right) + h^2\left(\frac{1}{\lambda_0^2} - \frac{2}{\lambda_0\lambda_\theta} + \frac{1}{\lambda_\theta}\right) = m^2c^2 + h^2\left(\frac{1}{\lambda_0^2} - \frac{2\cos\theta}{\lambda_0\lambda_\theta} + \frac{1}{\lambda_\theta}\right)$$

Now we cancel, multiply throughout by $\lambda_0\lambda_\theta/2mhc$ and factorize:

$$\lambda_\theta - \lambda_0 = \frac{h}{mc}\left(1 - \cos\theta\right).$$

This is Compton's equation for the wavelength shift of scattered X-rays, and numerous subsequent experiments supported Compton's model of scattering.

Interestingly, Einstein's expression for the energy of a quantum of light agrees with Compton's idea of a particle-like quantum. All of these equations involve the constant h, which turned out to not just be a free parameter, but a fundamental constant of nature.

The experiment of X-ray scattering lent credence to the idea that light needs to be thought of as not just a wave, but also as a particle. Today, we call the quanta of light photons, each with energy $h\nu$ and momentum $h\nu/c$. However, there was more to this "wave-particle duality" than just electromagnetic waves, as we will see in the double-slit experiment.

1.4.3 Double-Slit Experiment

In the classical double-slit experiment (Figure 1.7), coherent light that is shone on two narrow slits that are close together form an interference pattern. This is due to the phenomenon of interference: waves can "overlap" in space and form periodic patterns. (Figure 1.8) Interference is a defining feature of waves (which, as we must not forget, is a valid description of light), and it was already well-studied by the 1900s.

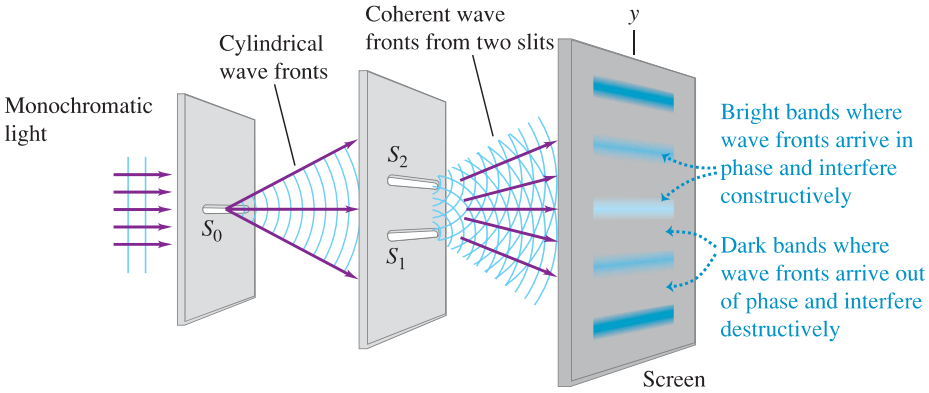

Figure 1.7: The setup of the classical double-slit experiment. Image taken from Young and Freedman's *University Physics* (13ed).

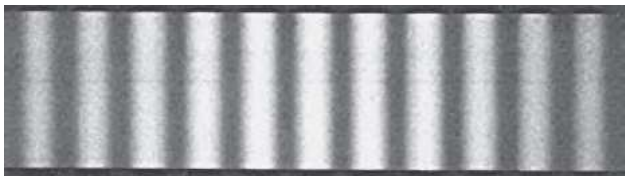

Figure 1.8: An image of the double-slit interference pattern. The light pattern is what would be seen on the screen. Image taken from Young and Freedman's *University Physics* (13ed).

People also used to think that electrons were small particles, and that they should behave like solid objects passing through two openings. Therefore, if we passed an electron beam through a double slit, we should get a bimodal distribution of electrons on the screen, where each mode was centered around the corresponding slit.

However, people soon discovered that electrons exhibited interference patterns too. In fact, if one electron was sent through the double slit at a time, we would build up a distribution on the screen that resembled the double-slit interference pattern. (Figure 1.9) In other words, electrons, which we know to be particles, exhibit distinctly wave-like behavior.

The fact that electrons could form interference patterns indicates that individual electrons also behave as waves, each with a characteristic wavelength. In practice, the wavelength can be found by measuring the distance between interference fringes. It turns out that the wavelength is given by the same relationship as that we saw for photons in the photoelectric and Compton effects:

$$\lambda = h/p$$

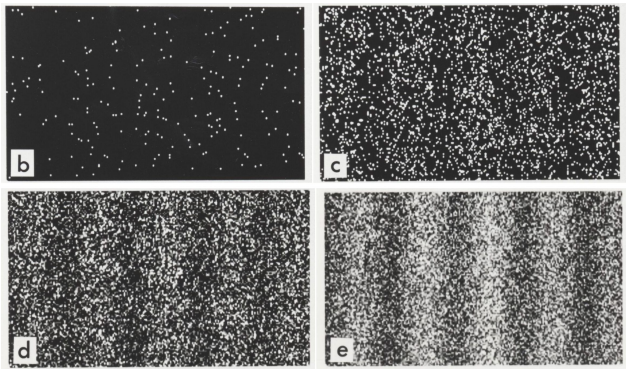

Figure 1.9: The result of a double-slit experiment with electrons. Single electrons were passed through a double slit, and each spot on the screen represents the incidence of one electron. As more electrons pass through the double slit (b-e), the distribution of electrons on the screen looks like the interference pattern in Figure 1.8. Experiment conducted by Akira Tonomura of Hitachi Research.

where λ is the wavelength of the electron, p is its momentum, and h is the Planck constant. This discovery further supported the idea that momentum-bearing particles can behave like waves, just like how electromagnetic waves can behave like momentum-bearing particles.

However, this idea of a "wavelength" does not explain the behavior of electrons satisfactorily. We know that electrons behave as particles most of the time: they have mass; they come in discrete packets; and they can be scattered off obstacles in distinctly un-wave-like ways. There seems to be no way to tell when electrons will act like particles and when they will instead behave like waves. This is where quantum mechanics comes in – it provides us with a way to describe particles as waves, while explaining their particle-like behavior.

Sidenote 1: It is a common misconception that electrons "interfere" with each other. However, as single electrons also produce this effect, that cannot be the case. It is also said that electrons "interfere" with themselves, but the truth of this statement cannot be determined without being more precise about what "interfering" with oneself means. The electrons do not pass through both slits at once, but their behavior is definitely influenced by the fact that there are multiple slits.

Sidenote 2: To be more chronologically accurate, the wave-like behavior of electrons and other particles in general was first hypothesized by Louis de Broglie in the form of "wave-particle duality". This came as a consequence of other experiments in quantum theory. The wave-like behavior of electrons was first experimentally verified by the Davisson-Germer experiment, which showed that electrons scattered off a metallic crystal exhibited behavior consistent with diffraction (another wave-like behavior). For simplicity, the phenomenon of interference is discussed here instead.

1.5 Quantum Mechanics: States in Quantum Mechanics

1.5.1 States and Wavefunctions

In its simplest form, quantum mechanics seeks to describe the wave-like behavior of particles. (The reverse problem, describing the particle-like behavior of waves, is the subject of quantum field theory.)

In classical mechanics, we can fully describe the state of a particle with its position and momentum. Using other known parameters such as the mass of the particle and the forces it experiences, we can fully predict the trajectory of the particle using Newton's second law.

In contrast, we can only describe a wave using its displacement at every point in space and time. There are some simplifications we can make for special cases: if we know that the wave is sinusoidal across space (x) and time (t), for example, we can express the displacement $y(x,t)$ as $y = a\cos(kx - \omega t) + b\sin(kx - \omega t)$ where k and ω are constants. Note that the peak of the wave has a constant value of $kx - \omega t$, so over a time period Δt, the peak moves by a length $\Delta x = \omega \Delta t/k$. In other words, the speed of this wave is ω/k.

In quantum mechanics, therefore, we must describe the wave nature of particles using a function that varies across space and time. We call this function the wavefunction, and we often write it as $\Psi(x,t)$ if there is one spatial dimension, or $\Psi(\mathbf{r},t)$, $\mathbf{r} = (x,y,z)$, if there are three spatial dimensions. This wavefunction Ψ is a complete description of a quantum mechanical state.

Physicists like to change their coordinate bases depending on the situation, and the functional form of $\Psi(\mathbf{r},t)$ depends on the coordinate system chosen. Since two wavefunctions with different functional forms may, in fact, describe the same state, it is useful to give each state a general, coordinate-independent representation. We write the state with a "ket", and the ket $|\Psi\rangle$ represents the state $\Psi(\mathbf{r},t)$ in the appropriate basis.

As we will see later, we can think of $|\Psi\rangle$ as a complex-valued vector. This allows us to do useful things in quantum mechanics without having to deal with complicated wavefunctions.

1.5.2 Review of Complex Numbers

Wavefunctions are, in general, complex-valued functions. Unlike the displacement of a physical wave $y(x,t)$, the value of a wavefunction at some point in space and time is not something you can observe directly.

A complex number is an extension of the real numbers. Instead of having just one unit value (1), we now have two unit values: the real unit (still 1), and the imaginary unit (denoted by i). The imaginary unit, in particular, has the property that $i^2 = -1$. We say that a complex number $z = a + ib$ has real part $\mathrm{Re}(z) = a \in \mathbb{R}$ and imaginary part $\mathrm{Im}(z) = b \in \mathbb{R}$, and all usual algebraic properties hold for complex numbers. In particular, we can add complex numbers $[(a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2)]$, subtract complex numbers and multiply complex numbers $[(a_1 + ib_1)(a_2 + ib_2) = a_1 \cdot a_2 + a_1 \cdot ib_2 + ib_1 \cdot a_2 + ib_1 \cdot ib_2 = (a_1a_2 - b_1b_2) + i(a_1b_2 + a_2b_1)]$. (Division is slightly more complicated, but it is still based on the same principles.) We denote the set of all complex numbers as $\mathbb{C}$. (Remember that the set of real numbers is $\mathbb{R}$.)

We can represent complex numbers in 2-dimensional space, with the real part on one axis and the imaginary part on the other. This 2-dimensional space is known as the complex plane, and a plot of this space is known as an Argand diagram. (Figure 1.10)

We say that the complex conjugate of z is $z^* = a - ib$. The conjugate is important because the product of a number and its conjugate is always a nonnegative real number: $z^*z = (a + ib)(a - ib) = a^2 + a \cdot (-ib) + (ib) \cdot a + (ib) \cdot (-ib) = a^2 + b^2 \in \mathbb{R}^+$.

From the Argand diagram, we can identify two other important quantities. The distance of z from the origin

O is r, and the angle of z from the positive real axis (in the counter-clockwise direction) is θ. We say that $|z| \equiv r \in \mathbb{R}^+$ is the magnitude of z, and that $\arg(z) \equiv \theta \in (-\pi, \pi]$ is its argument. By Pythagoras' theorem and trigonometry,

$$|z| = r = \sqrt{a^2 + b^2} \qquad \arg(z) = \theta = \tan^{-1}\left(\frac{b}{a}\right)$$
$$a = r\cos\theta \qquad b = r\sin\theta.$$

Crucially, note that $|z|^2 = z^* z$.

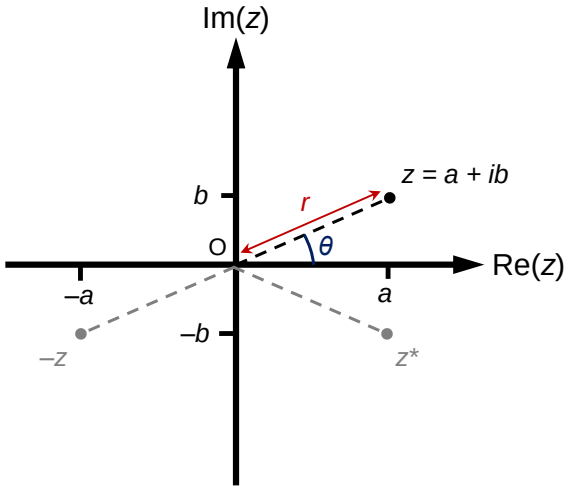

Figure 1.10: The complex plane. $z = a + ib \in \mathbb{C}$ is a complex number with $a, b > 0$. $-z$ (i.e. the negative of z) is its reflection about the origin; z^* (i.e. the complex conjugate of z) is its reflection about the real axis. $r = |z|$ is the magnitude of z, and $\theta = \arg(z)$ is its argument.

It turns out that we need not express z in terms of a and b – we can express z in terms of r and θ too. Leonhard Euler showed that

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

Recall that $z = a + ib = r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta)$. Using the above formula, $z = re^{i\theta}$. This complex exponential divides the complex number z into its magnitude (r) and its phase ($e^{i\theta}$).

The above formula is extremely important, as it allows us to make use of many useful properties of exponentials. For example, it allows us to take any complex number to the n^{th} power easily: $z^n = re^{in\theta}$, whereas doing the same in the $a + ib$ form is extremely tedious. Moreover, it is also simple to take the conjugate of a complex exponential: $z^* = re^{-i\theta}$.

Finally, we arrive at two different methods for dividing two complex numbers. We can use the complex conjugate:

$$\frac{z_0}{z} = \frac{z_0}{a + ib} = \frac{z_0}{a + ib}\frac{a - ib}{a - ib} = \frac{z_0(a - ib)}{a^2 + b^2},$$

or the complex exponential:

$$\frac{z_0}{z} = \frac{z_0}{re^{i\theta}} = \left(\frac{z_0}{r}\right)e^{-i\theta}.$$

(We can also use the complex exponential in multiplication.)

These are the basics necessary for understanding quantum mechanics.

1.5.3 Statistical Interpretation of the Wavefunction

The wavefunction alone is not a measurable quantity, but the wavefunction gives us information about the quantities we can measure.

There are different theories of how we should interpret the wavefunction, but the dominant interpretation is the Copenhagen interpretation. In this theory, quantum mechanics is inherently probabilistic, and the wavefunction gives us probability information about the state it represents. In particular, we can interpret $|\Psi(\mathbf{x}, t)|^2 = \Psi^*\Psi$ as the probability density of finding the particle in position $\mathbf{x}$ at time t. Like a wave, the wavefunction can interfere with itself, producing an interference pattern. In some locations, the wavefunction $\Psi = 0$, so there is zero probability of finding the electron at that position. This gives rise to the dark fringes of the interference pattern.

Probability densities are hard to work with, so we will instead consider a slightly different case where the space of possibilities is finite.

1.5.4 States as Vectors

Remember that wavefunctions are simply representations of a general state. In some instances, we don't need the full wavefunction to make predictions about the quantities we can measure, because there are only finitely many independent values that the quantity can take.

Consider, for instance, the polarization of light. We saw that light is an electromagnetic wave, and we know from Maxwell's equations that the electric field of the wave is perpendicular to its direction of propagation. In 3 dimensions, this means that the electric field can oscillate parallel to a plane that is normal to the direction of propagation. If we have a wave travelling in the x-direction, the electric field oscillates in the yz-plane.

In general, we can decompose any motion in the yz-plane into two separate components: a component in the y-direction, and a component in the z-direction. (Note that the electric field oscillates back and forth around the origin, so the negative and positive y-directions are indistinct.) These two directions are orthogonal, so we can treat them as independent states of electromagnetic waves. Then, any arbitrary electromagnetic wave travelling in the x-direction is a combination of these two states.

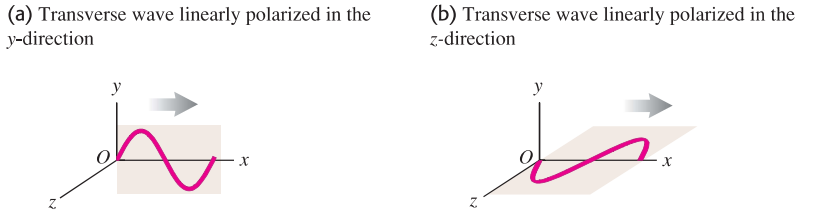

Figure 1.11: The two polarizations of light propagating in the x-direction. These two polarizations are independent of each other, which means that we can decompose any electromagnetic wave into two waves with perpendicular polarizations and the same direction of propagation. Image taken from Young and Freedman's *University Physics* (13ed).

Now, recall that we can think of light as photons as well. This means that photons also have a polarization, and they can be polarized in a direction perpendicular to their direction of travel. If we let their direction of travel be the x-axis, we can think of their polarizations as combinations of two parts: a component parallel to the y-axis, and a component parallel to the z-axis. We often say that a photon that is polarized in the

y-direction has a state of $|0\rangle$, and that a photon that is polarized in the z-direction has a state of $|1\rangle$.

We could have conceivably derived a wavefunction for the photon that describes its propagation, polarization and more. (In fact, the equations that make this possible were first derived by Paul Dirac, but the possibility of writing a photon wavefunction was not recognized until much later for various reasons, not least the development of quantum field theory.) However, such a wavefunction would be extremely complicated. Since we are only concerned with polarization and there are only two independent polarization states, we can describe the photon's state using a much more compact notation.

It may now seem counterintuitive that there are only two "independent" polarization states, because there seems to be infinitely many directions that a photon could be polarized in. However, from Cartesian geometry, we can write any other direction of polarization in terms of the two independent polarization states, $|0\rangle$ and $|1\rangle$. In figure 1.12, we see that all polarization states lie on the unit circle. The polarization state $|0\rangle$ lies along the y-axis, whereas the polarization state $|1\rangle$ lies along the z-axis. A photon that is polarized at an angle θ to the y-axis has state $|\phi\rangle$, and Cartesian geometry suggests that $|\phi\rangle$ is, in fact, a combination of $|0\rangle$ and $|1\rangle$, i.e. $|\phi\rangle = a|0\rangle + b|1\rangle$. In fact, we know that $a = \cos\theta$ and $b = \sin\theta$, so

$$|\phi\rangle = \cos\theta\,|0\rangle + \sin\theta\,|1\rangle.$$

One way to measure polarization is with a polarizer. If we have a polarizer oriented along the y-axis, photons with the polarization state $|0\rangle$ can pass through, whereas photons with the polarization state $|1\rangle$ cannot. Now, the Copenhagen interpretation postulates that:

$$\text{probability that a photon in state } |\phi\rangle \text{ passes through} = a^*a = \cos^2\theta$$
$$\text{probability that a photon in state } |\phi\rangle \text{ is blocked} = b^*b = \sin^2\theta.$$

There is no deeper reason behind why this is true. The above equations are effectively "guesses" of the Copenhagen interpretation, but these are very good guesses – its predictions agree with most, if not all, experiments conducted to date.

We see that writing the state $|\phi\rangle$ in terms of $|0\rangle$ and $|1\rangle$ tells us the probability that a photon in state $|\phi\rangle$ behaves as if it were in state $|0\rangle$ or $|1\rangle$. We call $|\phi\rangle$ a superposition state, and $|0\rangle$ & $|1\rangle$ pure states.

Moreover, we know that the sum of all probabilities must add to one. This gives us the condition that $a^*a + b^*b = 1$, which we see to be true for the state $|\phi\rangle$. This is, in fact, a general property of all quantum states: the magnitudes squared of the coefficients must sum to one. There are some caveats (such as the fact that a and b can also be complex numbers) which we will discuss in the future, but the general picture remains the same.

In summary, states in quantum mechanics are probabilistic. Particles behave like waves because wavefunctions exhibit wave behaviors like interference, with the qualification that the intensity distributions of typical waves are now probability distributions of particles across pure states.

In the double-slit experiment, we have no way of telling which slit the electron passed through if we only make measurements at the screen. Therefore, we can think of the electron as having passed through each slit with probability 1/2. An electron that passes through one slit has a wavefunction $\Psi_1(y,t)$ along the screen whose magnitude squared corresponds to the probability density of finding the electron at some position y, whereas an electron that passes through the other slit has another wavefunction $\Psi_2(y,t)$. Remember from above that we can combine pure states (represented by wavefunctions Ψ_1 and Ψ_2) to get a superposition state. If we let the wavefunction of the general electron be $\Psi(y,t) = a\Psi_1(y,t) + b\Psi_2(y,t)$, the probability of finding the electron in the first and second states are $a^*a = 1/2$ and $b^*b = 1/2$ respectively. One simple solution would be for $a = 1/\sqrt{2}$ and $b = 1/\sqrt{2}$, i.e.

$$\Psi(y,t) = \frac{1}{\sqrt{2}}\Psi_1(y,t) + \frac{1}{\sqrt{2}}\Psi_2(y,t).$$

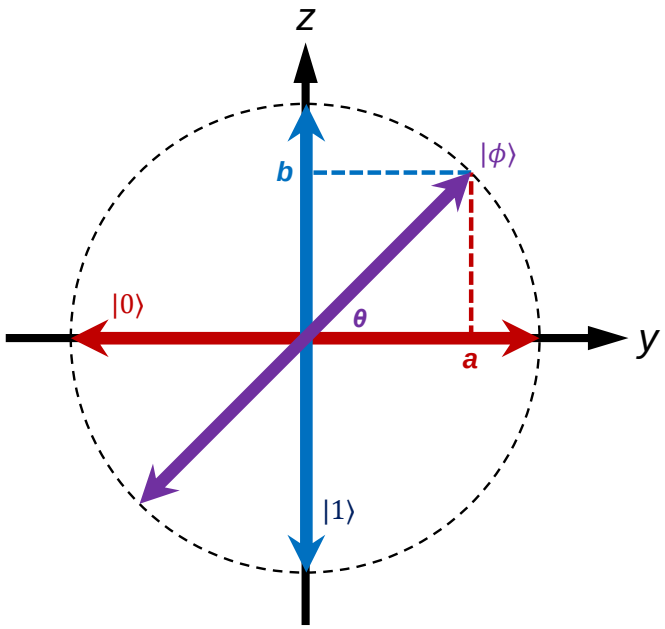

Figure 1.12: A superposition state $|\phi\rangle$. All states lie on the unit circle (dotted).

Now, under some approximations, we can derive simple expressions for the wavefunctions Ψ_1 and Ψ_2. Even though the derivation itself is beyond the scope of this lecture, knowing the wavefunctions Ψ_1 and Ψ_2 will be very useful. Referring to Figure 1.8, let the momentum of the electron be p, the distance from the slits to the screen be H, the width of slits S_1 and S_2 be d, and the distance between the slits S_1 and S_2 be D. If we take the origin ($y = 0$) to be the position along the screen that is equidistant from slits S_1 and S_2, the wavefunctions Ψ_1 and Ψ_2 take the form

$$\Psi_1(y,t) \propto \operatorname{sinc}\left(\frac{\pi p d(y + D/2)}{h\sqrt{(y + D/2)^2 + H^2}}\right) e^{ip\sqrt{(y+D/2)^2+H^2}/h}$$
$$\Psi_2(y,t) \propto \operatorname{sinc}\left(\frac{\pi p d(y - D/2)}{h\sqrt{(y - D/2)^2 + H^2}}\right) e^{ip\sqrt{(y-D/2)^2+H^2}/h}$$

where $\operatorname{sinc} x = \sin x/x$. (The time dependence of the wavefunction is neglected.)

Now we use the far-field approximation, which states that $H \gg y \gg D$:

$$\Psi_1(y,t) \propto \operatorname{sinc}\left(\frac{\pi p d y}{hH}\right) \exp\left[\frac{ipH}{h}\left(1 + \frac{y^2}{2H^2}\right)\right] \exp\left[\frac{ipyD}{2hH}\right]$$
$$\Psi_2(y,t) \propto \operatorname{sinc}\left(\frac{\pi p d y}{hH}\right) \exp\left[\frac{ipH}{h}\left(1 + \frac{y^2}{2H^2}\right)\right] \exp\left[\frac{-ipyD}{2hH}\right].$$

Therefore, using the fact that $e^{i\theta} + e^{-i\theta} = 2\cos\theta$,

$$\Psi(y,t) \propto \operatorname{sinc}\left(\frac{\pi p d y}{hH}\right) \exp\left[\frac{ipH}{h}\left(1 + \frac{y^2}{2H^2}\right)\right] \cos\left[\frac{ipyD}{2hH}\right].$$

The probability distribution of finding the electron at position y is hence

$$|\Psi(y,t)|^2 = \Psi^*\Psi \propto \text{sinc}^2\left(\frac{\pi p d y}{hH}\right)\cos^2\left[\frac{pyD}{2hH}\right]. \tag{1.1}$$

Now we can plot this as an intensity distribution, just like in Figure 1.9. We see that we get a similar interference pattern as before. (Figure 1.13) This demonstrates how wavefunctions can capture the wave-like behavior of particles like electrons.

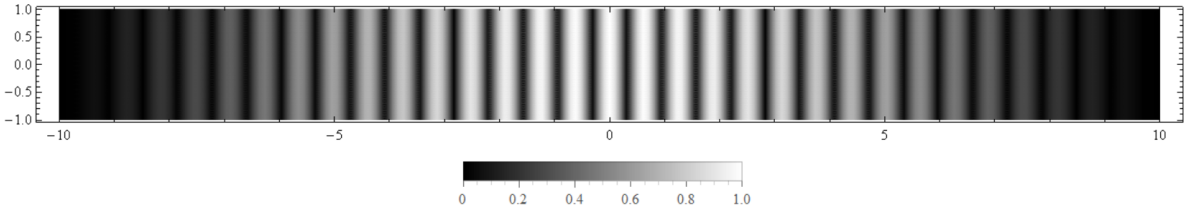

Figure 1.13: Interference pattern predicted by Equation 1.1. White represents a relative intensity of 1, and black represents a relative intensity of 0. The lengths are in arbitrary units. The probability that an electron is incident on the screen at a dark fringe is 0. The color is scaled to the square root of intensity.

In the next lecture, we will discuss more general properties of states, and other interesting things we can do with states without having to resort to wavefunctions like we did today.