

Physics 362K, Fall 2015
Quantum Physics II: Atoms and Molecules
© 2015 D. J. Heinzen

Contents

I.	Principles of Quantum Mechanics	2
A.	Measurement and wave-particle duality	2
1.	Blackbody radiation.....	2
2.	Photoelectric effect	6
3.	Momentum of photons, Compton scattering	10
4.	Wave-particle duality	14
5.	The probability amplitude	19
6.	Observables and uncertainty in quantum mechanics; Heisenberg's microscope	31
B.	Dirac's formulation of quantum mechanics.....	34
1.	Motivation in terms of experimental measurements	34
2.	Linear algebraic formulation for a discrete basis of states	38
3.	Continuous basis sets.....	57
4.	Time evolution of the state vector	64
5.	Multiple Observables.....	67

I. Principles of Quantum Mechanics

This section of the course notes is a review section. It covers much of the same ground as Physics 355, Modern Physics and Thermodynamics, and Physics 373, Quantum Physics I. For this reason, I will expect that you already know most of this material. This section of the course notes will function more as an alternative reference for concepts from previous courses than as a text containing new material. I may refer to particular equations or sub-sections of this section as we cover other topics. It may be helpful for you to study this section as a review, but this is not required.

I won't assume that you're completely familiar with everything in these notes. I've highlighted relevant sections as follows:

Light grey highlight: means you should have seen this topic before, but I won't assume you're completely familiar with it. If it is important to this course, we'll go back over it.

Dark grey highlight: means you probably haven't seen this topic before. If it is important to this course, we'll go over it. We probably won't cover some of these topics at all.

If you do want to study from these notes, it would be best to first make sure you understand the non-highlighted sections. If you find that I'm assuming you have covered some topic in this section, and you haven't, please let me know.

In 1925-1926, Schrödinger formulated his wave equation for electrons, and Heisenberg formulated “matrix mechanics”. Within a year, Heisenberg, Schrödinger, Jordan, Dirac, and others proved that these two theories are equivalent, and combined them into the theory that we now refer to as “quantum mechanics”. Dirac’s further refinement of the theory and notation, as expressed in his well-known textbook *Quantum Mechanics*,¹ has become the standard formulation of quantum theory. This section of notes will review the basics of this theory and its experimental foundation.

A. Measurement and wave-particle duality

1. Blackbody radiation

Blackbody radiation is the glow - electromagnetic radiation - given off by hot objects. This radiation is the same for any two objects that are completely black and at the same temperature T . By the late 1800s, it was already known that the total radiated intensity (power per unit surface area) of blackbody radiation obeys the *Stefan-Boltzmann law*:

$$I = \sigma T^4 \tag{1.1}$$

¹ *Quantum Mechanics*, 4th ed. (Oxford Univ. Press, 1958). The 1st edition was published in 1930.

where $\sigma = 5.67 \times 10^{-8} \frac{\text{W}}{\text{m}^2 \text{K}^4}$ is the *Stephan-Boltzmann constant*.² (1.2)

The Stephan-Boltzmann law is an idealization, since no real surface is perfectly black. The *emissivity* of a surface is the ratio of the actual emitted intensity to that given by the Stephan-Boltzmann law; a perfectly black surface has an emissivity of 1. Most surfaces which appear black to the eye have emissivity in the range 0.8 to 0.95; certain metal oxide surfaces have infrared emissivity greater than 0.99.

The best way to produce a nearly ideal blackbody is to drill a small hole in an oven at temperature T with blackened interior walls. A photon (quantum of EM radiation) that enters the hole may reflect from the walls of the oven once or twice, but it is extremely unlikely to come back out of the hole before being absorbed. Therefore the hole appears nearly perfectly black from the outside (at low temperature). The small size of the hole also means that the radiation field inside the oven is unperturbed by coupling to the lower temperature environment outside the oven. Blackbody radiation at temperature T emerges from the hole.

EM radiation inside an oven contains a wide spread of wavelengths λ . We can characterize the radiation by its *spectral density* $\tilde{u}(\lambda)$, which is its energy per unit volume per unit wavelength interval. In 1893 Wilhelm Wien derived a formula that constrains the form of spectral density $\tilde{u}(\lambda)$ of blackbody radiation. While his result did not provide the correct functional form of $\tilde{u}(\lambda)$, it did predict that $\tilde{u}(\lambda)$ must have a maximum at a wavelength λ_{max} that is inversely proportional to temperature:

$$\lambda_{\text{max}} = \frac{b}{T} \quad (1.3)$$

This is in accord with experience: a bright red object is hotter than a dull red object, and a blue flame is hotter than a red one. This result is known as *Wien's displacement law*. Accurate measurements give the value

$$b = 2.898 \times 10^{-3} \text{ m K} \quad (1.4)$$

In 1898, Ferdinand Kurlbaum carried out accurate measurements of the Stephan-Boltzmann constant, and in 1900, Otto Lummer and Ernst Pringsheim carried out the first accurate measurements of the spectrum of blackbody radiation $\tilde{u}(\lambda)$. As of 1900, no theory correctly explained their measured spectrum.

In 1900, Max Planck published a new theory of blackbody radiation.³ Although no one realized it at the time, this paper constituted the beginnings of quantum mechanics. Planck's paper contains these two critical passages:

² Throughout these notes, I will give modern values for the constants unless otherwise noted.

³ Annalen der Physik, vol. 4, p. 553 ff (1901). An English translation of this paper may be found here: <http://theochem.kuchem.kyoto-u.ac.jp/Ando/planck1901.pdf>

“Moreover, it is necessary to interpret U_N not as a continuous, infinitely divisible quantity, but as a discrete quantity composed of an integral number of finite equal parts. Let us call each such part the energy element ε ; consequently we must set

$$U_N = P\varepsilon$$

where P represents a large integer generally, while the value of ε is yet uncertain.”

Later in the paper, we find:

“If we apply Wien’s displacement law in the latter form to equation (6) for the entropy S , we then find that the energy element ε must be proportional to the frequency ν , thus: $\varepsilon = h\nu$ ”

Here, Planck introduced both quantization of energy and *Planck’s constant* h into physics:

$$h = 6.626 \times 10^{-34} \text{ J s.} \quad (1.5)$$

Planck goes on to derive the energy per unit volume per unit frequency interval $u(\nu)$ contained in a blackbody radiation field in equilibrium: ⁴

$$u_{Pl}(\nu) = \frac{8\pi h \nu^3}{c^3} \left(\frac{1}{e^{h\nu/k_B T} - 1} \right) \quad (1.6)$$

$$\text{where } c = 2.998 \times 10^8 \text{ m/s} = \text{speed of light} \quad (1.7)$$

$$\text{and } k_B = 1.381 \times 10^{-23} \text{ J/K} = \text{Boltzmann's constant.} \quad (1.8)$$

The Planck spectrum of eq. (1.6) is in precise agreement with experiments. Planck further derived the relatively accurate values $h = 6.55 \times 10^{-34} \text{ J s}$ and $k_B = 1.346 \times 10^{-23} \text{ J/K}$ by requiring that his theoretical expressions match the experimental results of Kurlbaum, Lummer, and Pringsheim.

From eq. (1.6), it follows that the Stephan-Boltzmann constant is ⁵

$$\sigma = \frac{2\pi^5 k_B^4}{15h^3 c^2}, \quad (1.9)$$

which agrees with the measured value given in equation (1.2).

Planck’s idea of energy quantization took about 20 years to be fully accepted. Planck’s paper does not provide any meaningful comment on the energy quantization, nor give any physical motivation for it. ⁶

Around 1900, Lord Rayleigh and James Jeans published an independent, classical theory of blackbody radiation. To begin the calculation, we imagine that radiation is confined to a cubic volume

⁴ Here ν is the frequency of the radiation in Hz. Since $\nu = c / \lambda$, $d\nu = -(c / \lambda^2) d\lambda$. It follows that the spectral energy density functions are related by $\tilde{u}(\lambda) = (c / \lambda^2) u(\nu)$.

⁵ Planck did not include this result in his paper, even though it is a relatively simple calculation.

⁶ Historians sometimes state that Planck does not quantize the field itself, but quantizes only oscillators to which the field is coupled. As I read the paper, it’s not clear whether that was his intent.

$V = L^3$ at temperature T , and that the radiation is in thermal equilibrium at this temperature. We further suppose that the electromagnetic field satisfies reflecting boundary conditions on the surface of the cube. It follows that the radiation can only exist in specific normal modes, each of which has a well-defined frequency ν . They calculated that the number of modes per unit frequency interval is

$$g(\nu) = \frac{8\pi V}{c^3} \nu^2 \quad (1.10)$$

Furthermore, the energy contained in a given mode is proportional to the square of its amplitude of oscillation. Therefore we can apply the equipartition theorem of classical thermodynamics, and it follows that each mode must carry an average energy $k_B T$. The energy per unit volume per unit frequency interval is

$$u_{cl}(\nu) = \frac{g(\nu)k_B T}{V} = \frac{8\pi k_B T}{c^3} \nu^2 \quad (1.11)$$

It is immediately apparent that there is a serious problem, because the total energy per unit volume is

$$\frac{U}{V} = \int_0^\infty u_{cl}(\nu) d\nu = \frac{8\pi k_B T}{c^3} \int_0^\infty \nu^2 d\nu = \infty \quad (1.12)$$

This is called the *ultraviolet catastrophe*: the energy per unit frequency interval increases without limit vs. frequency, and therefore the total energy in the field is infinite.

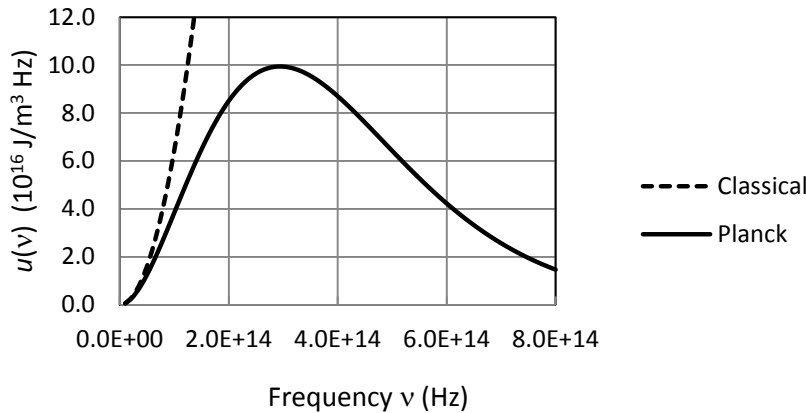


Figure 1.1. Comparison of Planck spectrum (solid line) with the classical spectrum (dashed line) for $T = 5000 \text{ K}$.

In Figure 1.1, I show a comparison of the Planck spectrum (eq. (1.6)) and the Rayleigh-Jeans classical spectrum (eq. (1.11)), for a temperature $T = 5000 \text{ K}$. At this temperature, the Planck spectrum peaks at a frequency $\nu = 5.6 \times 10^{14} \text{ Hz}$, which corresponds to a wavelength $\lambda = c / \nu = 540 \text{ nm}$ in the blue region of the spectrum.⁷ From the figure, you can see that the classical theory works for low

⁷ This wavelength differs from that of eq. (1.3) because that equation gives the wavelength at which $\tilde{u}(\lambda)$ peaks, rather than the wavelength corresponding to the frequency at which $u(\nu)$ peaks.

frequencies, but fails for high frequencies. That is because Planck's energy quantum is small at low frequencies; $h\nu \ll k_B T$. In this regime the quantization effect is not very important because the energy quantum is so small compared to the typical thermal energy of a mode. However, for large frequencies, $h\nu \gg k_B T$. In thermodynamics, the population of a particular state of energy E is proportional to its Boltzmann factor $\exp(-E/kT)$. In Planck's theory, the restriction $E = nh\nu$, with n an integer, causes the Boltzmann factor to become exponentially small when $h\nu \gg k_B T$ and $n \geq 1$. That is, the high frequency modes are "frozen out". By contrast, the energy per mode is $k_B T$ in the classical theory no matter how large the frequency is.

2. Photoelectric effect

The photoelectric effect is the emission of electrons by a surface illuminated by light. It was discovered in 1887 by Heinrich Hertz. In 1902, Philipp Lenard completed a careful study of the effect using the apparatus schematically illustrated in Fig. 1.2(a). The apparatus consisted of two electrodes, an anode and a cathode, contained in a vacuum tube. Lenard observed that a "photocurrent" flows between the two electrodes when the cathode is illuminated by ultraviolet light. He measured the photocurrent I as a function of the potential difference V between the two electrodes. He found that the photocurrent vs. potential difference behaved as shown in Fig. 1.2(b). For sufficiently high potential difference, the photocurrent I saturated at some value I_0 . If the voltage is reduced to low values, the photocurrent I decreases. If the voltage is reversed, he found that there is a negative voltage $-V_0$ at which the photocurrent falls to zero. V_0 was a few volts, and is called the *stopping potential*. Lenard also studied the current as a function of light intensity. He found that if the light intensity is increased, the stopping potential remains the same and the saturated photocurrent I_0 increases in proportion to the light intensity, as shown in Fig. 1.2(b).

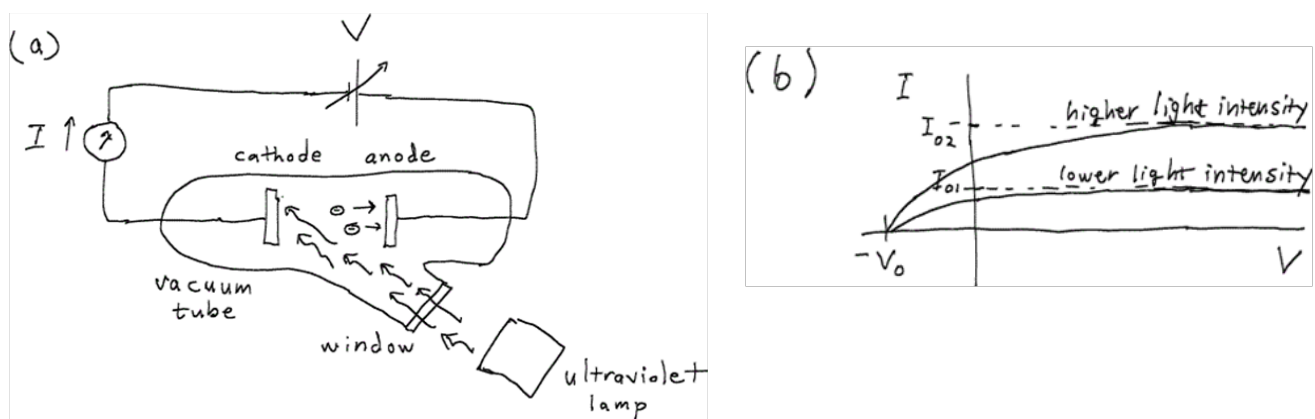


Fig. 1.2. (a) Lenard's apparatus for study of the photoelectric effect. (b) Dependence of the measured photocurrent I on potential difference V .

Part of the explanation of Fig. 1.2(b) is straightforward. There must be a distribution of kinetic energies of the ejected electrons, which ends abruptly at some maximum kinetic energy KE_{\max} . For $eV_0 > KE_{\max}$, none of the emitted electrons has enough kinetic energy to overcome the repulsive electrostatic potential energy eV_0 between the cathode and the anode, where

$$e = 1.602 \times 10^{-19} \text{ C} = \text{magnitude of the electron charge.} \quad (1.13)$$

As the voltage is increased above $-V_0$, some electrons begin to make it to the anode. Eventually, for large enough V all the electrons emitted from the surface make it to the anode. So I_0 must measure the total current of photoelectrons emitted from the surface.

In 1905, Albert Einstein published the paper “On a heuristic point of view about the creation and conversion of light.”⁸ In the first sections of this paper he analyzes blackbody radiation as an equilibrium process of absorption and emission of radiation by a collection of oscillators. He argues that Planck’s quantum hypothesis leads to a consistent theory of blackbody radiation and avoids the problem of the “ultraviolet catastrophe” (although he doesn’t call it that). The most important part of the paper starts with the sentence:

“The usual idea that the energy of light is continuously distributed over the space through which it travels meets with especially great difficulties when one tries to explain photo-electric phenomena, as was shown in the pioneering paper by Mr. Lenard.”

In trying to explain the photoelectric effect with older ideas, one does indeed run into "great difficulties." According to classical theory, each electron in the metal is bound into place by a harmonic restoring force of some resonant frequency. The light wave applies an oscillating force $-eE_0 \cos(2\pi\nu t)$ to each electron in the cathode surface, where E_0 and ν are the amplitude and frequency of the light wave. These electrons undergo a forced, non-resonant oscillation at frequency ν . Most of their energy will either be re-radiated as a reflected electromagnetic wave or absorbed by the metal as heat due to the coupling of the electron oscillators to each other. Electrons can be ejected only if their oscillation amplitude exceeds the limit of the restoring force (which is presumably anharmonic for large amplitudes). There are a number of difficulties with this theory:

- (i) It is hard to explain the fact that the stopping potential does not depend on light intensity. A greater light intensity should lead to more energetic electrons, which should require a greater potential to stop.
- (ii) It is hard to understand the magnitude of the stopping potential (a few Volts). The power intercepted by any given electron is extremely small, due to the tiny “area” of the beam that it intercepts, and that power is quickly dissipated by re-radiation or by coupling to the other electrons. It is basically impossible for any given electron to acquire a few electron volts worth of energy from the optical intensity available from a lamp.

⁸ *Ann. Physik* **17**, 132 (1905).

(iii) It is difficult to understand the linear dependence of the photocurrent on light intensity. According to classical theory, there should be some specific energy of oscillation that is sufficient to overcome the restoring force for the electrons. At low light intensity, all electrons should have energy well below this. When the light intensity is turned up to the point that the energy of a typical electron approaches the limit of the restoring force, there should be a sudden, nonlinear increase in the photocurrent.

Einstein goes on to propose an alternative theory of the photoelectric effect. He supposes that a light beam consists of “energy quanta” which are today called *photons*,⁹ and that the energy of a photon is

$$E = h\nu \quad (1.14)$$

just as Planck had assumed.¹⁰ He also assumes that each photon can be absorbed by a single electron, so that the electron acquires the full quantum of energy $h\nu$. This energy is sufficient to eject the electron from the surface. The emitted electrons have a distribution of energies because they can lose energy on their way out of the surface. So in the experiments, $h\nu$ is the maximum possible energy just prior to leaving the surface. In addition, Einstein understood that there would be an overall potential step for the electrons at the surface, which we denote today as the *work function* Φ . The maximum kinetic energy of the emitted electrons would then be

$$KE_{\max} = eV_0 = h\nu - \Phi \quad (1.15)$$

This is referred to as *Einstein’s photoelectric equation*. As Einstein notes,

“As far as I can see, our ideas are not in contradiction to the properties of the photoelectric action observed by Mr. Lenard. If every energy quantum of the incident light transfers its energy to electrons independently of all other quanta, the velocity distribution of the electrons, that is, the quality of the resulting cathode radiation, will be independent of the intensity of the incident light; on the other hand, *ceteris paribus*, the number of electrons leaving the body should be proportional to the intensity of the incident light.”

Einstein’s theory made a number of untested predictions, including a linear relationship between stopping potential V_0 and optical frequency ν . Robert Millikan decided to test Einstein’s theory. His experiments used line radiation from discharge lamps to provide monochromatic radiation at select wavelengths, and both a sodium and lithium surface. His results were published in 1916,¹¹ and I’ve reproduced the key result of the paper in Fig. 1.3. Einstein’s prediction of a linear relationship between stopping potential and photon frequency is clearly confirmed. In his paper, Millikan ends with a several page discussion, which basically amounts to a summary of the confusion that results from trying to

⁹ The term “photon” was coined in 1926 by the chemist Gilbert Lewis.

¹⁰ Actually Einstein does not use the constant h in his paper, but rather a combination of other constants. However in a section of the paper that discusses Planck’s theory, he essentially shows that this combination of constants is the same as Planck’s h .

¹¹ Phys. Rev. **7**, 355 (1916). This article can be viewed through Physical Review online, from a UT internet connection, at: http://prola.aps.org.ezproxy.lib.utexas.edu/pdf/PR/v7/i3/p355_1

reconcile theory of the photoelectric effect with the known facts of classical electromagnetism.¹² He seems somewhat skeptical of Einstein's theory himself, but concludes the paper with this second-to-last sentence:

"Einstein's photoelectric equation has been subjected to very searching tests, and it appears in every case to predict exactly the observed results."

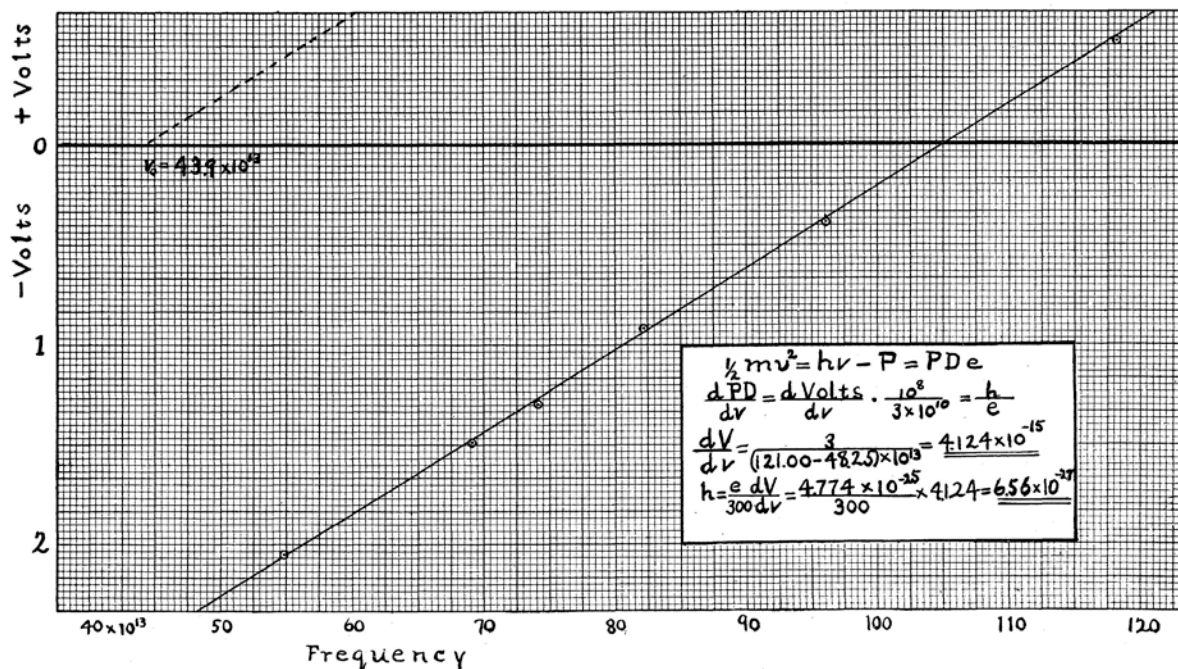


Fig. 1.3. Millikan's data for the stopping potential V_0 of the photoelectric effect as a function of optical frequency ν . The inset shows his calculation of Planck's constant h from his data in cgs units. His result for h deviates from the correct value by only 1%.

The photoelectric effect is used today to make sensitive and flexible photon detectors such as the position-sensitive detector (PSD) shown in Fig. 1.4.¹³ The PSD is enclosed in an evacuated glass envelope. The front piece of glass is coated on the inside with a thin layer of semiconductor which functions as a semi-transparent photocathode. When a photon is absorbed by the photocathode, an electron is emitted due to the photoelectric effect. This electron is accelerated towards a microchannel plate by an applied electric field. The microchannel plate consists of many parallel hollow tubes, each of

¹² Millikan would not have known this, but it turns out to be possible to explain the photoelectric effect with *semi-classical theory*. In this context, "semi-classical" means that the atoms are treated with quantum mechanics but the electromagnetic field is treated classically. For this reason, it is debatable whether the photoelectric effect proves that the energy of the electromagnetic field is quantized.

¹³ A photocathode and microchannel plate combined with a phosphor screen is also called an *image intensifier*. Similar devices, referred to as "intensified CCD cameras" are also available: these couple the output of the microchannel plate into a CCD array. An even more recent technology uses "on-chip electron multiplication" within a CCD array itself to produce a near quantum-limited PSD.

which is coated inside with a semi-conducting material. A voltage is applied lengthwise across the tubes. When the photo-electron strikes the inside of one of the tubes, its energy is high enough to eject several secondary electrons from the surface. Each of these secondary electrons is accelerated further down the tube by the applied field, and again strikes the inside wall. Each of those electron impacts liberates several more electrons, and this process continues until the end of the tube. The result is that each time a photoelectron strikes the inside of one of the tubes, a shower of $10^5 - 10^7$ secondary electrons is emitted from the output end of the tube. These secondary electrons are accelerated into a resistive anode, and the burst of charge is drained off at the four corners of the anode. Because of the electron multiplication inside the microchannel plate, the charge sent to the anode is large enough to be processed by conventional analog-to-digital converter electronics with low noise and uncertainty. The position of the secondary electron hit is computed from the relative charges drained from the four corners of the anode (more charge drains from corners close to the hit), and digitally recorded and/or output to a display device. The device provides the precise time and location of each photon detection event.

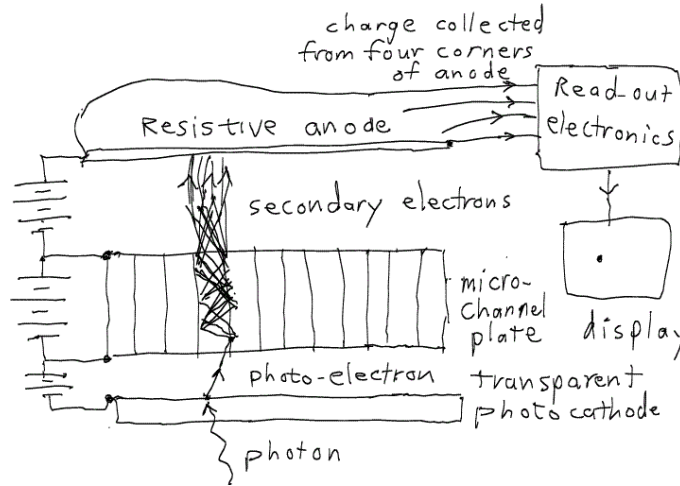


Fig. 1.4. Position-sensitive photon detector (PSD)

3. Momentum of photons, Compton scattering

The *wavenumber* of a wave is $k = 2\pi / \lambda$. According to classical electrodynamics, the wavenumber and frequency of an electromagnetic wave obey the relation

$$\omega = ck. \quad (1.16)$$

According to classical theory, electromagnetic waves carry both energy and momentum. For instance, suppose we have a pulse of radiation which contains a very narrow range of frequencies centered at frequency ω ,¹⁴ and that the total energy contained in this pulse is E . It can be shown that

¹⁴ In this course, we will always follow the convention that ν is the frequency in Hz, and $\omega = 2\pi\nu$ the angular frequency in rad/s. Similarly k is an angular wavenumber in units of rad/m.

according to classical electrodynamics, the momentum contained in this pulse is ¹⁵

$$p = E / c. \quad (1.17)$$

To make contact with quantum physics, let's further suppose that this pulse contains a single quantum of energy, so that

$$E = \hbar\omega, \quad (1.18)$$

$$\text{where } \hbar = \frac{h}{2\pi} = 1.0546 \times 10^{-34} \text{ J s} \quad (1.19)$$

Combining equations (1.16), (1.17), and (1.18), we find that the momentum of one photon should be

$$p = \hbar k = \frac{h}{\lambda} \quad (1.20)$$

Thus, we expect that the momentum of electromagnetic radiation should be quantized in units of $\hbar k$.

Quantized photon momentum was first verified in 1923 by Arthur Compton with the experiment illustrated in Fig. 1.5(a). ¹⁶ X-rays from a tube with a molybdenum target illuminated a sample of graphite, and some of the X-rays scattered from the graphite. X-rays scattered at angle θ relative to the incident rays were selected with lead collimating slits. The spectrum of these angle-selected rays was analyzed with a Bragg spectrometer consisting of a single calcite crystal and an X-ray detector. Since the spacing between the calcite crystal planes was accurately known, Compton was able to deduce the spectrum of the X-rays vs. wavelength from the spectrum of the X-rays vs. Bragg angle shown in Fig. 1.5(b). Over the region of the spectrum studied, ¹⁷ rays incident on the graphite sample consisted of the single K_α line of the molybdenum atom, as shown in the upper spectra of Fig. 1.5(b). (We will discuss X-ray lines later in this course.)

Compton observed that when θ is not zero, the scattered X-ray spectrum has two lines. One has the same wavelength λ_0 as the molybdenum K_α line, and the second has a larger wavelength λ . The splitting $\lambda - \lambda_0$ between these two lines increases with the angle θ . Compton proposed a theory of this splitting which yielded the result

$$\lambda - \lambda_0 = \lambda_c (1 - \cos(\theta)) \quad (1.21)$$

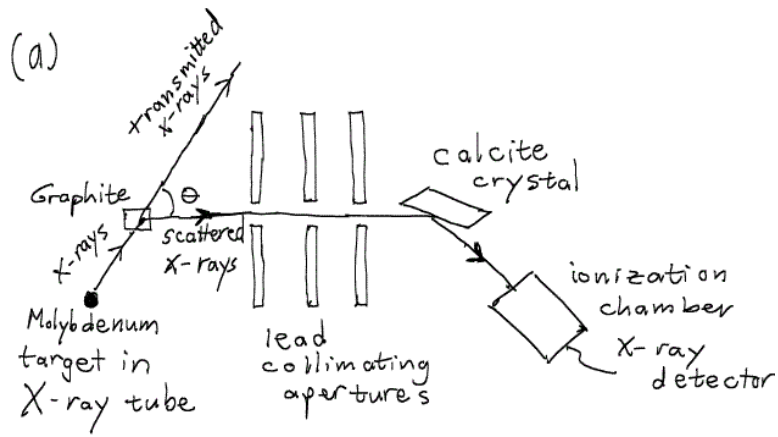
where

$$\lambda_c = \frac{h}{m_e c} = 2.43 \times 10^{-12} \text{ m} \quad (1.22)$$

¹⁵ See, for instance, Griffiths, *Introduction to Electrodynamics*, 3rd ed., section 9.3.2.

¹⁶ Phys. Rev. **21**, 483 (1923). This paper can be found here: <http://www.aip.org/history/gap/PDF/compton.pdf>

¹⁷ I am referring to the data of Fig. 1.5(b), which is the data shown in most textbooks. It is actually from a follow-up paper to the above: Phys. Rev. **22**, 409 (1923).



(b)

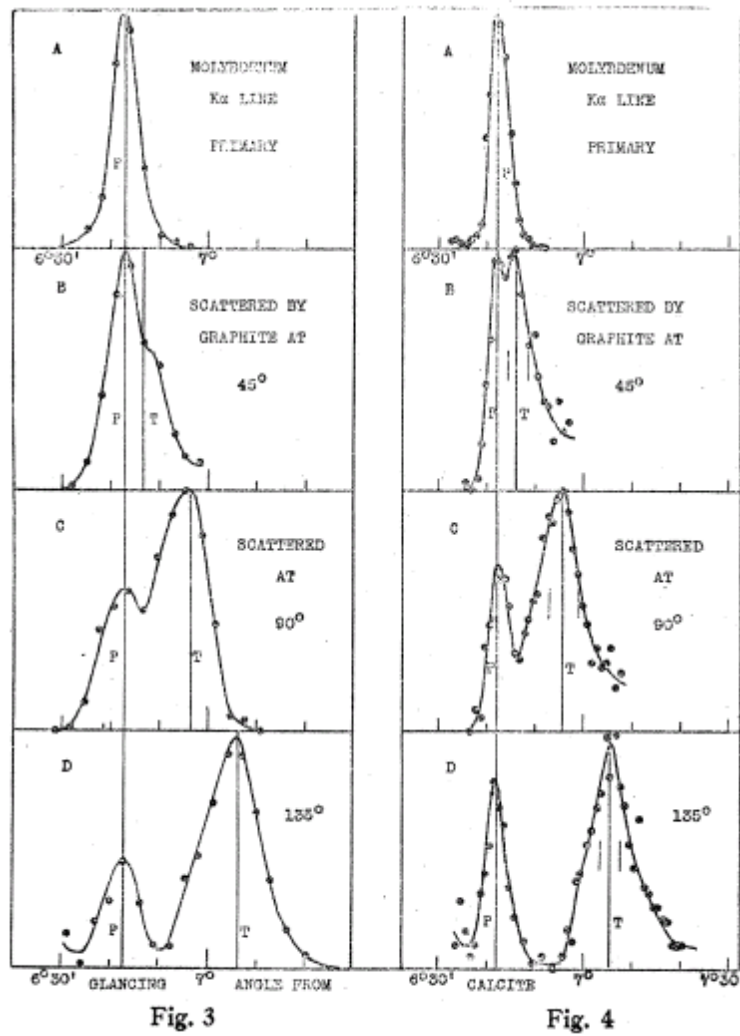


Fig. 1.5. (a) Arrangement of Compton's 1923 X-ray scattering experiment.
(b) Spectrum of the scattered X-rays for various angles θ .

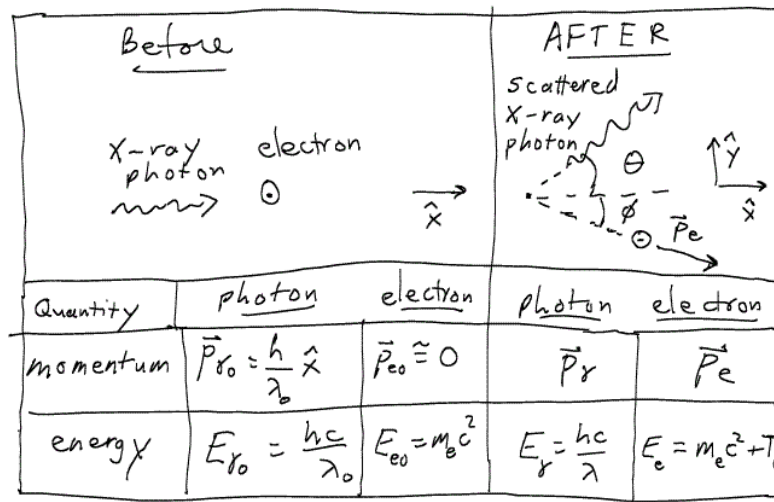


Fig. 1.6. Illustration of Compton scattering.

with

$$m_e = 9.109 \times 10^{-31} \text{ kg} = \text{mass of the electron.} \quad (1.23)$$

Compton verified that eq. (1.21) accurately described the experimental splitting. We now refer to λ_c as the *Compton wavelength of the electron*.

Compton's theory relies on a simple model of the scattering process, as shown in Fig. 1.6. We assume that individual X-ray photons scatter from individual electrons in the graphite. Before the scattering event, we assume that the X-ray photon carries momentum $\vec{p}_{\gamma 0} = (h / \lambda_0) \hat{x}$ and energy $E_{\gamma 0} = hc / \lambda_0$. An outer electron of a graphite atom has a kinetic energy on the order of a few eV. This turns out to be negligible relative to the other energies in the scattering problem. So, we can assume that before the scattering event, the electron has momentum $\vec{p}_{e0} \approx 0$ and energy $E_{e0} \approx m_e c^2$.

After the scattering event, the scattered photon has momentum p_{γ} directed at an angle θ relative to the incident X-ray beam propagation direction \hat{x} . We shall choose the \hat{y} direction so that this momentum vector \vec{p}_{γ} lies in the x - y plane. The energy of the scattered photon is $E_{\gamma} = hc / \lambda$. Due to the transfer of momentum from the photon, the electron acquires momentum \vec{p}_e directed at some angle ϕ relative to \hat{x} , and kinetic energy T_e .

We assume that momentum is conserved. The z -components of momentum are all zero. From the conservation of the x - and y -components of the momentum we obtain the equations

$$p_{\gamma 0} = p_{\gamma} \cos(\theta) + p_e \cos(\phi) \quad (1.24)$$

$$0 = p_{\gamma} \sin(\theta) - p_e \sin(\phi) \quad (1.25)$$

From which we find that

$$p_e^2 = p_{\gamma 0}^2 + p_\gamma^2 - 2p_{\gamma 0}p_\gamma \cos(\theta) \quad (1.26)$$

We also assume that energy is conserved, which gives

$$E_{\gamma 0} + m_e c^2 = E_\gamma + m_e c^2 + T_e = E_\gamma + \sqrt{m_e^2 c^4 + p_e^2 c^2} \quad (1.27)$$

since the total relativistic energy of the electron is

$$E_e = \sqrt{m_e^2 c^4 + p_e^2 c^2} = m_e c^2 + T_e. \quad (1.28)$$

Using the relation $E_\gamma = p_\gamma c$, we find from eq. (1.27) that

$$E_{\gamma 0} - E_\gamma = p_{\gamma 0} c - p_\gamma c = \sqrt{m_e^2 c^4 + p_e^2 c^2} - m_e c^2 \quad (1.29)$$

From which it follows that

$$p_e^2 = (p_{\gamma 0} - p_\gamma)^2 + 2m_e c (p_{\gamma 0} - p_\gamma) \quad (1.30)$$

From eqs. (1.30) and (1.26) it follows that

$$m_e c (p_{\gamma 0} - p_\gamma) = p_{\gamma 0} p_\gamma (1 - \cos(\theta)) \quad (1.31)$$

$$\text{and } \left(\frac{h}{p_\gamma} - \frac{h}{p_{\gamma 0}} \right) = \frac{h}{m_e c} (1 - \cos(\theta)) \quad (1.32)$$

so that finally

$$\lambda - \lambda_0 = \lambda_c (1 - \cos(\theta)) \quad (1.33)$$

So, we can explain the wavelength difference by treating the X-ray scattering as the scattering of a photon from an electron, taking into account the quantized energy and momentum of the photon. It is not possible to explain this wavelength shift with classical theory, so Compton's experiment provides powerful evidence for the photon.

Compton saw two lines in the scattered X-ray spectrum, a line with wavelength λ that we accounted for with the above argument, and an *unshifted* line of wavelength λ_0 . He correctly deduced that this unshifted line is due to scattering of the X-rays from the *inner* electrons of the carbon atoms, which are rather strongly bound to the carbon nucleus. Because of this stronger binding of the electron, the entire carbon atom recoils as a unit rather than the individual electron. We could account for this effect in the above analysis by substituting the mass of the whole atom for the mass of the electron. The resulting wavelength shift is not significant due to the much larger atomic mass.

4. Wave-particle duality

These developments introduced a paradox into physics. There is overwhelming evidence that light is an electromagnetic wave described by Maxwell's equation. Yet, light sometimes appears to be made of

discrete particles – photons. This behavior – sometimes a wave, sometimes a particle – is referred to as *wave-particle duality*. We now know that *all* “particles” – photons, electrons, protons, nuclei, atoms, pions, *etc.* exhibit wave-particle duality. This was first suggested by Louis deBroglie in 1924, and the wave behavior of electrons was first directly confirmed experimentally in 1927 by Clinton Davisson and Lester Germer, who observed Bragg diffraction of electron waves from a nickel crystal. It was also confirmed indirectly by Erwin Schrödinger’s successful wave-mechanical solution for the energy levels of the hydrogen atom in 1926. Other experiments have confirmed that wave-particle duality applies to all microscopic particles.

a) *Dispersion relation for matter waves, and the Schrödinger equation*

Waves for “particles” with non-zero rest mass m are called *matter waves* or *deBroglie waves*. The relation between momentum and wavelength is the same for massive particles and photons: $\lambda = h / p$. The relation between energy and frequency is also the same: $E = h\nu$. However, the momentum of a non-relativistic particle is mv and its energy is $(1/2)mv^2$, so the relations between frequency and momentum are different than for photons:

$$\lambda = \frac{h}{p} = \frac{h}{mv} \quad \text{or} \quad p = \frac{h}{\lambda} = \hbar k \quad (1.34)$$

$$E = \frac{p^2}{2m} = \hbar\omega \quad (1.35)$$

where λ is the *deBroglie wavelength*.

The *dispersion relation* of a wave is the relation between its frequency ω and wavenumber k . For example, the dispersion relation for an electromagnetic wave is $\omega = ck$, as given in equation (1.16). A wave which has a dispersion relation like this, with the form $\omega = \text{constant} \times k$, is called a *non-dispersive* wave. It has the property that a wave pulse will not change its shape as it propagates (*i.e.* doesn't "disperse"). Otherwise the wave is called *dispersive*. A wave pulse for a dispersive wave will change its shape as it propagates. Usually, it will get wider with increasing time and distance.

Eqs. (1.34) and (1.35) imply a definite dispersion relation for non-relativistic matter waves (for *free* particles, *i.e.* for the case $V(x) = 0$):

$$\omega = \frac{\hbar}{2m} k^2 \quad (1.36)$$

Thus, matter waves are *dispersive*.

In 1926, Schrödinger proposed that electrons are governed by what we now call the *Schrödinger equation*, and he used this equation to successfully derive the quantized energies of the hydrogen atom. While I am not familiar with the details of Schrödinger's work, it seems likely that he deduced this equation by noticing that the wave equation for electron matter waves has to have the dispersion relation (1.36). Considering first the case in which the electron is free, *i.e.* the potential $V(x) = 0$, this requirement leads to the equation

$$-\frac{\hbar^2}{2m_e} \frac{\partial^2 \psi}{\partial x^2} = i\hbar \frac{\partial \psi}{\partial t} \quad (1.37)$$

You can see this by substituting a wave solution $\psi = \psi_0 e^{i(kx - \omega t)}$ for the wave amplitude ψ . Also notice that unlike the case of electromagnetic waves, or waves on a string, use of the complex exponential form $e^{i(kx - \omega t)}$ rather than real form $\cos(kx - \omega t)$ is not just a mathematical convenience. There seems to be no way to write down a wave equation with the correct dispersion relation without using an inherently complex equation.

Once you have the equation (1.37) for free electrons, it's not such a leap to think of what the equation should be for electrons that move in a potential $V(x)$. Since $E = \hbar\omega$ is the energy of the electron, we

see that $i\hbar \frac{\partial}{\partial t} = \hbar\omega$ is the total energy. And since $\hbar k$ is the momentum of the electrons, we see that

$$-\frac{\hbar^2}{2m_e} \frac{\partial^2}{\partial x^2} = \frac{\hbar^2 k^2}{2m_e}$$

is the kinetic energy of the electrons. If the potential energy $V(x)$ is not zero, it then seems entirely reasonable to say that the kinetic energy plus the potential energy is equal to the total energy, *i.e.* that

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V(x)\psi = i\hbar \frac{\partial \psi}{\partial t} \quad (1.38)$$

which is just the Schrödinger equation in one dimension.

b) *Young's double slit experiment with photons and electrons*

We can learn more about wave-particle duality from the result of the experiment shown in Fig.

1.7(a). This is Young's double-slit interference experiment with an imaging photon or electron detector similar to the one described in Fig. 1.4. A video of an experiment with photons is here:

<http://www.youtube.com/watch?v=MbLzh1Y9POQ> . A video of an experiment with electrons is here: <http://www.youtube.com/watch?v=oxknfn97vFE&NR=1> .

Fig. 1.7(b) shows a few frames from the first video. In each experiment, the wave (light or electrons) is sent through a double slit, and an imaging detector is placed behind the slits. We see that particles are detected one-by-one at seemingly random positions. Yet over time, the intensity of the particles (number detected per unit area) shows a double-slit interference pattern. These experiments show wave-particle duality in a particularly direct way, with individual particle detection events visible. It is remarkable that *no matter how spread out the electromagnetic wave is – it could be a meter squared – when energy is absorbed by the detector, it is absorbed within a tiny area by one electron.*

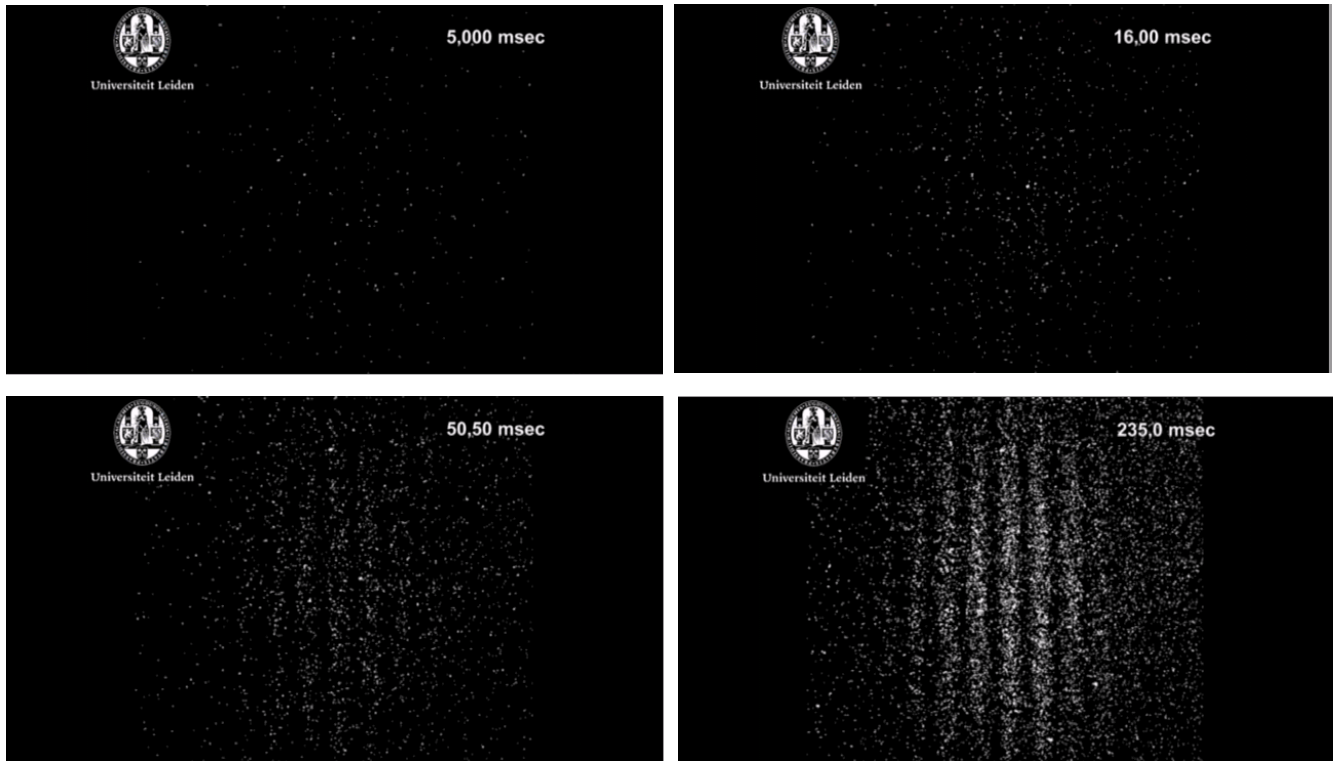
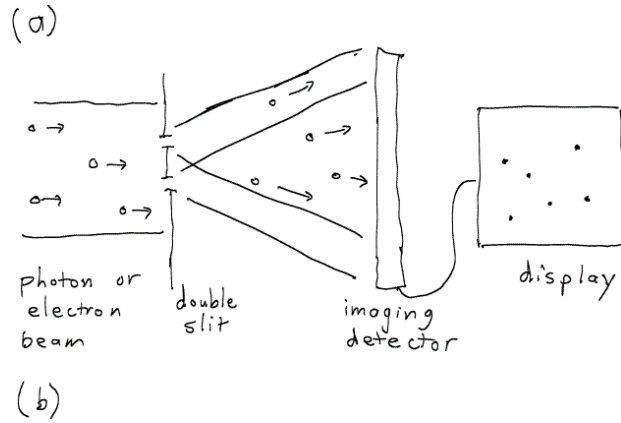


Fig. 1.7. a) Double slit experiment with an imaging detector (PSD). b) Images showing buildup of interference fringe pattern from many individual photon detections. Images are screen-captures from the video <http://www.youtube.com/watch?v=MbLzh1Y9POQ>, produced by David Dysktra, Steven Busch, Wouter Peters, and Martin van Exter and the University of Leiden.

c) *Polarization experiment with photons*

In order to further illustrate wave-particle duality, let me discuss another experiment, as illustrated in Fig. 1.8. Suppose that we shine light through a polarizer and then through a polarizing beam-splitter cube (PBSC). The polarizer transmits light polarized along some axis \hat{e} in the x - y plane, and absorbs light polarized at 90 degrees to that axis. The PBSC transmits light polarized along the \hat{y} -direction and reflects light polarized along the \hat{x} -direction. We also suppose that we have photon detectors (with

photoelectron multiplication) in both the transmitted beam (\hat{y} -detector) and reflected beam (\hat{x} -detector). Finally, we suppose the polarizer has its transmission axis \hat{e} oriented at an angle θ with respect to the \hat{x} -direction.

We know how to analyze this experiment with classical electromagnetism. The wave transmitted through the first polarizer and incident onto the PBSC has an electric field $\vec{E}_i = \hat{e}E_0 \cos(kz - \omega t)$, where \hat{e} is a unit vector directed along the transmission axis of that polarizer. The power incident on the PBSC is

$$P_i = \frac{1}{2} \epsilon_0 c E_0^2 A \quad (1.39)$$

$$\text{where } \epsilon_0 = 8.854 \times 10^{-12} \text{ F/m} \quad (1.40)$$

is the *electric permittivity constant*, and we assume that the beam has a uniform intensity over area A . We can express the electric field incident on the PBSC as a superposition of a field polarized along \hat{x} and a field polarized along \hat{y} :

$$\vec{E}_i = \hat{e}E_0 \cos(kz - \omega t) = \cos(\theta)\hat{x}E_0 \cos(kz - \omega t) + \sin(\theta)\hat{y}E_0 \cos(kz - \omega t) \quad (1.41)$$

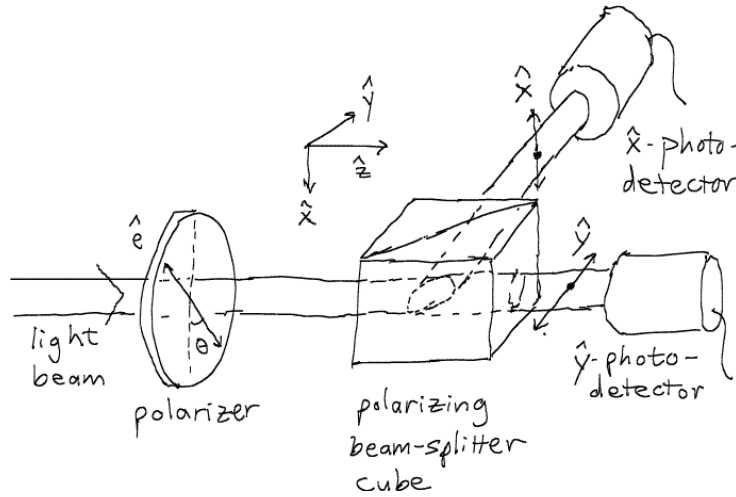


Fig. 1.8. Experiment with polarization of photons

Only the y -component of the field is transmitted through the PBSC, so the transmitted field is

$$\vec{E}_t = \sin(\theta)\hat{y}E_0 \cos(kz - \omega t) \quad (1.42)$$

Similarly, only the x -component of the field is reflected, so the reflected field is

$$\vec{E}_r = \cos(\theta)\hat{x}E_0 \cos(ky - \omega t) \quad (1.43)$$

The power falling on the \hat{y} -detector is

$$P_t = \frac{1}{2} \varepsilon_0 c \sin^2(\theta) E_0^2 = \sin^2(\theta) P_i \quad (1.44)$$

and the power falling on the \hat{x} -detector is

$$P_r = \frac{1}{2} \varepsilon_0 c \cos^2(\theta) E_0^2 = \cos^2(\theta) P_i \quad (1.45)$$

This result is called *Malus' law*.

If we look carefully at the result of this experiment, we find a problem with the above interpretation: the photodetectors don't put out a perfectly smooth photocurrent. They put out individual current pulses, one for each absorbed photon. Therefore, we need to think about this experiment from a quantum point of view. In that point of view, the light beam incident on the PBSC is made of photons, each of which is " \hat{e} -polarized." If we imagine that $\theta = 0$ or $\theta = 90^\circ$, then it is clear that all photons should be reflected or transmitted through the PBSC, respectively. That is what actually happens in the experiment. However, what if θ has some intermediate value? Then, what happens is that each photon ends up in one of the other of the two detectors at random. The probability to end up in the \hat{x} -detector is $\cos^2(\theta)$, and the probability to end up in the \hat{y} -detector is $\sin^2(\theta)$. We *never* observe a "fraction of a photon" arriving at a detector. This relative probability for reflection and transmission gives rise to the time-averaged distribution of power between the two detectors given by eqs. (1.44) and (1.45).

In this example of wave-particle duality, we see that the division of the average intensity of the beam obeys the laws of wave physics (Maxwell's equations for polarized light beams), but that when the beam is detected, it appears to be particles (photons).

5. The probability amplitude

a) General formulation

The paradox of wave-particle duality is "resolved" in quantum mechanics by simply accepting what the experiments tell us: photons, electrons, and all other "particles" behave as *both particles and waves*. We could have invented a new word to describe these entities, but we haven't – we just refer to them as particles. So, in quantum physics, you should imagine that the word "particle" has quotation marks around it, to remind you of its dual nature.

The experiments of Figs. 1.7 and 1.8 tell you something important: *quantum theory is not deterministic*. Let me discuss this point for the photon polarization experiment. You can only predict the *probability* that a given photon will hit a detector. Pretty much everyone has trouble with this concept – it just seems that there should be some way to describe how each photon "decides" which way to go at the PBSC. The answer to this kind of objection was emphasized particularly by Heisenberg: the only valid theory is one that is verified experimentally. The theory we have given agrees with experiment. A hypothesis for how the photon "decides" which way to go, that cannot be tested experimentally, cannot

be a *required* element of the theory.¹⁸

We can explain the results of the photon polarization experiment by supposing that each photon has polarization

$$\hat{e} = \cos(\theta)\hat{x} + \sin(\theta)\hat{y} \quad (1.46)$$

Then, we take

$$c_x = \cos(\theta) = \text{the probability amplitude for the photon to be polarized along } \hat{x}, \quad (1.47)$$

$$\text{and } c_y = \sin(\theta) = \text{the probability amplitude for the photon to be polarized along } \hat{y}. \quad (1.48)$$

The probability amplitude has the following meaning:

The magnitude squared of the probability amplitude gives the probability for the photon to be measured to have that property.

Therefore, the probability for each photon to be detected in the \hat{x} -detector is $|c_x|^2 = \cos^2(\theta)$, and the probability to be detected in the \hat{y} -detector is $|c_y|^2 = \sin^2(\theta)$. This theory describes exactly the result of the experiment.

You might think we don't need the probability amplitude. Why not just say that the beam contains a fraction $\cos^2(\theta)$ of \hat{x} -polarized photons, and a fraction $\sin^2(\theta)$ of \hat{y} -polarized photons? The answer is that this wouldn't work if we rotated the PBSC so that its transmission and reflection polarization axes lie along some other directions (\hat{x}', \hat{y}') . What we find experimentally is that the photons again divide on the PBSC according to new probability amplitudes squared: $|c'_x|^2 = \cos^2(\theta')$ and $|c'_y|^2 = \sin^2(\theta')$, where the photon polarization vector has been expressed in the new "measurement basis" (\hat{x}', \hat{y}') as $\hat{e} = \cos(\theta')\hat{x}' + \sin(\theta')\hat{y}'$. This is different than any result we would get by considering the beam to be composed of \hat{x} - and \hat{y} -polarized photons. No matter how we modify the experiment, what happens is that the *amplitudes* for the photons are governed by the classical wave theory of light, as expressed by

¹⁸ As far as I am aware, the only testable hypotheses for how particles "decide" which way to go are given by *hidden variable theories*. In such theories, it is supposed that each particle has some additional variable λ that we, at present, do not know how to calculate from the starting conditions of the experiment. It is this variable that "tells" the particles which way to go. According to such theories, nature is still deterministic and we're just not clever enough to determine λ . In 1964, John Bell noticed that hidden variable theories make predictions for correlations of particle detection events that differ from the predictions of "standard" quantum mechanics. Experiments have measured these correlations and ruled out hidden variable theories. Indeterminacy is still a fundamental part of quantum mechanics.

Maxwell's equations, and the probabilities for the detection of the photons are given by the amplitudes squared. We can never regard any individual photon as having any definite polarization, except for the polarization state \hat{e} in which it has been prepared.

The same concept describes the results of Young's double slit experiment. We suppose that each particle has *probability amplitudes* ψ_1 and ψ_2 to go through slits 1 and 2, respectively, and that the total probability amplitude to the right of the slits is the linear superposition of amplitudes $\psi = \psi_1 + \psi_2$. The probability amplitudes ψ_1 and ψ_2 obey wave equations. The probability to detect the particles at the PSD is the magnitude squared of the resultant amplitude at the position of the detector: $|\psi|^2 = |\psi_1 + \psi_2|^2$. The probabilistic interpretation of the amplitude ψ explains the random particle-by-particle detection events, and the wave interference between ψ_1 and ψ_2 explains the observed fringes.

[Dark grey shading of the following material means I'm not expecting you to know this from previous courses.]

So far, it may appear that “particles” behave as particles only at detectors, and everywhere else behave as waves. Unfortunately (or fortunately, if you love difficult puzzles), nature is more complicated than that. You can see this from the Compton scattering experiment. There is no detector at the point at which the scattering occurs, yet we cannot understand this experiment as the scattering of an electromagnetic wave from an electron deBroglie wave. We cannot even understand it as the scattering of an electromagnetic wave from a point electron – in that case there would be no shift in the wavelength of the scattered X-rays. It is essential to view this experiment as the scattering of a photon from an electron, both considered as particles. So in some cases, “particles” behave as particles *when they interact with other particles*.

In order to understand experiments like Compton scattering, we are forced to extend the concept of the probability amplitude. The reason is that the photon that is scattered from the electron is different than the photon that is incident on the electron, and that difference (in frequency) cannot be described by classical physics. In other words, the incident photon has been “destroyed” and a new, scattered photon “created.” So, our theory must incorporate *probability amplitudes for particles to be created and destroyed*. The theory that incorporates creation and destruction of photons, electrons, and positrons is *quantum electrodynamics* (QED). QED is the simplest form of *quantum field theory*, and was first used to calculate the spontaneous emission rate of an atom by Eugene Wigner and Victor Weisskopf in 1929, and the cross-section for Compton scattering by Oskar Klein and Yoshio Nishina, also in 1929. In QED, photons, electrons, and positrons give rise to a corresponding *quantum field*. For instance, the electromagnetic field is a quantum field composed of photons. This means that the classical potentials Φ and \vec{A} and the corresponding classical electromagnetic fields \vec{E} and \vec{B} are only approximations. In reality, these have a “grainy” quality to them because they are made up of individual photons.

Quantum electrodynamics was fully worked out in the late 1940s and the 1950s by Richard Feynman, Julian Schwinger, Sin-Itiro Tomonaga, and others. It is a complicated theory, and we will not

be able to go into much detail. However, there is a wonderful small book by Feynman that presents the main ideas in an understandable form,¹⁹ and I highly recommend it. I'll give a brief description along these lines here. Since we have discussed the Compton scattering experiment, I'll show qualitatively how that problem is solved in QED. The process involves the destruction of a photon of wavevector \vec{k}_0 and the creation of a photon of different wavevector \vec{k} by interaction with an electron. The wavelength shift is the difference $\lambda - \lambda_0 = (2\pi / k) - (2\pi / k_0)$. Note, however, that this calculation will describe more than just the wavelength shift that we already derived. It will also describe the intensity and angular distribution of the scattered X-rays, or to put it another way, the differential cross section for Compton scattering.

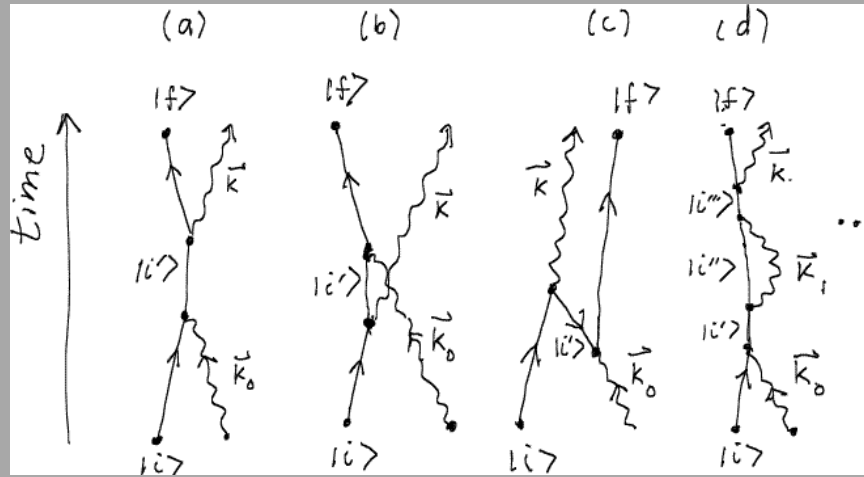


Fig. 1.9. Feynman diagrams representing probability amplitudes for Compton scattering.

The electron will start in some well-defined state $|i\rangle$ (in this case some electronic “orbital” of an electron in a graphite solid), and end in some different state $|f\rangle$ due to the transfer of momentum from the absorbed photon. Suppose for the moment that we detect *both* the final state of the scattered photon, and the final state of the electron. In that case, we want to calculate the probability that the photon will be detected in state $|\vec{k}\rangle$ and the electron detected in state $|f\rangle$ at some time t_1 , given that the system was prepared with the photon in the state $|\vec{k}_0\rangle$ and the electron in the state $|i\rangle$ at some earlier time t_0 . The calculation follows along lines similar to the double-slit experiment or the photon polarization experiment. We cannot say for certain what properties the photons and electron have at any intermediate time, just like we can't say a particular photon is “ \hat{x} -polarized” before it hits the PBSC. All we can do is to calculate *probability amplitudes*. And, if there is more than one way to go from the initial state to the final state, *those probability amplitudes add*, just as they did for the Young's double slit experiment. The probability to get the measurement result is the magnitude squared of the probability amplitude for that

¹⁹ *QED, The Strange Theory of Light and Matter* (Princeton Univ. Press. 1988).

result. So what we need to do is to figure out all the ways the electron and photon can get from their initial state to their final state, and add together the probability amplitudes for all those different ways.

Here is one sequence of events that could lead to a transition from the initial state to the final state: the electron could evolve freely for some time, then absorb the photon of wavevector \vec{k}_0 and make a transition to a different state $|i'\rangle$, then evolve freely for some additional time, then emit the photon of wavevector \vec{k} and make a transition to the final state $|f\rangle$, and then evolve freely for some additional time. There will be some probability amplitude for this process ψ_a that we can calculate. This will be the product of the probability amplitudes for the free evolution of the electrons and photons and the probability amplitudes for photon absorption and emission. The probability amplitudes for free evolution will obey wave equations like the Schrödinger equation or Maxwell's equations. The resultant probability amplitude ψ_a will also involve integrals over all possible intermediate states $|i'\rangle$ of the electron, as well as all possible photon emission and absorption times. (This is just the superposition of amplitudes again.) Calculating the expression for this amplitude is a little complicated, but this kind of calculation occurs so frequently in QED that there is a set of rules about how to calculate them, once you specify the sequence of events. It is convenient to summarize the sequence of events by a *Feynman diagram*, as shown in Fig. 1.9(a). The straight lines represent the free evolution of an electron in a given state, and the squiggly lines represent the free evolution of a photon in a certain state. The intersection points between straight and squiggly lines represent a photon absorption or emission event. Time increases in the upward direction. A skilled practitioner of QED can just look at such a diagram and write out its probability amplitude.

Now, it turns out that this is not the only possible process. It could be that the electron first emits the \vec{k} photon and then absorbs the \vec{k}_0 photon. This may seem strange, but it can happen. The diagram for this case is shown in Fig. 1.9(b); the probability amplitude for that diagram is ψ_b . Even stranger things can happen. For example, the electron can first emit the \vec{k} photon, then run backwards in time, absorb the \vec{k}_0 photon, and then run forwards in time again. This process is represented by the diagram of Fig. 1.9(c), with probability amplitude ψ_c . If you don't like that way of looking at it, there is another one. The backwards-time-traveling electron is the same thing as its forward-time-traveling antiparticle, the positron. So you can think of the process as a photon of wavevector \vec{k}_0 and a electron in state $|i\rangle$ traveling forward in time, followed by an interaction that destroys the photon and creates an electron-positron pair, with the electron in state $|f\rangle$. At a later time, the positron and the electron in state $|i\rangle$ simultaneously annihilate each other and emit a photon of wavevector \vec{k} . Then the photon and the electron propagate freely until time t_1 .

There are still more possibilities. For example, it is possible for the electron to emit and then later reabsorb a third photon. That process is represented by diagram 1.9(d) with probability amplitude ψ_d . It

is possible for this same process to happen, but with the time order of the various photon absorption and emission events changed, so we will have another set of diagrams for those processes. And the possibilities go on and on – in fact, the set of diagrams is an infinite series. For each diagram in the infinite series, there is a corresponding probability amplitude. The overall probability amplitude is

$$\psi(f, \vec{k}) = \psi_a + \psi_b + \psi_c + \psi_d + \psi_e + \dots \quad (1.49)$$

and the probability to detect the electron in state $|f\rangle$ and the photon in state $|\vec{k}\rangle$ is just

$$P(f, \vec{k}) = |\psi(f, \vec{k})|^2 = |\psi_a + \psi_b + \psi_c + \psi_d + \psi_e + \dots|^2 \quad (1.50)$$

You may wonder how the wavelength shift comes into this picture. It is not apparent from the diagrams themselves, but the formulas for the Feynman diagram amplitudes $\psi_{a,b,\dots}$, contain a factor that makes an amplitude large if energy and momentum are conserved, and small if they are not conserved. The approximate conservation of energy and momentum causes the probability distribution $|\psi(f, \vec{k})|^2$ to be strongly peaked at the value of $k = |\vec{k}| = 2\pi / \lambda$ that satisfies Compton's formula (1.33). In practice, these peaks are nearly δ -functions. That is because the initial and final of the system are relatively long-lived, which makes the energy uncertainty in the overall process small. They are also relatively extended in space, which makes the momentum uncertainty in the overall process small. (In the *intermediate* states of these diagrams, energy and momentum conservation can be pretty strongly violated because those intermediate states can be very short-lived and localized.)

In the Compton scattering experiment, the final state of the electron is not detected. How do we account for this? The answer is that we *sum the probabilities for the measurement result to be $|f, \vec{k}\rangle$ over the unobserved final states $|f\rangle$ of the electron*:

$$P(\vec{k}) = \sum_f P(f, \vec{k}) = \sum_f |\psi(f, \vec{k})|^2 \quad (1.51)$$

This gives us the probability to detect a photon of wavevector \vec{k} . The reason eq. (1.51) is correct is that even though we didn't measure the final state of the electron, in principle we *could have*. The probability for the photon to have a certain final state should not depend on whether we decided to put an electron-state-detector into the experiment or not. Since the probability to obtain the result $|f, \vec{k}\rangle$ is just $|\psi(f, \vec{k})|^2$, it follows that the probability to get a photon in state $|\vec{k}\rangle$, regardless of what happened to the electron, is the sum given by eq. (1.51).

This kind of calculation may seem very difficult, and in general it is. However, calculations in QED often converge very quickly. For example, the Klein-Nishina formula for the Compton scattering cross-section only takes into account the first two diagrams 1.9(a) and 1.9(b), and it accurately describes the

experimental cross-section.

It is very important to understand the difference between eq. (1.50) and eq. (1.51), so I'll discuss that further. Suppose that we have a number of different ways to get from one quantum state to another as shown in Fig. 1.10(a). And further, suppose that the initial state defined by the values α and β of two simultaneously measurable observables, and the final state by the values α' and β' of these same observables. For example, the paths of Fig. 1.10(a) could represent the path from a particle source, through the two different slits of a Young's double slit experiment, and to a final point on the detector.

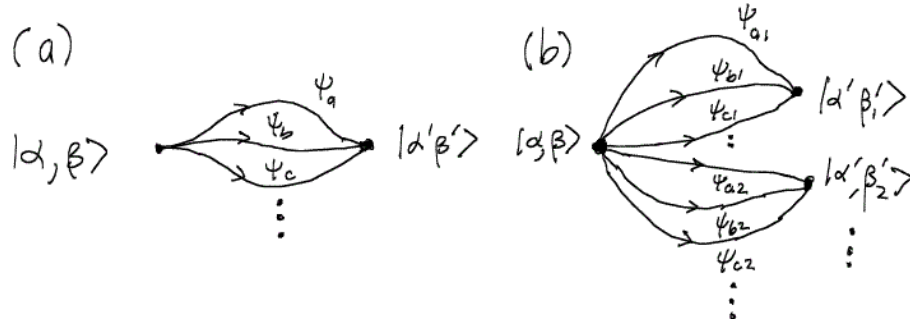


Fig. 1.10. (a) Different paths that start at a single quantum state $|\alpha, \beta\rangle$ and end at a single quantum state $|\alpha', \beta'\rangle$. (b) Paths that start at a single quantum state $|\alpha, \beta\rangle$ and end at a number of different quantum states $|\alpha', \beta'_1\rangle, |\alpha', \beta'_2\rangle, \dots$

The variables α and α' could represent the position of the particle emission on the source and the position of particle detection on the detector. The variables β and β' could represent a component of spin of the particle at the source and detector. (There would have to be some possibility of “spin-flips” in the experiment for this variable β to be relevant.) Or, the two paths could represent diagrams 1.9(a) and 1.9(b) of the Compton scattering experiment. In that case, the variables α and α' could represent the initial and final states of the photon, and the variables β and β' could represent the initial and final states of the electron.

To find the probability to obtain the measurement result (α', β') , add the amplitudes ψ_a, ψ_b, \dots of all possible ways to transition from the initial to the final state, and take the magnitude squared of the resultant amplitude:

$$P(\alpha', \beta') = |\psi(\alpha', \beta')|^2 = |\psi_a + \psi_b + \dots|^2 \quad (1.52)$$

Suppose, however, that we are only going to measure the probability that the system will be found to have value α' , and that we will not make any observation of the second simultaneously observable variable β' . Two states with the same value of α' , but different values of β' are *different states*. That is, $|\alpha', \beta'_1\rangle \neq |\alpha', \beta'_2\rangle$, if $\beta'_1 \neq \beta'_2$.

To find the probability to obtain the measurement result α' , *irrespective of the value of an unobserved but simultaneously measurable quantity β' , add the probabilities to obtain each of the results (α', β'_i) :*

$$P(\alpha') = \sum_i P(\alpha', \beta'_i) = |\psi_1|^2 + |\psi_2|^2 + \dots = |\psi_{a1} + \psi_{b1} + \dots|^2 + |\psi_{a2} + \psi_{b2} + \dots|^2 + \dots \quad (1.53)$$

An experiment in which *amplitudes* add, like those described by equation (1.52) will exhibit *interference*. This interference is associated with the cross terms $\psi_a^* \psi_b$, $\psi_b^* \psi_a$, etc. in the probability $|\psi_a + \psi_b + \dots|^2$. The fringes in a double-slit experiment are an example of this interference. On the other hand, an experiment in which probabilities add will not have interference (among those outcomes for which the probabilities add).

To summarize this point:

Amplitudes to go from the same initial state to the same final state *interfere* with each other.

Amplitudes to go from the same initial state to different final states *do not interfere* with each other.

These rules are consistent with all of the experiments we have discussed so far, and they are universal rules in quantum mechanics. I should concede that we haven't been very specific about what "simultaneously observable" means, but we'll come back to that below.

[Grey shading of the following material means I'm not expecting you to know this from previous courses.]

One final thing I would like to point out is that if you accept QED as the correct theory, then Einstein's theory of the photoelectric effect makes perfect sense. Of course one electron will absorb all the energy of one photon. That is the lowest-order Feynman diagram for a photon-electron interaction, and therefore it is the diagram with the highest probability amplitude.

b) *Probability amplitudes for particles in classical fields*

Since the classical electromagnetic field is only an approximation, this raises the question of when that approximation is OK. Roughly speaking, the classical electromagnetic approximation will tend to be good when fields are composed of many photons and bad when they are not. Quantization effects also tend to be more pronounced for higher frequency photons than for lower frequency ones, due to the larger quanta of energy and momentum.

After all of this discussion about QED, it may seem surprising that it's just fine to treat the hydrogen atom as an electron moving in the classical Coulomb potential of the proton. Fig. 1.11 shows Feynman's QED picture of the hydrogen atom. In QED, the proton and electron are constantly exchanging huge numbers of photons. These photons are responsible for the electrostatic force between the proton and the electron. Because of their large number and relatively low frequency, the photons produce a field which is very close the classical Coulomb electric field. QED calculations of the hydrogen atom level energies

agree with the experimentally measured levels of hydrogen to within the experimental accuracy of roughly 1 part in 10^{12} . This is one of the reasons we believe that QED is the complete and accurate theory of photons, electrons, and positrons. However, we can also calculate the energy levels of hydrogen with classical fields. We find that these calculated level positions differ from the experimentally measured levels by only about 1 part in 10^6 . So, unless you are striving for very high accuracy (which is unusual for realistic problems), the classical field approximation is fine for most calculations of electronic wavefunctions in atomic, molecular, and condensed matter physics.

In this class, we will spend most of our time studying the *quantum mechanics of massive particles* (e.g. *electrons, protons, etc.*) moving in classical fields. The connection between the probability amplitude and experiments in this regime is as follows:

If massive “particles” (electron, proton, etc.) move in fields that can be approximated as classical fields, then at the position of *detection*, massive “particles” behave as particles. At all other positions, massive “particles” behave as waves. The entity that is “waving” is the *probability amplitude* or *wavefunction* $\psi(\vec{r})$. The wavefunction $\psi(\vec{r})$ is found by solving the Schrödinger equation or its relativistic extension. At the position of a detector, the probability to detect a particle is proportional to $|\psi|^2$.

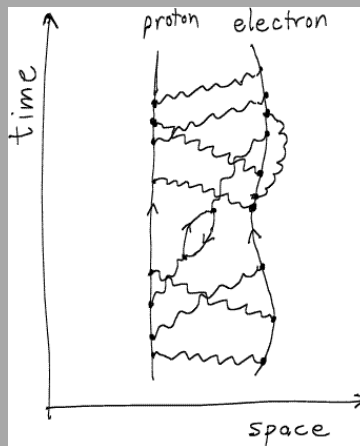


Fig. 1.11. Feynman diagram for the time evolution of a hydrogen atom.

This is, as we’ve discussed, an approximation. However, it is a pretty good one for many problems of interest. Also, the mathematical tools that we will develop apply to all of “quantum physics”, including QED. For this reason, most physicists use the term “quantum mechanics” to refer to all of quantum theory including QED. I’ll do that as well, qualifying the term as in the above statement if necessary.

c) *Quantum optics*

The experiment of Fig. 1.7 with photons suggests that we should have statements analogous to the bold-faced statements above for photons. But there is a big difference between photons and massive

particles like electrons and photons. In the regime we discussed above, electrons are never created or destroyed. In contrast, optical photons are frequently created and destroyed by their interactions with matter. As a result, we have to consider amplitudes for scattering processes like those for Compton scattering given in Figs. 1.9(a) and (b) - even for optical photons. Especially if only a few photons are present, we are going to require QED to properly describe them, and that is a completely different situation than exists for electrons in atomic, molecular, and condensed matter physics. However, we require a version of QED in which only photons are created and destroyed, not electrons. This restricted version of QED is much simpler than the full-blown QED of Feynman *et al.* It is also referred to as *Quantum Optics*.²⁰

But still, what about the experiment of Fig. 1.7? In that experiment, photons are “destroyed” only at the detector (except for those absorbed by the double-slit aperture, which we are ignoring). In that situation, can we make statement analogous to those we made above for massive particles?

The answer is yes, with a caveat. In QED, we introduce *complex* (*i.e.* involving $i = \sqrt{-1}$) field operators \vec{E}^+ and \vec{E}^- . The dependence of these field operators on space and time is given by Maxwell’s equations. The expectation value $\langle \vec{E}^+ + \vec{E}^- \rangle$ has the physical interpretation of the classical electric field \vec{E} . And, the probability to detect a photon is proportional to the expectation value $\langle \vec{E}^- \vec{E}^+ \rangle$. These are rigorously correct statements, so you can always use the operators \vec{E}^+ and \vec{E}^- if you have to. However, because the properties of the operator combination $\vec{E}^+ + \vec{E}^-$ are so similar to those of the classical field \vec{E} , if no photon absorption or emission is occurring in intermediate states, as in Fig. 1.7, then everything usually works out fine if you think of the complex field \mathcal{E} as a “wavefunction” for the photon, where the physical electric field is given by $\vec{E} = \text{Re}(\mathcal{E}(\vec{r}, t)\hat{e})$. For example, a plane wave field would have $\mathcal{E} = \mathcal{E}_0 \exp(i(\vec{k} \cdot \vec{r} - i\omega t))$. It is usually valid to say that the probability to detect a photon is proportional to $|\mathcal{E}|^2$. It is also valid to think of the components of the polarization vector \hat{e} as probability amplitudes for polarization states. In fact, we have already described a number of experiments in these terms, and those descriptions were perfectly correct.

But, be aware that there are some pitfalls for the unwary if you try to go too far with this oversimplified theory. We will not have time in this class to discuss the subtleties involved.

Usually, when photons interact with transparent macroscopic matter, it is perfectly fine to treat their interactions using volume-averaged properties of that matter such as refractive index. This is true even though scattering of photons is involved. If we think of an optical photon scattering event as a low-photon-energy case of Compton scattering, then the scattering of optical photons from electrons in a transparent material is completely dominated by *elastic scattering*, in which the scattered photon has the

²⁰ Accessible introductions can be found in: Rodney Loudon, *The quantum theory of light*, 3rd ed. (Oxford, 2000), or C. Gerry and P. Knight, *Introductory Quantum Optics* (Cambridge Univ. Press, 2005).

same wavelength λ_0 as the incident photon, *and* the electron starts and ends in the same state $|f\rangle = |i\rangle$. In that case, the amplitudes for a photon scattering from the electrons in the media add up in the same way as they do for a classical electromagnetic wave scattering from those electrons. It is just the addition-of-amplitudes case illustrated in Fig. 1.10(a). Therefore, when you think of the laws governing division of optical beam intensities at dielectric surfaces, diffraction gratings, polarizing beams splitters, and so on, you can think of them equally well as the division of classical fields obeying the volume-averaged or “macroscopic” form of Maxwell’s equations, *or* as an identical division of the probability amplitude for photons. Again, we have already explained the photon polarization experiment in those terms, and that explanation was perfectly fine. You can see how all of this works out in more depth in the book by Feynman.

d) *The quantum measurement problem*

Quantum theory has proven to be 100% successful at describing the properties of atoms, molecules, light, and subatomic particles. Yet, there is a logical problem with this theory that has troubled us since the earliest days of quantum mechanics. That is, what does it mean to say that a particle is “detected”? It is as if we are supposed to draw a kind of “line” and say, before this line the system is quantum and only has a certain probability amplitude function, as in Fig. 1.12(a). But after this line the “measurement” has occurred and the photon, photoelectron, or other particle has been “detected,” and has a sharply defined position as in Fig. 1.12(b). This transition from 1.12(a) to 1.12(b) is referred to as the *collapse of the wavefunction*.

In the PSD of Fig. 1.4, it is clear where this line is to be drawn: a little ways into the channel multiplier, just past the point where a single photoelectron has been converted into a few electrons or so. Before this point, we have only one or a few electrons, and we are still in the quantum world. Much past this point, we have a shower of hundreds or thousands of electrons – so many that we can treat them as a classical current pulse. From a practical point of view, essentially all problems in quantum mechanics will have some “logical” dividing line of the kind we discussed above. Yet this seems unsatisfactory from a fundamental point of view, because the dividing line is arbitrary. Shouldn’t a single theory describe the whole apparatus, and not require such an arbitrary division? The lack of a satisfactory answer to this question is called the *quantum measurement problem*. Generally speaking, physicists are divided as to whether this is a real problem or not. It’s inadvisable to spend too much time thinking about it, because you’re virtually certain to get nowhere, just like everyone else who has tried to solve it.

In section B, we will consider a generalization of the measurement process like that shown in Fig. 1.13(a). A beam of particles in some state $|\psi\rangle$ will be split into secondary beams by some “A-splitter” such that all the particles in one secondary beam are in state $|a_1\rangle$ and all the particles in the other secondary beam in state $|a_2\rangle$. A particle which is detected in the “ a_1 beam” will be said to have been “measured to be in state $|a_1\rangle$ ”, and similarly for a_2 . For the photon polarization experiment, the A-splitter would be the PBSC, and the states $|a_1\rangle$ and $|a_2\rangle$ would be \hat{x} -polarization and \hat{y} -polarization.

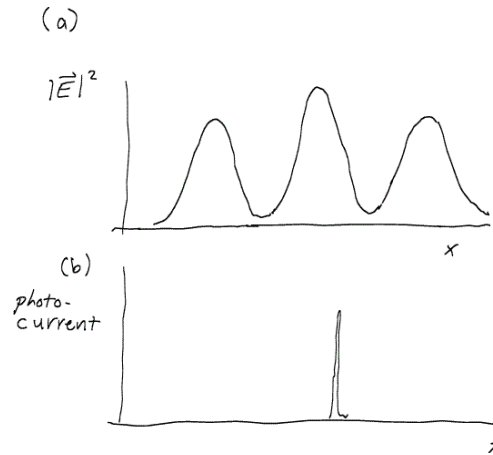


Fig. 1.12. (a) Average intensity for many photons hitting the detector vs. position x for a Young's double slit experiment. (b) Photocurrent for one detected photon vs. position x , for a typical detection event.

I show Fig. 1.13 to emphasize that the “measurement” takes place *only* when the particle is actually detected in the particle detector, not when the beam is divided by the A-splitter. For, if we wish, we could take out the detectors, replace them by mirrors, and put the beams into an “A-combiner” which is just an A-splitter installed backwards, as shown in Fig. 1.13(b). Then, what happens is the secondary beams will recombine so as to produce some *new state* $|\psi'\rangle$. This new state will be sensitive to the difference $\Delta\phi$ between the phases accumulated by the two probability amplitudes for states $|a_1\rangle$ and $|a_2\rangle$ in the interval between division and recombination of the beam. In optics, we call this an *interferometer*. For the case of the polarization experiment, the reconstructed state $|\psi'\rangle$ would be some state of definite polarization, possibly elliptical, that depends on $\Delta\phi$. To put it another way, measurements in quantum mechanics obey a “Yogi Berra” principle – *it ain't over until it's over*. That is:

In quantum theory, until it's over, *all you have are amplitudes*.

You *never* say “this particle has property a_i ”, unless it is in the pure state $|\psi\rangle = |a_i\rangle$.

If $|\psi\rangle \neq |a_i\rangle$, you can only say “this particle has *probability amplitude* ψ_i to have property a_i .”

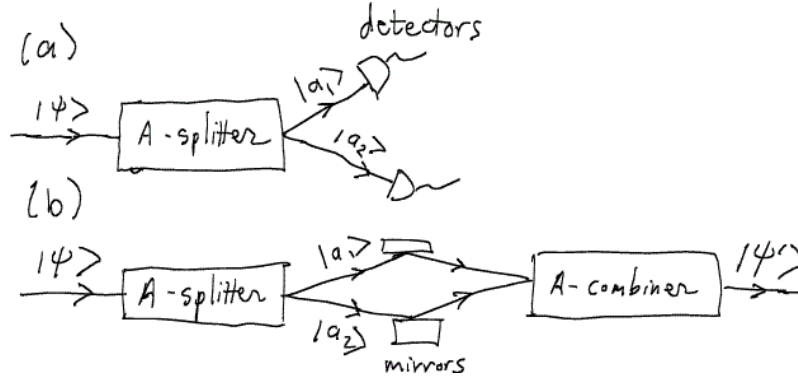


Fig. 1.13. Illustration of the possibility of coherent recombination of a divided beam.

6. Observables and uncertainty in quantum mechanics; Heisenberg's microscope

In quantum mechanics, an *observable* is any quantity that can be measured. This concept of an observable is much less straightforward than in classical physics because of the following two intimately connected facts:

i) The size of an atom is about $0.3 \text{ nm} = 0.3 \times 10^{-9} \text{ m}$ because quantum effects set that size to be a small multiple of

$$1 \text{ Bohr} = a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2} = 0.5292 \times 10^{-10} \text{ m} \quad (1.54)$$

Quantum effects must be involved because \hbar appears in eq. (1.54).

ii) When a measurement is made on an atomic-scale object, the disturbance of that object by the measurement cannot be neglected.

This point is so important that it is worth going over in some detail. Consider the “Heisenberg’s microscope” shown in Fig. 1.14(a). Our goal is to measure the position x of a particle of mass m . In order to do this, we scatter photons of wavelength λ , frequency $\omega = 2\pi c / \lambda$, and wavevector $k = 2\pi / \lambda$ from the object, and view the scattered photons with a high magnification microscope. In our updated version of the microscope, we’ll suppose that the scattered photons are detected with a position sensitive photon detector. Each photon transmitted through the lens will be detected at some specific point by the PSD (shown as an individual dot in Fig. 1.14(a)). However, since photons exhibit wave-particle duality, the intensity pattern recorded by the PSD will be the diffraction pattern of the circular aperture of the objective lens. (We assume here that the magnification of the optical system is large enough that the finite size of the secondary electron shower is negligible.) That pattern is just the usual “Airy disk”, and we’ll define the “size” Δx of that pattern to be the diameter of the first minimum of the Airy disk referred to the object plane. (In other words, we take the size of the Airy disk on the actual PSD and divide by the magnification of the microscope.) According to wave-optical theory, the size of the disk is

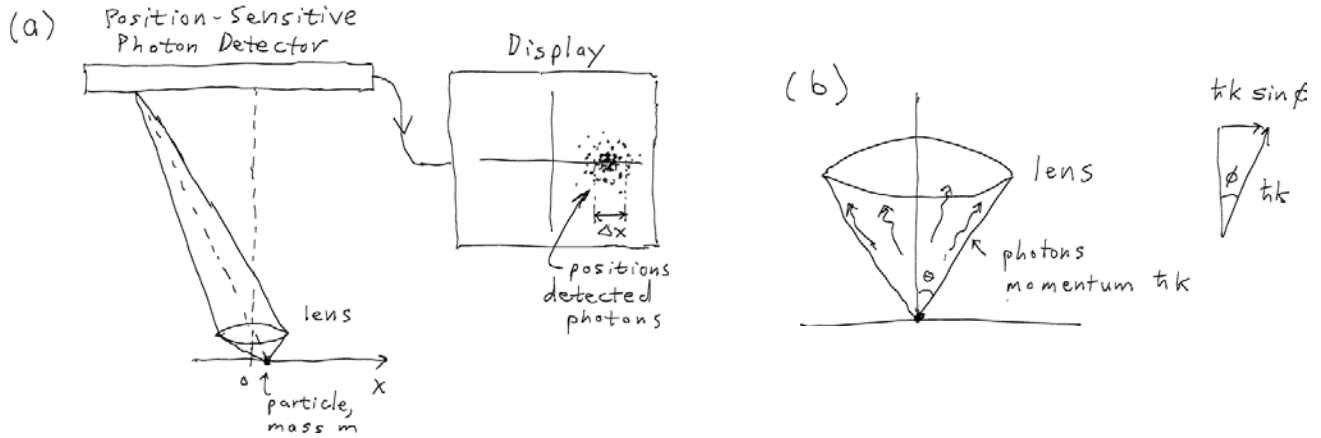


Fig. 1.14. Updated Heisenberg microscope. (a) Experimental arrangement. Note that each dot in the “display” is the position of one detected photon. The scatter of dots represents the positions of many detected photons, not the spread of the secondary electrons. (b) Detail of photon emission into lens.

$$\Delta x = 1.22 \frac{\lambda}{\sin(\theta)} \quad (1.55)$$

where θ is the half-angle subtended by the objective lens. We suppose that the particle is free (for example it is at rest inside of a vacuum chamber) and initially at rest. According to the rules of quantum physics, we don't know the angle ϕ at which the first photon will be emitted - only the probability amplitude that it will be emitted into a given angle ϕ . When that photon is emitted, it carries a momentum $\hbar k$, and x -component of momentum $p_x = \hbar k \sin(\phi)$. By conservation of momentum, the particle must recoil, *i.e.* it picks up a momentum $p_{\text{recoil}} = -\hbar k \sin(\phi)$. Ignoring factors of 2 and such, the uncertainty in this recoil momentum is

$$\Delta p_{\text{recoil}} \approx \hbar k \theta = \frac{h}{\lambda} \theta \quad (1.56)$$

since $\sin(\phi) \sim \phi$, and the uncertainty in ϕ is $\Delta \phi \sim \theta$. From eqs. (1.55) and (1.56), it follows that

$$\Delta x \Delta p_{\text{recoil}} \sim \left(\frac{\lambda}{\theta} \right) \left(\frac{h}{\lambda} \theta \right) = h \quad (1.57)$$

This result shows that in order to measure the position of the particle with accuracy Δx , we are forced to transfer a random momentum $\Delta p_{\text{recoil}} \sim \frac{h}{\Delta x}$ to the particle. Equation (1.56) provides us with a quantitative measure of the minimum perturbation to the system: we can't perform an independent measurement of the particle's momentum without picking up the uncertainty Δp_{recoil} . We see that it is impossible to *simultaneously* measure the position and momentum of the particle with uncertainties below the limit set by the *Heisenberg uncertainty principle*. A precise statement of the Heisenberg uncertainty principle is

$$(\Delta x)_{rms} (\Delta p_x)_{rms} \geq \frac{\hbar}{2} \quad (1.58)$$

where $(\Delta x)_{rms} = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$ and $(\Delta p_x)_{rms} = \sqrt{\langle (p_x - \langle p_x \rangle)^2 \rangle}$ are the rms uncertainties (standard deviations) of x and p_x , respectively, and the brackets $\langle \rangle$ denote an average over the probability distribution. You might think of the idea of recording many more than one photon, since it then would be possible to determine the center of the pattern in Fig. 1.14(a) to better than Δx . It turns out that this doesn't allow you to circumvent eq. (1.58) because the additional position information gained from more photons comes at the expense of an increase in Δp_x . No matter what you do, you can't think of an experiment than can measure *both* x and p_x better than the limit set by eq. (1.58).

If you set m to the mass of an electron, and Δx to be the size of an atom, you will discover that Δp_x is of the order of magnitude of the momentum of the electron in an atom. This means that there is no possibility of measuring classical-like trajectories of electrons inside atoms. Therefore we cannot *require* our theory of an atom to provide us with such trajectories, since a theory can be established as valid only by comparison with experiment. Because of this, the theory of an atom will involve many fewer variables than a classical theory, because we don't need to specify the particle trajectories. Uncertainty will be an inherent part of the theory.

These same quantum effects resolve one of the mysteries of classical physics: why are atoms stable in the first place, and what determines their size? In classical physics, if you take almost any starting initial conditions for the N electrons of a neutral atom and crank away at Newton's laws of motion on a computer, you will discover that the atom quickly sheds off the outer electrons one by one. The positive energy change for the ionized electrons is compensated by a decrease in the energy of inner electrons as their orbits shrink closer and closer to the nucleus. This continues until only one electron remains. Even the last electron eventually spirals into the nucleus, because its periodic orbit causes it to radiate electromagnetic waves and lose energy.

In quantum mechanics, the size of the atom is set by the uncertainty principle. The electrons would like to lower their energy as much as possible. They can minimize their potential energy by getting closer to the nucleus. But this localization in position comes at the expense of added momentum, and therefore kinetic energy, because of the uncertainty principle. The minimum energy state is reached through a compromise between these two effects, and turns out to have a finite extension of a few times the a_0 . Thus, the stability and finite size of atoms is intimately connected with quantum uncertainty.

B. Dirac's formulation of quantum mechanics

1. Motivation in terms of experimental measurements

In this section, I will review Dirac's formulation of quantum mechanics. Before reviewing the main features of the theory, I'd like to start by connecting the form the theory takes with the experiments as much as possible. We will suppose that the photon polarization experiment of Fig. 1.8 can be generalized as illustrated in Fig. 1.15. We suppose that a beam of particles has been prepared in a *pure state*. A pure state is one which is prepared with the maximum degree of specification possible within the limits set by quantum mechanics. We'll describe further what it takes to prepare a pure state below. We also suppose that a device can in principle be built to measure some property A of that system, which means that A is an observable. As was the case for the photon polarization experiment of Fig. 1.8, we suppose that the apparatus has an "A-splitter" that divides the input beam into multiple secondary beams according to the possible measurement results a_1, a_2, a_3, \dots of observable A , and that each beam is directed into a particle detector. If a particle is detected in the detector corresponding to result a_i , we say that the particle has been *measured to have the value a_i for the observable A* . We assume the splitter is ideal, meaning that it puts out one beam for every possible measurement result and has no internal loss of particles. We also suppose that the splitter is "classically describable".²¹ For the case of massive particles, this is equivalent to the requirement that the A-splitter act on the particles only with classically describable fields.

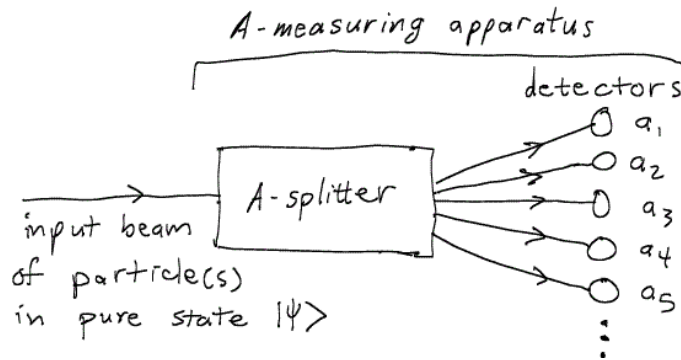


Fig. 1.15. Generalized A-measuring apparatus.

²¹ If this condition were violated, the particle P in state $|\psi\rangle$ could couple to some other quantum system O within the A-splitter without any absorption of the particle. In this case, information about the quantum state of the particle P is transferred to system O . That information can be then extracted by a measurement on system O . If such coupling and measurement is intentional, it is called a *quantum non-demolition (QND) measurement*. Quantum mechanics is perfectly capable of describing such a measurement strategy, but the system O would have to be explicitly included in the formalism to give correct results for all possible measurements on the secondary beams produced by the A-splitter. Even if the coupling is unintentional, there is a transfer of information to system O which would make the theory we develop here incorrect. Such a possibility is excluded by this "classical" restriction on the A-splitter.

Here are three examples of such an experiment:

(i) Figure 1.8 showed an example for the measurement of photon polarization. The A -splitter is the polarizing beam-splitter cube, which is “classically describable”. A pure state would be any state of well-defined polarization, such as that produced by a polarizer.

(ii) We could modify the experiment of Fig. 1.14 so that we would send the particle itself through a lens to the detector, rather than bouncing photons from it. For an electron this would just be an electron microscope. In this case, the quantity measured is the position x , which is continuous rather than discrete. Experimentally, we’d measure position with a position-sensitive detector in place of a discrete array of detectors. A pure state would be any state described by a Schrödinger wavefunction $\psi(x)$. The A -splitter would be the lens. For an electron microscope, it would be a lens that acts on the electrons with a classically describable electric field.

(iii) The Stern-Gerlach experiment is of this kind. The quantity measured is the z -component of the atomic magnetic moment. The A -splitter here would be the Stern-Gerlach magnet, which acts on the atoms with a classically describable magnetic field gradient. A pure state could be prepared by putting another Stern-Gerlach magnet in front of the “measuring magnet”, possibly in a different orientation, and sending only one of its output beams into the measurement apparatus. Such a state-preparation device is the spin analog of an optical polarizer.

Following Dirac, we are now in a position to start building up the formalism. First, we use the symbol $|\psi\rangle$ to denote a pure state of the system. This symbol is called a *ket*. We postulate that there exist certain states $|a\rangle$, such that a measurement of A for a particle in that state is *certain* to yield the result a .

We shall call $|\psi\rangle$ the *state vector*, the special states $|a\rangle$ *eigenstates*, and the possible measurement results a *eigenvalues*, in anticipation of the mathematically equivalent concepts that we’ll come to in a bit.

For photon polarization, we could prepare an eigenstate by rotating the polarizer in Fig. 1.8 to produce either of the two polarizations \hat{x} or \hat{y} . For the microscope experiment, we could prepare an eigenstate by localizing the particle to an extremely narrow wavepacket, such that $|\psi(x)|^2 = \delta(x - x_0)$. We could prepare an eigenstate for the Stern-Gerlach experiment by orienting the “preparation” magnet to be either parallel or anti-parallel to the “measurement” magnet.

We assume that *an arbitrary state $|\psi\rangle$ may be expressed as a superposition of the eigenstates of A :*

$$|\psi\rangle = s_1 |a_1\rangle + s_2 |a_2\rangle + s_3 |a_3\rangle + \cdots \quad (1.59)$$

for an observable with discrete eigenvalues, or

$$|\psi\rangle = \int s(a) |a\rangle da \quad (1.60)$$

for an observable with continuous eigenvalues, where the s_i is a complex number, or $s(a)$ is a complex function. This is referred to as the *expansion postulate*.

The numbers s_i or $s(a)$ are *probability amplitudes*. As in our experimental examples, *the probability to obtain result a is proportional to the magnitude squared of the probability amplitude*. It follows that

$$P(a_i) = \frac{|s_i|^2}{\sum_j |s_j|^2} \text{ is the probability to obtain the measurement result } a_i, \text{ or} \quad (1.61)$$

$$P(a)da = \frac{|s(a)|^2 da}{\int |s(a')|^2 da'} \text{ is the probability for a measurement of } A \text{ to yield a} \quad (1.62)$$

result in an interval of width da centered at value a

for the discrete and continuous eigenvalue cases, respectively. The probability to obtain *some* result is obtained by summing over all possible measurement results,

$$\sum_i P(a_i) = \sum_i \frac{|s_i|^2}{\sum_j |s_j|^2} = \frac{\sum_i |s_i|^2}{\sum_j |s_j|^2} = 1 \quad (1.63)$$

$$\text{or } \int P(a)da = \frac{\int |s(a)|^2 da}{\int |s(a')|^2 da'} = 1 \quad (1.64)$$

for the discrete and continuous cases, respectively. That is, the particle must be detected *somewhere*, since we have assumed there is no loss of the particle until it hits one of the detectors.

Equations (1.59) through (1.62) are the logical extension of experiments such as examples (i) through (iii) above to the general case of Fig. 1.15. What eq. (1.59) means physically is that when a beam of particles in the state $|\psi\rangle$ is sent through the A -splitter, its “linear combination of pure-state beams”

$\sum_i s_i |a_i\rangle$ divides into a set of secondary beams, each with a definite property a_i and amplitude s_i . We cannot attribute a definite property a_i to any given particle in the beam prior to the splitter (except for the special case $|\psi\rangle = |a_i\rangle$); we can only specify the amplitude s_i for the particle to have that property. The probability that the particle is actually detected with that property (*i.e.* by the detector on the i th beam) is proportional to $|s_i|^2$. These properties of the theory coincide with what we see in the example experiments. Furthermore, this theory accurately describes what happens if we “rotate” the A -splitter.

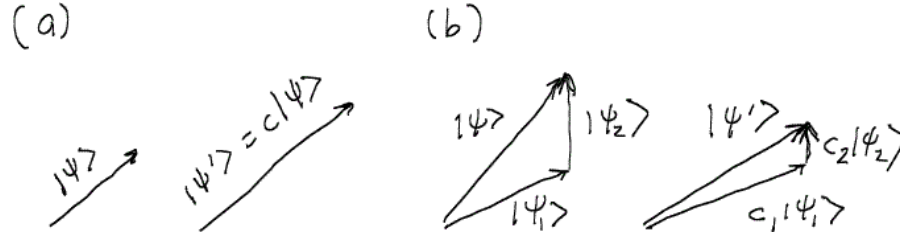


Fig. 1.16. (a) A state vector $|\psi\rangle$ and a state vector $|\psi'\rangle = c|\psi\rangle$, where c is an arbitrary complex constant, represent the *same* state. (b) A state vector $|\psi\rangle = |\psi_1\rangle + |\psi_2\rangle$ and $|\psi'\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle$, where c_1 and c_2 are arbitrary complex constants with $c_1 \neq c_2$, represent *different* states.

In quantum mechanics, the length and phase of a vector $|\psi\rangle$ have no physical significance. In other words, we can multiply $|\psi\rangle$ by any complex number $\alpha = |\alpha|e^{i\phi}$ to obtain a new vector

$$|\psi'\rangle = \alpha|\psi\rangle = s'_1|a_1\rangle + s'_2|a_2\rangle + s'_3|a_3\rangle + \dots, \text{ with } s'_1 = \alpha s_1, s'_2 = \alpha s_2, \dots \quad (1.65)$$

for the discrete case, and similarly for the continuous case. By doing this, the length of the vector is changed by the factor $|\alpha|$ and its phase by an amount ϕ , as illustrated in Fig. 1.16(a). However, $|\psi'\rangle$ and $|\psi\rangle$ represent the *same* physical state. That must be so, since the only way to distinguish states is by measurement, and the probability to obtain each result a_i is the same for the vector $|\psi'\rangle$ as it is for $|\psi\rangle$.

On the other hand, the *relative* lengths or phases of two vectors *do* have physical significance, as illustrated in Fig. 1.16(b). For, if a given vector $|\psi\rangle$ is expressed as a superposition of other vectors $|\psi_1\rangle, |\psi_2\rangle, \dots$ and if each vector is multiplied by $\alpha_i = |\alpha_i|e^{i\phi_i}$, then we have

$$|\psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle + c_3|\psi_3\rangle + \dots \Rightarrow |\psi'\rangle = \alpha_1 c_1|\psi_1\rangle + \alpha_2 c_2|\psi_2\rangle + \alpha_3 c_3|\psi_3\rangle + \dots \quad (1.66)$$

In this case, unless all constants α_i are the same, $|\psi'\rangle$ and $|\psi\rangle$ represent *different* states. For example, consider the polarization state $|\psi\rangle = |\hat{x}\rangle + |\hat{y}\rangle$ and $|\psi'\rangle = \alpha|\hat{x}\rangle + |\hat{y}\rangle$. The state $|\psi\rangle$ is a state of linear polarization at an angle of 45° to the x -axis, and the probability for a photon to be detected in the x -detector is $1/2$. If α is real and not equal to 1, the state $|\psi'\rangle$ is a state of linear polarization, but of some different orientation than that of $|\psi\rangle$, and the probability for a photon to be detected in the x -detector is not equal to $1/2$. If the phase ϕ is not equal to a multiple of π , the polarization state is elliptical. If $|\alpha| = 1$, this difference between $|\psi'\rangle$ and $|\psi\rangle$ would not be apparent in the experiment of Fig. 1.8, but it would be apparent if we put a waveplate in front of the polarizing beam-splitter cube of that experiment.

The *norm* of a vector is

$$\text{Norm}(|\psi\rangle) = \sum_i |s_i|^2 \text{ for the discrete case;} \quad (1.67)$$

$$\text{Norm}(|\psi\rangle) = \int |s(a)|^2 da \text{ for the continuous case} \quad (1.68)$$

Often, we shall find it convenient to use *normalized vectors*, defined by $\text{Norm}(|\psi\rangle) = 1$. In this case the measurement probability formulas (1.61) and (1.62) simplify to

$$P(a_i) = |s_i|^2 \text{ is the probability to obtain the measurement result } a_i, \text{ or} \quad (1.69)$$

$$P(a)da = |s(a)|^2 da \text{ is the probability for a measurement of } A \text{ to yield a} \quad (1.70)$$

result in an interval of width da centered at value a

for the discrete and continuous eigenvalue cases, respectively. To normalize any vector, just multiply it by the inverse of the square root of its norm. As discussed above, that new vector will represent exactly the same state as the un-normalized one, and then you can work with the simplified probability formulas (1.69) and (1.70) rather than (1.61) and (1.62). We will often normalize state vectors for this reason, but we won't avoid working with non-normalized vectors altogether.

2. Linear algebraic formulation for a discrete basis of states

a) *Bra space, inner product*

I will first develop Dirac's formulation of quantum mechanics for the case in which the possible measurement outcomes a form a discrete set $\{a_1, a_2, \dots\}$. The continuous eigenvalue case will be treated below. Let's put eq. (1.59) and the usual expression for a Cartesian coordinate vector next to each other:

$$|\psi\rangle = s_1 |a_1\rangle + s_2 |a_2\rangle + s_3 |a_3\rangle + \dots \quad (1.71)$$

$$\vec{s} = s_x \hat{x} + s_y \hat{y} + s_z \hat{z} \quad (1.72)$$

Looking at these two equations, we see the analogy between them. The individual kets $|a_i\rangle$ can be thought of as “basis vectors”, and the complex numbers s_i can be thought of as the “components” of the state vector $|\psi\rangle$ in the basis $\{|a_i\rangle\}$. The “vector” space of quantum mechanics is quite different than Cartesian vector space, though. First, it has one axis for each possible measurement outcome rather than three axes. Second, the “component” of the vector along each axis is a complex number rather than a real number. Finally, the “component” of the vector does not have the interpretation of a physical displacement. Rather it is a dimensionless quantity – *the probability amplitude for the system to be in a particular state*.

We shall introduce a second vector space called the *bra space*. Bra vectors will be written as $\langle\psi|$. If $|\psi\rangle$ represents a given state in the ket space, then $\langle\psi|$ represents the *same* state in the bra space. The purpose of the bra space is to allow us to define the *inner product* or *scalar product* as

$\langle a_i|\psi\rangle \equiv$ *probability amplitude* that a measurement of A

of a particle in the state $|\psi\rangle$ will yield the result a_i . (1.73)

and $\langle\psi|a_i\rangle \equiv \langle a_i|\psi\rangle^*$ (1.74)

Using eq. (1.59), we can just read off the inner product:

$\langle a_i|\psi\rangle = s_i$ (1.75)

Also, if the state $|\psi\rangle$ happens to be one of the states $|a_j\rangle$, then

$\langle a_i|a_j\rangle = \delta_{ij}$ (1.76)

That is because the probability amplitude is 0 if $i \neq j$. And if $i = j$, then $\langle a_i|a_j\rangle$ must be non-zero,

because $|\langle a_i|a_j\rangle|^2$ is proportional to the probability to detect the particle in the i th detector, which is

equal to 1. We shall *always* use the convention that the vectors $|a_i\rangle$ are normalized, which means that

$|\langle a_i|a_j\rangle|^2 = 1$ if $i = j$. Together with eq. (1.74), this means that $\langle a_i|a_i\rangle = 1$.

From eqs. (1.59) and (1.73) to (1.76) we deduce that

$\langle\psi| = s_1^* \langle a_1| + s_2^* \langle a_2| + s_3^* \langle a_3| + \dots$ (1.77)

Suppose we have the vector $|\psi\rangle$ (eq. (1.59)) and a second vector

$|\phi\rangle = p_1|a_1\rangle + p_2|a_2\rangle + p_3|a_3\rangle + \dots$ (1.78)

Their inner product is given by

$$\begin{aligned} \langle\psi|\phi\rangle &= (s_1^* \langle a_1| + s_2^* \langle a_2| + s_3^* \langle a_3| + \dots)(p_1|a_1\rangle + p_2|a_2\rangle + p_3|a_3\rangle + \dots) \\ &= \sum_i \sum_j s_i^* p_j \langle a_i|a_j\rangle = \sum_i \sum_j s_i^* p_j \delta_{ij} = \sum_i s_i^* p_i \end{aligned} \quad (1.79)$$

As shown in Fig. 1.17, the inner product plays a very similar role in quantum mechanics to the dot product's role in ordinary Cartesian vectors: it is the "projection" of one vector onto another. The two vector spaces we have introduced, the bra and ket spaces, in which each vector has finite length and finite scalar product with all other vectors, is called a *Hilbert space*. The theory of operations in Hilbert

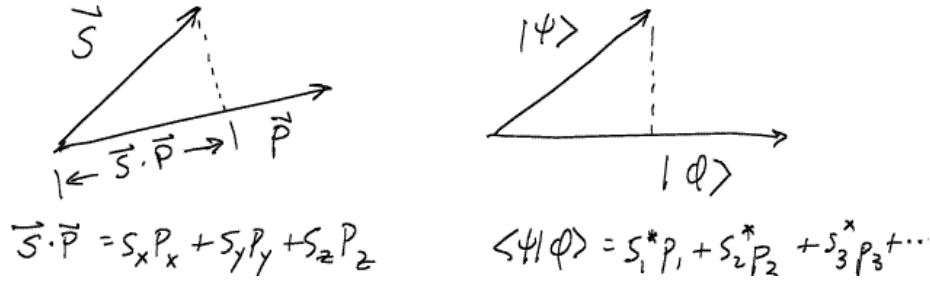


Fig. 1.17. (a) Dot product of two Cartesian vectors. (b) Inner product of a bra and ket vector.

space had already been worked out by mathematicians prior to the development of quantum mechanics. In physics, we also use the term *state space* for the space of all possible quantum states of a particle.

The inner product of a vector with itself is its norm:

$$\langle \psi | \psi \rangle = |s_1|^2 + |s_2|^2 + |s_3|^2 + \dots = \text{norm}(|\psi\rangle) \quad (1.80)$$

b) **Linear operators**

A linear operator O is a mapping of the ket space onto itself

$$O|\psi\rangle = |\gamma\rangle \quad (1.81)$$

with the property that if $O|\psi_1\rangle = |\gamma_1\rangle$ and $O|\psi_2\rangle = |\gamma_2\rangle$, then

$$O(c_1|\psi_1\rangle + c_2|\psi_2\rangle) = c_1|\gamma_1\rangle + c_2|\gamma_2\rangle \quad (1.82)$$

O associates with every possible ket $|\psi_i\rangle$ another ket $|\gamma_i\rangle$, and it does that linearly (*i.e.* it obeys eq. (1.82)).

c) **Projection operators; the operator “one”**

A *projection operator* is the operator

$$P_i = |a_i\rangle\langle a_i| \quad (1.83)$$

When applied to a state vector, the resulting vector is

$$P_i|\psi\rangle = |a_i\rangle\langle a_i|\psi\rangle = s_i|a_i\rangle \quad (1.84)$$

That is, P_i “projects out” that part of the vector $|\psi\rangle$ “pointing in the $|a_i\rangle$ -direction.” P_i is an operator because it associates a vector $P_i|\psi\rangle$ with every vector $|\psi\rangle$. An analogous operator in Cartesian vector

space would be $P_x = \hat{x}\hat{x}$. When applied to a vector \vec{S} , the result would be $P_x \vec{S} = \hat{x}\hat{x} \cdot \vec{S} = S_x \hat{x}$.

One of the most useful tricks in quantum mechanics is the operator “1”:

$$1 = \sum_i |a_i\rangle\langle a_i| \quad (1.85)$$

To see that this is indeed equal to 1:

$$1 \times |\psi\rangle = \sum_i |a_i\rangle\langle a_i|\psi\rangle = \sum_i s_i |a_i\rangle = |\psi\rangle \quad (1.86)$$

In geometrical terms, what we are doing is to take the inner product of $|\psi\rangle$ with each of the basis vectors, which yields that component of $|\psi\rangle$. Then we multiply back by each basis vector. It just amounts to expressing $|\psi\rangle$ in terms of the $\{|a_i\rangle\}$ basis vectors. It turns out that we can perform many operations in quantum mechanics just by inserting or taking out this operator.

Equation (1.85) says that the sum of all projection operators is equal to 1. Eq. (1.85) is called the *closure relation*, because it corresponds to the assumption that the eigenvectors $\{|a_i\rangle\}$ form a complete orthonormal set; the space is “closed” to include only those vectors that can be represented as linear superpositions of the eigenvectors $\{|a_i\rangle\}$.

d) **Matrix representations of bras, operators, and kets**

Once we have written bras and kets as vectors in a basis, we can express operators as matrices in a basis. For example, choosing the basis $\{|a_i\rangle\}$, we define a *matrix element of O* as

$$O_{ij} = \langle a_i | (O | a_j \rangle) \quad (1.87)$$

We can rewrite eq. (1.81) in this basis using the trick of the operator “1” three times:

$$\begin{aligned} 1 \times O \times 1 \times |\psi\rangle &= 1 \times |\gamma\rangle \\ \sum_k |a_k\rangle\langle a_k| O \sum_j |a_j\rangle\langle a_j|\psi\rangle &= \sum_\ell |a_\ell\rangle\langle a_\ell|\gamma\rangle \\ \sum_i \sum_j |a_i\rangle O_{ij} s_j &= \sum_\ell |a_\ell\rangle g_\ell \end{aligned} \quad (1.88)$$

where $|\gamma\rangle = \sum_i g_i |a_i\rangle$. Taking the inner product of this equation with $\langle a_i|$ gives

$$\sum_i \sum_j \delta_{ik} O_{kj} s_j = \sum_\ell \delta_{i\ell} g_\ell, \text{ or}$$

$$\sum_j O_{ij} s_j = g_i \quad (1.89)$$

which is just the usual equation for a vector as a product of a matrix times another vector. If we like we can write this out explicitly in matrix notation as

$$\begin{bmatrix} O_{11} & O_{12} & \cdots & O_{1n} \\ O_{21} & O_{22} & & O_{2n} \\ \vdots & & \ddots & \\ O_{n1} & O_{n2} & & O_{nn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \quad (1.90)$$

where we have assumed the number of possible measurement results for the operator A is n . Thus, we can express O a matrix, and the kets $|\psi\rangle$ and $|\gamma\rangle$ as column vectors.

The matrix representation of the bra vector $\langle\psi|$ is a *row matrix*:

$$\begin{bmatrix} s_1^* & s_2^* & \cdots & s_n^* \end{bmatrix} \quad (1.91)$$

The inner product can also be written as a matrix equation:

$$\langle\psi|\phi\rangle = \begin{bmatrix} s_1^* & s_2^* & \cdots & s_n^* \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = s_1^* p_1 + s_2^* p_2 + \cdots + s_n^* p_n \quad (1.92)$$

Always, the usual rule for multiplication of matrices – rows by columns – applies. Here is another example of a matrix expression for an object:

$$\begin{aligned} \langle\phi|(O|\psi\rangle) &= \begin{bmatrix} p_1^* & p_2^* & \cdots & p_n^* \end{bmatrix} \begin{bmatrix} O_{11} & O_{12} & \cdots & O_{1n} \\ O_{21} & O_{22} & & O_{2n} \\ \vdots & & \ddots & \\ O_{n1} & O_{n2} & & O_{nn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} p_1^* & p_2^* & \cdots & p_n^* \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \\ &= p_1^* g_1 + p_2^* g_2 + \cdots + p_n^* g_n \end{aligned} \quad (1.93)$$

The product of two operators OM is defined by the relation

$$(OM)|\psi\rangle \equiv O(M|\psi\rangle) \quad (1.94)$$

for all vectors $|\psi\rangle$. The components of the operator OM are give as

$$\langle a_i | OM | a_j \rangle = \langle a_i | O \times 1 \times M | a_j \rangle = \sum_k \langle a_i | O | a_k \rangle \langle a_k | M | a_j \rangle = \sum_k O_{ik} M_{kj} \quad (1.95)$$

This shows that the matrix for the product of two operators O and M is just the project of the matrix of O

with the matrix of M :

$$OM = \begin{bmatrix} O_{11} & O_{12} & \cdots & O_{1n} \\ O_{21} & O_{22} & & O_{2n} \\ \vdots & & \ddots & \\ O_{n1} & O_{n2} & & O_{nn} \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & & M_{2n} \\ \vdots & & \ddots & \\ M_{n1} & M_{n2} & & M_{nn} \end{bmatrix} \quad (1.96)$$

e) Action of an operator on a bra; Hermitian operators

We defined an operator O by its action on the ket vectors in eq.(1.81). We define the action of an operator O on a bra $\langle\phi|$ by setting

$$(\langle\phi|O)|\psi\rangle = \langle\phi|(O|\psi\rangle) \equiv \langle\phi|O|\psi\rangle \quad (1.97)$$

for all kets $|\psi\rangle$. With this definition, eq. (1.97) must be true for all basis kets $|\psi\rangle \rightarrow |a_i\rangle$, so that

$$(\langle\phi|O)|a_i\rangle \equiv \langle\phi|O|a_i\rangle \text{ is just the } i\text{th component of the bra } \langle\phi|O. \text{ It follows that the bra } \langle\phi|O \text{ is given by}$$

$$\langle\phi|O = \sum_i \langle\phi|O|a_i\rangle \langle a_i| \quad (1.98)$$

With this definition, we can dispense with the parentheses in any inner product expression $\langle\phi|O|\psi\rangle$, as indicated by eq. (1.97). O may either “operate to the right” or “operate to the left” – the two things are equivalent. The object $\langle\phi|O|\psi\rangle$ is sometimes also called a “matrix element” even if $|\psi\rangle$ and $|\phi\rangle$ are not both basis vectors. This is sloppy terminology, but I need to point it out because most people use it, and I will probably too, at times.

The bra $\langle\phi|O$ is *not* generally the bra that represents that same state as the ket $O|\phi\rangle$! If it were, we should have

$$\langle\phi|O|\psi\rangle = \langle\psi|O|\phi\rangle^* \quad (1.99)$$

but this is true only for certain operators O and not in the general case.

The *dual* of an object is the corresponding object in the opposite space. Thus, the dual of the ket $|\psi\rangle$ is the bra $\langle\psi|$. Similarly the dual of the bra $\langle\psi|$ is the ket $|\psi\rangle$. The dual of the ket $c_1|a_1\rangle + c_2|a_2\rangle$ is the bra $c_1^*\langle a_1| + c_2^*\langle a_2|$.

We have seen that the object $\langle\beta|O$ is *not* necessarily the dual of the object $O|\beta\rangle$. We define a new operator O^\dagger by the condition that

$$\langle \beta | O^\dagger = \text{dual of } O | \beta \rangle. \quad (1.100)$$

for every possible β . The operator O^\dagger is called the *Hermitian conjugate* or *adjoint* of O . By eq. (1.97), it follows that

$$\langle \delta | O | \beta \rangle = \langle \beta | O^\dagger | \delta \rangle^* \quad (1.101)$$

I will leave it as an exercise to show that when O is expressed in a basis, eq. (1.100) is equivalent to the definition that *the matrix of O^\dagger is the complex-conjugate transpose of the matrix of O* :

$$O_{ij}^\dagger = O_{ji}^* \quad (1.102)$$

$$\text{or } O^\dagger = \begin{bmatrix} O_{11}^* & O_{21}^* & \cdots & O_{n1}^* \\ O_{12}^* & O_{22}^* & & O_{n2}^* \\ \vdots & & \ddots & \vdots \\ O_{1n}^* & O_{2n}^* & \cdots & O_{nn}^* \end{bmatrix} \quad (1.103)$$

It is also not difficult to show that

$$(AB)^\dagger = B^\dagger A^\dagger \quad (1.104)$$

A *Hermitian operator* is one that is equal to its Hermitian conjugate:

$$O_{Herm} = O_{Herm}^\dagger \quad (1.105)$$

From eqs. (1.101) and (1.105), we see that only a Hermitian operator has the property of eq.(1.99); *i.e.*

$$\langle a_i | O_{Herm} | a_j \rangle = \langle a_j | O_{Herm} | a_i \rangle^* \quad (1.106)$$

The *adjoint* or *Hermitian conjugate* of an arbitrary object is obtained by

- (i) reversing the order of all factors
- (ii) replacing any number with its complex conjugate
- (iii) replacing any operator with its Hermitian conjugate
- (iv) replacing any bra with the corresponding ket, and
- (v) replacing any ket with the corresponding bra

If object α is the adjoint of object β , then object β is the adjoint of object α . In other words, adjoint objects always occur in pairs. Here are some examples of such adjoint pairs:

$$O \leftrightarrow O^\dagger \quad cA|\psi\rangle \leftrightarrow c^* \langle \psi | A^\dagger \quad A|\psi\rangle \langle \phi | B \leftrightarrow B^\dagger |\phi\rangle \langle \psi | A^\dagger \quad AB|\gamma\rangle \leftrightarrow \langle \gamma | B^\dagger A^\dagger$$

A complex number is a multiplicative factor that can occur anywhere in an expression. The same is true

for an inner product, since it is just a complex number. However, the order of all other objects – operators, stand-alone bras, and kets – must be strictly preserved. If the order of any of these factors is changed, the resulting object is different than the original one. For example, the operator product BA is not generally the same object as the operator product AB . $|\psi\rangle\langle\phi|$ is an operator, whereas $\langle\phi|\psi\rangle$ is a complex number; that is, reversing the order of factors in $|\psi\rangle\langle\phi|$ produces an object which is not even of the same kind as the original object.

We wrote the condition for Hermiticity of an operator as $O_{Herm} = O_{Herm}^\dagger$. This is equivalent to saying that a Hermitian operator is one that is *self-adjoint*, i.e. it is equal to its adjoint.

[Light grey shading of the following material means that you should have seen this material in your previous courses, but I won't assume that you know it very well. If we need this material for this course, I'll go back over it.]

f) *Change of basis, unitary operators*

Let me adopt the notation $\{|a, i\rangle\} = \{|a_i\rangle\}$ ($i = 1, 2, \dots, n$) for our “original” basis vectors. If we wish, we can always construct a new set of basis vectors $\{|b, i\rangle\}$ ($i = 1, 2, \dots, n$) from the set $\{|a, i\rangle\}$. a stands for “basis vector set a ”, b stands for “basis vector set b ”, and i labels which of the n vectors I'm talking about. For example, $|b, 3\rangle$ is the third basis vector of the new basis b .

To define the new basis vectors, we just need to specify their component in terms of the original basis vectors; *i.e.* we must specify all of the inner products: $\langle a, i | b, j \rangle$. Then

$$|b, j\rangle = 1 \times |b, j\rangle = \sum_k |a, k\rangle \langle a, k | b, j\rangle \quad (1.107)$$

The basis set $\{|a, i\rangle\}$ is an orthonormal set by definition (See eq. (1.76)). We will assume that the transformation coefficients $\langle a, i | b, j \rangle$ are always restricted so that the basis set $\{|b, i\rangle\}$ is also orthonormal; $\langle b, i | b, j \rangle = \delta_{ij}$. It follows that there is an operator $1 = \sum_j |b, j\rangle \langle b, j|$ in the new basis. It also follows that we can express any vector $|\psi\rangle$ as

$$|\psi\rangle = \sum_k s_k^{(a)} |a, k\rangle = \sum_k s_k^{(a)} \times 1 \times |a, k\rangle = \sum_k \sum_j s_k^{(a)} |b, j\rangle \langle b, j | a, k\rangle \quad (1.108)$$

But we can also expand $|\psi\rangle$ in the new basis, so that

$$|\psi\rangle = \sum_i s_i^{(b)} |b, i\rangle \quad (1.109)$$

In this section, we are adopting the *passive* view of a basis transformation. In this view $|\psi\rangle$ is always the *same* vector. The numbers $s_i^{(a)}$ are the expansion coefficients of $|\psi\rangle$ in the original $\{|a,i\rangle\}$ basis (identical to the s_i of the previous sections), and the numbers $s_i^{(b)}$ are the expansion coefficients of $|\psi\rangle$ in the new $\{|b,i\rangle\}$ basis. Taking the inner product of eq. (1.108) with the bra $\langle b,i|$, we obtain

$$s_i^{(b)} = \langle b,i|\psi\rangle = \sum_k s_k^{(a)} \langle b,i|a,k\rangle \quad (1.110)$$

Thus the relation between the components of $|\psi\rangle$ in the new and old basis is

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}^{(b)} = \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1n} \\ U_{21} & U_{22} & & U_{2n} \\ \vdots & & \ddots & \vdots \\ U_{n1} & U_{n2} & \cdots & U_{nn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}^{(a)} \quad (1.111)$$

$$\text{where } U_{ij} = \langle b,i|a,j\rangle \quad (1.112)$$

is the inner product of the i th b basis vector with j th a basis vector. The superscripts a and b specify that the two vectors are written in terms of the original (a) and new (b) basis sets, respectively. For example,

$$\begin{bmatrix} 1/\sqrt{3} \\ 0 \\ i/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix}^{(a)} = (s_1^{(a)}|a,1\rangle + s_2^{(a)}|a,2\rangle + s_3^{(a)}|a,3\rangle + s_4^{(a)}|a,4\rangle) = \frac{1}{\sqrt{3}}(|a,1\rangle + i|a,3\rangle - |a,4\rangle) \quad (1.113)$$

and similarly for a vector labeled with a superscript b .

We stated above that we assume that the coefficients $\langle b,i|a,k\rangle$ are restricted so as to make the new basis $\{|b,i\rangle\}$ orthonormal. We'll skip the proof, but it is not that difficult to show that this places the following restriction on the *transformation matrix* U :

$$U^\dagger U = 1 \quad (1.114)$$

The *inverse* of a matrix U^{-1} is that matrix which has the property

$$U^{-1}U = 1 \quad (1.115)$$

Therefore another way to write the condition of eq. (1.114) is

$$U^\dagger = U^{-1} \quad (1.116)$$

A matrix having the property (1.114) or (1.116) is said to be *unitary*. Transformations of the form of eq. (1.111) are called *unitary transformations*.

A Cartesian vector analog of a unitary transformation is shown in Fig. 1.18. We can express a vector \vec{S} in either the (a) or (b) coordinate system:

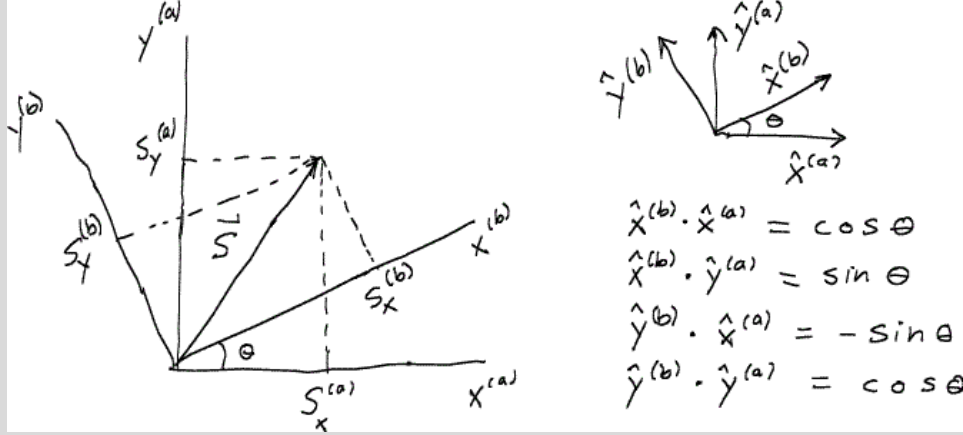


Fig. 1.18. Cartesian vector analog of a unitary transformation in the passive view.

$$\vec{S} = S_x^{(a)} \hat{x}^{(a)} + S_y^{(a)} \hat{y}^{(a)} = S_x^{(b)} \hat{x}^{(b)} + S_y^{(b)} \hat{y}^{(b)} \quad (1.117)$$

Again, since we are adopting the passive view, \vec{S} is the *same* vector; it is just expressed in two different coordinate systems. The components of the matrix of the transformation are given by the projections of the *b* basis vectors onto the *a* basis vectors:

$$U = \begin{bmatrix} \hat{x}^{(b)} \cdot \hat{x}^{(a)} & \hat{x}^{(b)} \cdot \hat{y}^{(a)} \\ \hat{y}^{(b)} \cdot \hat{x}^{(a)} & \hat{y}^{(b)} \cdot \hat{y}^{(a)} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \quad (1.118)$$

We can check that the transformation matrix is unitary:

$$\begin{aligned} U^\dagger U &= \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & -\cos(\theta)\sin(\theta) + \cos(\theta)\sin(\theta) \\ -\cos(\theta)\sin(\theta) + \cos(\theta)\sin(\theta) & \cos^2(\theta) + \sin^2(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \quad (1.119)$$

as required. Finally, we can determine the relation between the components of the vector \vec{S} in the two bases from

$$\begin{bmatrix} S_x^{(b)} \\ S_y^{(b)} \end{bmatrix} = \begin{bmatrix} S_x^{(a)} \\ S_y^{(a)} \end{bmatrix}^{(b)} = U \begin{bmatrix} S_x^{(a)} \\ S_y^{(a)} \end{bmatrix}^{(a)} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} S_x^{(a)} \\ S_y^{(a)} \end{bmatrix}^{(a)} = \begin{bmatrix} S_x^{(a)} \cos(\theta) + S_y^{(a)} \sin(\theta) \\ -S_x^{(a)} \sin(\theta) + S_y^{(a)} \cos(\theta) \end{bmatrix}^{(b)} \quad (1.120)$$

If we change the basis vectors, then the matrix expression for an operator also changes. The matrix elements of O in the *b* basis are

$$\begin{aligned}
O_{ij}^{(b)} &= \langle b, i | O | b, j \rangle = \sum_k \sum_\ell \langle b, i | a, k \rangle \langle a, k | O | a, \ell \rangle \langle a, \ell | b, j \rangle \\
&= \sum_k \sum_\ell U_{ik} O_{k\ell}^{(a)} U_{j\ell}^* = \sum_k \sum_\ell U_{ik} O_{k\ell}^{(a)} U_{\ell k}^\dagger
\end{aligned} \tag{1.121}$$

Equation (1.121) just gives the components of the equivalent matrix equation

$$O^{(b)} = U O^{(a)} U^\dagger \tag{1.122}$$

So, to find the matrix expression for O in the $\{|b, i\rangle\}$ basis, we just multiply the matrix expression for O in the $\{|a, i\rangle\}$ basis on the left by U and on the right by U^\dagger .

g) *Active view of a unitary transformation*

The *active* view of a unitary transformation is an alternative to the passive view that we discussed above. In the active view, we suppose that the basis vectors $\{|i\rangle \equiv |a, i\rangle\}$ are held fixed. In this basis, a given vector can be expressed as

$$|\psi^{(a)}\rangle = \sum_i s_i^{(a)} |a, i\rangle \tag{1.123}$$

If we wish, we can apply a unitary transformation to every vector:

$$|\psi^{(b)}\rangle = U |\psi^{(a)}\rangle \tag{1.124}$$

In the active view, $|\psi^{(b)}\rangle$ is a new vector, different from $|\psi^{(a)}\rangle$. U can be any unitary operator, subject only to the unitarity requirement $U^\dagger = U^{-1}$. A *unitary transformation* is the result of the application of the U to every vector.

Let's suppose that we have some other linear operator $O^{(a)}$. We know that $O^{(a)}$ is defined by a mapping of vectors

$$O^{(a)} |\psi^{(a)}\rangle = |\varphi^{(a)}\rangle \tag{1.125}$$

We define $O^{(b)}$ as the operator which has the *same* mapping of the unitarily transformed vectors $|\psi^{(b)}\rangle$:

$$O^{(b)} |\psi^{(b)}\rangle = |\varphi^{(b)}\rangle \tag{1.126}$$

It follows that

$$O^{(b)} U |\psi^{(a)}\rangle = O^{(b)} |\psi^{(b)}\rangle = |\varphi^{(b)}\rangle = U |\varphi^{(a)}\rangle = U O^{(a)} |\psi^{(a)}\rangle \tag{1.127}$$

$$\text{or } O^{(b)} U = U O^{(a)} \tag{1.128}$$

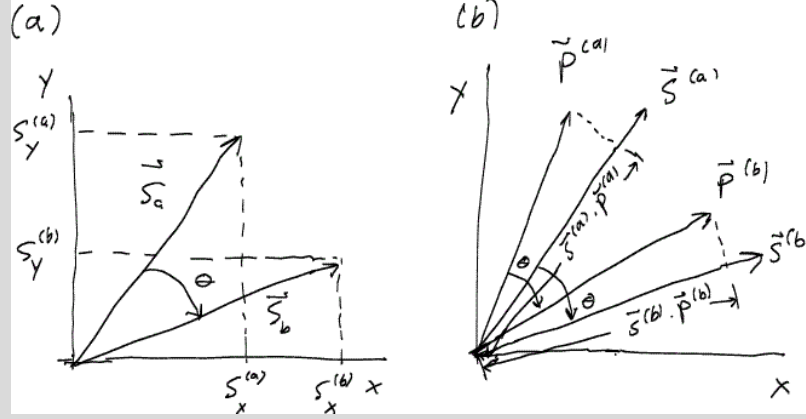


Fig. 1.19. (a) Cartesian vector analog of a unitary transformation in the active view. (b) Cartesian vector analog of the preservation of relationships between vectors by a unitary transformation.

Multiplying on the right by $U^\dagger = U^{-1}$, we obtain

$$O^{(b)} = U O^{(a)} U^\dagger \quad (1.129)$$

Thus, in the active view, we have exactly the same mathematical relationships between the transformed vectors and operators as in the passive view.

A unitary transformation preserves the relationships between all vectors. This is a trivial statement in the passive view, since the vectors are regarded as the same before and after the transformation. In the active view, the mathematically equivalent statement is to say that the projections of all vectors onto each other are unchanged, which we can easily show:

$$\langle \psi^{(b)} | \varphi^{(b)} \rangle = \langle \psi^{(a)} | U^\dagger U | \varphi^{(a)} \rangle = \langle \psi^{(a)} | 1 | \varphi^{(a)} \rangle = \langle \psi^{(a)} | \varphi^{(a)} \rangle \quad (1.130)$$

It may again be helpful to consider the Cartesian vector analog of the active unitary transformation, as shown in Fig. 1.19. In the passive view the vector \vec{S} remained the same, and the basis vectors rotated forward by angle θ . In the active view of the equivalent transformation (corresponding to the same matrix U), the basis vectors remain the same, and the vector $\vec{S}^{(a)}$ rotates backwards by an angle θ to produce the new vector $\vec{S}^{(b)}$. By inspection of Figs. 1.18 and 1.19, you can see that the relationships between the components of $\vec{S}^{(a)}$ and $\vec{S}^{(b)}$ are the same in the two pictures.

Fig. 1.19(b) illustrates the preservation of the relationship between vectors by the unitary transformation. Two vectors, $\vec{S}^{(a)}$ and $\vec{P}^{(a)}$ have a certain relative orientation before the transformation. The transformed vectors $\vec{S}^{(b)}$ and $\vec{P}^{(b)}$ have the same relative orientation, since they were both rotated through the same angle in vector space. One way we can see this is to note that \vec{S} has the same dot product with \vec{P} before and after the transformation, and that the same is true with every other pair of vectors. This is the Cartesian vector analogy of eq. (1.130).

h) Observables; eigenvalues and eigenvectors

Let's connect back to the point we made near the beginning of section B.1, that the special states $|a\rangle$ and the values a are equivalent to the mathematical concepts of an eigenvector and an eigenvalue. Recall that our goal is to describe measurements of the type illustrated in Fig. 1.15. The theory to do that has to account for the principle of superposition of probability amplitudes, and for the possibility of a “rotated” measurement apparatus. For example, our theory has to correctly account for the results of the photon polarization experiment of Fig. 1.8, regardless of whether we measure the probability for photons to be polarized along two perpendicular directions $\{\hat{x}^{(a)}, \hat{y}^{(a)}\}$, or along some other perpendicular directions $\{\hat{x}^{(b)}, \hat{y}^{(b)}\}$. The way we do this is to make the following association between mathematical objects and measurement results:

Each observable A of a system is associated with a corresponding Hermitian linear operator A .

The eigenvalues a_i of the operator A give the possible measurement results of observable A , and an eigenstate $|a_i\rangle$ is a state such that a measurement of A for a particle in that state is certain to give the result a_i .

To these two associations we add the *expansion hypothesis* and the probabilistic interpretation of the probability amplitude that we've already introduced, and which I'll repeat here for completeness:

An arbitrary quantum state $|\psi\rangle$ can be expressed as a linear superposition of eigenstates

$$|\psi\rangle = \sum_i s_i |a_i\rangle. \quad s_i \text{ is the probability amplitude to obtain the measurement result } a_i \text{ in a measurement}$$

of A . The probability to obtain the result a_i is $|s_i|^2$.

These statements constitute fundamental hypotheses of quantum mechanics, whose justification is the resulting theory gives results that agree with experiment. Let me partially justify them by further expanding on them and considering how they are consistent with the experimental results we've already looked at.

A Hermitian operator O has the following properties:

- (a) $O^\dagger = O$
- (b) All eigenvalues of O are real.
- (c) Two eigenvectors belonging to different eigenvalues are orthogonal.
- (d) The eigenvectors of O form a complete, orthogonal set of basis states. (Some clarification of this statement is needed if some eigenvectors are degenerate; we will make this clarification below.)

These statements can be proven mathematically, but we'll skip that and focus on a more physics-oriented discussion instead.

An eigenvector $|a_i\rangle$ of A with corresponding eigenvalues a_i satisfies the eigenvalue equation

$$A|a_i\rangle = a_i|a_i\rangle \quad (1.131)$$

Let's take a look at the matrix form for A in the basis of its eigenvectors. We have

$$A_{ij} = \langle a_i | A | a_j \rangle = a_j \langle a_i | a_j \rangle = a_j \delta_{ij} \quad (1.132)$$


Hence the matrix for the operator A is

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix} \quad (1.133)$$

The matrix form of eq. (1.131) is

$$A|a_i\rangle = \begin{bmatrix} a_1 & \cdots & 0 & \cdots & 0 \\ & \ddots & & & \\ 0 & & a_i & & 0 \\ & & & \ddots & \\ 0 & \cdots & 0 & \cdots & a_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_i \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_i |a_i\rangle \quad (1.134)$$

All $s_j = 0$, except for $s_i = 1$



The eigenvalue equation is mathematically trivial in this basis, because A is already diagonal. The eigenvector for the i th eigenvalue is just a vector with all components zero, except for the i th component equal to 1. The eigenvalues of A give the possible results of a measurement. Furthermore, since an eigenvector in this basis has non-zero probability amplitude s_i only for the corresponding eigenvector $|a_i\rangle$, that eigenvector represents a state that is certain to yield the corresponding eigenvalue a_i in a measurement of A . The eigenvalues must be real numbers because they represent a value for a physical quantity such as a component of a magnetic moment. A diagonal matrix with only real matrix elements is Hermitian, so A is a Hermitian operator.

Let me appeal back to Fig. 1.15, and describe in physics terms why the eigenstates are orthogonal. In the measurement apparatus, the pure state $|\psi\rangle$ time-evolves through the A -splitter which divides the $|\psi\rangle$ state beam into secondary beams. Since each of these secondary beams hits a particular detector, each secondary beam is by definition in the corresponding eigenstate $|a_i\rangle$. After the A -splitter these eigenstates are clearly orthogonal, because the beams are physically separated. If a particle is in the i th beam, its probability to be detected in that beam is 1, and its probability and therefore its probability

amplitude to be detected in any other beam is zero. You might be concerned that this orthogonality somehow arises in the A -splitter itself. But, particle motion through the A -splitter must itself obey the laws of quantum mechanics, and we have assumed that the A -splitter does not absorb any particles or make any measurement on its own, without the subsequent detectors. We will see later that the time evolution of $|\psi\rangle$ as it moves through the A -splitter is equivalent to a unitary transformation of $|\psi\rangle$ (active view).²² Unitary transformations preserve the orthonormality of any set of vectors. Therefore, since the eigenvectors $|a_i\rangle$ are orthogonal after the A -beamsplitter, they are orthogonal at all points in the experiment.

Let me also give a physics argument for completeness of the basis of states $|a_i\rangle$. (Mathematicians, please avert your eyes.) We supposed that the A -measuring apparatus outputs one secondary beam for every possible measurement result. The goal of the theory is to predict the probability amplitude for all possible measurement results. It follows that the set of eigenstates $\{|a_i\rangle\}$ is complete.

The above equations were all expressed in the basis $\{|a, i\rangle \equiv |a_i\rangle\}$. If we wish, we can write the eigenvalue equation in some other basis $\{|b, i\rangle\}$ related to $\{|a, i\rangle\}$ by a unitary transformation U . In this basis, the matrix expression $A^{(b)}$ for the eigenvalue equation will be less trivial than eq. (1.134), because $A^{(b)}$ will not be a diagonal matrix. However, we can calculate the eigenvectors and eigenvalues of $A^{(b)}$. The eigenvalues and eigenvectors of $A^{(b)}$ and $A^{(a)}$ will be identical, since they represent the *same* quantities, just expressed in a different basis. (We are adopting the passive view here.) The eigenvalues will be the same set of numbers in the two bases. The components of the eigenvectors in the b basis will be different than those in the a basis. However, if we wish, we can use these components to construct the unitary matrix $U^{-1} = U^\dagger$ that will transform all vectors and operators back to the a basis. When we do this, we will find that the matrix $A^{(b)}$ is re-expressed as the *diagonal matrix* $A^{(a)}$. This process is called *diagonalizing* A .

Since the eigenvalues and eigenvectors of A are the same in any basis related to the basis $\{|a, i\rangle \equiv |a_i\rangle\}$ by a unitary transformation, our interpretation of the eigenvectors and eigenstates of A holds in any orthonormal basis that spans the same space as the $\{|a_i\rangle\}$ basis. It also follows that any operator that represents a physical observable must be Hermitian, since we can always write down the corresponding eigenvalue equation in the basis of its eigenstates as in eq. (1.134).

²² Any absorption of particles inside the A -splitter would correspond to non-unitary evolution, since a unitary transformation conserves the norm of any vector. The lack of a coupling of the particle P to another system O inside the A -splitter is also crucial for this argument. If P did couple to a system O , only the *composite system* $(P + O)$ must undergo unitary evolution.

Often, we will not know the eigenvalues and eigenvectors of an observable in advance, but will know its matrix in some basis that spans the space of possible quantum states. I will comment more on how such matrices are determined below. In this case, we can find the possible measurement results and the corresponding eigenvectors by a matrix diagonalization. I'll illustrate that with a simple example. The spin angular momentum operator for a spin-1/2 particle is

$$\vec{S} = S_x \hat{x} + S_y \hat{y} + S_z \hat{z} = \frac{\hbar}{2} (\sigma_x \hat{x} + \sigma_y \hat{y} + \sigma_z \hat{z}) \quad (1.135)$$

where σ_x, σ_y , and σ_z are the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{(z)} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}^{(z)} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^{(z)} \quad (1.136)$$

These matrices were first determined based on physics considerations that we'll come to later; the point here is just that we have specific matrix expressions for our observables S_x, S_y , and S_z . The matrices are written in the basis of eigenstates of S_z , since the operator $S_z = \frac{\hbar}{2} \sigma_z$ is diagonal in this basis. Therefore, I've put the superscript z on the matrices to clarify that. In this basis, the eigenvectors and eigenvalues of S_z are

$$S_z |s_{z1}\rangle = +\frac{\hbar}{2} |s_{z1}\rangle \quad \text{with eigenvalue } s_{z1} = +\frac{\hbar}{2} \quad \text{and eigenvector } |s_{z1}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(z)} \quad (1.137)$$

$$S_z |s_{z2}\rangle = -\frac{\hbar}{2} |s_{z2}\rangle \quad \text{with eigenvalue } s_{z2} = -\frac{\hbar}{2} \quad \text{and eigenvector } |s_{z2}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}^{(z)} \quad (1.138)$$

Now, we have an operator $S_y = \frac{\hbar}{2} \sigma_y$ that is Hermitian, represents an observable, and whose matrix is non-diagonal. What are the possible measurement results (eigenvalues) for S_y , and what are the eigenvectors corresponding to those eigenvalues? We can answer these questions by diagonalizing S_y . To do so, we begin by writing down the eigenvalue equation for S_y

$$S_y |s_{yi}\rangle = s_{yi} |s_{yi}\rangle \quad (1.139)$$

$$\begin{pmatrix} 0 & \frac{-i\hbar}{2} \\ \frac{i\hbar}{2} & 0 \end{pmatrix}^{(z)} \begin{pmatrix} c_{1i} \\ c_{2i} \end{pmatrix}^{(z)} = s_{yi} \begin{pmatrix} c_{1i} \\ c_{2i} \end{pmatrix}^{(z)} \quad (1.140)$$

$$\begin{pmatrix} -s_{yi} & \frac{-i\hbar}{2} \\ \frac{i\hbar}{2} & -s_{yi} \end{pmatrix} \begin{pmatrix} c_{1i} \\ c_{2i} \end{pmatrix}^{(z)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^{(z)} \quad (1.141)$$

We expect two eigenvalues s_{yi} ($i=1,2$) since the matrix is 2×2 . Each corresponding eigenvector has two unknown amplitudes c_{1i} and c_{2i} in the original basis. (Note the switch in notation here: c is standing for the same thing that I've used the letter s for earlier, and I'm using the letter s to represent an eigenvalue, previously represented by the letter a .) The linear matrix eq. (1.141) has a solution only if the determinant of the matrix of coefficients is zero,

$$\det \begin{pmatrix} -s_{yi} & \frac{-i\hbar}{2} \\ \frac{i\hbar}{2} & -s_{yi} \end{pmatrix} = s_{yi}^2 - \left(-\left(\frac{i\hbar}{2} \right)^2 \right) = s_{yi}^2 - \left(\frac{\hbar}{2} \right)^2 = 0 \quad (1.142)$$

There are two solutions to this equation, $s_{y1} = +\frac{\hbar}{2}$, and $s_{y2} = -\frac{\hbar}{2}$. Therefore, the possible results for a measurement of S_y are the same as those for a measurement of S_z . That shouldn't be surprising because the choice of a z -axis in physical space is arbitrary. We could have chosen any axis as the z -axis. So it makes sense that the possible measurement results for *any* component of \vec{S} are $+\frac{\hbar}{2}$ and $-\frac{\hbar}{2}$.

To find the eigenvectors that go with the eigenvalue s_{y1} , we just substitute the value of s_{y1} back into eq. (1.141). We get

$$\frac{\hbar}{2} \begin{pmatrix} -1 & -i \\ i & -1 \end{pmatrix} \begin{pmatrix} c_{11} \\ c_{21} \end{pmatrix}^{(z)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^{(z)} \quad (1.143)$$

The first row of this matrix equation reads

$$\frac{\hbar}{2} (-c_{11} - ic_{21}) = 0 \quad (1.144)$$

The second row gives the same equation. Eq. (1.144) gives the relationship between c_{11} and c_{21} that must be satisfied for the eigenvector $|s_{y1}\rangle$. We therefore cannot determine both c_{11} and c_{21} uniquely. But that is not a problem, because an eigenvector always has one unconstrained complex variable – an arbitrary complex number that could multiply that vector, exactly as we discussed in eq. (1.65). Therefore we only need to get the *relative amplitude and phase* of c_{11} and c_{21} correct to specify the eigenvector. We could do this, for example, if we take $c_{11} = 1$ and $c_{21} = i$. However, the resulting

eigenvector is not normalized, and we will usually normalize our vectors. We can do this by choosing instead $c_{11} = 1/\sqrt{2}$ and $c_{21} = i/\sqrt{2}$. Thus, the eigenvector that is the solution to the eigenvalue equation for eigenvalue s_{y1} is

$$|s_{y1}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}^{(z)} \quad (1.145)$$

If we repeat this process for the second eigenvector, we obtain

$$|s_{y2}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}^{(z)} \quad (1.146)$$

We can check that the eigenvectors are orthogonal

$$\langle s_{y1} | s_{y2} \rangle = \frac{1}{2} {}^{(z)} \begin{pmatrix} 1 & -i \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix}^{(z)} = \frac{1}{2} (1 + i^2) = 0 \quad (1.147)$$

and they are, as required since they are eigenvectors of a Hermitian operator belonging to different eigenvalues.

We can now switch basis if we prefer, to the eigenvectors of S_y . We could write down the matrix form for S_y in this basis immediately – it's just the diagonal matrix with the two eigenvalues on the diagonal. However, for purposes of illustration, I'll work out the unitary transformation matrix U and use that instead. U is given by

$$U_{ij} = \langle s_{yi} | s_{zj} \rangle \quad (1.148)$$

$$\text{For example, } U_{12} = \langle s_{y1} | s_{z2} \rangle = \frac{1}{\sqrt{2}} {}^{(z)} \begin{pmatrix} 1 & -i \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}^{(z)} = \frac{-i}{\sqrt{2}} \quad (1.149)$$

Reading off the remaining three coefficients gives

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \quad (1.150)$$

The matrix expressions for $|s_{y1}\rangle$ and $|s_{y2}\rangle$ in the new basis are

$$|s_{y1}\rangle = U \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix}^{(z)} = \frac{1}{2} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix}^{(z)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{(y)} \quad (1.151)$$

$$|s_{y2}\rangle = U \begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix}^{(z)} = \frac{1}{2} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix}^{(z)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}^{(y)} \quad (1.152)$$

And the matrix expression for S_y in the new basis is

$$S_y^{(y)} = U S_y^{(z)} U^\dagger = \frac{1}{2} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}^{(z)} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = \frac{\hbar}{4} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} 1 & -1 \\ i & i \end{pmatrix} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1.153)$$

Again, it's not surprising that everything looks the same when we choose the basis of eigenstates of S_y rather than S_z .

Not all matrices can be diagonalized. In order to be diagonalizable, the eigenvectors of a matrix must form a complete, orthogonal set. Such a matrix is called a *normal matrix*. It is possible to prove that a matrix N is normal if and only if $N^\dagger N = N N^\dagger$. Both Hermitian matrices and unitary matrices are normal, and therefore can be diagonalized. We will mostly be concerned with these two types of matrices in quantum mechanics, so most matrices we encounter will be diagonalizable. However, we occasionally work with non-Hermitian, non-unitary matrices. In case the question arises, those matrices are diagonalizable if they are normal.

Let me rewrite the eigenvalue equation for a general operator O :

$$O|o_{Ri}\rangle = o_{Ri}|o_{Ri}\rangle \quad (1.154)$$

This equation is also called the *right eigenvalue equation*, and the eigenvalues are also called *right eigenvalues*, and the eigenvectors *eigenkets*. We can also write down a *left eigenvalue equation*

$$\langle o_{Li}|O = \langle o_{Li}|o_{Li} \quad (1.155)$$

In the general case, the set of right eigenvalues $\{o_{Ri}\}$ is identical to the set of left eigenvalues $\{o_{Ri}\}$, so that

$$O|o_{Ri}\rangle = o_i|o_{Ri}\rangle \quad \text{and} \quad \langle o_{Li}|O = \langle o_{Li}|o_i \quad (1.156)$$

but the left eigenvector $\langle o_{Li}|$ is *not* generally the dual of the right eigenvector $|o_{Ri}\rangle$. Therefore, you should not assume that $O|o_i\rangle = o_i|o_i\rangle$ necessarily implies $\langle o_i|O = \langle o_i|o_i$. The correct correspondence is

$$O|o_i\rangle = o_i|o_i\rangle \Leftrightarrow \langle o_i|O^\dagger = \langle o_i|o_i^* \quad (1.157)$$

This follows from the fact that the second equation is just the Hermitian conjugate of the first. However, for normal operators it is true that the left and right eigenvectors are the dual of each other. So in that case it is true that

$$O_{normal}|o_i\rangle = o_i|o_i\rangle \Leftrightarrow \langle o_i|O_{normal} = \langle o_i|o_i \quad (1.158)$$

This would be true, for example, for Hermitian operators. This can also be seen from eq. (1.157), taking into account that fact that for a Hermitian operator $O = O^\dagger$ and the eigenvalues are real.

i) *Expectation value*

For a normalized vector, the magnitude squared $|s_i|^2$ of the probability amplitude gives the probability to find result a_i in a measurement of A . This provides complete information about the distribution of measurement outcomes of A , and is all the information that quantum mechanics has to give about those outcomes. However, sometimes we will only want the average value of measurements of A . This is called the *expectation value* of A . It is given by

$$\langle A \rangle \equiv \frac{\langle \psi | A | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\sum_i |s_i|^2 a_i}{\langle \psi | \psi \rangle} \quad (1.159)$$

In other words, it is the average over all possible measurement results a_i weighted by their probability of occurrence.

j) *Infinite-dimensional basis sets*

In most cases of practical interest, the number n of possible measurement results is actually infinite, even if the possible measurement results are discrete as shown in Fig. 1.15. We have sidestepped this issue so far, either by assuming n to be finite or ignoring the issue altogether. To have a mathematically consistent theory that accounts for this possibility, we need a theory with an infinite-dimensional basis set, in which operators given by $\infty \times \infty$ dimensional matrices. This raises issues that are of interest to mathematicians but of no interest to us. Any problem of practical interest will have the overwhelming share of its norm concentrated in a finite number of basis states. Therefore we may truncate the state space to have a finite number of dimensions, and calculate everything with that to as good an approximation as we like. A larger number of included states in the truncated basis will give more accuracy.

3. Continuous basis sets

a) *Illustration with the position basis $|x\rangle$*

Here, we consider the case in which the possible results of a measurement form a continuous set. Since the position x is the most common example, we'll use the symbol x for the measurement variable; just keep in mind that this discussion is general and could apply to any observable with continuous eigenvalues. We shall symbolize the corresponding position operator by X .

For continuous eigenvalues, a number of mathematical issues arise. These issues were not fully resolved until the 1930s, mostly by John Von Neumann. Again, we can ignore all of these issues in

practice. For the most part, we can just use continuous basis sets in ways which are known to be OK based on a more rigorous analysis. Alternatively, we can approximate a continuous basis set by a set of basis vectors with a very small but finite spacing. The corresponding theory is identical to the finite-dimensional theory we discussed above. The smaller the spacing (and the greater the number of basis vectors), the more accurate the calculation. In fact, a good way to think of the continuous eigenvalue case is the continuous limit of the closely spaced, discrete eigenvalue case.

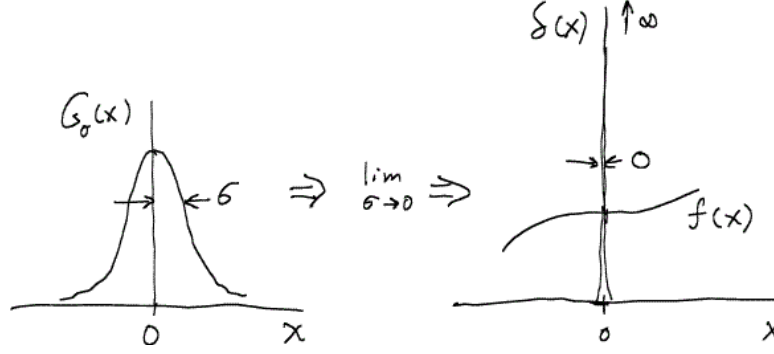


Fig. 1.20. Dirac δ -function as the zero-width limit of a function $G_\sigma(x)$.

In the continuous case, we have basis kets $|x\rangle$ and bras $\langle x|$. The operator 1, or closure relation, is given by:

$$1 = \int dx |x\rangle \langle x| \quad (1.160)$$

In the continuous limit, the sum over finitely spaced eigenvalues is replaced by an integral over the continuously space ones. The orthormality condition becomes

$$\langle x' | x \rangle = \delta(x' - x) \quad (1.161)$$

where $\delta(x)$ is the *Dirac delta function*. That function is defined by

$$\delta(x) = \lim_{\sigma \rightarrow 0} G_\sigma(x) \quad (1.162)$$

where $G_\sigma(x)$ is pretty much any reasonably behaved, normalized function of width σ that is peaked at $x = 0$ and symmetric about $x = 0$. Part of “reasonably behaved” is that $G_\sigma(x)$ should fall reasonably quickly with distance from the origin. ($|G_\sigma(x)| \sim 1/x^n$, with $n \geq 2$, is certainly sufficient.) Normalized

means $\int_{-\infty}^{\infty} G_\sigma(x) dx = 1$. One suitable choice for $G_\sigma(x)$ is a normalized Gaussian, $G_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$. We

can understand what the δ -function is graphically by just looking at the limiting form of the function, as in Fig. 1.20.

Important properties of the δ -function include

$$\int_a^b \delta(x - x_0) dx = \begin{cases} 1, & \text{if } a < x_0 < b \\ 0, & \text{otherwise} \end{cases} \quad (1.163)$$

$$\int_a^b f(x) \delta(x - x_0) dx = \begin{cases} f(x_0), & \text{if } a < x_0 < b \\ 0, & \text{otherwise} \end{cases} \quad (1.164)$$

$$\delta(ax) = \frac{1}{|a|} \delta(x) \quad (1.165)$$

Properties (1.163) and (1.165) can be derived from the definition (1.162). Property (1.164) follows from an argument illustrated in Fig. 1.20. If we integrate $f(x)\delta(x)$ over x , the integrand is non-zero over an infinitesimal width interval centered at $x = 0$ due to the δ -function term in the product. Only values of x in that interval contribute to the integral. Since $f(x)$ is a smooth function, it will be constant over that interval, with value $f(0)$, and so can be removed from the integrand. The integral over the remaining delta function just gives a factor of 1. This argument, with the peak of the δ -function shifted to $x = x_0$, gives eq. (1.164).

An arbitrary quantum state $|\psi\rangle$ can be expanded in the position basis using the operator 1

$$|\psi\rangle = 1 \times |\psi\rangle = \int dx |x\rangle \langle x|\psi\rangle \quad (1.166)$$

The inner product $\langle x|\psi\rangle$ is a complex number. Actually, because x is a continuous variable, $\langle x|\psi\rangle$ is a complex function of x . Referring back to eq. (1.73), $\langle x|\psi\rangle$ the *probability amplitude for a particle in state $|\psi\rangle$ to be measured to have value x in a measurement of X* . $|\langle x|\psi\rangle|^2$ is proportional to the probability to find the particle at position x . In the continuous limit, this is an infinitesimal number, since the number of possible positions is infinite. Considering the continuous limit of the discrete case, we have

$$|\langle x|\psi\rangle|^2 dx = \text{the probability to find the particle in an interval of width } dx \text{ centered at } x. \quad (1.167)$$

Here, we are considering the state vector to be normalized, which in the continuous limit means

$$\int |\langle x|\psi\rangle|^2 dx = 1 \quad (1.168)$$

But, looking at the above paragraph and equations, you should notice that you have already seen the object $\langle x|\psi\rangle$ many times before. It is

$$\langle x|\psi\rangle = \psi(x) = \text{the wavefunction of the particle.} \quad (1.169)$$

The Hermitian conjugate of this equation is

$$\langle \psi | x \rangle = \psi^*(x) \quad (1.170)$$

The matrix element of an operator O is

$$\langle x | O | x' \rangle = O(x, x') \quad (1.171)$$

In the general case, O is a complex function of the two variables x' and x , because the matrix element is a complex number, and x' and x are continuous variables. An operator which is expressed as $O(x, x')$ is said to be in *kernel form*.

Let's look at some important special cases. The kernel form for the operator X is

$$\langle x | X | x' \rangle = \langle x | x' | x' \rangle = x' \langle x | x' \rangle = x' \delta(x - x') \quad (1.172)$$

If an operator V that is a function of X can be written as a Taylor series expansion in powers of X , then

$$V(X)|x\rangle = (c_0 + c_1 X + c_2 X^2 + \dots)|x\rangle = (c_0 + c_1 x + c_2 x^2 + \dots)|x\rangle = V(x)|x\rangle \quad (1.173)$$

since $X^n |x\rangle = X^{n-1} X |x\rangle = X^{n-1} x |x\rangle = x X^{n-2} X |x\rangle = \dots = x^n |x\rangle$. The kernel form of $V(X)$ is

$$\langle x | V(X) | x' \rangle = \langle x | V(x') | x' \rangle = V(x') \langle x | x' \rangle = V(x') \delta(x - x') \quad (1.174)$$

We can also evaluate “matrix elements” of the form

$$\begin{aligned} \langle \psi | V(X) | \phi \rangle &= \langle \psi | \times 1 \times V(X) \times 1 \times | \phi \rangle = \int dx \int dx' \langle \psi | x \rangle \langle x | V(X) | x' \rangle \langle x' | \phi \rangle \\ &= \int dx \int dx' \psi^*(x) V(x) \delta(x' - x) \phi(x') = \int dx \psi^*(x) V(x) \phi(x) \end{aligned} \quad (1.175)$$

b) *The momentum basis*

I would like to introduce an operator D , which I can do by giving it in kernel form

$$\langle x | D | x' \rangle = \delta'(x - x') \quad (1.176)$$

In this equation δ' is the *derivative of the δ -function*, defined by

$$\delta'(x) = \lim_{\sigma \rightarrow 0} \frac{d}{dx} G_\sigma(x) \quad (1.177)$$

Given any smooth function $f(x)$, the most important properties of δ' are

$$\int_a^b \delta'(x - x') f(x') dx' = \begin{cases} f'(x) \equiv -\frac{df(x')}{dx'} \Big|_{x'=x}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases} \quad (1.178)$$

$$\delta'(-x) = -\delta'(x) \quad (1.179)$$

For example, we can evaluate the following matrix element

$$\langle x|D|\psi\rangle = \langle x|D \times 1 \times |\psi\rangle = \int dx' \langle x|D|x'\rangle \langle x'|\psi\rangle = \int dx' \delta'(x-x') \psi(x') = \psi'(x) \quad (1.180)$$

Let me introduce a new continuous basis $\{|k\rangle\}$, by giving all of its components in the basis $\{|x\rangle\}$.

$$\langle x|k\rangle = \frac{1}{\sqrt{2\pi}} e^{ikx} \quad (1.181)$$

The Hermitian conjugate of this equation gives the components of the vector $|x\rangle$ in the basis $\{|k\rangle\}$.

$$\langle k|x\rangle = \frac{1}{\sqrt{2\pi}} e^{-ikx} \quad (1.182)$$

Let's check that the new basis is orthonormal

$$\langle k'|k\rangle = \langle k'| \times 1 \times |k\rangle = \int_{-\infty}^{\infty} dx \langle k'|x\rangle \langle x|k\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx e^{-ik'x} e^{ikx} = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx e^{i(k-k')x} \quad (1.183)$$

The integral on the right does not converge (one of those mathematical issues), but we can define an integral $G_{\sigma}(\Delta k)$ that is close to it as

$$\begin{aligned} \langle k'|k\rangle &\approx G_{\sigma}(\Delta k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx e^{(i\Delta kx - \sigma|x|)} = \frac{1}{2\pi} \left(\int_{-\infty}^0 dx e^{(i\Delta kx + \sigma x)} + \int_0^{\infty} dx e^{(i\Delta kx - \sigma x)} \right) \\ &= \frac{1}{2\pi} \left(\frac{1}{i\Delta k + \sigma} - \frac{1}{i\Delta k - \sigma} \right) = \frac{1}{\pi\epsilon} \left(\frac{\sigma^2}{\sigma^2 + (\Delta k)^2} \right) \end{aligned} \quad (1.184)$$

where $\Delta k = k' - k$. In fact, the two integrals are identical in the limit as $\sigma \rightarrow 0$. But, $G_{\sigma}(\Delta k)$ meets all the requirements of a function whose $\sigma \rightarrow 0$ limit is a δ -function. (It is a normalized Lorentzian function rather than a normalized Gaussian function.) Since that is the limit we require, we see that

$$\langle k'|k\rangle = \lim_{\sigma \rightarrow 0} G_{\sigma}(\Delta k) = \delta(k' - k) \quad (1.185)$$

so the basis $\{|k\rangle\}$ is orthonormal.

The wavefunction of the state $|k\rangle$ is $\psi_k(x) = \langle x|k\rangle = \frac{1}{\sqrt{2\pi}} e^{ikx}$. We recognize this as an *infinite plane wave state*. Such a state has an exactly defined wavevector k , and we know from the fundamental relation $p = \hbar k$ that this must be a state of exactly defined momentum *i.e.* an eigenstate of momentum. Such a momentum eigenstate is written as $|p\rangle$, and we set

$$|p\rangle = \frac{1}{\sqrt{\hbar}} |k\rangle \quad (1.186)$$

The factor $1/\sqrt{\hbar}$ is required to make $\{|p\rangle\}$ an orthonormal basis set, *i.e.* so that

$$\langle p' | p \rangle = \frac{1}{\hbar} \langle k' | k \rangle = \frac{1}{\hbar} \delta(k' - k) = \delta(\hbar(k' - k)) = \delta(p' - p) \quad (1.187)$$

where we have used the property (1.165) of the δ -function.

The operator P , whose eigenstates are $|p\rangle$, is

$$P = -i\hbar D \quad (1.188)$$

We can check this directly

$$\begin{aligned} P|p\rangle &= 1 \times P \times 1 \times |p\rangle = \int dx \int dx' |x\rangle \langle x| P |x'\rangle \langle x'| p\rangle = -i\hbar \int dx \int dx' |x\rangle \langle x| D |x'\rangle \frac{1}{\sqrt{\hbar}} \langle x'| k\rangle \\ &= -i\hbar \int dx \int dx' |x\rangle \delta'(x - x') \frac{1}{\sqrt{\hbar}} e^{ikx'} = -i\hbar \int dx |x\rangle \frac{d}{dx} \left(\frac{1}{\sqrt{\hbar}} e^{ikx} \right) \\ &= -i\hbar(ik) \int dx |x\rangle \left(\frac{1}{\sqrt{\hbar}} e^{ikx} \right) = \hbar k \int dx |x\rangle \langle x| p\rangle = \hbar k \times 1 \times |p\rangle \end{aligned} \quad (1.189)$$

That is,

$$P|p\rangle = \hbar k |p\rangle = p |p\rangle \quad (1.190)$$

The *momentum space wavefunction* of the particle in a state $|\psi\rangle$ is its probability amplitude to be in a state of momentum p , *i.e.*

$$\varphi(p) = \langle p | \psi \rangle \quad (1.191)$$

Alternatively, we can express the momentum space wavefunction as

$$\varphi_k(k) = \langle k | \psi \rangle = \sqrt{\hbar} \langle p | \psi \rangle = \sqrt{\hbar} \varphi(p) \quad (1.192)$$

If we know the position-space wavefunction $\psi(x)$ of a particle, we can use that to find $\varphi_k(k)$:

$$\varphi_k(k) = \langle k | \times 1 \times | \psi \rangle = \int dx \langle k | x \rangle \langle x | \psi \rangle = \frac{1}{\sqrt{2\pi}} \int dx e^{-ikx} \psi(x) \quad (1.193)$$

A similar calculation allows us to find $\psi(x)$ if we know $\varphi_k(k)$. Summarizing, we find that the two wavefunctions are related by

$$\varphi_k(k) = \frac{1}{\sqrt{2\pi}} \int e^{-ikx} \psi(x) dx \quad (1.194)$$

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int e^{ikx} \phi_k(k) dk \quad (1.195)$$

Equations (1.194) and (1.195) just say that $\phi_k(k)$ and $\psi(x)$ are *Fourier transforms* of each other. If we choose p rather than k as the momentum variable, these equations read

$$\phi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{-ipx/\hbar} \psi(x) dx \quad (1.196)$$

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{ipx/\hbar} \phi(p) dp \quad (1.197)$$

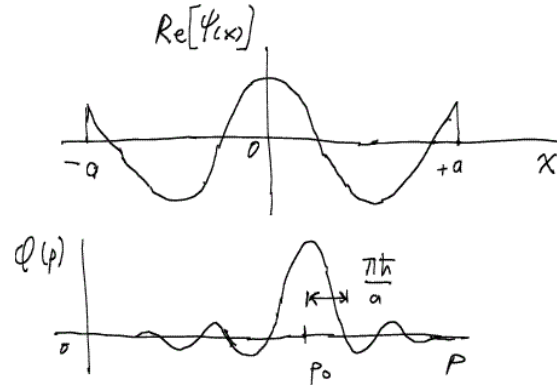
The normalization of the wavefunctions obeys

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx = \int_{-\infty}^{\infty} |\phi_k(k)|^2 dk = \int_{-\infty}^{\infty} |\phi(p)|^2 dp \quad (1.198)$$

so if one of the wavefunctions is normalized they all are. This must be the case, since all three expressions are just the norm of $|\psi\rangle$, expressed in different bases.

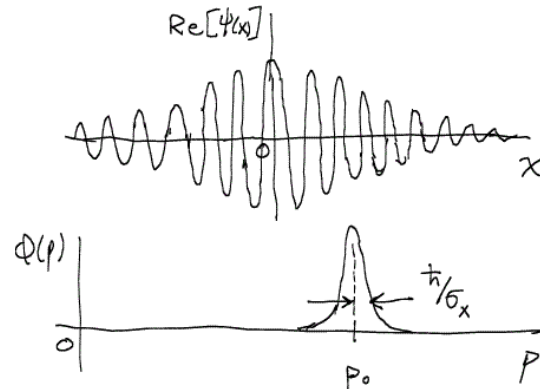
Here are a few examples of such normalized wavefunction pairs:

$$\psi(x) = \begin{cases} \frac{1}{\sqrt{2a}} e^{ip_0 x/\hbar}, & \text{if } |x| > a \\ 0, & \text{otherwise} \end{cases}$$



$$\phi(p) = \sqrt{\frac{1}{\pi a \hbar}} \frac{\sin[(p - p_0)a/\hbar]}{(p - p_0)/\hbar}$$

$$\psi(x) = \frac{1}{(2\pi\sigma_x^2)^{1/4}} e^{ip_0 x/\hbar} \exp\left(\frac{-x^2}{4\sigma_x^2}\right)$$

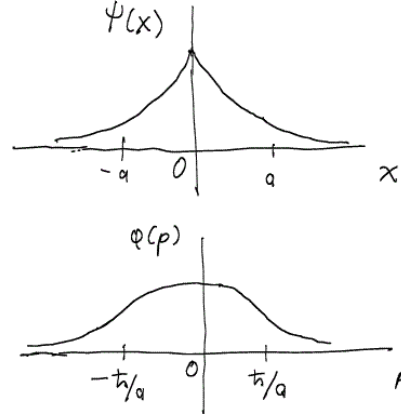


$$\phi(p) = \left(\frac{4\sigma_x^2}{2\pi\hbar^2}\right)^{1/4} e^{-\sigma_x^2(p-p_0)^2/\hbar^2}$$

(minimum uncertainty wavepacket)

$$\psi(x) = \frac{1}{\sqrt{a}} e^{-|x|/a}$$

$$\phi(p) = \sqrt{\frac{2\hbar}{\pi a}} \frac{(\hbar/a)}{p^2 + (\hbar/a)^2}$$



4. Time evolution of the state vector

a) Schrödinger equation

So far, everything we have done must be regarded as the expressions for state vectors and operators *at some given time*. The time evolution of a state vector is given by the Schrödinger equation

$$H|\Psi(t)\rangle = i\hbar \frac{d}{dt} |\Psi(t)\rangle \quad (1.199)$$

where H is the *Hamiltonian operator*, which is the operator for the total energy. This is a generalized version of Schrödinger's wave equation, because it applies to all situations, not just to a particle moving in a potential. This statement must be regarded as another postulate, justified because it gives agreement with experiment. It reduces to Schrödinger's wave equation for a particle moving in a potential, as we'll now show.

The Hamiltonian for a particle of mass m moving in a one dimensional potential $V(x)$ is

$$H = \frac{P^2}{2m} + V(X) \quad (1.200)$$

where the two terms represent the kinetic and potential energy of the particle. Using the operator 1 in the position basis three times, we find that

$$\begin{aligned} P^2 |\Psi(t)\rangle &= \int dx' |x'\rangle \int dx'' \langle x' | P | x'' \rangle \int dx''' \langle x'' | P | x''' \rangle \langle x''' | \Psi(t) \rangle \\ &= \int dx |x'\rangle \int dx'' (-i\hbar \langle x' | D | x'' \rangle) \int dx''' (-i\hbar \langle x'' | D | x''' \rangle) \langle x''' | \Psi(t) \rangle \\ &= -\hbar^2 \int dx |x'\rangle \int dx'' \delta'(x' - x'') \int dx''' \delta'(x'' - x''') \Psi(x''', t) \\ &= -\hbar^2 \int dx |x'\rangle \int dx'' \delta'(x' - x'') \frac{d}{dx''} \Psi(x'', t) = -\hbar^2 \int dx' |x'\rangle \frac{d^2}{dx'^2} \Psi(x', t) \end{aligned} \quad (1.201)$$

Also

$$\begin{aligned}
V(X)|\Psi(t)\rangle &= \int dx |x'\rangle \int dx'' \langle x'|V(X)|x''\rangle \langle x''|\Psi(t)\rangle \\
&= \int dx |x'\rangle \int dx'' V(x'') \delta(x' - x'') \Psi(x'', t) = \int dx' |x'\rangle V(x') \Psi(x', t)
\end{aligned} \tag{1.202}$$

The Schrödinger equation for this case becomes

$$H|\Psi(t)\rangle = \left(\frac{P^2}{2m} + V(X) \right) |\Psi(t)\rangle = \int dx' |x'\rangle \left(\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx'^2} + V(x') \right) \Psi(x', t) \right) = i\hbar \frac{d}{dt} |\Psi(t)\rangle \tag{1.203}$$

If we project this equation onto the bra $\langle x|$, we obtain

$$\begin{aligned}
&\int dx' \langle x|x'\rangle \left(\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx'^2} + V(x') \right) \Psi(x', t) \right) \\
&= \int dx' \delta(x - x') \left(-\frac{\hbar^2}{2m} \left(\frac{d^2}{dx'^2} + V(x') \right) \Psi(x', t) \right) = i\hbar \frac{d}{dt} \langle x|\Psi(t)\rangle
\end{aligned} \tag{1.204}$$

which gives us Schrödinger's wave equation

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right) \Psi(x, t) = i\hbar \frac{d}{dt} \Psi(x, t) \tag{1.205}$$

b) Schrödinger equation in a basis of energy eigenstates ; Time evolution operator

Let's write the eigenvalue equation for the Hamiltonian operator

$$H|\psi_n\rangle = E_n |\psi_n\rangle \tag{1.206}$$

Since H is the Hermitian operator representing the energy of the system, its eigenvalues E_n are real and represent the possible results of a measurement of the energy of the system, and its eigenvectors form a complete set. Therefore it is possible to expand an arbitrary state vector $|\psi\rangle$ as a linear combination of these eigenstates

$$|\Psi(t)\rangle = \sum_n c_n(t) |\psi_n\rangle \tag{1.207}$$

We have accounted for the time dependence of the state vector by an explicit time dependence of the probability amplitudes $c_n(t)$.

We can substitute the form (1.207) for the state vector into the Schrödinger equation (1.199), and obtain

$$H|\Psi(t)\rangle = i\hbar \frac{d}{dt}|\Psi(t)\rangle \quad (1.208)$$

$$\sum_k c_k(t) H|\psi_k\rangle = i\hbar \sum_k \frac{dc_k(t)}{dt} |\psi_k\rangle$$

Taking the inner product of this equation with $\langle\psi_n|$ gives

$$\begin{aligned} \sum_k c_k(t) \langle\psi_n| H |\psi_k\rangle &= \sum_k c_k(t) E_n \langle\psi_n|\psi_k\rangle = \sum_k c_k(t) E_n \delta_{nk} = c_n(t) E_n \\ &= i\hbar \sum_k \frac{dc_k(t)}{dt} \langle\psi_n|\psi_k\rangle = i\hbar \sum_k \frac{dc_k(t)}{dt} \delta_{nk} = i\hbar \frac{dc_n(t)}{dt} \end{aligned} \quad (1.209)$$

$$\frac{dc_n(t)}{dt} = -\frac{i}{\hbar} E_n c_n(t)$$

The solution to this equation is

$$c_n(t) = c_n(0) \exp\left(-i \frac{E_n t}{\hbar}\right) \quad (1.210)$$

And so the solution to the Schrödinger equation can be written as

$$|\Psi(t)\rangle = \sum_n c_n(0) e^{-i \frac{E_n t}{\hbar}} |\psi_n\rangle \quad (1.211)$$

This shows that the general solution to the time-dependent Schrödinger equation (TDSE) can be written as a linear combination of the vectors $|\psi_n\rangle$ that are solutions of the *time-independent Schrödinger equation* (TISE) (1.206), with amplitudes in the particularly simple form (1.210). The state vectors $|\psi_n\rangle$

are called *stationary states*, because the vector $|\Psi(t)\rangle = e^{-i \frac{E_n t}{\hbar}} |\psi_n\rangle$ is a solution of the TDSE that does not change with time. (Recall that $|\psi_n\rangle$ and $e^{-i \frac{E_n t}{\hbar}} |\psi_n\rangle$ represent the *same* state, since one is just the other times a complex number.)

Consider the operator H^n . If we apply this operator to one of the vectors $|\psi_n\rangle$, we obtain

$$H^n |\psi_k\rangle = H^{n-1} H |\psi_k\rangle = H^{n-1} E_k |\psi_k\rangle = E_k H^{n-1} |\psi_k\rangle = \dots = (E_k)^n |\psi_k\rangle \quad (1.212)$$

Next, consider the operator $e^{-iHt/\hbar}$. If we apply this operator to one of the vectors $|\psi_n\rangle$, we obtain

$$e^{-iHt/\hbar} |\psi_k\rangle = \sum_n \frac{1}{n!} \left(-\frac{iHt}{\hbar}\right)^n |\psi_k\rangle = \sum_n \frac{1}{n!} \left(-\frac{it}{\hbar}\right)^n H^n |\psi_k\rangle = \sum_n \frac{1}{n!} \left(-\frac{it}{\hbar}\right)^n (E_k)^n |\psi_k\rangle = e^{-iE_k t/\hbar} |\psi_k\rangle \quad (1.213)$$

If we apply this operator to $|\Psi(0)\rangle$, we obtain

$$e^{-iHt/\hbar} |\Psi(0)\rangle = \sum_n c_n(0) e^{-iHt/\hbar} |\psi_n\rangle = \sum_n c_n(0) e^{-iE_n t/\hbar} |\psi_n\rangle \quad (1.214)$$

i.e. $e^{-iHt/\hbar} |\Psi(0)\rangle = |\Psi(t)\rangle$

The operator

$$U(t, 0) = e^{-iHt/\hbar} \quad (1.215)$$

is the *time-development operator*. It is the operator which, applied to the state vector $|\Psi(0)\rangle$ at time zero, gives the state vector $|\Psi(t)\rangle$ at time t . Its Hermitian conjugate is

$$\begin{aligned} U^\dagger(t, 0) &= \left(e^{-iHt/\hbar} \right)^\dagger = \sum_n \frac{1}{n!} \left(\left(\frac{-iHt}{\hbar} \right)^n \right)^\dagger = \sum_n \frac{1}{n!} \left(\frac{+it}{\hbar} \right)^n \left((HH \cdots H)^n \right)^\dagger \\ &= \sum_n \frac{1}{n!} \left(\frac{+iHt}{\hbar} \right)^n = e^{+iHt/\hbar} \end{aligned} \quad (1.216)$$

The time-development operator is *unitary*, since

$$\begin{aligned} U^\dagger(t, 0)U(t, 0) &= e^{+iHt/\hbar} \times 1 \times e^{-iHt/\hbar} = \sum_k e^{+iHt/\hbar} |\psi_k\rangle \langle \psi_k| e^{-iHt/\hbar} \\ &= \sum_k e^{+iE_k t/\hbar} |\psi_k\rangle \langle \psi_k| e^{-iE_k t/\hbar} = \sum_k |\psi_k\rangle \langle \psi_k| = 1 \end{aligned} \quad (1.217)$$

Therefore, it is possible to view the time development of any state vector as *a unitary transformation produced by the time-development operator*. It follows that the inner product of any two state vectors $|\Psi(t)\rangle$ and $|\Phi(t)\rangle$ that represent physically realizable states (*i.e.* obey the Schrödinger equation) is independent of time. For example, the norm of any vector is independent of time. Also, if two state vectors are orthogonal at any time, they are orthogonal at all times.²³ This point of view can be very useful for formal arguments. Sometimes it's not so useful for practical calculations because $e^{-iHt/\hbar}$ can be tricky to calculate.

5. Multiple Observables

a) Expressions for operators corresponding to an observable

So far, we have focused on a system whose state can be expressed as a linear combination of the eigenstates $\{|a_i\rangle\}$ of one observable A . However, there are generally multiple possible observables even for very simple systems. For a particle moving in a one-dimensional potential, we can measure the particle's position X , momentum P , and functions of those variable such as P^2 , $V(X)$, or energy E . For a particle moving in a three-dimensional potential we have position and momentum observables in three

²³ This completes the “physics proof” of the orthogonality of the basis state vectors that we gave on p. 51.

dimensions, X, Y, Z, P_x, P_y, P_z , as well as functions of those variables. Particles may also possess observables associated with their spin.

This brings us to a question we have not completely answered yet: given an observable A , how do we write down the operator expression for A ? One answer, if we know all possible measurement results a_i , is

$$A = \sum_i |a_i\rangle a_i \langle a_i| \quad \text{or} \quad A = \int |a\rangle a \langle a| da \quad (1.218)$$

for the discrete or continuous eigenvalue cases, respectively. If you use this expression for A to work out its matrix in the basis of eigenstates of A , then you'll see this is the same thing as eq. (1.133). Eq. (1.218) would allow us, for example, to write down an operator for a position observable X, Y , or Z .

However, eq. (1.218) cannot be the whole answer, because there are different possible sets of basis states $\{|a_i\rangle\}, \{|b_i\rangle\}, \{|c_i\rangle\}, \dots$ corresponding to different observables $A, B, C \dots$. To do any useful calculations we are going to have to express every object in one basis, and we won't necessarily know how the vectors $|a_i\rangle$ are related to the basis vectors we want to work with. To give a specific example, we may want to work with the position basis $\{|x\rangle\}$. In that case, the expression of the form (1.218) for $P_x = \int |p_x\rangle p_x \langle p_x| dp_x$ doesn't do us any good at all unless we know how to express the vectors $|p_x\rangle$ in terms of the basis vectors $\{|x\rangle\}$.

To express a momentum operator in a position basis, we use the method of section 1.3.b, *i.e.* we write

$$\begin{aligned} P_x &= -i\hbar D_x \quad \text{with} \quad \langle x|D_x|x'\rangle = \delta'(x-x') \\ P_y &= -i\hbar D_y \quad \text{with} \quad \langle y|D_y|y'\rangle = \delta'(y-y') \\ P_z &= -i\hbar D_z \quad \text{with} \quad \langle z|D_z|z'\rangle = \delta'(z-z') \end{aligned} \quad (1.219)$$

Therefore, we now have defined operator expressions for X, Y, Z, P_x, P_y , and P_z in a common basis.

We can use these expressions for X, Y, Z, P_x, P_y , and P_z to write down the operator for (almost) any observable that corresponds to a classical quantity $f(x, y, z, p_x, p_y, p_z)$. The rule is: *replace the classical position and momentum variables with the corresponding quantum operators*. For example, the z -component of orbital angular momentum of a particle in classical physics is $L_z = xp_y - yp_x$. Therefore, in quantum mechanics, the operator for the z -component of angular momentum is $L_z = XP_y - YP_x$. Another example is the energy of a particle of mass m moving in a one-dimensional potential.

Classically it is $H = \frac{p_x^2}{2m} + V(x)$. Therefore, the operator for the total energy of a particle moving in one-

dimensional potential is $H = \frac{P_x^2}{2m} + V(X)$.

Other operators of practical interest correspond to components of spin angular momentum or magnetic moment. Spin is not a function of the position or momentum of the particle, and therefore spin operators cannot be expressed in terms of the position basis vectors $\{|x\rangle\}$. Instead, we introduce the operators directly using eq. (1.218). We get the form of the operators by assuming that *spin-angular momenta obey the same algebraic rules as quantum orbital angular momenta*. That is, the possible values of the z -component of the spin angular momentum take on the values

$$S_z = M_s \hbar \quad \text{with} \quad M_s = -S, -S+1, -S+2, \dots, S \quad (1.220)$$

and the magnitude squared of the spin angular momentum must be

$$|\vec{S}|^2 = S(S+1)\hbar^2, \quad (1.221)$$

where S is a non-negative integer or half-integer

$$S = 0, \text{ or } \frac{1}{2}, \text{ or } 1, \text{ or } \frac{3}{2}, \dots \quad (1.222)$$

The value of S for any given particle must be determined experimentally. For electrons, neutrons, and protons, $S = 1/2$. The operator for S_z follows from eqs. (1.218) and (1.220), so the basis for the spin states of a particle has $2S+1$ dimensions. The operator forms for S_x and S_y in the basis of eigenstates of S_z are determined from the requirements that they have the same eigenvalues as S_z , that $|\vec{S}|^2 = S_x^2 + S_y^2 + S_z^2$ has the single eigenvalue given by eq. (1.221), and that the operators S_x, S_y , and S_z obey the same commutation relations as orbital angular momentum operators:

$$S_x S_y - S_y S_x = i\hbar S_z \quad S_y S_z - S_z S_y = i\hbar S_x \quad S_z S_x - S_x S_z = i\hbar S_y \quad (1.223)$$

For $S = 1/2$, these requirements determine the form of the operators given by eqs. (1.135) and (1.136).

Operators for spin magnetic moment are proportional to the operator \vec{S} .

With one exception discussed below, these rules allow us to write down the operators in a basis for nearly all observables that we'll want to work with.

b) Commuting and non-commuting observables; generalized uncertainty relation

In quantum mechanics, we define the *commutator* $[A, B]$ of two operators as

$$[A, B] = AB - BA \quad (1.224)$$

If $[A, B] = 0$, we say that “ A and B commute”, and it follows that the order of A and B in an operator

expression is of no importance. However if $[A, B] \neq 0$, we say that “ A and B are non-commuting operators”, and their order of appearance does matter: $AB \neq BA$. For instance from eq. (1.223) we see that the components of the spin angular momentum do not commute; *e.g.* $[S_x, S_y] \neq 0$. We can also show that X and P_x do not commute:

$$\begin{aligned}
[X, P_x] &= XP_x - P_x X = -i\hbar(XD_x - D_x X) = -i\hbar \int dx \int dx' |x\rangle \langle x| (XD_x - D_x X) |x'\rangle \langle x'| \\
&= -i\hbar \int dx \int dx' |x\rangle (x - x') \langle x| D_x |x'\rangle \langle x'| = -i\hbar \int dx \int dx' |x\rangle (x - x') \delta'(x - x') \langle x'| \\
&= -i\hbar \int dx |x\rangle \frac{d}{dx'} [(x - x') \langle x'|]_{x'=x} = -i\hbar \int dx |x\rangle [-1 \langle x'|]_{x'=x} = i\hbar \int dx |x\rangle \langle x| \\
\Rightarrow [X, P_x] &= i\hbar
\end{aligned} \tag{1.225}$$

In section 1.A.6, we showed that it is not possible to measure *both* the position X and the momentum P_x of a particle with perfect accuracy; the product of their uncertainties must obey the Heisenberg uncertainty principle. The physical reason for this is that it is impossible to precisely measure X without disturbance to the momentum P_x , and vice-versa.

It turns out that *it is not possible to simultaneously measure any two non-commuting observables with perfect precision*. We can generalize Heisenberg’s uncertainty principle as follows. We define the uncertainties in A and B to be

$$\Delta A = \sqrt{\langle (A - \langle A \rangle)^2 \rangle} \quad \Delta B = \sqrt{\langle (B - \langle B \rangle)^2 \rangle} \tag{1.226}$$

where $\langle \rangle$ denotes an expectation value. The deviation of a measurement result from the average result is $A - \langle A \rangle$. The square of that quantity $(A - \langle A \rangle)^2$ is always a non-negative number; $\langle (A - \langle A \rangle)^2 \rangle$ is the *mean-square* deviation of A from $\langle A \rangle$. ΔA is the *root-mean-square* (rms) deviation of many measurement results from the average result.

Now, consider the ket

$$|\phi\rangle = (A + i\lambda B)|\psi\rangle \tag{1.227}$$

where λ is an arbitrary real number. The norm of $|\phi\rangle$ is always greater than or equal to zero:

$$\begin{aligned}
\langle \phi | \phi \rangle &= \langle \psi | (A - i\lambda B)(A + i\lambda B) | \psi \rangle = \langle \psi | A^2 | \psi \rangle + \langle \psi | i\lambda (AB - BA) | \psi \rangle + \lambda^2 \langle \psi | B^2 | \psi \rangle \geq 0 \\
\langle A^2 \rangle + i\lambda \langle [A, B] \rangle + \lambda^2 \langle B^2 \rangle &\geq 0
\end{aligned} \tag{1.228}$$

Since $\langle \phi | \phi \rangle$, λ , $\langle A^2 \rangle$, and $\langle B^2 \rangle$ are all real, $i\langle [A, B] \rangle$ is real. Therefore eq. (1.228) is an ordinary

quadratic equation for λ with real coefficients. The only way eq. (1.228) can be true for all λ is if its discriminant (" $b^2 - 4ac$ ") is less than or equal to zero:

$$(i\langle[A, B]\rangle)^2 - 4\langle A^2\rangle\langle B^2\rangle \leq 0 \Rightarrow \langle A^2\rangle\langle B^2\rangle \geq \left(\frac{i}{2}\langle[A, B]\rangle\right)^2 \quad (1.229)$$

For the state $|\psi\rangle$, it follows that for $A' = A - \langle A\rangle$ and $B' = B - \langle B\rangle$,

$$\langle A'^2\rangle\langle B'^2\rangle \geq \left(\frac{i}{2}\langle[A, B]\rangle\right)^2 \quad (1.230)$$

i.e. that

$$\Delta A \Delta B \geq \frac{1}{2} |\langle[A, B]\rangle| \quad (1.231)$$

This is the *generalized uncertainty principle*. It gives the minimum uncertainty product for a simultaneous measurement of A and B . For example, if we use the commutation relation (1.225), we find the usual statement of the uncertainty principle

$$\Delta X \Delta P_x \geq \frac{\hbar}{2} \quad (1.232)$$

If two observables commute, then $\Delta A \Delta B \geq 0$; there is no lower limit to the product of the uncertainties with which A and B can be simultaneously measured. In principle, both could be measured simultaneously measured with perfect accuracy. We say that A and B are *compatible observables*.

There is a problem with the prescription we have gave above for the conversion of classical expressions into quantum operators – it doesn't work for the product of two non-commuting operators. For example, suppose we have a quantity whose classical expression is $O = xp_x$. According to the above prescription, the quantum operator for this observable should be XP_x . But the operator XP_x is not Hermitian:

$$(XP_x)^\dagger = P_x^\dagger X^\dagger = P_x X = XP_x + P_x X - XP_x = XP_x - [X, P_x] = XP_x - i\hbar \neq XP_x \quad (1.233)$$

By the same argument, any product of two non-commuting Hermitian operators is non-Hermitian. Since XP_x is not Hermitian it cannot represent an observable. Also, note that in classical physics we could just as well write the term as $p_x x$. Following our rule, we would then write the operator as $P_x X$. In quantum mechanics, $P_x X \neq XP_x$, so we get a different operator depending on the arbitrary choice of order of X and P_x .

When we encounter a product of two non-commuting operators in a classical expression, we have an addendum onto our rule for the construction of operators: *take the symmetric combination of the two*

product expressions:

$$\text{If (classical physics) } O = AB \text{ and (quantum physics) } [A, B] \neq 0, \text{ then } O = \frac{1}{2}(AB + BA) \quad (1.234)$$

If A and B are Hermitian, the operator $\frac{1}{2}(AB + BA)$ is Hermitian and well-defined, so that solves the problem. The rule (1.234) might strike you as somewhat arbitrary, and, well – it is. You can regard it as yet another postulate, justified because the resulting theory agrees with experiment.

c) **Complete set of commuting operators (C.S.C.O.)**

The following theorem is very important in quantum mechanics: *if a number of diagonalizable operators all commute with each other, they have a common set of eigenvectors, i.e. they can all be diagonalized simultaneously by the same basis.* We'll prove this for the case of two operators A and B . We will write our basis vectors temporarily as $|\alpha_i\rangle$, where i labels which vector we are referring to, and α_i represents eigenvalues of A and/or B (i.e. α_i may represent a set of eigenvalues). By assumption A and B commute, so $AB = BA$. We can write this equation in components

$$\sum_k \langle \alpha_i | A | \alpha_k \rangle \langle \alpha_k | B | \alpha_j \rangle - \sum_k \langle \alpha_i | B | \alpha_k \rangle \langle \alpha_k | A | \alpha_j \rangle = 0 \quad (1.235)$$

Also by assumption, A is diagonalizable. Therefore we can choose a basis $\{|\alpha_i\rangle\}$ that diagonalizes A . It follows that we have

$$\langle \alpha_i | A | \alpha_k \rangle = a_k \langle \alpha_i | \alpha_k \rangle = a_k \delta_{ik} \quad (1.236)$$

where a_k is the eigenvalue for the basis vector $|\alpha_k\rangle$. Eq. (1.235) becomes

$$(a_i - a_k) \langle \alpha_i | B | \alpha_k \rangle = 0 \quad (1.237)$$

This equation says that $\langle \alpha_i | B | \alpha_k \rangle$ can be non-zero only if $a_i = a_k$. We suppose that we re-arrange the basis vectors $\{|\alpha_i\rangle\}$ in order of increasing eigenvalue a_i . It may be that some values of a_i occur more than once. We say that this eigenvalue is *g-fold degenerate*, where g is the number of times that the specific eigenvalue occurs. The g eigenvectors belonging to the same eigenvalue are said to constitute a *degenerate subspace of A* . For example, we could have $a_1 = a_2$, and $a_3 = a_4 = a_5$, with $a_3 > a_1$, and $a_6 > a_5$. The first eigenvalue would be two-fold degenerate, and the second eigenvalue three-fold degenerate. In this instance, the matrices of A and B will look like this:

$$A = \begin{bmatrix} a_1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & a_1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & a_3 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & a_3 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & a_3 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & a_6 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad B = \begin{bmatrix} B_{11} & B_{12} & 0 & 0 & 0 & 0 & \cdots \\ B_{21} & B_{22} & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & B_{33} & B_{34} & B_{35} & 0 & \cdots \\ 0 & 0 & B_{43} & B_{44} & B_{45} & 0 & \cdots \\ 0 & 0 & B_{53} & B_{54} & B_{55} & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & B_{66} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (1.238)$$

With the basis $\{|\alpha_i\rangle\}$ so ordered, eq. (1.237) says that the matrix of B must have *block-diagonal form* as shown above. The dimension of each block is $g \times g$ where g is the degeneracy of the corresponding eigenvalue of A .

It is possible to diagonalize B by diagonalizing each block separately. That is because the eigenvalue equation for B , in this basis, separates into uncoupled eigenvalue equations for each block. Since the blocks are uncoupled, and we have assumed that B is diagonalizable, it follows that each block must be separately diagonalizable. So, we then diagonalize each block of B , obtaining a new set of basis states in which B is diagonal. Because each block involve basis states belonging to a single eigenvalue of A , A remains diagonal in the new basis. To say this another way, within each g -fold degenerate subspace of A , the matrix of A is a multiple of an $g \times g$ unit matrix. A unit matrix is invariant with respect to changes in basis, so A remains diagonal in the subspace when we diagonalize B in that subspace.

This argument shows that A and B can be simultaneously diagonalized. The argument can be extended to any number of mutually commuting operators A, B, C, \dots (That is, operators for which $[A, B] = [B, C] = [A, C] = \dots = 0$.) If a set of Hermitian operators A, B, C, \dots mutually commute, then it is possible to find a set of vectors $\{|\alpha_i\rangle\}$ which are simultaneous eigenvectors of all of the commuting operators A, B, C, \dots

Now, this theorem helps us to answer an important question: when we have multiple observables, what do we do with the expansion postulate, eq. (1.59) or (1.60)? Obviously, we can't describe every possible state with only one eigenvalue of one observable. Do we have a new expansion postulate, involving eigenstates of multiple observables? And if so, do we use all observables, or just some? What about observables that don't commute with each other, and therefore can't have a state in which both observables have perfectly defined values?

The answer is: *the general state vector $|\psi\rangle$ can be written as a linear combination of the simultaneous eigenvectors of a complete set of commuting observables (C.S.C.O) $\{A, B, C, \dots\}$*

$$|\psi\rangle = s_1 |\{abc...\}_1\rangle + s_2 |\{abc...\}_2\rangle + s_3 |\{abc...\}_3\rangle + \dots \quad (1.239)$$

$|\{abc...\}_i\rangle$ is a “common eigenvector” means that

$$\begin{aligned} A|\{abc...\}_i\rangle &= a_i |\{abc...\}_i\rangle \\ \text{and } B|\{abc...\}_i\rangle &= b_i |\{abc...\}_i\rangle \\ \text{and } C|\{abc...\}_i\rangle &= c_i |\{abc...\}_i\rangle \dots \end{aligned} \quad (1.240)$$

where a_i is the value of a from the i th set of eigenvalues $\{abc...\}_i$, and similarly for b_i, c_i, \dots . We know it is possible to find such a set of common eigenvectors by the theorem we proved above.

A complete set of commuting observables is one in which every operator in the set commutes with every other operator in the set, *and* in which the specification of a full set of eigenvalues $\{abc...\}_i$ *uniquely* specifies every eigenvector, *and* for which the number of such operators is the minimum needed to uniquely specify each vector. It is perhaps easiest to illustrate this concept with an example. Suppose we have a 4-dimensional Hilbert space, and that we have determined the matrices for four mutually commuting operators A, B, C , and D in some basis. Then, suppose that we simultaneously diagonalize A, B, C , and D . In the new basis of eigenvectors $\{|1\rangle, |2\rangle, |3\rangle, |4\rangle\}$, we find that

$$A = a_0 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \quad (1.241)$$

$$B = b_0 \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1.242)$$

$$C = c_0 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \quad (1.243)$$

$$D = d_0 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad (1.244)$$

In this example, the set of observables $\{A, B, D\}$ or $\{A, C, D\}$ is a CSCO. In the basis of simultaneous eigenstates of A , B , and D , we have the four eigenvectors

$$\begin{aligned} \{|1\rangle, |2\rangle, |3\rangle, |4\rangle\} &= \left\{ |\{a, b, d\}_1\rangle, |\{a, b, d\}_2\rangle, |\{a, b, d\}_3\rangle, |\{a, b, d\}_4\rangle \right\} \\ &= \left\{ |a_0, -b_0, d_0\rangle, |2a_0, -b_0, d_0\rangle, |2a_0, b_0, d_0\rangle, |2a_0, b_0, -d_0\rangle \right\} \end{aligned} \quad (1.245)$$

Therefore, the set $\{A, B, D\}$ meets our requirements. The operators mutually commute with each other. Their simultaneous eigenvectors are each *uniquely* labeled by their eigenvalues $\{a, b, d\}_i$. And, the set $\{A, B, D\}$ has the minimum number of operators needed to accomplish this: it is not possible to throw any operator out of the set and still have each eigenvector uniquely labeled. The alternative set $\{A, C, D\}$ also meets these requirements.

Most of the concepts developed above generalize in a straightforward way to a C.S.C.O. with more than one observable. For example, the unit operator becomes

$$1 = \sum_i |\{a, b, \dots\}_i\rangle \langle \{a, b, \dots\}_i| \quad (1.246)$$

The general quantum mechanical state vector is a superposition of the simultaneous eigenvectors of a C.S.C.O., as given in eq. (1.239). This extension of the expansion postulate is justified because it gives a theory that agrees with experiment. However that's not very surprising. We assumed throughout sections 1.B.1-5 that $\{X\}$ is a C.S.C.O in one dimension. The C.S.C.O $\{X, Y, Z\}$ in three dimensions is a natural generalization. Additions of observables to a C.S.C.O. are generally of a similar nature.

A C.S.C.O is not unique. For a particle moving in a three-dimensional potential, we could choose $\{X, Y, Z\}$ or $\{H, \vec{L}^2, L_z\}$ or $\{H, \vec{L}^2, L_x\}$ as our C.S.C.O., where H is the Hamiltonian and $\vec{L} = \vec{R} \times \vec{P}$ the orbital angular momentum operator. A C.S.C.O. for a specific problem can be chosen to make that problem easiest to solve.

A C.S.C.O can vary according to what we consider to be our “system”. The C.S.C.O for a hydrogen atom could be as large as the set $\{X_{cm}, Y_{cm}, Z_{cm}, H, \vec{L}^2, L_z, S_z, I_z\}$, where X_{cm}, Y_{cm} , and Z_{cm} are the coordinates of the center-of-mass of the atom, H, \vec{L}^2, L_z the Hamiltonian, orbital angular momentum squared, and z -component of the orbital angular momentum for the motion of the electron relative to the center of mass, S_z the z -component of the electron spin angular momentum, and I_z the z -component of the proton spin-angular momentum. However, for many problems some observables will be completely uncoupled from other observables. For example, often there is no physical coupling between the center of mass coordinate and the other coordinates of the hydrogen atom. If we only wish to study quantities relating to the internal structure of the atom, it is perfectly fine to define the “system” to

consist only of the internal degrees of freedom of the atom; *i.e.* the C.S.C.O. of the “hydrogen atom” would be $\{H, \vec{L}^2, L_z, S_z, I_z\}$. Now, it turns out that the nuclear spin, the electron spin, and the orbital motions of the electron are always coupled together in the hydrogen atom. However, this coupling is small, especially for the nucleus. So it may be that, to a good approximation, we can neglect that coupling in some problem we want to solve. If we neglect the coupling of the electron to the nuclear spin, and only care about the electronic degrees of freedom, then our “hydrogen atom” would have a C.S.C.O given by $\{H, \vec{L}^2, L_z, S_z\}$. Or, maybe we can even neglect the coupling of the electron orbital motion to the electron spin, and we only care about the electron orbital motion. In this case our C.S.C.O. would be given by $\{H, \vec{L}^2, L_z\}$. However, note that *this is only valid if the coupling of the observables included in the C.S.C.O. to other observables is negligible for the problem of interest.*

Now, we are finally in a position to clarify two points we discussed earlier. First, in section 1.A.5.a, we used the term “simultaneously observable” when we discussed the conditions under which *probability amplitudes* add and the conditions under which *probabilities* add. “Simultaneously observable” means that the observables are both members of a C.S.C.O.

In section 1.B.2.h, we stated that the eigenstates of a Hermitian operator A form a complete, orthonormal set. We can now provide a necessary clarification of this point. If the operator A has an eigenvalue with degeneracy g , then the eigenvectors for this eigenvalue are not uniquely determined by the eigenvalue equation for A . We can find sets of g eigenvectors that span the degenerate subspace, but are not orthogonal to each other. However, the fact that the operator A has a degenerate eigenvalue means that it cannot be a C.S.C.O. by itself. There must be one or more additional operators in the C.S.C.O. When we simultaneously diagonalize all operators in the C.S.C.O, the resulting set of simultaneous eigenvectors will be complete and orthonormal. If we assume this is always done when degenerate eigenvalues occur, then the eigenvectors of any Hermitian operator A always form a complete, orthonormal set of states.