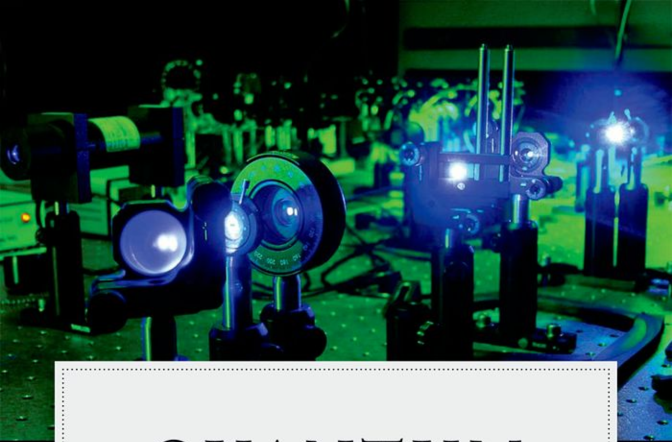


QUANTUM MECHANICS

THEORY AND EXPERIMENT

MARK BECK

QUANTUM MECHANICS

This page intentionally left blank

QUANTUM MECHANICS

Theory and Experiment

Mark Beck

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data
Beck, Mark (Mark K.), 1963–
Quantum mechanics : theory and experiment / Mark Beck.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-19-979812-4 (hardcover : alk. paper)
1. Quantum theory—Textbooks. 2.
Mechanics—Textbooks. I. Title.
QC174.12.B43 2012
530.12—dc23 2011042970

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

For Annie, Marisol, and Lupe, with love.

This page intentionally left blank

Contents

LAB 2 • “PROOF” OF THE EXISTENCE OF PHOTONS 449

LAB 3 • SINGLE-PHOTON INTERFERENCE 463

LAB 4 • QUANTUM STATE MEASUREMENT 475

LAB 5 • TESTING LOCAL REALISM 483

Table of Symbols

Symbol	Description	First used on page
$\langle x \rangle$	Mean of x	3
Δx^2	Variance of x	4
Δx	Standard deviation of x	4
$P(x)$	Probability of x	6
$p(x)$	Probability density of x	8
$P(x, y)$	Joint probability of x and y	9
$P(x \mid y)$	Conditional probability of x given y	9
$\boldsymbol{u}_x$	Unit vector pointing in x-direction	9
$\doteq$	"is represented by"	10
$\bar{\mathbf{M}}$	Matrix $\bar{\mathbf{M}}$	13
M_{ij}	Element in the i^{th} row and j^{th} column of $\bar{\mathbf{M}}$	13
$\bar{\mathbf{1}}$	Identity matrix	14
δ_{ij}	Kronecker delta	14
$\boldsymbol{\mathcal{E}}$	Electric field	21
$\boldsymbol{k}$	Wave vector	21
λ	Wavelength	21
f	Frequency	21

(*continued*)

Symbol	Description	First used on page
ω	Angular frequency	21
c	Speed of light in vacuum	21
$\boldsymbol{\varepsilon}$	Unit polarization vector (Jones vector)	22
n	Index of refraction	26
$\boldsymbol{S}$	Poynting vector	27
$\boldsymbol{\mathcal{B}}$	Magnetic field	28
$\mathrm{PA}_{\mathrm{HV}}$	Polarization analyzer (subscript denotes orientation)	28
$\overline{\mathbf{J}}$	Jones matrix	33
τ_c	Coherence time	45
l_c	Coherence length	45
$\left\|\psi\right\rangle$	Quantum state (ket)	48
$\left\langle\psi\right\|$	Quantum state (bra)	49
$\left\langle\psi_1\middle\|\psi_2\right\rangle$	Inner product (bracket)	50
$\hat{O}$	Operator	65
$\hat{R}(\theta,\boldsymbol{u}_z)$	Rotation operator	66
$\hat{R}_p(\theta)$	Polarization rotation operator	66
$\hat{O}^\dagger$	Adjoint operator	68
$\hat{U}$	Unitary operator	69
$\hat{P}_\psi = \left\|\psi\right\rangle\left\langle\psi\right\|$	Projection Operator	69
$O_{ij} = \left\langle\psi_i\middle\|\hat{O}\middle\|\psi_j\right\rangle$	Matrix element	72
$\hat{\wp}_{HV}$	Polarization operator in *HV*-basis	91
$\left\langle\hat{O}\right\rangle$	Expectation value of O	95
$\left[\hat{A},\hat{B}\right]$	Commutator of $\hat{A}$ and $\hat{B}$	97
$\boldsymbol{\mu}$	Magnetic dipole moment	111
h	Plank's constant	114
$\hbar = h/2\pi$	h-bar	114

Symbol	Description	First used on page
SA_z	Spin analyzer (subscript denotes orientation)	114
$\lvert +z \rangle$	Spin-up along z-axis state of a spin-1/2 particle	115
$\hat{S}_z$	z-component of spin operator	116
$\bar{\sigma}$	Pauli Spin Matrix	116
g	g factor	124
$\hat{\boldsymbol{S}}$	Spin operator	128
$\hat{\boldsymbol{L}}$	Orbital angular momentum operator	128
$\hat{\boldsymbol{J}}$	Total angular momentum operator	128
$\hat{J}^2$	Total angular momentum squared operator (“J-squared”)	129
$\hat{J}_+, \hat{J}_-$	Raising and Lowering Operators	131
$\otimes$	Direct product	154
$\hat{\rho}$	Density operator	163
$\lvert \phi^\pm \rangle, \lvert \psi^\pm \rangle$	Bell States	172
$\hat{U}(t)$	Unitary time evolution operator	193
$\hat{H}$	Hamiltonian operator	194
Ω	Larmor frequency	197
Ω_R	Generalized Rabi frequency	210
$\lvert x \rangle$	Eigenstate of position	215
$\psi(x)$	Wave function	216
$\delta(x)$	Delta function	217
$\hat{T}(D)$	Translation operator	221
$\lvert p \rangle$	Eigenstate of momentum	227
$\tilde{\psi}(p)$	Wave function in momentum representation	227

(*continued*)

Symbol	Description	First used on page
$\mathfrak{F}\left\{\psi\left(x\right)\right\}$	Fourier transform of $\psi\left(x\right)$	237
$\Psi\left(x,t\right)$	Time dependent wave function	242
$\boldsymbol{j}\left(\boldsymbol{r},t\right)$	Probability flux	249
κ	Spring constant	275
$\hat{a}, \hat{a}^\dagger$	Annihilation and creation operators	277
$\hat{n}$	Number operator	277
$\left\|n\right\rangle$	Fock, or number, state	278
$H_n\left(x\right)$	Hermite Polynomial	283
$\left\|\alpha\right\rangle$	Coherent state	287
$P_l^{m_l}\left(x\right)$	Associated Legendre function	308
$Y_l^{m_l}\left(\theta,\phi\right)$	Spherical harmonic	309
e	Electron charge	315
ε_0	Permitivity of free space	315
$L_j^k\left(\rho\right)$	Associated Laguerre polynomial	317
a_0	Bohr radius	319
$j_l\left(\rho\right)$	Spherical Bessel function	330
β_{nl}	Zero of spherical Bessel function	330
m_e^*	Effective mass of electron	332
m_h^*	Effective mass of hole	332
$E_n^{(j)}$	j^{th}-order energy	337
$\left\|\psi_n^{(j)}\right\rangle$	j^{th}-order eigenstate	337
α	Fine structure constant	350
μ_0	Permeability of free space	351
$\bar{\lambda}_c$	Compton wavelength	353
$\hat{\mathbf{d}}$	Electric dipole moment operator	370
$\boldsymbol{n}\left(E\right)$	Density of states	376

Symbol	Description	First used on page
$\hat{A}_H(t)$	Heisenberg picture operator	384
$\hat{\boldsymbol{\mathcal{E}}}(\boldsymbol{r},t)$	Electric field operator	388
$\hat{X}_\theta$	Field quadrature amplitude operators	389
$\hat{S}(r)$	Squeezing operator	394
$\hat{\boldsymbol{\mathcal{E}}}^{(+)}(\boldsymbol{r},t)$	Postive frequency part of the electric field operator	401
$\hat{I}(\boldsymbol{r},t)$	Intensity operator	402
$:\cdots:$	Normal ordering	411
$\mathcal{T}$	Time ordering	411
$g^{(2)}(\tau)$	Degree of second-order coherence	412

This page intentionally left blank

Preface

Physics is both a theoretical and an experimental science. We need theories to develop an understanding of nature, and experiments to verify our theories. Nowhere in physics is the interplay between theory and experiment more important than in quantum mechanics, because the theory is frequently counterintuitive (if not outright mind-boggling), making experimental verification all the more important.

The goal of this book is to explicate some of the theoretical and experimental aspects of quantum mechanics, at the level of a junior or senior undergraduate. (I assume a previous exposure to quantum mechanics at the level of a "Modern Physics" course.) The book contains 17 chapters that describe the theoretical underpinnings of quantum mechanics, and 5 laboratories that allow one to observe experimental confirmation of aspects of the theory. The experiments include: "Proving" that light contains photons, single-photon interference, and tests of local realism.

The experiments all examine the behavior of single photons and photon pairs, so in order to coordinate the laboratories and the text, the text introduces the formalism of quantum mechanics using photon polarization. This has several advantages, in addition to connecting with the labs. Polarization is a two-dimensional system, so the mathematics is straightforward, making it a good starting point. The quantum description of polarization also has strong analoges in the classical description (presented in chap. 2), which makes understanding the physics easier. Once the formalism of quantum mechanical states and operators has been introduced using polarization, the text goes on to describe spin systems, time evolution, continuous variable systems (particle in a box, harmonic oscillator, hydrogen atom, etc.), and perturbation theory. Along the way important topics such as quantum measurement (chap. 5) and entanglement (chap. 8) are discussed. The text also includes introductions to quantum field theory (chap. 16) and quantum information (chap. 17).

While most of the text follows sequentially as presented, the material in chap. 17 (Quantum Information) may be covered at any point after the material in chap. 8 (Two-Particle Systems and Entanglement).

Some of the end-of-chapter problems are marked with an *. These problems illustrate important ideas or prove relationships that are only alluded to in the text, or are simply more challenging problems. A solutions manual is available for instructors.

Some of the chapters have complements, which serve as appendices for the chapters. Most of the material in these complements is supplementary, and instructors can skip them without impairing their ability to cover material in later chapters. The exceptions to this are complements 2.A and 10.A. Also, complement 8.C needs to be covered before discussing hyperfine structure in sec.14.4.

The first two laboratories can be performed after the material in chapt. 2 has been covered. Lab 3 *can* be performed after chap. 3, but it will be better appreciated after covering the material in chap. 5. Lab 4 requires complement 5.A, and lab 5 requires complement 8.B. The book website (www.oup.com/us/QuantumMechanics) and my website (http://www.whitman.edu/~beckmk/QM/) contain supplementary information (equipment lists, etc.) regarding the laboratories.

I'd like to acknowledge the people from whom I learned quantum mechanics, especially Mike Raymer, Carlos Stroud, Ian Walmsley, Joe Eberly, and Leonard Mandel. I'd also like to acknowledge the books that have been most influential in shaping the way that I think about quantum mechanics: *Introduction to Quantum Mechanics* by David J. Griffiths, *A Modern Approach to Quantum Mechanics* by John S. Townsend (from which I borrowed the idea for the "Experiments" in chaps. 3 and 6), and *Quantum Mechanics* by Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloë (from which I borrowed the idea of chapter complements).

Funding for the development of the laboratories came from the National Science Foundation and Whitman College. A number of people contributed to the development of the laboratories, including David Branning, Alex Carlson, Robert Davies, Vinsunt Donato, Enrique Galvez, Ashifi Gogo, Jesse Lord, Morgan Mitchell, Matt Neel, Larry North, Matt Olmstead, Will Snyder, and Jeremy Thorn.

I'd also like to thank the people who read and commented on parts of the manuscript: Andrew Dawes, Peter Doe, John Essick, Warren Grice, Kurt Hoffman, Doug Hundley, Doug Juers, Shannon Mayer, Fred Moore, Sarah Nichols, Mike Raymer, Jay Tasson, and Steven van Enk.

Finally, I'd like to thank the people at Oxford University Press and TNQ, especially my editor Phyllis Cohen, for their hard work in making this project a success.

M.B.
Walla Walla, WA
August 2011

QUANTUM MECHANICS

This page intentionally left blank

CHAPTER 1

Mathematical Preliminaries

Before beginning a discussion of quantum mechanics, it's useful to review some topics from mathematics and classical physics. In this chapter we'll go over some areas of probability theory and linear algebra that we'll find useful later. You may be familiar with the concepts presented here, but I suggest that you read through this chapter, if only to familiarize yourself with the notation and terminology we'll be using.

1.1 PROBABILITY AND STATISTICS

1.1.1 Moments of Measured Data

We want to measure a quantity x, which is a property of some object. So, we break out our x-meter and make measurements, getting N values of x: $x_{m\,1}$, $x_{m\,2}$, etc. We're using the subscript m here to indicate that these are measured values. Collectively we'll refer to these measurements as $x_{m\,i}$, where $i = 1, 2, \ldots, N$.

We can calculate the average of x, $\langle x \rangle$, by adding up the measured values and dividing by the number of measurements:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_{m\,i}. \tag{1.1}$$

We'll use the bracket symbol $\langle \ldots \rangle$ to denote an average; it represents the average of whatever is inside the brackets. We'll also refer to $\langle x \rangle$ as the mean of x. We can calculate the average of x^n, $\langle x^n \rangle$, by simply using the definition of the average

$$\langle x^n \rangle = \frac{1}{N} \sum_{i=1}^{N} \left(x_{m\,i} \right)^n. \tag{1.2}$$

Another name for $\langle x^n \rangle$ n^{th}-order moment of x.

Remember, the brackets $\langle \ldots \rangle$ mean that we average *whatever quantity* is inside the brackets. We calculate this average by adding up the values and dividing by the number of measurements. The average of a function of x, $f(x)$, is thus

$$\langle f(x) \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_{m\,i}). \tag{1.3}$$

The average is linear. This means that

$$\langle A_1 f_1(x) + A_2 f_2(x) \rangle = A_1 \langle f_1(x) \rangle + A_2 \langle f_2(x) \rangle, \tag{1.4}$$

Where A_1 and A_2 are constants.

Frequently, we're interested in how far a typical measured value might be from the mean value. Let's define the deviation from the mean as

$$\delta x = x - \langle x \rangle. \tag{1.5}$$

The average of this quantity is 0, so it is not a good measure of the fluctuations:

$$\langle \delta x \rangle = \langle x - \langle x \rangle \rangle = \langle x \rangle - \langle \langle x \rangle \rangle = \langle x \rangle - \langle x \rangle = 0. \tag{1.6}$$

Here we have used both the linearity property [eq. (1.4)] and the fact that once a quantity has been averaged it becomes a constant. The average of a constant is the constant. $\langle \delta x \rangle$ is not a good measure of the fluctuations, because roughly half the time it's positive, and half the time it's negative, so it averages zero. To overcome this problem, we can average the square of δx. Calculating this, we see that

$$\begin{aligned} \langle (\delta x)^2 \rangle &= \langle (x - \langle x \rangle)^2 \rangle \\ &= \langle (x^2 - 2x\langle x \rangle + \langle x \rangle^2) \rangle \\ &= \langle x^2 \rangle + \langle (-2\langle x \rangle x) \rangle + \langle (\langle x \rangle^2) \rangle \\ &= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2 . \end{aligned} \tag{1.7}$$

The variance of x, Δx^2, is defined as

$$\Delta x^2 \equiv \langle (\delta x)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2. \tag{1.8}$$

The standard deviation of x, Δx, is the square root of the variance:

$$\Delta x \equiv \sqrt{\Delta x^2} = \left(\left\langle x^2\right\rangle - \langle x\rangle^2\right)^{1/2}. \tag{1.9}$$

The standard deviation is a measure of how far a particular measurement is likely to be from the mean, and we will use it to quantify the uncertainty in a series of measurements.[1]

EXAMPLE 1.1

Measurements of x yield the values

$$x_{m\,i} = 9, 5, 25, 23, 10, 22, 8, 8, 21, 20 \quad i = 1, ..., 10. \tag{1.10}$$

Calculate the mean and standard deviation of this data.

To calculate the mean of x we use eq. (1.1):

$$\langle x\rangle = \frac{1}{10}(9+5+25+23+10+22+8+8+21+20) = 15.1. \tag{1.11}$$

In order to calculate the standard deviation, we first need to calculate $\left\langle x^2\right\rangle$. Using eq. (1.2) we find

$$\left\langle x^2\right\rangle = \frac{1}{10}\left(9^2+5^2+25^2+23^2+10^2+22^2+8^2+8^2+21^2+20^2\right) = 281.3. \tag{1.12}$$

Substituting these results into eq. (1.9) yields

$$\Delta x = \left[281.3 - (15.1)^2\right]^{1/2} = 7.3. \tag{1.13}$$

Using the standard deviation as the uncertainty of the measurements, we say that $x = 15.1 \pm 7.3$.

1.1.2 Probability

Let's look at the data from example 1.1 [eq. (1.10)] a little differently, by creating a histogram of the data. We break the full range of the data into M segments, known as bins. The bins have equal widths, and we'll label them as x_j $(j = 1, 2, ..., M)$. Note that there's no subscript m on x_j, because we're labeling an x value corresponding to a bin (which is a value that *could* have been measured), not a particular measured value. We sort individual measurements into bins, then count the number of measurements in each bin, $N(x_j)$. This process, known as histogramming, is best illustrated with an example.

---

[1] Frequently the variance will be defined with a normalization factor of 1/(N–1) in front of the sum in eq. (1.3) instead of 1/N, which makes the normalization factor for the standard deviation $1/\sqrt{N-1}$. For details on why this is the case, see secs. 8.1 and 10.2 of ref. [1.1]. The choice of $1/\sqrt{N-1}$ is common in calculators and spreadsheets. In the limit of large N the difference is insignificant. In this book we'll assume a 1/N normalization.

EXAMPLE 1.2

Create a histogram of the data from example 1.1.

The data values range from 5 to 25; let's break this interval into 7 equal-width bins. The first bin contains measurements which yield x values of 5, 6, and 7, and we'll label it bin $x_1 = 6$, corresponding to the center value. Bin $x_2 = 9$ contains values 8, 9 and 10, and so on.

Examining the data in eq. (1.10), we find that one measurement yields a value of 5, 6, or 7. Thus, the number of measurements in bin $x_1 = 6$ is $N(6) = 1$. Four of our measurements fall in the second bin (holding values 8, 9, and 10) so $N(9) = 4$. Continuing on in this vein, we can find the rest of the histogram values. A plot of the histogram is shown in fig. 1.1.

Histograms allow us to estimate the probability that we will obtain a particular measurement. The probability $P(x_j)$ that a measurement will fall into a particular bin x_j is simply the ratio of the number of measurements in that bin to the total number of measurements. In other words

$$P(x_j) = \frac{N(x_j)}{N}. \tag{1.14}$$

The probability distribution corresponding to the histogram calculated in example 1.2 is plotted in fig. 1.1. As can be seen, the probability distribution is a scaled version of the histogram.

Summing all the histogram values must yield the total number of measurements N. If the histogram has M bins then

$$\sum_{j=1}^{M} P(x_j) = \frac{\sum_{j=1}^{M} N(x_j)}{N} = \frac{N}{N} = 1. \tag{1.15}$$

Our probability is properly normalized, which means that the sum of the probabilities is 1.

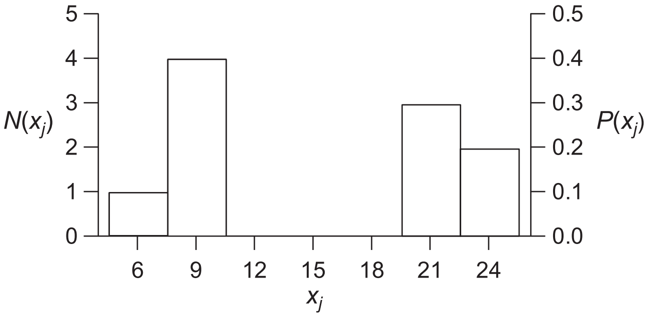


Fig 1.1 The histogram calculated in example 1.2 is plotted on the left axis, while the corresponding probability distribution is plotted on the right axis.

Equation (1.14) is only an estimate of the probability, and it's a fairly coarse estimate at that if the number of measurements N is small. Furthermore, with small N it is necessary to use fairly wide bins, so the x resolution of $P\left(x_j\right)$ is not very good. Conversely, if N is large, it is possible to obtain a fairly accurate and high-resolution estimate of $P\left(x_j\right)$.

One thing we can do with a probability distribution is simply plot it. This allows us to visualize our data, and see which measured values are likely to occur, and which are not very likely. For example, fig. 1.1 tells us that measurements of x are clumped into regions, which is more information than simply saying that $x = 15.1 \pm 7.3$ (from example 1.1). We can also use the probabilities to directly calculate moments, without having to go back to the original data. Given the probabilities $P\left(x_j\right)$ for $j = 1, 2, \ldots, M$, the mean of x is given by

$$\langle x \rangle = \sum_{j=1}^{M} x_j P\left(x_j\right). \tag{1.16}$$

Thus, to calculate the mean, weight the value by its probability, and then sum over all possible values. In general the mean of a function of x, $f\left(x\right)$, is given by

$$\langle f\left(x\right) \rangle = \sum_{j=1}^{M} f\left(x_j\right) P\left(x_j\right). \tag{1.17}$$

EXAMPLE 1.3

Calculate the mean and standard deviation of the data from example 1.1, using the corresponding probability distribution.

Figure 1.1 shows the histogram of the data from example 1.1. We can use this histogram to estimate the probability distribution using eq. (1.14). This probability distribution is plotted on the right axis of fig. 1.1. Using these probabilities, eq. (1.16) tells us that

$$\langle x \rangle = \left(6\right)\left(0.1\right) + \left(9\right)\left(0.4\right) + \left(21\right)\left(0.3\right) + \left(24\right)\left(0.2\right) = 15.3, \tag{1.18}$$

and eq. (1.17) says that

$$\langle x^2 \rangle = \left(6\right)^2 \left(0.1\right) + \left(9\right)^2 \left(0.4\right) + \left(21\right)^2 \left(0.3\right) + \left(24\right)^2 \left(0.2\right) = 283.5. \tag{1.19}$$

The standard deviation is then

$$\Delta x = \left[283.5 - \left(15.3\right)^2\right]^{1/2} = 7.0. \tag{1.20}$$

Note that the calculation of $\langle x \rangle$ using the estimated probability distribution in example 1.3 is not in perfect agreement with the direct calculation in example 1.1, although the two determinations of $\langle x \rangle$ agree to well within one standard deviation. The agreement is not perfect because the probability distribution in example 1.3 is just an estimate of the true distribution. In the limit that the number of measurements is very large, and the resolution of $P\left(x_j\right)$ is very fine, the estimated distribution will approach the true distribution, and the mean calculated using the probabilities will approach the mean determined directly from the data.

1.1.3 Continuous Probability Distributions

So far we've talked about probabilities that are determined at discrete values of x, for example, $P\left(x_j\right)$ for $j = 1, 2, \ldots, M$. When talking about real data we always have a finite measurement resolution, so measured probability distributions will always be discrete. Theoretically, however, it is possible to discuss probability distributions of a continuous variable.

Consider the position of a particle x. Position is a continuous variable and, in principle, the particle can be anywhere. Since it can be *anywhere*, the probability that it will be in any *particular* place is zero. For example, the probability of finding the particle at *exactly* $x = 9.999\ldots$ is zero—it doesn't make sense to talk about it. What it does make sense to talk about is the probability of finding the particle within some range of positions, say between $x = 9.99$ and $x = 10.00$. If the range is small, the probability of finding the particle will be proportional to the range (e.g., a particle is twice as likely to be found within a 2 μm interval than within a 1 μm interval). If we take dx to be a small–length interval (in the sense of a differential), then the probability $P\left(x\right)$ that the particle will be found between x and $x+dx$ is

$$P\left(x\right) = p\left(x\right)dx, \tag{1.21}$$

where $p\left(x\right)$ is called the probability density of x.[2] Continuous distributions are normalized by integrating the probability density over its entire range. If a particle can be found anywhere between $x = -\infty$ and $x = \infty$, the normalization condition is

$$\int_{-\infty}^{\infty} p\left(x\right)dx = 1. \tag{1.22}$$

For discrete probabilities we calculate the mean by weighting a value by its probability, and summing over all possible values. For continuous distributions we do the same thing, except that the sum is replaced by an integral. We thus have

$$\left\langle x\right\rangle = \int_{-\infty}^{\infty} xp\left(x\right)dx, \tag{1.23}$$

and

$$\left\langle f\left(x\right)\right\rangle = \int_{-\infty}^{\infty} f\left(x\right)p\left(x\right)dx. \tag{1.24}$$

If we wish to calculate the standard deviation of x, we do it the same way as we did in the discrete case, using eq. (1.9).

---

[2] The probability on the left-hand side of eq. (1.21) is dimensionless, while the interval dx has units of length (m). For the units in eq. (1.21) to work out, then p (x) must have units of inverse length (m^{-1}). A probability density always has the inverse units of its argument.

1.1.4 Joint and Conditional Probabilities

Sometimes we're interested in the probability of more than one thing. For example, we want to know the probability that a particle has a particular value of x *and* a particular value of y (i.e., it has both of these values simultaneously). We denote this probability by $P(x,y)$, and refer to it as the joint probability of x and y.

The joint probability density $p(x,y)$, has the property that integrating with respect to one variable yields the probability density of the other. For example,

$$p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy. \tag{1.25}$$

Here we say that $p(x)$ is the marginal probability density of x. Averages are obtained by integrating with respect to both variables:

$$\langle f(x,y) \rangle = \int_{-\infty}^{\infty} f(x,y)\,p(x,y)\,dxdy. \tag{1.26}$$

In some situations we want to know the probability of obtaining x, given a particular value of y. We denote this probability by $P(x \mid y)$, and call it a conditional probability, because it represents the probability of x conditioned on y. The joint and conditional probabilities are related by

$$P(x,y) = P(x \mid y)P(y). \tag{1.27}$$

For more details see ref. [1.1], sec. 3.7.

1.2 LINEAR ALGEBRA

1.2.1 Vectors and Basis Sets

We'll use what you already know about vectors from first-year physics to develop a vocabulary for linear algebra, and then we'll build on that.

We can express a vector $\boldsymbol{a}$ (we'll denote vectors with bold type), that "lives" in the x-y plane as

$$\boldsymbol{a} = a_x \boldsymbol{u}_x + a_y \boldsymbol{u}_y, \tag{1.28}$$

where $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$ are dimensionless unit vectors that point in the positive x- and y-directions respectively.[3] *Any* vector in the x-y plane can be expressed by giving its components (coordinates) in the $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$ directions; in eq. (1.28) the components are a_x and a_y. The vectors $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$ are called basis vectors, and they make up a basis set. This basis set contains two vectors, because it is a 2-D vector space (an N-dimensional vector space must have N basis vectors).

---

[3] We are not using the more familiar notion of $\hat{\boldsymbol{i}}$ and $\hat{\boldsymbol{j}}$, or $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$, to denote the unit vectors because we will reserve the caret symbol ^ for something else.

Vectors are a shorthand notation. The symbol $\boldsymbol{a}$ is a convenient way to denote an object that has multiple components. To specify the 2-D vector $\boldsymbol{a}$ we need to specify two numbers, its components a_x and a_y. Equation (1.28) is one particular notation for doing this, but there are others. For example, we could specify $\boldsymbol{a}$ as a row vector

$$\boldsymbol{a} \doteq \left(a_x, a_y\right), \tag{1.29}$$

or a column vector

$$\boldsymbol{a} \doteq \begin{pmatrix} a_x \\ a_y \end{pmatrix}. \tag{1.30}$$

As long as we know what the components of $\boldsymbol{a}$ are, we know what the vector $\boldsymbol{a}$ is. There are different notations for $\boldsymbol{a}$, but they all represent the same vector.[4]

In terms of column vectors, the basis vectors in the x-y plane can be written

$$\boldsymbol{u}_x \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \boldsymbol{u}_y \doteq \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.31}$$

The column vector equivalent of eq. (1.28) is thus

$$\boldsymbol{a} \doteq a_x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a_x \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_y \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix}. \tag{1.32}$$

The row vector equivalent can be written out in a similar manner.

One nice thing about the notation of eq. (1.28) is that it explicitly contains the basis vectors $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$; there is no ambiguity about what the basis vectors are. However, the row and column vectors of eqs. (1.29) and (1.30) do not explicitly reference the basis vectors. The reader needs to know what basis is being used, because if the basis changes the row and column vectors change. This is best illustrated with a specific example.

Let's say that in the x-y coordinate system

$$\boldsymbol{a} = 2\boldsymbol{u}_x + 2\boldsymbol{u}_y \doteq \begin{pmatrix} 2 \\ 2 \end{pmatrix}. \tag{1.33}$$

This vector is displayed graphically in fig. 1.2(a). We can also express $\boldsymbol{a}$ in the x'-y' coordinate system, which is rotated from the x-y system by 45°, as shown in fig. 1.2(b). In this coordinate system

$$\boldsymbol{a} = 2\sqrt{2}\boldsymbol{u}'_x \doteq \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}. \tag{1.34}$$

---

[4] In eqs. (1.29) and (1.30) we didn't use $=$, but rather the symbol $\doteq$. We'll use $\doteq$ to denote "is represented by" as opposed to "is equal to". The reasons for doing this will be detailed below. We're borrowing this notation from ref. [1.2].

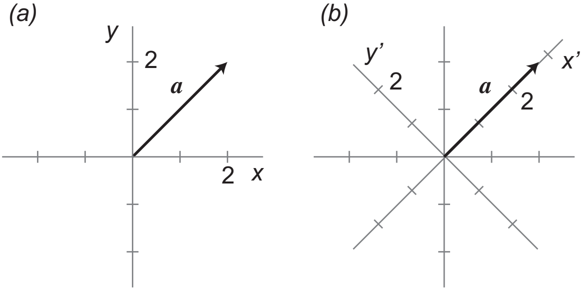


Fig 1.2 The vector $\boldsymbol{a}$ represented in: (a) the *x-y* coordinate system, (b) the *x′-y′* coordinate system.

As can be seen in fig. 1.2, $\boldsymbol{a}$ has not changed—it has the same length and points in the same direction. In eqs. (1.33) and (1.34) we have merely expressed $\boldsymbol{a}$ using two different basis sets. If we use the unit vector notation there is no ambiguity about what coordinate system we're using. If we use the column vector notation, however, there is no indication of what basis we're talking about. If we simply see $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$, how do we know whether this is expressed in the x-y basis or the x'-y' basis? From now on, if there is potential for confusion, we will attempt to remove this ambiguity by placing a subscript on the vectors to indicate which basis they are being expressed in. For example, we will write these vectors as $\begin{pmatrix} 2 \\ 2 \end{pmatrix}_{x,y}$ and $\begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}_{x',y'}$.

The difference between a vector, and its representation as a row or column vector, is why we're using the symbol $\doteq$ to mean "is represented by". The vector itself is always the same (i.e., it is independent of the representation), but it is represented differently depending upon which basis we use.

1.2.2 The Inner Product

We know that we can write the dot product of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as

$$\boldsymbol{a} \bullet \boldsymbol{b} = a_x b_x + a_y b_y + a_z b_z. \tag{1.35}$$

The dot product is a way of "multiplying" two vectors. It's also called the scalar product because the result of this operation is a scalar, not a vector. An operation which combines two vectors to produce a scalar is generally referred to as an inner product. To compute the inner product we multiply the components of the two vectors, and then add up all the products. In terms of row and column vectors, the inner product of $\boldsymbol{a}$ and $\boldsymbol{b}$ would be written as

$$\boldsymbol{a}\bullet\boldsymbol{b} \doteq \left(a_x, a_y, a_z\right)\begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = a_x b_x + a_y b_y + a_z b_z. \tag{1.36}$$

Since we'll be dealing with vector spaces that are more general than three-dimensional real space, it won't always be convenient to use the subscripts *x*, *y*, and *z* to denote the coordinates. Instead, we'll often use numbers to denote the coordinates. For example

$$\boldsymbol{a} \doteq \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}. \tag{1.37}$$

We've written the inner product of eq. (1.36) in terms of the product of a row vector and a column vector, with the row vector written on the left. This is the way we have to write inner products in terms of row and column vectors. We cannot directly take the inner product of two row or two column vectors. If we have two column vectors, for example, we must first convert one to a row vector before taking the inner product. There are two steps to doing this: first, we write the elements of the column vector as a row vector; then we take the complex conjugate of each of the elements. The row vector corresponding to the column vector in eq. (1.37) is

$$\boldsymbol{a} \doteq \left(a_1^*, a_2^*, a_3^*\right). \tag{1.38}$$

We convert a row vector to a column vector in the same manner. You're probably used to thinking about real vectors that live in real space, so you wouldn't ordinarily think about taking the complex conjugate. However, in quantum mechanics the vectors we'll deal with are in general complex, and we have to keep this in mind.

Why do we take the complex conjugate when switching to a row vector from a column vector, or vice versa? Imagine taking the inner product of a vector with itself. To do this we need two versions of the vector, one a row and the other a column. The inner product is then

$$\left(a_1^*, a_2^*, a_3^*\right)\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = a_1^* a_1 + a_2^* a_2 + a_3^* a_3 = \left|a_1\right|^2 + \left|a_2\right|^2 + \left|a_3\right|^2. \tag{1.39}$$

The complex conjugate ensures that we get a real number when we take the inner product of a vector with itself. The square root of the inner product of a vector with itself is called the norm of the vector. The norm is a measure of the "length" of the vector, which is why we desire that it always be a positive number. A vector is normalized if its norm (or consequently its norm squared) is equal to 1.

Ordinarily we think of the components of a vector as being the elements of its column vector representation. With this in mind, the inner product of two vectors, $\boldsymbol{a}$ and $\boldsymbol{b}$, in a general N-dimensional space can be written as

$$\left(a_1^*, \cdots, a_N^*\right)\begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} = \sum_{i=1}^{N} a_i^* b_i. \tag{1.40}$$

If we already know the row vector representation, we don't need to take the complex conjugate *again* to compute the inner product. We only need to do this when switching from a column vector to a row vector.

In ordinary three-dimensional (3-D) space we think of vectors as being orthogonal if they make a 90° angle with respect to each other. Equivalently, we say that two vectors are orthogonal if their dot product is 0. In the more general vector spaces we will be talking about in quantum mechanics, two vectors are orthogonal if their inner product is equal to 0.

Again, thinking about 3-D space, we almost always work in coordinate systems where our basis vectors are mutually orthogonal, and we also find it convenient to use normalized basis vectors. If we have a basis set in which all of the vectors are mutually orthogonal, and normalized, we have an orthonormal basis set. For example, the unit vectors $\boldsymbol{u}_x$, $\boldsymbol{u}_y$, and $\boldsymbol{u}_z$ in 3-D are an orthonormal basis set.

1.2.3 Matrices

In our discussion of matrices we'll focus on square matrices (i.e., matrices which have the same number of rows and columns), because 99% of the time in quantum mechanics that's what we deal with. I'll assume that you know how to take the determinant of such a matrix. I'll also assume that you know how to multiply a matrix and a vector, or two matrices. However, to be clear on notation we'll present some expressions, which you might not be familiar with, that describe these processes.

We'll use M_{ij} to denote the element in the i^{th} row and j^{th} column of the matrix $\bar{\mathbf{M}}$ (we'll use bold with an overbar to denote a matrix). Recall that the product of a matrix and column vector yields another column vector. In N-dimensions, the components of $\boldsymbol{b}$, which comes from the operation $\boldsymbol{b} = \bar{\mathbf{M}}\boldsymbol{a}$, can be written as the inner product of the rows of $\bar{\mathbf{M}}$ with $\boldsymbol{a}$:

$$b_i = \sum_{j=1}^{N} \mathrm{M}_{ij} a_j. \tag{1.41}$$

Since knowing the components determines the vector, eq. (1.41) expresses the multiplication of a matrix and a column vector. Similarly, we can express the elements of the matrix product $\bar{\mathbf{M}} = \bar{\mathbf{A}}\bar{\mathbf{B}}$ in terms of the inner product of the i^{th} row of the first matrix with the j^{th} column of the second:

$$\mathrm{M}_{ij} = \sum_{k=1}^{N} \mathrm{A}_{ik} \mathrm{B}_{kj}. \tag{1.42}$$

You're probably already aware, and in the problems you'll confirm, that in general $\bar{\mathbf{A}}\bar{\mathbf{B}} \neq \bar{\mathbf{B}}\bar{\mathbf{A}}$. The order in which we multiply matrices *matters*—matrix multiplication is not commutative. More typically, we'll say that matrices don't always commute. This has important implications in quantum mechanics, as we'll discuss later.

1.2.4 Eigenvalues and Eigenvectors

The following is a mathematical problem that arises frequently, often as the result of a physical problem: We are given a matrix $\bar{\mathbf{M}}$, and we want to find the vectors $\boldsymbol{x}$ and constants λ that are solutions of the equation

$$\bar{\mathbf{M}}\boldsymbol{x} = \lambda \boldsymbol{x}. \tag{1.43}$$

In other words, there are some special vectors such that when we multiply them by the matrix $\bar{\mathbf{M}}$, give us back a scaled version of themselves. The vectors that are solutions to eq. (1.43) are called the eigenvectors of $\bar{\mathbf{M}}$, and the corresponding constants are called the eigenvalues of $\bar{\mathbf{M}}$.[5]

How do we solve this problem? Start by rewriting eq. (1.43) as

$$\bar{\mathbf{M}}\boldsymbol{x} - \lambda \boldsymbol{x} = 0. \tag{1.44}$$

In order to factor out the vector $\boldsymbol{x}$, we need to introduce the identity matrix $\bar{\mathbf{1}}$, which has 1s on the diagonal, and 0s everywhere else. In 3-D, for example, the identity is given by

$$\bar{\mathbf{1}} \doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.45}$$

Another way to express the identity matrix is to say that its elements are equal to the Kronecker delta δ_{ij}, which is defined as

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. \tag{1.46}$$

The identity matrix is useful because multiplying it times any vector (or matrix) simply returns the same vector (or matrix). For example, in 2-D

$$\bar{\mathbf{1}}\boldsymbol{x} \doteq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \doteq \boldsymbol{x}. \tag{1.47}$$

Using this fact, we can insert the identity into eq. (1.44) and obtain

---

[5] More generally this problem is expressed in terms of linear operators, as opposed to matrices, as we'll discuss in later chapters.

$$\bar{\mathbf{M}}\boldsymbol{x} - \lambda\boldsymbol{x} = \bar{\mathbf{M}}\boldsymbol{x} - \lambda\bar{\mathbf{1}}\boldsymbol{x} = \left(\bar{\mathbf{M}} - \lambda\bar{\mathbf{1}}\right)\boldsymbol{x} = 0. \tag{1.48}$$

Instead of continuing on generally, we'll specialize to 2-D, so we can explicitly write eq. (1.48) as

$$\begin{pmatrix} M_{11} - \lambda & M_{12} \\ M_{21} & M_{22} - \lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1.49}$$

This is equivalent to a system of 2 equations

$$\begin{aligned} \left(M_{11} - \lambda\right)x_1 + M_{12}x_2 &= 0\,, \\ M_{21}x_1 + \left(M_{22} - \lambda\right)x_2 &= 0\,, \end{aligned} \tag{1.50}$$

in 3 unknowns: λ, x_1 and x_2. There is a solution to this problem if and only if the determinant of the matrix in eq. (1.49) is 0:

$$\begin{vmatrix} M_{11} - \lambda & M_{12} \\ M_{21} & M_{22} - \lambda \end{vmatrix} = \left(M_{11} - \lambda\right)\left(M_{22} - \lambda\right) - M_{12}M_{21} = 0. \tag{1.51}$$

We'll just outline the rest of the procedure before doing a specific example. Begin by noticing that eq. (1.51) is 2nd-order in λ. This means that there are 2 solutions, which we'll call λ_a and λ_b. Once these eigenvalues have been found, each will have its own corresponding eigenvector, $\boldsymbol{x_a}$ or $\boldsymbol{x_b}$. To find the eigenvectors, substitute an eigenvalue into the original equation, then solve for the corresponding eigenvector. For example:

$$\begin{pmatrix} M_{11} - \lambda_a & M_{12} \\ M_{21} & M_{22} - \lambda_a \end{pmatrix}\begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = 0. \tag{1.52}$$

This is equivalent to two linear equations in two unknowns, x_{a1} and x_{a2}, so we can solve for them

$$\begin{aligned} \left(M_{11} - \lambda_a\right)x_{a1} + M_{12}x_{a2} &= 0\,, \\ M_{21}x_{a1} + \left(M_{22} - \lambda_a\right)x_{a2} &= 0\,. \end{aligned} \tag{1.53}$$

In a general N-dimensional problem there will be N eigenvalues and N eigenvectors.

There's one last trick. The equations that determine the eigenvectors [e.g., eq. (1.53)] are *not* independent (the solutions for the eigenvalues ensure that this is the case). This means that there is no unique solution to the problem; we can solve for x_{a2} in terms of x_{a1}, or vice versa, but neither is uniquely determined. That this must be so is seen in the fact that if $\boldsymbol{x}_a$ is a solution to our original problem, eq. (1.43), then any constant times $\boldsymbol{x}_a$ is also a solution—there are an infinite number of $\boldsymbol{x}_a$'s corresponding to λ_a. We are thus free to impose one more constraint in order to obtain a unique solution. Typically we choose to normalize the eigenvectors. This means

$$\left(x_{a1}^*, x_{a2}^*\right)\begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = \left|x_{a1}\right|^2 + \left|x_{a2}\right|^2 = 1. \tag{1.54}$$

It’s probably easiest to learn how to find eigenvalues and eigenvectors using an example.

EXAMPLE 1.4

Find the eigenvalues and eigenvectors of

$$\bar{\mathbf{M}} \doteq \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}. \tag{1.55}$$

First we find the eigenvalues by subtracting λ from each of the diagonal terms, and setting the determinant equal to 0 [eq. (1.51)]:

$$\begin{vmatrix} 0-\lambda & 1 \\ -2 & 3-\lambda \end{vmatrix} = 0, \tag{1.56}$$

which yields

$$-\lambda(3-\lambda) + 2 = \lambda^2 - 3\lambda + 2 = 0. \tag{1.57}$$

Solving this quadratic equation yields two eigenvalues:

$$\lambda_{\pm} = \frac{-(-3) \pm \sqrt{(-3)^2 - 4(1)(2)}}{2(1)}, \tag{1.58}$$

$$\lambda_+ = 2, \ \lambda_- = 1. \tag{1.59}$$

To find the eigenvector corresponding to $\lambda_+ = 2$, we know that [eq. (1.52)]

$$\begin{pmatrix} 0-2 & 1 \\ -2 & 3-2 \end{pmatrix}\begin{pmatrix} x_{+1} \\ x_{+2} \end{pmatrix} = 0. \tag{1.60}$$

This yields 2 equations in 2 unknowns:

$$\begin{aligned} -2x_{+1} + x_{+2} &= 0, \\ -2x_{+1} + x_{+2} &= 0. \end{aligned} \tag{1.61}$$

Notice that these are the same equation, so we solve it to obtain $x_{+2} = 2x_{+1}$. An eigenvector corresponding to $\lambda_+ = 2$ is thus:

$$\boldsymbol{x}_+ \doteq \begin{pmatrix} x_{+1} \\ 2x_{+1} \end{pmatrix} = x_{+1}\begin{pmatrix} 1 \\ 2 \end{pmatrix}. \tag{1.62}$$

We can normalize this using

$$\left(x_{+1}\right)^2\left(1^2+2^2\right)=1,\ x_{+1}=\frac{1}{\sqrt{5}}. \tag{1.63}$$

So, one eigenvalue-eigenvector pair is

$$\lambda_+ = 2,\ \boldsymbol{x}_+ \doteq \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}. \tag{1.64}$$

The eigenvector corresponding to $\lambda_- = 1$ is found in a similar fashion; it is

$$\boldsymbol{x}_- = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{1.65}$$

1.3 References

[1.1] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics with Applications, 7th ed.* (Pearson Prentice Hall, Upper Saddle River, NJ, 2004).

[1.2] J.J. Sakurai and J. Napolitano, *Modern Quantum Mechanics, 2nd ed.* (Addison-Wesley, Boston, MA, 2011).

1.4 PROBLEMS

Note: In several of these problems you will be asked to calculate the variance, or the standard deviation. If you use your calculator or a spreadsheet, make sure you know whether it uses $1/N$ or $1/(N-1)$. If it uses $1/(N-1)$, convert to $1/N$.

Data A (x_i): 10, 13, 14, 14, 6, 8, 7, 9, 12, 14, 13, 11, 10, 7, 7
Data B (x_i,y_i): (3,4), (5,8), (4,4), (8,5), (3,5), (4,5), (5,8), (8,5), (8,4), (3,4), (3,8), (4,8)

1.1 Calculate the mean and variance of Data A directly from the data.

1.2 Create a histogram of Data A, and from it estimate the probability distribution. Use min and max values for the data range of 5 and 14, and use 5 bins (5 and 6 go in the first bin; label it x_i=5.5, etc.).

1.3 Use the probability distribution from problem 1.2 to calculate the mean of

Data A.

1.4 Calculate the variance from the probability distribution of problem 1.2 using two different expressions:

(a) $\left\langle (\delta x)^2 \right\rangle = \left\langle \left(x - \langle x \rangle\right)^2 \right\rangle$, and

(b) $\left\langle (\delta x)^2 \right\rangle = \left\langle x^2 \right\rangle - \langle x \rangle^2$.

Do all the values for the variance (these two, and that from problem 1.1) agree? Why or why not?

1.5 For Data B, determine $P(x)$ and $P(y)$ for each of the measured values of x and y.

1.6* For Data B, determine $P(x=5, y=8)$ and $P(x=5 \,|\, y=8)$. Verify that $P(x,y) = P(x \,|\, y) P(y)$ for $x = 5$ and $y = 8$.

1.7 Find the constant c that normalizes the probability density $p(x) = c \dfrac{1}{(5-3x)^2 + 4}$ over the range $-10 \le x \le 10$. Compute $\langle x \rangle$ and Δx for this distribution.

1.8 Find the constant c that normalizes the probability density

$$p(x) = \begin{cases} ce^{-\alpha x} & x \ge 0 \\ 0 & x < 0 \end{cases},$$

where α is a positive, real constant. Compute $\langle x \rangle$ and Δx for this distribution.

1.9 Compute $\langle x \rangle$ and Δx for the (already normalized) probability density $p(x) = \dfrac{1}{\beta\sqrt{\pi}} e^{-(x-\alpha)^2/\beta^2}$, where α and β are positive, real constants.

1.10 Compute

$$\begin{pmatrix} 5 & 0 & 4 \\ 3 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} =$$

Verify that eq. (1.41) properly computes this product.

1.11 Compute

$$\begin{pmatrix} 5 & 0 & 4 \\ 3 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 7 \\ 8 & 0 & 1 \\ 3 & 0 & 0 \end{pmatrix} =$$

Verify that eq. (1.42) properly computes this product.

1.12* Compute

$$\begin{pmatrix} 5 & 0 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 3 & 7 \end{pmatrix} =$$

Are they the same?

1.13 Find the determinant

$$\begin{vmatrix} 4 & 2 \\ 1 & 8 \end{vmatrix} =$$

1.14 Find the determinant

$$\begin{vmatrix} 5 & 0 & 4 \\ 3 & 1 & 0 \\ 2 & 0 & 2 \end{vmatrix} =$$

1.15 For the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

(a) Find the eigenvalues and normalized eigenvectors.
(b) Prove that the eigenvectors are orthogonal.

1.16* For the matrix

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

(a) Find the eigenvalues and normalized eigenvectors.
(b) Prove that the eigenvectors are orthogonal.

CHAPTER 2

Classical Description of Polarization

This chapter describes the classical theory of the polarization of an electromagnetic wave. Here we'll learn how to represent polarization as a 2-dimensional (2-D) vector, and how optical elements which affect the polarization are represented as matrices. By combining these concepts, we'll be able to calculate how a combination of optical elements will transform the polarization of a wave.

We're exploring classical polarization now, because in chapter 3 we'll be talking about the quantum mechanics of polarization. It's easier to understand the quantum mechanics if you already know the corresponding classical physics.

2.1 POLARIZATION

2.1.1 The Polarization Vector

An electromagnetic wave, such as a light wave, consists of propagating electric and magnetic fields. The polarization of the wave is determined by the direction of the electric field vector. The electromagnetic field is a transverse wave, so the electric field $\boldsymbol{\mathcal{E}}$ is perpendicular to the direction of propagation. Assume that a wave is propagating in vacuum in the z-direction, so that its wave vector is given by $\boldsymbol{k} = k\boldsymbol{u}_z$. The magnitude of the wave vector is related to the wavelength λ, frequency f, and angular frequency ω of the wave by

$$k = \frac{2\pi}{\lambda} = \frac{2\pi f}{c} = \frac{\omega}{c}, \tag{2.1}$$

where c is the speed of light in vacuum.[1] The electric field can be written as

---

[1] The angular frequency is often referred to simply as the frequency. You should be able to tell which frequency is meant from the context.

$$\mathcal{E} = \mathcal{E}_x \boldsymbol{u}_x + \mathcal{E}_y \boldsymbol{u}_y, \tag{2.2}$$

where

$$\mathcal{E}_x = \mathcal{E}_{0x} \cos(kz - \omega t) \text{ and } \mathcal{E}_y = \mathcal{E}_{0y} \cos(kz - \omega t + \phi). \tag{2.3}$$

Here $\mathcal{E}_{0x}$ and $\mathcal{E}_{0y}$ are the amplitudes of the fields in the x- and y-directions.[2] The phase shift ϕ represents the fact that these two field components do not necessarily oscillate with the same phase. The polarization of the wave is determined by the relative magnitudes of $\mathcal{E}_{0x}$ and $\mathcal{E}_{0y}$, and by the phase shift.

In eq. (2.3) we have written the field in terms of cosine functions because it is a real, physical quantity. However, it is more convenient to write the field in terms of complex exponentials, with the understanding that we can always take the real part whenever we are interested in finding the true field. With this convention the components of the field become

$$\mathcal{E}_x = \mathcal{E}_{0x} e^{i(kz-\omega t)} \text{ and } \mathcal{E}_y = \mathcal{E}_{0y} e^{i(kz-\omega t+\phi)}, \tag{2.4}$$

and the total field is

$$\mathcal{E} = \mathcal{E}_{0x} e^{i(kz-\omega t)} \boldsymbol{u}_x + \mathcal{E}_{0y} e^{i(kz-\omega t+\phi)} \boldsymbol{u}_y. \tag{2.5}$$

The amplitude $\mathcal{E}_0$ of the field $\mathcal{E}$ is

$$\mathcal{E}_0 = \left(\mathcal{E}_{0x}{}^2 + \mathcal{E}_{0y}{}^2\right)^{1/2}. \tag{2.6}$$

In terms of this amplitude, the field is

$$\mathcal{E} = \mathcal{E}_0 e^{i(kz-\omega t)} \left[\frac{\mathcal{E}_{0x}}{\mathcal{E}_0} \boldsymbol{u}_x + \frac{\mathcal{E}_{0y}}{\mathcal{E}_0} e^{i\phi} \boldsymbol{u}_y \right]. \tag{2.7}$$

We define the polarization vector $\boldsymbol{\varepsilon}$ to be equal to the quantity in brackets in eq. (2.7):

$$\boldsymbol{\varepsilon} = \left[\frac{\mathcal{E}_{0x}}{\mathcal{E}_0} \boldsymbol{u}_x + \frac{\mathcal{E}_{0y}}{\mathcal{E}_0} e^{i\phi} \boldsymbol{u}_y \right]. \tag{2.8}$$

In general this is a complex vector; it is also a unit vector:

$$|\boldsymbol{\varepsilon}| = \left[\boldsymbol{\varepsilon}^* \bullet \boldsymbol{\varepsilon}\right]^{1/2} = \left[\left(\frac{\mathcal{E}_{0x}}{\mathcal{E}_0}\right)^2 + \left(\frac{\mathcal{E}_{0y}}{\mathcal{E}_0}\right)^2 \right]^{1/2} = 1. \tag{2.9}$$

The electric field can be written in terms of the polarization vector as

$$\mathcal{E} = \mathcal{E}_0 e^{i(kz-\omega t)} \boldsymbol{\varepsilon}. \tag{2.10}$$

---

[2] In Chapter 1 we defined $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$ to be unit vectors that point along the x- and y-directions.

In eq. (2.10) there is no x- or y-dependence—the field is constant over any plane perpendicular to the propagation direction (the z-axis). A field of this form is called a plane wave.

An optical detector does not respond directly to the electric field, but instead to the power incident on it. This is because electric fields oscillate very rapidly in the optical region of the spectrum. Detectors cannot follow these oscillations, and instead average over many oscillation periods. The power is proportional to the intensity I, which we'll define to be equal to the square magnitude of the field:

$$I \equiv |\mathcal{E}|^2 = \mathcal{E}^* \bullet \mathcal{E}. \tag{2.11}$$

The intensity doesn't fluctuate at optical frequencies.[3] This fact can be seen in

$$I = \mathcal{E}^* \bullet \mathcal{E} = \left(\mathcal{E}_0 e^{-i(kz-\omega t)} \boldsymbol{\varepsilon}^*\right) \bullet \left(\mathcal{E}_0 e^{i(kz-\omega t)} \boldsymbol{\varepsilon}\right) = \mathcal{E}_0^{\ 2}, \tag{2.12}$$

which has no oscillatory time dependence. The intensity follows slow (compared to the detector response time) fluctuations in the field amplitude $\mathcal{E}_0$, but not oscillations at optical frequencies. Detectors effectively respond to the intensity defined in eq. (2.11).

2.1.2 Linear Polarization

Consider the special case of a zero phase shift ($\phi = 0$) between the x- and y-components of the field. In this case the polarization vector is real, and is equal to

$$\boldsymbol{\varepsilon} = \left[\frac{\mathcal{E}_{0x}}{\mathcal{E}_0} \boldsymbol{u}_x + \frac{\mathcal{E}_{0y}}{\mathcal{E}_0} \boldsymbol{u}_y\right]. \tag{2.13}$$

This vector describes a line which makes an angle of θ,

$$\theta = \tan^{-1}\left(\frac{\mathcal{E}_{0y}}{\mathcal{E}_{0x}}\right), \tag{2.14}$$

with respect to the x-axis. Because the polarization vector points along a line, the field is said to be linearly polarized.

Figure 2.1 shows a snapshot of a linearly polarized (along $\boldsymbol{u}_y$) wave propagating in the z-direction. As time advances the whole wave slides along the z-axis, so that in any given plane perpendicular to the z-axis the electric field oscillates back and forth. To observe simulations of fields with different polarizations propagating in time, see ref. [2.1].

2.1.3 Circular Polarization

Now consider the special case of $\phi = \pi/2$, and $\mathcal{E}_{0x} = \mathcal{E}_{0y} = \mathcal{E}_0 / \sqrt{2}$. The polarization vector is then

---

[3] In this text we are defining the intensity by eq. (2.11). Other texts define the intensity I' to be equal to the average power per unit area of an electromagnetic wave. These two definitions are proportional to each other, $I' = (1/2)\, c\varepsilon_0 I$, where ε_0 is the permittivity of free space, so they behave in the same manner.

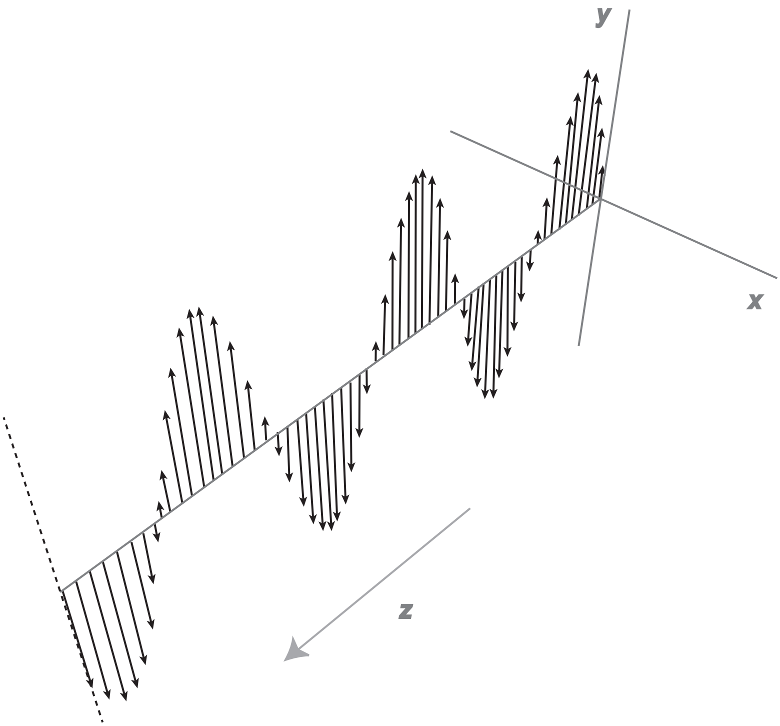


Fig 2.1 A linearly polarized wave propagating in the *z*-direction.

$$\boldsymbol{\varepsilon} = \frac{1}{\sqrt{2}}\left(\boldsymbol{u}_x + i\boldsymbol{u}_y\right). \tag{2.15}$$

How can we make sense of this complex vector? Remember that to find the real field, we need to take the real part. However, the polarization vector is constant in both time and space, so taking its real part just leaves a constant factor, which yields no information about the oscillations of the field. We need to go back to the real, physical, temporally and spatially varying fields described in eq. (2.3). With the assumptions we're making here, these fields become

$$\mathcal{E}_x = \frac{\mathcal{E}_0}{\sqrt{2}}\cos\left(kz - \omega t\right), \tag{2.16}$$

$$\mathcal{E}_y = \frac{\mathcal{E}_0}{\sqrt{2}}\cos\left(kz - \omega t + \frac{\pi}{2}\right) = -\frac{\mathcal{E}_0}{\sqrt{2}}\sin\left(kz - \omega t\right). \tag{2.17}$$

These fields are equal-amplitude, sinusoidally oscillating waves that are phase shifted with respect to each other. To examine how the fields oscillates in time, in a plane perpendicular to the propagation direction, we set $z = 0$, and obtain

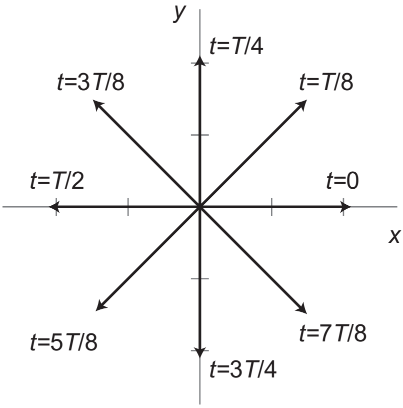


Fig 2.2 The direction of the electric field vector at different times for a left-circuarly polarized wave that is propagating out of the page.

$$\mathcal{E}_x = \frac{\mathcal{E}_0}{\sqrt{2}}\cos\left(-\omega t\right) = \frac{\mathcal{E}_0}{\sqrt{2}}\cos\left(\omega t\right), \tag{2.18}$$

$$\mathcal{E}_y = -\frac{\mathcal{E}_0}{\sqrt{2}}\sin\left(-\omega t\right) = \frac{\mathcal{E}_0}{\sqrt{2}}\sin\left(\omega t\right). \tag{2.19}$$

Figure 2.2 shows the total electric field vector, obtained from eqs. (2.18) and (2.19), at different times. The time in fig. 2.2 is expressed as fractions of the oscillation period $T = 2\pi/\omega$. The electric field sweeps out a circle as it rotates counterclockwise, so the field is circularly polarized. The field makes one full rotation during each oscillation period T.

Circular polarization comes in two flavors: left-circular and right-circular. Figure 2.2 represents left-circularly polarized light: point your left thumb toward the source (into the page), and your fingers will curl in the direction of rotation. This convention is exactly *opposite* to what you might expect: your thumb must point toward the source, which is opposite the direction of propagation. Figure 2.3 shows a 3-D representation of a propagating left-circularly polarized wave.

2.1.4 Elliptical and Random Polarization

Linear and circular polarizations are special cases. If the wave has a well-defined polarization, but neither of these special cases is satisfied, the wave is elliptically polarized. The orientation and eccentricity of the ellipse are determined by the ratio of the amplitudes of the x- and y-components, and the phase shift.

If the polarization fluctuates randomly in time, the polarization is said to be random, or the field is said to be unpolarized. Because this is a random process, we can't specify

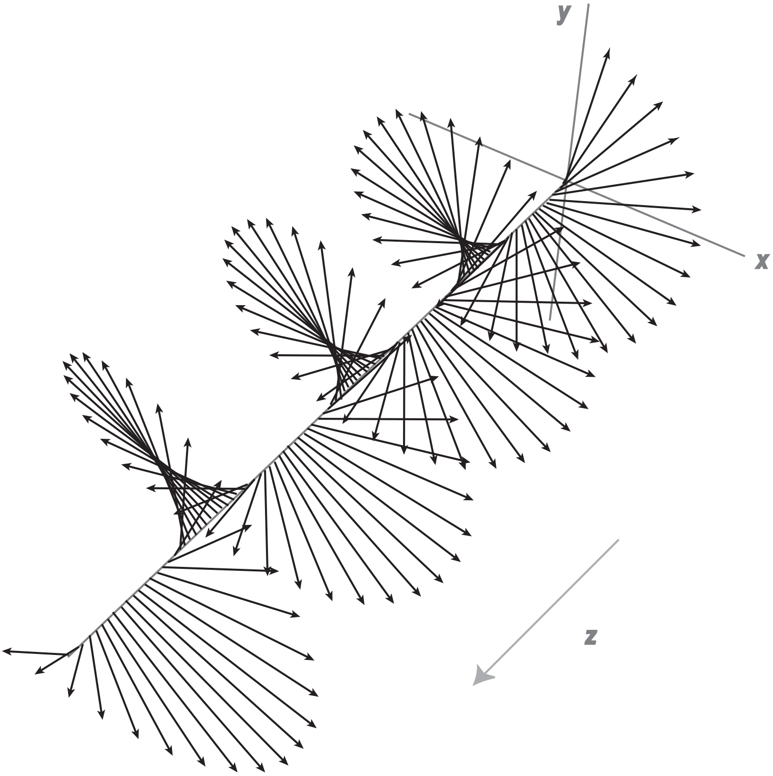


Fig 2.3 A left-circularly polarized wave propagating in the *z*-direction.

exactly the parameters of the wave; however, we can describe the statistics of the fluctuations. One common way to do this is to use the Stokes parameters (ref. [2.2], sec. 8.13).

2.2 BIREFRINGENCE

2.2.1 Birefringent Materials

There are certain materials, called birefringent materials, which are extremely useful in modifying the polarization of a wave. Before getting into details, we'll start with a brief review of wave propagation in media.

A medium such as glass or water has an index of refraction n associated with it. The index of refraction is determined by the speed of light in the medium v, and is given by $n = c/v$. Inside a medium of index of refraction n, eq. (2.1) becomes:

$$k = \frac{\omega}{v} = \frac{n\omega}{c} = \frac{n2\pi}{\lambda}. \tag{2.20}$$

The phase shift between two points on a wave, separated by a distance z, is

$$\phi = kz = \frac{n2\pi z}{\lambda}. \tag{2.21}$$

The phase shift depends on the index of refraction. For a "normal" material such as glass, the index of refraction depends on the frequency (wavelength) of the wave (a property known as dispersion), but not on the polarization or direction of propagation; such a material is said to be isotropic.

Birefringent materials are anisotropic—the index of refraction depends on the polarization. Such materials are crystalline in structure, having one or more preferred directions in space. In uniaxial crystals light polarized along a particular axis (called the optic axis) has one index of refraction, while light polarized perpendicular to the optic axis has a different index. In biaxial crystals there are three indices of refraction, associated with polarization in three different directions.

What causes this difference in the indices of refraction for different polarizations? Remember that crystalline structures have a very specific and ordered arrangement of atoms in the material. Imagine, for example, that the atoms are arranged in sheets, and the sheets are stacked on top of each other. It's not difficult to imagine that fields polarized parallel or perpendicular to the sheets will behave differently in such a crystal. For more details, see ref. [2.2], sec. 8.4.

We can always decompose the polarization of a wave into two orthogonal components. Because birefringent materials have preferred directions in space, it is natural to choose components associated with these directions. The direction associated with the lower index of refraction is referred to as the fast axis, because light polarized along that axis propagates at a faster speed. Conversely, the orthogonal direction is referred to as the slow axis; light polarized in this direction propagates at a slower speed.

2.2.2 Poynting Vector Walk-Off

At some point in your life you've probably come across a piece of calcite crystal and seen that when you place it on top of something, it creates two separated images (fig. 2.4). This is the phenomenon of "double refraction," which occurs in birefringent materials.

You might think that it's the two indices of refraction that lead to these two images. The explanation would then be that the two different indices cause rays to refract at two different angles because of Snell's law. This does happen, but it's not the whole story. Snell's law states

$$n_i \sin\theta_i = n_r \sin\theta_r, \tag{2.22}$$

where i refers to the incident beam and r refers to the refracted beam. The angles refer to the direction of the wave vector $\boldsymbol{k}$ with respect to the surface normal. If the incident beam is normal to the surface $\theta_i = 0$, then $\theta_r = 0$, independent of the indices of refraction, and the beam passes into the medium without bending. Yet, in a birefringent material the light *can* bend, even when Snell's law seems to indicate that it shouldn't. If $\boldsymbol{k}$ doesn't bend, what *does*?

The energy flow bends. Energy flows along the direction of the Poynting vector $\boldsymbol{S}$, which is given in terms of the electric and magnetic fields as

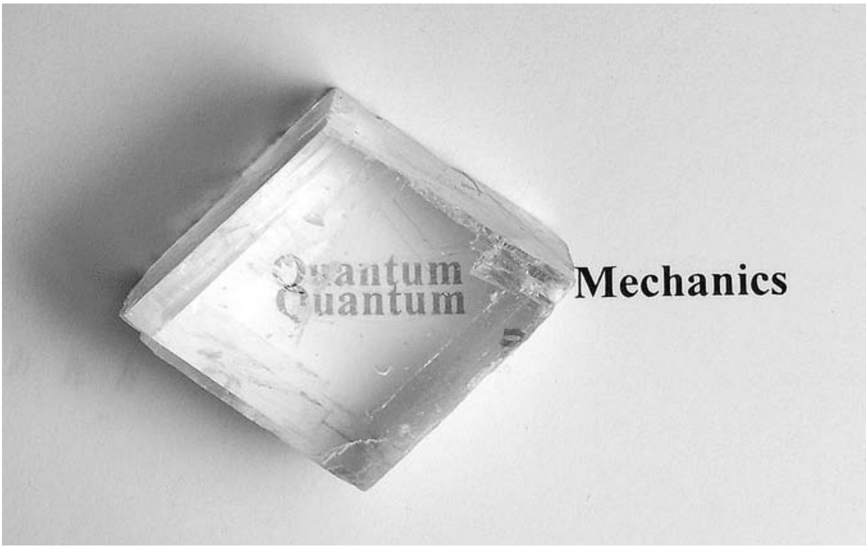


Fig 2.4 Double refraction creating 2 images in calcite.

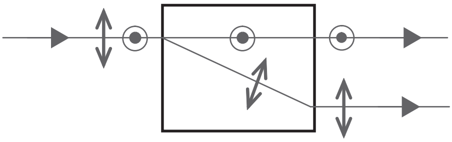


Fig 2.5 A beam of light containing both vertical and horizontal polarizations is split into two beams by Poynting vector walk-off. Here ⊙ denotes vertical polarization, and ↕ denotes horizontal polarization.

$$\mathcal{S} = \frac{1}{\mu}\left(\mathcal{E} \times \mathcal{B}\right), \tag{2.23}$$

where μ is the permeability of the medium. In an isotropic medium, $\boldsymbol{k} \parallel \mathcal{S}$ and $\boldsymbol{k} \perp \mathcal{E}$. In an anisotropic medium, light of one polarization (called the ordinary wave) satisfies $\boldsymbol{k} \parallel \mathcal{S}$ and $\boldsymbol{k} \perp \mathcal{E}$, but for light of the orthogonal polarization (called the extraordinary wave) these relations do not necessarily hold.

For an ordinary wave incident normally on a boundary, both $\boldsymbol{k}$ and $\mathcal{S}$ are transmitted straight through without bending. For an extraordinary wave, $\boldsymbol{k}$ doesn't bend, but $\mathcal{S}$ (and hence the beam itself) does. This phenomena is called Poynting vector walk-off. A general wave containing both polarization components will split into two pieces traveling in different directions, as shown in fig. 2.5. If the crystal faces are parallel, then the two outgoing beams emerge from the crystal displaced from each other, but parallel.

In this textbook we will often refer to a beam displacing polarizer (BDP, also called a beam displacing prism), such as in fig. 2.5, as a polarization analyzer, PA. In the case of fig. 2.5, the beam is split into vertical and horizontal components, so we will refer to it is as a PA_{HV}. If we were to rotate this device by 45° the beam would be analyzed into +45° and –45° linear polarization components, and we would refer to it as a PA_{45}.

2.3 MODIFYING THE POLARIZATION

Frequently we need to modify the polarization of a wave, and there are a number of optical elements which are useful for this purpose.

2.3.1 Linear Polarizers

A linear polarizer has a preferred direction, known as the polarization axis or transmission axis. This axis is typically oriented along a line perpendicular to the wave propagation direction. A linear polarizer projects the incident electric field onto its transmission axis, transmitting only the component of the field along this axis.

We can explicitly write down the electric field transmitted by a linear polarizer. Consider an electromagnetic wave incident on such a polarizer, with an electric field of the form

$$\mathcal{E}_i = \mathcal{E}_0 e^{i(kz-\omega t)}\boldsymbol{\varepsilon}, \tag{2.24}$$

where $\boldsymbol{\varepsilon}$ represents any arbitrary polarization. The transmission axis points along the direction of the unit vector $\boldsymbol{u}_\theta$, where θ is the angle between the transmission axis and the horizontal axis. (It's common to refer to the x-direction as horizontal, and the y-direction as vertical.) The transmitted field is

$$\mathcal{E}_t = \left(\mathcal{E}_i \cdot \boldsymbol{u}_\theta\right)\boldsymbol{u}_\theta = \mathcal{E}_0 e^{i(kz-\omega t)}\left(\boldsymbol{\varepsilon}\cdot\boldsymbol{u}_\theta\right)\boldsymbol{u}_\theta, \tag{2.25}$$

The output field is linearly polarized along $\boldsymbol{u}_\theta$, and the amplitude is reduced by a factor of $\left|\boldsymbol{\varepsilon}\cdot\boldsymbol{u}_\theta\right|$, as shown in in fig. 2.6.

For linearly polarized light that makes an angle ψ with the horizontal axis ($\boldsymbol{\varepsilon} = \boldsymbol{u}_\psi$), the amplitude of the field is reduced by a factor of $\left|\cos\left(\theta-\psi\right)\right|$. The intensity is proportional to the square of the field, so the intensity of the light decreases by a factor of $\cos^2\left(\theta-\psi\right)$. This is known as Malus's law. For a circularly polarized beam incident on a linear polarizer, the field amplitude is reduced by factor of $1/\sqrt{2}$, and the intensity is reduced by factor of 1/2, independent of the direction the polarizer axis.

2.3.2 Polarizing Beam Splitters

A device that splits a beam into orthogonal polarization components (usually horizontal and vertical linear polarizations) is called a polarizing beam splitter (PBS). A piece of birefringent material such as that shown in fig. 2.5 can be used as a PBS. Polarizing

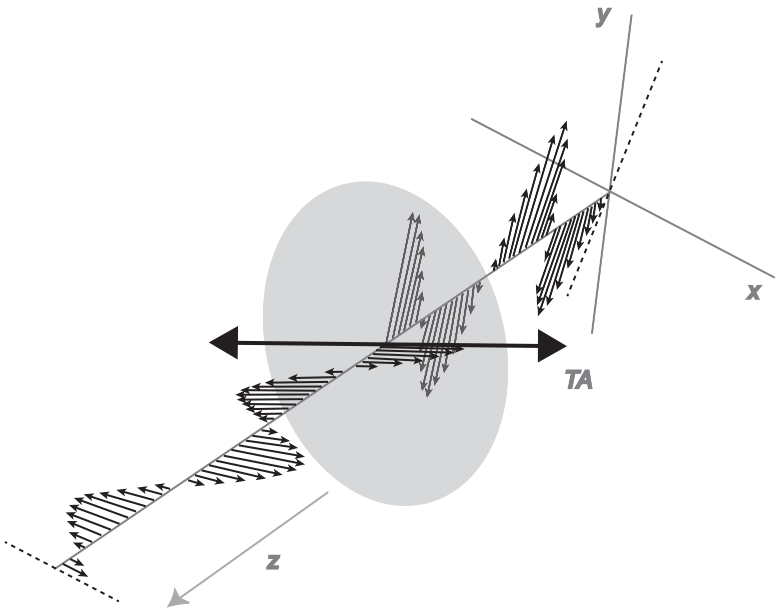


Fig 2.6 A linearly polarized wave propagating through a linear polarizer. The transmission axis of the polarizer is denoted by TA.

beam splitters can also be made using thin-film coatings; these are useful because the two beams emerge at right angles to each other.

2.3.3 Wave Plates

When propagating through a birefringent material of length l, the component of the polarization along the fast axis accumulates a phase of

$$\phi_f = \frac{n_f 2\pi l}{\lambda}, \tag{2.26}$$

where n_f is index of refraction along the fast axis. The component of the polarization along the orthogonal axis accumulates a phase of

$$\phi_s = \frac{n_s 2\pi l}{\lambda}, \tag{2.27}$$

where n_s is index of refraction along the slow axis. The wave thus acquires a relative phase shift between these two components of

$$\Delta\phi = \phi_s - \phi_f = \frac{\left(n_s - n_f\right)2\pi l}{\lambda}. \tag{2.28}$$

As described in sec. 2.1, the relative phase shift between two orthogonal polarization components is an important parameter in determining the polarization of a wave.

A wave plate is an optical element that uses a phase shift between orthogonal field components to transform the polarization of a wave. One important type of wave plate is called a quarter-wave plate. In a quarter-wave plate the phase shift between the fast and slow axes is $\Delta\phi = 2\pi/4 = \pi/2$, which corresponds to a quarter of a wavelength shift.[4] The other most common wave plate is a half-wave plate, in which the relative phase shift corresponds to half of a wavelength. Since the phase shift and the indices of refraction depend on wavelength, wave plates are designed to work properly at specific wavelengths.

Consider a vertically polarized beam incident on a quarter-wave plate. If the fast axis of the wave plate is parallel to the input polarization, there is only one relevant index of refraction, n_f, and the wave acquires a phase shift, but no alteration of its polarization. However, if the fast axis of the quarter-wave plate makes an angle of 45° from the horizontal, half the light is polarized along the fast axis and the other half is polarized along the slow axis. These two polarizations acquire a $\pi/2$ relative phase shift. In Sec. 2.1.3 we learned that a wave whose two components have equal amplitudes and a relative phase shift of $\pi/2$ is circularly polarized. Thus, a quarter-wave plate can transform a linearly polarized beam into a circularly polarized beam, as shown in fig. 2.7. It can also change a circularly polarized beam into a linearly polarized beam. The primary utility of a half-wave plate is that it can rotate linear polarization through an arbitrary angle; the rotation angle is determined by the orientation of the wave plate (proof of this comes below).

2.4 JONES VECTORS AND JONES MATRICES

2.4.1 Jones Vectors

It is conventional to take the horizontal and vertical linear polarization vectors to be $\boldsymbol{\varepsilon}_H = \boldsymbol{u}_x$ and $\boldsymbol{\varepsilon}_V = \boldsymbol{u}_y$. These vectors then serve as basis vectors, which we can use to describe other polarization vectors. For example, referring to eqs. (2.13) and (2.14), the polarization vector for a field making an angle of 45° with respect to the horizontal axis can be written as

$$\begin{aligned} \boldsymbol{\varepsilon}_{+45} &= \frac{1}{\sqrt{2}}\boldsymbol{u}_x + \frac{1}{\sqrt{2}}\boldsymbol{u}_y \\ &= \frac{1}{\sqrt{2}}\left(\boldsymbol{\varepsilon}_H + \boldsymbol{\varepsilon}_V\right). \end{aligned} \tag{2.29}$$

If the polarization of an arbitrarily polarized wave is written in the horizontal/vertical basis as a column vector, it is referred to as a Jones vector. Table 2.1 contains a list of common Jones vectors.

---

[4] For a quarter-wave plate, the phase shift is in general $\Delta\phi = 2\pi j + \pi/2$, where j is an integer. A wave plate in which $j = 0$ is referred to as a 0-order wave plate.

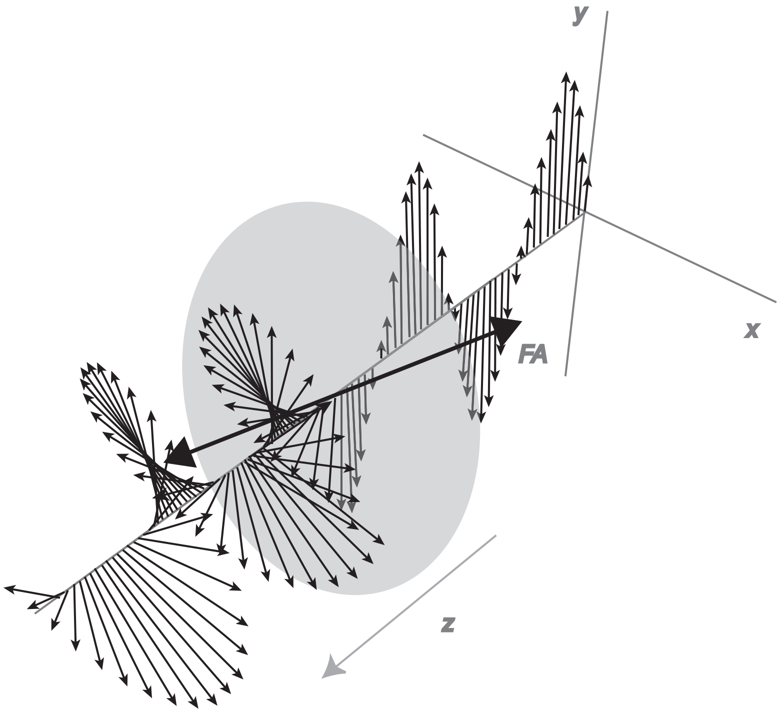


Fig 2.7 A linearly polarized wave is transformed to a circularly polarized wave by a quarter-wave plate. The fast axis of the wave plate is denoted by FA.

2.4.2 Jones Matrices

Certain types of optical elements, such as linear polarizers or wave plates, can modify the polarization of a wave. The mathematical objects that change one vector into another are matrices, so these polarization-changing elements are represented by matrices, called Jones matrices. I'll denote Jones matrices with the symbol $\overline{\mathbf{J}}$; table 2.2 lists some common Jones matrices. Examples 2.1 and 2.2 illustrate how Jones vectors and matrices can be used to describe the polarization behavior of beams interacting with optical elements.

EXAMPLE 2.1

What does a half-wave plate, whose fast axis makes an angle of θ with respect to horizontal, do to a beam incident with horizontal polarization?

From the tables, the Jones vector for the incident beam is

$$\boldsymbol{\varepsilon}_H = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{2.30}$$

and the Jones matrix for the wave plate is

$$\bar{\mathbf{J}}_{\lambda/2\,\theta} = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}. \tag{2.31}$$

The output polarization is

$$\begin{aligned} \boldsymbol{\varepsilon} &= \bar{\mathbf{J}}_{\lambda/2\,\theta}\boldsymbol{\varepsilon}_H \\ &= \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos 2\theta \\ \sin 2\theta \end{pmatrix}. \end{aligned} \tag{2.32}$$

Physically what does this polarization represent? From table 2.1, we see that this represents a linearly polarized beam whose polarization makes an angle of 2θ with respect to the horizontal.[5]

Table 2.1 Common Jones vectors.

Polarization	Polarization Vector	Jones Vector
Horizontal	$\boldsymbol{\varepsilon}_H$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
Vertical	$\boldsymbol{\varepsilon}_V$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
+45° linear	$\boldsymbol{\varepsilon}_{+45} = \frac{1}{\sqrt{2}}(\boldsymbol{\varepsilon}_H + \boldsymbol{\varepsilon}_V)$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
–45° linear	$\boldsymbol{\varepsilon}_{-45} = \frac{1}{\sqrt{2}}(\boldsymbol{\varepsilon}_H - \boldsymbol{\varepsilon}_V)$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$
Linear at angle θ w.r.t. horizontal	$\boldsymbol{\varepsilon}_\theta = \cos\theta\boldsymbol{\varepsilon}_H + \sin\theta\boldsymbol{\varepsilon}_V$	$\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$
Left circular	$\boldsymbol{\varepsilon}_L = \frac{1}{\sqrt{2}}(\boldsymbol{\varepsilon}_H + i\boldsymbol{\varepsilon}_V)$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}$
Right circular	$\boldsymbol{\varepsilon}_R = \frac{1}{\sqrt{2}}(\boldsymbol{\varepsilon}_H - i\boldsymbol{\varepsilon}_V)$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$

---

[5] Note that we used $=$, not $\doteq$, in these expressions. That's because Jones vectors and matrices are always represented in the same basis, so the vector and the representation truly are equal.

Table 2.2 Common Jones matrices, apart from overall, constant phase factors.

Optical Element	Symbol	Jones Matrix
Horizontal polarizer	$\overline{\mathbf{J}}_H$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
Vertical polarizer	$\overline{\mathbf{J}}_V$	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
Linear polarizer, transmission axis θ w.r.t. horizontal	$\overline{\mathbf{J}}_\theta$	$\begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}$
Quarter-wave plate, fast axis at ±45°	$\overline{\mathbf{J}}_{\lambda/4\pm45}$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & \mp i \\ \mp i & 1 \end{pmatrix}$
Quarter-wave plate, fast axis θ w.r.t. horizontal	$\overline{\mathbf{J}}_{\lambda/4\,\theta}$	$\begin{pmatrix} \cos^2\theta + i\sin^2\theta & (1-i)\sin\theta\cos\theta \\ (1-i)\sin\theta\cos\theta & \sin^2\theta + i\cos^2\theta \end{pmatrix}$
Half-wave plate, fast axis θ w.r.t. horizontal	$\overline{\mathbf{J}}_{\lambda/2\,\theta}$	$\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$

This example shows that if a linearly polarized beam strikes a half-wave plate whose fast axis makes an angle of θ with the polarization, the output beam will have its polarization rotated by 2θ. Thus, by rotating a half-wave plate, we can rotate the polarization of a linearly polarized beam into any other linear polarization.

EXAMPLE 2.2

What does a linear polarizer, whose transmission axis makes an angle of +45° with respect to horizontal, do to the polarization of a beam incident with vertical polarization?

The output polarization is given by

$$\begin{aligned} \boldsymbol{\varepsilon} &= \overline{\mathbf{J}}_{\theta=+45}\boldsymbol{\varepsilon}_V \\ &= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] \\ &= \frac{1}{\sqrt{2}}\boldsymbol{\varepsilon}_{+45}. \end{aligned} \tag{2.33}$$

The output is polarized along +45°, as expected, but what's the extra factor of $1/\sqrt{2}$? Recall that the amplitude of the electric field is reduced after passing through a linear polarizer. In this example, the amplitude of the polarization vector is reduced by $1/\sqrt{2}$, meaning that the electric field is reduced by the same amount. Its intensity is reduced by $\left(1/\sqrt{2}\right)^2 = 1/2$.

It is standard practice to normalize the polarization vector at the input of an optical system to have a magnitude of 1. On output the polarization vector will in general be a unit polarization vector, multiplied by a complex constant. The amplitude of this constant determines the change in amplitude of the electric field, and the phase determines the phase shift of the field.

2.4.3 Ordering

Suppose that a wave with input polarization $\boldsymbol{\varepsilon}_i$ passes through a series of elements that modify its polarization. In order it experiences elements whose Jones matrices are $\overline{\mathbf{J}}_1, \overline{\mathbf{J}}_2, \ldots, \overline{\mathbf{J}}_n$. After the first element its polarization has been changed to $\boldsymbol{\varepsilon}_1$, where

$$\boldsymbol{\varepsilon}_1 = \overline{\mathbf{J}}_1 \boldsymbol{\varepsilon}_i. \tag{2.34}$$

This polarization serves as the input to the second element, and after it the polarization becomes

$$\boldsymbol{\varepsilon}_2 = \overline{\mathbf{J}}_2 \boldsymbol{\varepsilon}_1 = \overline{\mathbf{J}}_2 \left(\overline{\mathbf{J}}_1 \boldsymbol{\varepsilon}_i\right) = \overline{\mathbf{J}}_2 \overline{\mathbf{J}}_1 \boldsymbol{\varepsilon}_i. \tag{2.35}$$

Generalizing, the polarization $\boldsymbol{\varepsilon}_n$ after all n elements is

$$\boldsymbol{\varepsilon}_n = \overline{\mathbf{J}}_n \cdots \overline{\mathbf{J}}_2 \overline{\mathbf{J}}_1 \boldsymbol{\varepsilon}_i. \tag{2.36}$$

Note the ordering of the matrices. The first element encountered is written to the right, closest to the input polarization vector; the second element encountered is written to the left of the first, and so on. While at first glance the matrices in eq. (2.36) may appear to be written in reverse order, they are actually ordered properly. The Jones matrices corresponding to the different optical elements are ordered from right to left.

We learned in Sec. 1.2 that the *order* in which we multiply matrices matters. For example, in a system where the beam first travels through a half-wave plate oriented at an angle θ=45°, and then goes through a horizontal polarizer, the combination yields

$$\overline{\mathbf{J}}_H \overline{\mathbf{J}}_{\lambda/2\,45} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{2.37}$$

If the order is reversed we get instead

$$\overline{\mathbf{J}}_{\lambda/2\,45}\overline{\mathbf{J}}_H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{2.38}$$

which is clearly different. Not only is the math different, but the physics is as well. The output polarization depends on the order in which a beam passes through polarization-changing elements.

The properly ordered Jones matrices for a series of polarization-changing elements can be multiplied together to produce an effective Jones matrix for the entire system. The output polarization is determined by multiplying the input polarization vector by this effective Jones matrix.

2.5 POLARIZATION INTERFEROMETER

Consider the arrangement of two PA_{HV}'s shown in fig. 2.8. The first PA_{HV} splits the incident beam (assumed to be polarized at +45°) into separate horizontal and vertical components, while the second PA_{HV} recombines them. For the top beam, the first PA_{HV} acts like a vertical polarizer, so its Jones matrix is $\overline{\mathbf{J}}_V$. That same PA_{HV} acts as a horizontal polarizer, $\overline{\mathbf{J}}_H$, to the bottom beam. The second PA_{HV} recombines the beams, so it effectively adds the Jones matrices for the two paths

$$\overline{\mathbf{J}} = \overline{\mathbf{J}}_V + \overline{\mathbf{J}}_H = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \overline{\mathbf{1}}. \tag{2.39}$$

This is the identity matrix, which implies that this combination of PA_{HV}'s has no effect on the beam.

However, a closer look reveals that the two polarization components in fig. 2.8 have different path lengths through the system. This means that there is a phase shift between the horizontal and vertical polarizations. In the problems you will verify that the matrix

$$\overline{\mathbf{J}}_\phi = \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix} \tag{2.40}$$

induces a phase shift ϕ on the horizontal polarization, relative to the vertical polarization. The effective Jones matrix corresponding to Fig 2.8 is thus

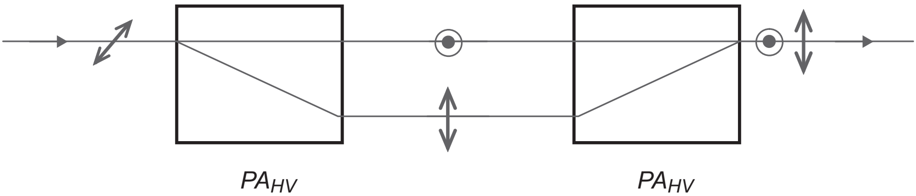


Fig 2.8 A +45° polarized beam of light is split into two beams and then recombined by a pair of PA_{HV}'s.

$$\bar{\mathbf{J}} = \bar{\mathbf{J}}_V + \bar{\mathbf{J}}_\phi \bar{\mathbf{J}}_H = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.41}$$

Figure 2.8 represents a type of polarization interferometer; the two polarizations are separated, phase shifted, and recombined. The fact that there is a relative phase shift between the two polarizations means that the output polarization is in general different from the input polarization.

Unfortunately, the polarization interferometer of fig. 2.8 has a flaw. It will work well enough with a continuous wave (CW) laser source like a helium-neon laser, but it won't work with all sources. Imagine that we put a very short pulse of light into this interferometer—the pulse gets split into two pulses that travel through the interferometer. If the path length difference between the arms is longer than the length of the pulse, then the two pulses won't get to the second $\mathrm{PA}_{\mathrm{HV}}$ at the same time; they won't overlap, so they won't interfere. Even if we don't use pulses, we have to worry about the coherence length of the light source. In order for two light waves to interfere, they must be coherent with respect to each other (See complement 2.A). If the path length difference between the two beams is longer than the coherence length, then the two beams will not interfere. Some very interesting light sources, most notably the single-photon source used in the labs described at the end of this book, have very short coherence lengths.

In order to see interference with short coherence length sources, it is necessary to modify the polarization interferometer of fig. 2.8 so that the two arms are nearly equal in length. This can be done using the design shown in fig. 2.9. Here, a half-wave plate whose fast axis is oriented at $\theta = 45^{\circ}$ from the horizontal is inserted between the two $\mathrm{PA}_{\mathrm{HV}}$'s. This half-wave plate flips the polarizations of the two beams, making the behavior of the two arms symmetric, and allowing the path lengths of the two arms to be equalized. In general the path lengths of the two arms will not be perfectly matched. Indeed, it is possible to tilt one of the $\mathrm{PA}_{\mathrm{HV}}$'s, which has the effect of slightly changing the optical path lengths, allowing one to easily adjust the relative phase shift ϕ between the two arms. Note that the relative phase shift ϕ is *proportional* to the tilt angle of the $\mathrm{PA}_{\mathrm{HV}}$ in fig. 2.9, not *equal* to the tilt angle.

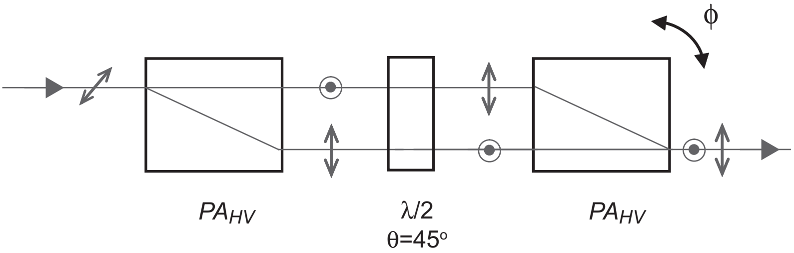


Fig 2.9 A half-wave plate inserted in a polarization interferometer to equalize the path lengths of the two arms. The relative phase ϕ between the two arms can be adjusted by tilting one of the $\mathrm{PA}_{\mathrm{HV}}$'s.

The Jones matrix for the interferometer of fig. 2.9 can be calculated as follows. The upper beam effectively consists of a vertical polarizer, a half-wave plate that flips the polarization to horizontal, and a phase shift on this now horizontally polarized beam. The bottom beam consists of a horizontal polarizer and a half-wave plate that flips its polarization to vertical. The two beams are then added together:

$$\begin{aligned}\overline{\mathbf{J}} &= \overline{\mathbf{J}}_{\phi}\overline{\mathbf{J}}_{\lambda/2\,\theta=45^{\circ}}\overline{\mathbf{J}}_{V} + \overline{\mathbf{J}}_{\lambda/2\,\theta=45^{\circ}}\overline{\mathbf{J}}_{H} \\ &= \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & e^{i\phi} \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & e^{i\phi} \\ 1 & 0 \end{pmatrix}.\end{aligned} \tag{2.42}$$

If the phase shift is adjusted to 0 this interferometer simply behaves as the half-wave plate in its center: horizontally polarized incident light emerges vertically polarized, and vice versa.

The polarization interferometer as described thus far does nothing to the intensity of the beam. This can be attributed to energy (power/intensity) conservation. The beam enters the interferometer, is split in two, and then recombined—all of the power at the input is passed to the output. To change the intensity of the beam we need to insert some form of polarizer.

The experimental arrangement shown in fig. 2.10 incorporates a polarizing beam splitter (in the form of a PA_{45}) at the end, and constitutes a full polarization interferometer. The PA_{45} splits the beam into +45° and –45° components, which provide two different paths for the output power, and allows the intensity of each of the output beams to be modulated.

To show that this interferometer does indeed modulate the output intensity, let's compute the effective Jones matrix for the +45° polarized output beam. Starting at the left of fig. 2.10, the input beam first passes through the polarization interferometer analyzed in eq. (2.42), and then passes through a polarizer oriented at +45°. The Jones matrix for this combination is

$$\overline{\mathbf{J}} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & e^{i\phi} \\ 1 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & e^{i\phi} \\ 1 & e^{i\phi} \end{pmatrix} \tag{2.43}$$

If the input beam is polarized at +45°, the polarization vector of the output beam is

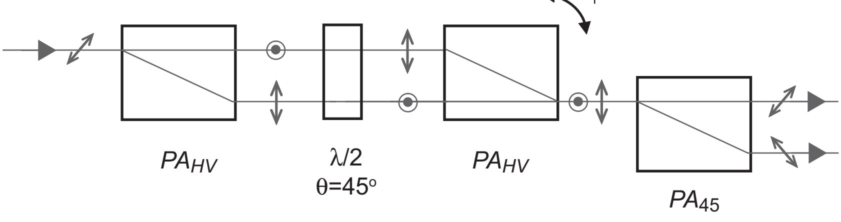


Fig 2.10 A complete polarization interferometer.

$$\begin{aligned}\boldsymbol{\varepsilon} &= \overline{\mathbf{J}}\boldsymbol{\varepsilon}_{+45} \\ &= \frac{1}{2}\begin{pmatrix} 1 & e^{i\phi} \\ 1 & e^{i\phi} \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{2\sqrt{2}}\begin{pmatrix} 1+e^{i\phi} \\ 1+e^{i\phi} \end{pmatrix} \\ &= \frac{1}{2}\left(1+e^{i\phi}\right)\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{2}\left(1+e^{i\phi}\right)\boldsymbol{\varepsilon}_{45}. \end{aligned} \tag{2.44}$$

The output beam is polarized at +45°, as expected. The output intensity is given by the input intensity I_i times the square magnitude of this vector:

$$\begin{aligned} I &= I_i\left|\frac{1}{2}\left(1+e^{i\phi}\right)\right|^2 \\ &= \frac{I_i}{4}\left(1+e^{i\phi}\right)\left(1+e^{-i\phi}\right) \\ &= \frac{I_i}{4}\left(2+e^{i\phi}+e^{-i\phi}\right) \\ &= \frac{I_i}{2}\left(1+\cos\phi\right). \end{aligned} \tag{2.45}$$

When the phase is adjusted to $\phi = 0$, $I = I_i$, the beams interfere constructively, and all of the incident light is output through the +45° output port of the PA_{45}. When the phase is adjusted to $\phi = \pi$, so that $I = 0$, the beams interfere destructively, and to conserve energy the light must be emerging from the –45° output port instead.

Equation (2.45) clearly shows that the apparatus in fig. 2.10 acts as an interferometer; the beams interfere producing an intensity that is modulated depending on the relative phase between them [compare eq. (2.45) with eq. (2.A.9) in complement 2.A]. While the two beams inside the first part of the interferometer have orthogonal polarizations, and would not normally interfere, the final PA_{45} projects the horizontal and vertical polarizations onto the +45° and –45° axes, where they do interfere.

2.6 References

[2.1] The Optices Project, http://webtop.msstate.edu/

[2.2] E. Hecht, *Optics, 4th ed*. (Addison Wesley, San Francisco, 2002).

2.7 PROBLEMS

2.1 A left-circularly polarized beam is incident on a linear polarizer whose transmission axis makes an angle of θ from the horizontal. By what factor is the intensity of this beam reduced? Does your answer depend on the angle θ?

2.2 (a) Find the Jones matrix that corresponds to a beam passing through (in this order): a half-wave plate with its fast axis +22.5° from the horizontal, a quarter-wave plate with its fast axis vertical, and a quarter-wave plate with its fast axis at +45° from the horizontal.

(b) If horizontally polarized light enters this arrangement, what polarization comes out?

2.3* (a) Find the eigenvalues and eigenvectors for the Jones matrix corresponding to a polarizer oriented with its transmission axis at +45°.

(b) Physically, what do these eigenvalues and eigenvectors correspond to? Does this make sense?

2.4 Prove that if a right-circularly polarized beam passes through a half-wave plate, the outgoing beam becomes left-circularly polarized, *independent of the orientation of the fast axis of the wave plate*. The outgoing beam also picks up an unimportant overall phase shift.

2.5 Prove that the Jones matrix for a quarter-wave plate with its fast axis at a general angle θ ($\overline{\mathbf{J}}_{\lambda/4\,\theta}$ in table 2.2) agrees with $\overline{\mathbf{J}}_{\lambda/4\,+45}$ for $\theta = 45^{\circ}$, to within an overall phase factor.

2.6* Show that the $\overline{\mathbf{J}}_{\phi}$ matrix [eq. (2.40)] adds a phase shift to a horizontally polarized beam, but does nothing to a vertically polarized beam.

2.7 (a) Compute the Jones matrix of a half-wave plate whose fast axis makes an angle of θ_1 from the horizontal, followed by a horizontal polarizer, followed by a half-wave plate whose fast axis makes an angle of θ_2 from the horizontal.

(b) Under what circumstances is this combination equivalent to a linear polarizer whose transmission axis makes an angle of θ from the horizontal?

2.8 (a) Compute the Jones matrix of a half-wave plate whose fast axis makes an angle of θ_1 from the horizontal, followed by a vertical polarizer, followed by a half-wave plate whose fast axis makes an angle of θ_2 from the horizontal.

(b) Under what circumstances is this combination equivalent to a linear polarizer whose transmission axis makes an angle of θ from the vertical?

2.9* Use your answers to problems 2.7 and 2.8 to explain how you can construct a PA_{45} out of two half-wave plates and a PA_{HV}.

2.10 (a) Compute the Jones matrix of a quarter-wave plate whose fast axis makes an angle of +45° from the horizontal, followed by a horizontal polarizer, followed by a quarter-wave plate whose fast axis makes an angle of –45° from the horizontal.

(b) If a right-circularly polarized wave is incident on this combination, what will be the output polarization and intensity of the transmitted beam?

(c) If a left-circularly polarized wave is incident on this combination, what will be the output polarization and intensity of the transmitted beam?

2.11 Repeat problem 2.10, replacing the horizontal polarizer with a vertical polarizer.

2.12* Use your answers to problems 2.10 and 2.11 to explain how you can construct a PA_C (a polarization analyzer that splits a beam into left- and right-circularly polarized states) out of two quarter-wave plates and a PA_{HV}.

2.13 Show that the matrix in eq. (2.42) cannot affect the amplitude of a beam. In other words, show that for an arbitrary input unit polarization vector, the output vector is also a unit vector.

2.14 Analyze the intensity of the –45° output port of fig. 2.10. Show that the interferometer conserves energy.

2.15* You're given the arrangement of half-wave plates and polarization analyzers shown in fig. 2.11. The angles refer to the angle made by the fast axis of the wave plate w.r.t. the horizontal; θ_1 is arbitrary. You wish this device to have absolutely no effect on the polarization of your beam: the polarization of the output is always exactly the same as that of the input, no matter what the input polarization is. How should θ_3 and ϕ be set in order to ensure that this is the case?

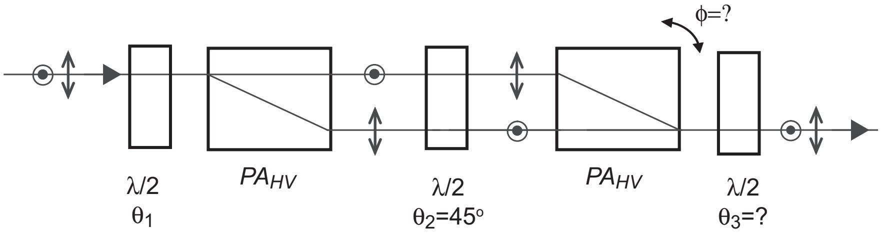


Fig 2.11 Arrangement of half-wave plates and polarization analyzers for problem 2.15.

COMPLEMENT 2.A

Coherence and Interference

Here we'll examine the phenomenon of interference, and show how it is affected by a property of waves called coherence. We're just scratching the surface of coherence theory here; for more details see ref. [2.A.1].

2.A.1 Interference

Consider the Mach-Zehnder interferometer shown in fig. 2.A.1. A beam of light is split into transmitted and reflected waves at the first beam splitter (BS), and mirrors then redirect the light to the second BS where the waves recombine. By translating one of the mirrors, the path length difference between the two arms Δl can be adjusted, which leads to a relative phase shift of

$$\phi = \frac{2\pi\Delta l}{\lambda}. \tag{2.A.1}$$

The electric field that emerges traveling to the right is the sum of the transmitted component of $\mathcal{E}_1$ and the reflected component of $\mathcal{E}_2$:

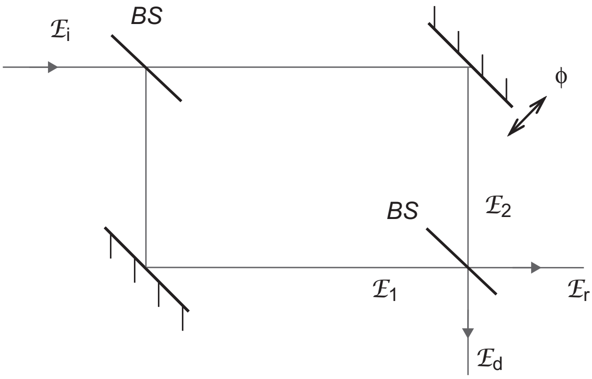


Fig 2.A.1 A Mach-Zehnder interferometer.

$$\mathcal{E}_r = t_1\mathcal{E}_1 + r_2\mathcal{E}_2 e^{i\phi}. \tag{2.A.2}$$

Here t_1 is the field transmission coefficient for $\mathcal{E}_1$, r_2 is the field reflection coefficient for $\mathcal{E}_2$, and we have incorporated the relative phase shift ϕ between the fields into the $\mathcal{E}_2$ term. In general the coefficients t_1 and r_2 are complex numbers, as they incorporate phase shifts induced on reflection or transmission from the beam splitters. The electric field from the downward port from the interferometer is

$$\mathcal{E}_d = r_1\mathcal{E}_1 + t_2\mathcal{E}_2 e^{i\phi}, \tag{2.A.3}$$

where r_1 is the field reflection coefficient for $\mathcal{E}_1$, and t_2 is the field transmission coefficient for $\mathcal{E}_2$.

The electric field is not a conserved quantity, it is energy (power/intensity) that must be conserved, and the intensity is given by the square magnitude of the field. You'll show in problem 2.A.1 that one way to satisfy energy conservation is for the field reflection and transmission coefficients to satisfy[6]

$$r_1 = -r_2 = r,\ t_1 = t_2 = t, \text{ and } r^2 + t^2 = 1. \tag{2.A.4}$$

For 50/50 beam splitters, which split the beams equally, $r = t = 1/\sqrt{2}$. The first beam splitter behaves identically to the second, so the fields at the first beam splitter transform as

$$\mathcal{E}_1 = r\mathcal{E}_i = \frac{1}{\sqrt{2}}\mathcal{E}_i, \tag{2.A.5}$$

$$\mathcal{E}_2 = t\mathcal{E}_i = \frac{1}{\sqrt{2}}\mathcal{E}_i, \tag{2.A.6}$$

where $\mathcal{E}_i$ is the electric field at the input of the interferometer. The equations for the two output fields are then

$$\mathcal{E}_r = \frac{1}{\sqrt{2}}\left(\mathcal{E}_1 - \mathcal{E}_2 e^{i\phi}\right) = \frac{\mathcal{E}_i}{2}\left(1 - e^{i\phi}\right) \tag{2.A.7}$$

$$\mathcal{E}_d = \frac{1}{\sqrt{2}}\left(\mathcal{E}_1 + \mathcal{E}_2 e^{i\phi}\right) = \frac{\mathcal{E}_i}{2}\left(1 + e^{i\phi}\right). \tag{2.A.8}$$

The intensity leaving the downward output port is

$$\begin{aligned} I_d &= \left|\mathcal{E}_d\right|^2 \\ &= \frac{\left|\mathcal{E}_i\right|^2}{4}\left(1 + e^{i\phi}\right)\left(1 + e^{-i\phi}\right) \\ &= \frac{I_i}{4}\left(2 + 2\cos\phi\right) \\ &= \frac{I_i}{2}\left(1 + \cos\phi\right), \end{aligned} \tag{2.A.9}$$

---

[6] Relations of this sort can be derived rigorously from Maxwell's equations.

where I_i is the input intensity. Notice that for phase shifts of $\phi = m2\pi$, where m is an integer, the two waves interfere constructively at the downward output port, and all of the light emerges from this port. Energy conservation dictates that in such circumstances no light is emitted from the right output. It is also possible to adjust the phase so that all of the light is emitted from the right output.

You might find it interesting to compare eq. (2.A.9) to eq. (2.45).

2.A.2 Coherence

Assume that the path length difference between the two arms of our Mach-Zehnder interferometer is nearly 0. Furthermore, assume the interferometer is well constructed, and isolated from vibrations. The relative phase of the two beams when they recombine at the second beam splitter is then constant, resulting in stable interference. The two waves striking the second beam splitter are coherent with respect to each other.

Now imagine that we vibrate one of the mirrors rapidly, with a vibration amplitude of larger than a wavelength. The phase shift is no longer constant, so the intensities of the two output ports fluctuate rapidly. If the timescale of the fluctuations is shorter than the response time of our detector (e.g., an eye), the detector will average over the fluctuations. Mathematically, we can model this by assuming that the relative phase between the two arms is a random variable. The average intensity leaving the downward output port of the interferometer can then be obtained from eq. (2.A.9) by averaging over ϕ

$$\begin{aligned}\langle I_d \rangle &= \frac{I_i}{2}\langle (1+\cos\phi)\rangle \\ &= \frac{I_i}{2}\left(1+\langle \cos\phi \rangle\right).\end{aligned} \tag{2.A.10}$$

If the path length difference fluctuates by more than a wavelength, then the relative phase fluctuates by more than 2π, and the cosine term averages to 0, leaving

$$\langle I_d \rangle = \frac{I_i}{2}. \tag{2.A.11}$$

The average intensity emerging from the right output port is the same.

The net result is that half of the light leaves from the right output, and the other half leaves downward. There is no interference—the intensity does not depend on the relative phase between the two arms. The two waves reaching the second beam splitter are incoherent with respect to each other, meaning that their relative phase varies randomly and rapidly, by more than 2π.

If the path length difference between the two arms fluctuates rapidly, but the fluctuation amplitude is less than a full wavelength, then the cosine does not average to 0. The interference pattern remains, but the modulation depth of the output intensity (i.e., the visibility of the interference) is decreased. In this case the waves in the two arms of the interferometer are said to be partially coherent. The visibility V of an interference pattern is defined as

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{2.A.12}$$

where $I_{\max}$ and $I_{\min}$ are the maximum and minimum intensities, measured as the phase of the interferometer is scanned.

If the fields are coherent, we analyze the interferometer by saying that the *fields* split and recombine at the beam splitters as described above. If the fields are incoherent, we can treat the interferometer more simply by saying that the *intensities* split and recombine at the beam splitters. The intensity reflection and transmission coefficients are given by $R = |r|^2$ and $T = |t|^2$, and satisfy $R + T = 1$. There is no phase associated with the intensity of a beam, so the relative phase of the two arms plays no role if the beams are incoherent.

2.A.3 Coherence Length

Now imagine that the Mach-Zehnder interferometer in fig. 2.A.2 is well-designed so that the mirrors do not vibrate, and the two arms are adjusted so that the path lengths are perfectly equal and $\phi = 0$. Further, imagine that there is some form of phase shifting device placed in the beam *before* it enters the interferometer. This device varies the phase of the input field with random phase shifts of over 2π in amplitude. The timescale on which this phase variation occurs is called the coherence time τ_c. Associated with the coherence time is a coherence length l_c, which is the distance that light travels in one coherence time: $l_c = c\tau_c$.

Since we're assuming that the path lengths of the two arms are perfectly equal, at the second beam splitter the two fields $\mathcal{E}_1$ and $\mathcal{E}_2$ are perfect copies of each other. Each of these fields contains the random phase variation imposed on the input field, but the phases of the two fields fluctuate *together*. The phase randomization does not affect the relative phase of the two fields, so they are coherent with each other, and will interfere at the second beam splitter as described above [e.g, eq. (2.A.9)].

Now assume that the path length difference between the two arms is fairly large, much larger than the coherence length ($\Delta l >> l_c$). When the two fields $\mathcal{E}_1$ and $\mathcal{E}_2$ reach the second beam splitter they are no longer perfect copies of each other, because one has been delayed by a time much longer than the coherence time. Each field fluctuates randomly, and this time they do *not* fluctuate together—the relative phase between the two beams fluctuates randomly. In this case the two beams are incoherent with each other, and will not interfere at the second beam splitter.

The coherence time of a light source is determined by the inverse of the bandwidth Δf (frequency spread) of the source

$$\tau_c \approx \frac{1}{2\pi\Delta f}. \tag{2.A.13}$$

This relation is only approximate because in general the coherence time depends on the details of the source spectrum, not just its bandwidth.

The bottom line is that sources with large bandwidths have short coherence times, and correspondingly short coherence lengths. Since the path length difference between the two arms of an interferometer must be shorter than the coherence length in order to see interference, interference from large bandwidth sources can only be observed if the path lengths of the two arms are very closely matched.

2.A.4 REFERENCES

[2.A.1] E. Hecht, *Optics, 4th ed*. (Addison Wesley, San Francisco, 2002), Ch. 12.

2.A.5 PROBLEMS

2.A.1* For a Mach-Zehnder interferometer, show that eq. (2.A.4) is consistent with energy conservation.

2.A.2 For a Mach-Zehnder interferometer, are the relationships

$$r_1 = r_2 = r,\ t_1 = t_2 = t, \text{ and } |r|^2 + |t|^2 = 1, \tag{2.A.14}$$

consistent with energy conservation?

2.A.3 A helium-neon laser has a central wavelength of $\lambda = 633\ nm$, and a bandwidth of approximately 1 GHz. What is its coherence length?

2.A.4 A green LED has a central wavelength of $\lambda = 530\ nm$, and a wavelength spread of $\Delta\lambda = 10\ nm$. What is its coherence length?

2.A.5 The measured intensity emerging from a Mach-Zehnder interferometer is found to be

$$I = I_0\left(0.8 + 0.2\cos\phi\right). \tag{2.A.15}$$

What is the corresponding visibility?

CHAPTER 3

Quantum States

Now that we know something about classical polarization, it's time to learn about the quantum mechanics of polarization. Quantum mechanically the polarization of a beam is described by a quantum state. The quantum state is a vector, but you'll see that it's a very different sort of vector than is used in classical physics.

Before going into specifics, we need to emphasize an important difference between the optical beams that will be used in the experiments described in this chapter and those described in the last. In chapter 2 the beams were classical electromagnetic waves, and we described them in terms of the electric field $\mathcal{E}$. In this chapter the beams are made of individual photons. You should have been exposed to the concept of a photon before; one way to think of a photon is as a quantum (an indivisible unit) of electromagnetic energy. But be careful! People often think of photons as being "particles" of light. While conceptualizing a photon in this way is occasionally useful, it's an overly simple picture that can easily lead to confusion.

While it is possible to describe a classical electromagnetic field using quantum mechanics, there is no way to describe a purely quantum field using classical electromagnetism. In this chapter we are interested in the behavior of individual photons. We cannot completely describe individual photons using Maxwell's equations, and must describe them using their quantum states. This is a subtle, but important, point that is explored further in labs 1 and 2.

3.1 STATE VECTORS

We'll approach this discussion observationally—we'll describe a series of experiments and observations made about them. Then we'll apply a mathematical description to them. These are not just *gedanken* (thought) experiments, these are experiments you can perform yourself in the laboratories described at the end of this book.

3.1.1 Experiment 1

Imagine that we have a source that produces a beam of light. We determine, possibly using the techniques described in lab 2, that the beam consists of individual photons that we can count. As shown in fig. 3.1, this beam is incident on a polarization analyzer oriented to separate the beam into horizontally and vertically polarized components—a PA_{HV}. N vertically polarized photons are transmitted by this polarization analyzer and horizontally polarized photons are blocked. The vertically polarized photons are incident on a second PA_{HV}, and all of these photons are transmitted to its vertical output port.

Classically we would say that the incident beam contains both horizontal and vertical polarization components that the first polarization analyzer splits apart. The second analyzer takes a vertically polarized input beam, and transmits 100% of the beam out of its vertical output port. Thus, what we already know about the polarization of classical waves also applies to the polarization of individual photons, at least in this simple situation.

Classically we described a vertically polarized wave in terms of its polarization vector $\boldsymbol{\varepsilon}_V$. Quantum mechanically we say that a vertically polarized photon is in state $\left|V\right\rangle$; similarly, a horizontally polarized photon would be in state $\left|H\right\rangle$. It is conventional to denote an arbitrary quantum state as $\left|\psi\right\rangle$. In this notation, which is originally due to Dirac, the state is a vector quantity; hence $\left|\psi\right\rangle$ is often referred to as the state vector.[1] Furthermore, $\left|\psi\right\rangle$ is a particular type of state vector that we refer to as a "ket," for reasons that will soon become transparent.

Don't confuse $\left|V\right\rangle$ with $\boldsymbol{\varepsilon}_V$. State vectors "live" in an abstract space called a Hilbert space, and, as such, don't really "point" in a particular direction. This distinction is subtle for polarization, and it may often sound like $\left|V\right\rangle$ and $\boldsymbol{\varepsilon}_V$ are the same. However, when we get to other quantum systems (e.g., electron spin) it will become more obvious that Hilbert space vectors and physical vectors are not the same.

Back to experiment 1, which is re-depicted in our new quantum mechanical notation in fig. 3.2. Here we say that the source produces photons in some polarization state $\left|\psi\right\rangle$. We don't know what that state is, but it doesn't really matter. The first PA_{HV} splits the beam into vertically and horizontally polarized photons, and we keep only those that

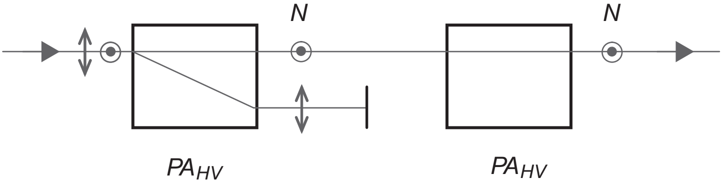


Fig 3.1 Experiment 1.

---

[1] Note that in addition to polarization, photons have other properties as well, such as energy and propagation direction. The full state vector should also describe these properties, but for the moment we're only interested in polarization.

are vertically polarized. We say that the PA_{HV} in combination with the beam block prepares a beam of photons in state $\left|V\right\rangle$; the PA_{HV} and the beam block serve as a state-preparation device. We know that the photons have been successfully prepared in state $\left|V\right\rangle$, because when we analyze them with the second PA_{HV} we find that they remain in this state with 100% certainty. If we wish to prepare photons in state $\left|H\right\rangle$, we would simply block the vertical output of the first PA_{HV} instead, and the horizontally polarized photons from this analyzer would be transmitted through the horizontal port on the second analyzer with 100% certainty.

3.2 BASIS STATES

We can write an arbitrary classical polarization vector in the $\boldsymbol{\varepsilon}_H$ and $\boldsymbol{\varepsilon}_V$ polarization basis. Experiment 1 suggests that we should be able to do the same thing with quantum mechanical polarization states. Unless we get evidence to the contrary, it's reasonable to assume that the vectors $\left|H\right\rangle$ and $\left|V\right\rangle$ form a basis that can represent other vectors in the Hilbert space of polarization states. We can express a general polarization state as a linear combination of $\left|H\right\rangle$ and $\left|V\right\rangle$ states:

$$\left|\psi\right\rangle = c_H\left|H\right\rangle + c_V\left|V\right\rangle, \tag{3.1}$$

where the constants c_H and c_V are complex numbers. A state of this form is said to be a superposition of the $\left|H\right\rangle$ and $\left|V\right\rangle$ states.

3.2.1 Orthogonality and the Inner Product

The vectors $\boldsymbol{\varepsilon}_H$ and $\boldsymbol{\varepsilon}_V$ are orthogonal (i.e., there is no component of $\boldsymbol{\varepsilon}_H$ along $\boldsymbol{\varepsilon}_V$). Experiment 1 indicates that this is true for the state vectors as well. When a photon in state $\left|V\right\rangle$ is incident on the second polarization analyzer it comes out in the same state; there is no $\left|H\right\rangle$ component to $\left|V\right\rangle$. Similarly, if we had repeated experiment 1 by sending $\left|H\right\rangle$ photons into the second analyzer, we would find that they have no $\left|V\right\rangle$ component.

The inner product of orthogonal vectors is 0. However, before we talk about the inner product in Hilbert space, we need to define a new type of vector called a "bra." For every ket vector $\left|\psi\right\rangle$ we can associate a corresponding bra vector $\left\langle\psi\right|$ (think of associating a row vector with a column vector). The vector space of bras is often referred to as the dual space to the vector space of kets.

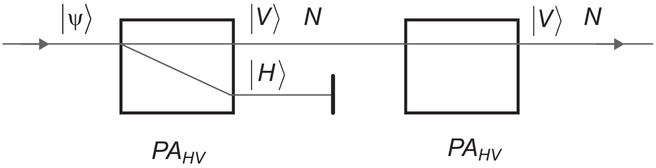


Fig 3.2 Experiment 1 re-depicted in terms of quantum state vectors.

We can form an inner product by placing a row vector to the left of a column vector, a combination which results in a scalar. The analogy here is that we place a bra to the left of a ket and form a "bracket" (get it now?), which is an inner product:

$$\langle \psi_1 | \psi_2 \rangle = c, \tag{3.2}$$

where c is a complex number. Since $|H\rangle$ and $|V\rangle$ are orthogonal, we must have

$$\langle H | V \rangle = \langle V | H \rangle = 0. \tag{3.3}$$

Furthermore, it is convenient to deal with orthonormal basis sets, so

$$\langle H | H \rangle = \langle V | V \rangle = 1. \tag{3.4}$$

As stated above in eq. (3.1), an arbitrary polarization state can be written as $|\psi\rangle = c_H |H\rangle + c_V |V\rangle$. For a given state, how does one determine the coefficients in this linear combination? They can be found by projecting the basis states onto $|\psi\rangle$ using the inner product. For example:

$$\begin{aligned} \langle H | \psi \rangle &= \langle H | \left(c_H |H\rangle + c_V |V\rangle \right) \\ &= \langle H | c_H |H\rangle + \langle H | c_V |V\rangle \\ &= c_H \langle H | H \rangle + c_V \underbrace{\cancel{\langle H | V \rangle}}_{=0} \\ &= c_H . \end{aligned} \tag{3.5}$$

Similarly, $\langle V | \psi \rangle = c_V$.

Given the ket $|\psi\rangle$ expressed in a particular basis, how do we find the corresponding bra $\langle \psi |$? We replace all the kets by their corresponding bras, and replace all the constants by their complex conjugates, that is,

$$|\psi\rangle = c_H |H\rangle + c_V |V\rangle \quad \Leftrightarrow \quad \langle \psi | = c_H^* \langle H | + c_V^* \langle V |. \tag{3.6}$$

It is then straightforward to see that

$$\langle \psi | H \rangle = c_H^* = \langle H | \psi \rangle^* . \tag{3.7}$$

This is a general property of the inner product of state vectors; reversing the order of the product yields the complex conjugate.

Since we generally work with normalized states, we'll impose the condition

$$\begin{aligned} \langle \psi | \psi \rangle &= \left(c_H^* \langle H | + c_V^* \langle V | \right) \left(c_H |H\rangle + c_V |V\rangle \right) \\ &= c_H^* c_H \langle H | H \rangle + c_H^* c_V \underbrace{\cancel{\langle H | V \rangle}}_{=0} + c_V^* c_H \underbrace{\cancel{\langle V | H \rangle}}_{=0} + c_V^* c_V \langle V | V \rangle \\ &= |c_H|^2 + |c_V|^2 = 1 . \end{aligned} \tag{3.8}$$

Table 3.1 Analogies between vectors in three different vector spaces.

Property	Real (2-D) Space	Row/Column Vectors	QM (Hilbert Space)
Basis vectors, first type	$\boldsymbol{u}_x, \boldsymbol{u}_y$	Column vectors $\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}$	Kets $\vert H\rangle, \vert V\rangle$
Basis vectors, second type	No difference $\boldsymbol{u}_x, \boldsymbol{u}_y$	Row vectors $(1\quad 0),(0\quad 1)$	Bras $\langle H\vert, \langle V\vert$
General vectors, first type	$\boldsymbol{a}=a_x\boldsymbol{u}_x+a_y\boldsymbol{u}_y$	$\begin{pmatrix}a_1\\a_2\end{pmatrix}=a_1\begin{pmatrix}1\\0\end{pmatrix}+a_2\begin{pmatrix}0\\1\end{pmatrix}$	$\vert\psi\rangle=\psi_H\vert H\rangle+\psi_V\vert V\rangle$
General vectors, second type	No difference $\boldsymbol{a}=a_x\boldsymbol{u}_x+a_y\boldsymbol{u}_y$	$(a_1^*\quad a_2^*)=$ $a_1^*(1\quad 0)+a_2^*(0\quad 1)$	$\langle\psi\vert=\psi_H^*\langle H\vert+\psi_V^*\langle V\vert$
Inner product	Dot product $\boldsymbol{a}\bullet\boldsymbol{b}$	$(a_1^*\quad a_2^*)\begin{pmatrix}b_1\\b_2\end{pmatrix}$	Bracket $\langle\psi_1\vert\psi_2\rangle$
Orthogonality	$\boldsymbol{u}_x\bullet\boldsymbol{u}_y=0$	$(0\quad 1)\begin{pmatrix}1\\0\end{pmatrix}=0$	$\langle H\vert V\rangle=0$
Normalization	$\boldsymbol{u}_x\bullet\boldsymbol{u}_x=1$	$(1\quad 0)\begin{pmatrix}1\\0\end{pmatrix}=1$	$\langle H\vert H\rangle=1$
Finding coefficients (Use inner product with basis vectors.)	$a_x=\boldsymbol{u}_x\bullet\boldsymbol{a}$	$a_1=(1\quad 0)\begin{pmatrix}a_1\\a_2\end{pmatrix}$	$\psi_H=\langle H\vert\psi\rangle$

For a normalized vector, the sum of the square magnitudes of its coefficients in an orthonormal basis add up to 1. We'll soon use this fact to give a physical interpretation to the coefficients.

Table 3.1 illustrates some analogies between vectors in different vector spaces.

3.3 OTHER STATES

Experiment 1 allowed us to establish the existence of basis vectors and orthogonality. Let's apply these concepts to help us understand some other states.

3.3.1 Experiment 2

Imagine that we prepare a beam of photons in state $\left|V\right\rangle$, and then send that beam through a polarization analyzer whose axes are oriented at ±45° (a PA$_{45}$). As shown in fig. 3.3, if N vertically polarized photons are incident on the PA$_{45}$, half are transmitted from the +45°

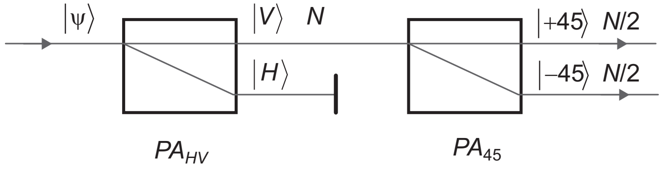


Fig 3.3 Experiment 2.

port in state $|+45\rangle$, and the other half are transmitted from the –45° port in state $|-45\rangle$. If we repeat this experiment by sending $|H\rangle$ photons into the PA_{45} we obtain the same result.

When we say that half of the photons emerge from each of the output ports, we don't mean that *exactly* half of the photons emerge from each port, we mean that on *average* half of the photons emerge from each port. The transmission of photons through the PA_{45} is a random process. It is not possible to say for certain which port any given photon will emerge from, but the probability of exiting either of the output ports is 50%. If we were to repeat experiment 2 many times, we would find that the average number of photons emitted from each port is $N/2$, with a standard deviation of approximately $\sqrt{N/2}$.

This observation is consistent with what we would obtain using classical waves: a wave polarized along $\boldsymbol{\varepsilon}_V$ would be split by a PA_{45} into equal $\boldsymbol{\varepsilon}_{+45}$ and $\boldsymbol{\varepsilon}_{-45}$ components.[2] This suggests that $|V\rangle$ can be written as a linear combination of $|+45\rangle$ and $|-45\rangle$. However, might there be another explanation? Before we write $|V\rangle$ in terms of $|+45\rangle$ and $|-45\rangle$, we need to be sure that there is no other explanation for experiment 2.

Maybe it is the case that the polarization components of a photon along ±45° are completely *independent* of the polarization components along the horizontal and vertical axes. If that were true, then the PA_{45} would split the beam into $|+45\rangle$ and $|-45\rangle$ because of the properties of state $|\psi\rangle$, not because of the properties of state $|V\rangle$. How would we test this possibility?

Think generally about how we test whether properties are independent. Suppose we have a classical particle that has momentum $\boldsymbol{p}$. First we measure the x-component of its momentum and obtain the value p_x. We subsequently measure the y-component of its momentum and obtain the value p_y. In classical physics p_x and p_y are independent quantities, so the fact that we measured p_y for the y-component is due to the properties of $\boldsymbol{p}$, and has nothing to do with the fact that we measured p_x for the x-component. We can verify that p_x and p_y are independent by going back and measuring p_x again after the p_y measurement. We obtain the same value for p_x again, because the measurement of p_y in no way influences the measured value of p_x.

3.3.2 Experiment 3

Similarly, for individual photons we can test whether the polarization components along ±45° are independent of the horizontal and vertical components by alternating measurements along the different axes. Such a series of measurements is illustrated as

---

[2] Of course, there is the important distinction that a classical wave is split deterministically, while photons are split randomly.

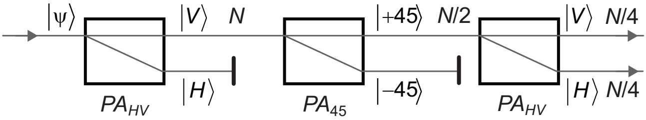


Fig 3.4 Experiment 3.

experiment 3, in fig. 3.4. We see that the second $\mathrm{PA_{HV}}$ splits the $|{+45}\rangle$ photons into $|V\rangle$ and $|H\rangle$, which indicates that the polarization components of a photon along $\pm 45^{\circ}$ are *not* independent of its polarization components along the horizontal and vertical axes; this is also true classically. If they were independent we would expect all $N/2$ photons incident on the second $\mathrm{PA_{HV}}$ to emerge in state $|V\rangle$, because the $\mathrm{PA_{45}}$ would not have disturbed the fact that they were initially prepared in state $|V\rangle$.

3.4 PROBABILITIES

Experiment 3 compels us to write $|{+45}\rangle$ as a linear combination of $|H\rangle$ and $|V\rangle$:

$$|{+45}\rangle = c_H |H\rangle + c_V |V\rangle. \tag{3.9}$$

But how do we determine the values of the coefficients c_H and c_V? Looking at the output of the second $\mathrm{PA_{HV}}$ in fig. 3.4 shows that the state $|{+45}\rangle$ is split into equal amounts of $|H\rangle$ and $|V\rangle$, so it is reasonable to assume that the magnitudes of c_H and c_V are the same. If we insist that $|{+45}\rangle$ is normalized, then eq. (3.8) yields the constraint that $|c_H|^2 + |c_V|^2 = 1$. This leaves us with

$$|c_H|^2 = |c_V|^2 = \frac{1}{2}, \tag{3.10}$$

or

$$|c_H| = |c_V| = \frac{1}{\sqrt{2}}. \tag{3.11}$$

The above is not just math, there's some important physics involved. We have assigned a *physical meaning* to the coefficients in a linear combination such as eq. (3.9). We have said that the square magnitude of the coefficient multiplying a particular polarization state is equal to the probability that the photon will be measured to have that polarization. With this in mind, the coefficients c_H and c_V are often referred to as "probability amplitudes," a term originated by Richard Feynman. The probability is the magnitude squared of the probability amplitude.

Remember, however, that the coefficients are complex numbers. Now that we know what their magnitudes are, we need to determine their phases. Using eq. (3.11), we can rewrite eq. (3.9) as

$$|{+45}\rangle = \frac{1}{\sqrt{2}} e^{i\phi_H} |H\rangle + \frac{1}{\sqrt{2}} e^{i\phi_V} |V\rangle = \frac{1}{\sqrt{2}} e^{i\phi_H} \left(|H\rangle + e^{i(\phi_V - \phi_H)} |V\rangle \right). \tag{3.12}$$

The constant phase factor $e^{i\phi_H}$ in front is an overall, or absolute phase factor. It has a magnitude of one, and in this experiment it does not affect any measurable quantities, so we can safely take it to be equal to 1 ($\phi_H = 0$).[3] This leaves

$$\left|+45\right\rangle = \frac{1}{\sqrt{2}}\left(\left|H\right\rangle + e^{i\phi_V}\left|V\right\rangle\right), \tag{3.13}$$

and we need more information to determine the phase ϕ_V.

If we repeat experiment 3, but select the $\left|-45\right\rangle$ beam instead of the $\left|+45\right\rangle$ beam, we'd get the same result: it would be made of 50% $\left|H\right\rangle$ and 50% $\left|V\right\rangle$. Thus, the analysis above also applies to this state, so

$$\left|-45\right\rangle = \frac{1}{\sqrt{2}}\left(\left|H\right\rangle + e^{i\phi'_V}\left|V\right\rangle\right). \tag{3.14}$$

The phase ϕ'_V must be different from ϕ_V, otherwise $\left|+45\right\rangle$ and $\left|-45\right\rangle$ would be the same state.

If we repeat experiment 1 (fig. 3.2) using PA$_{45}$ measurements instead of PA$_{\text{HV}}$ measurements, we find that the states $\left|+45\right\rangle$ and $\left|-45\right\rangle$ are orthogonal. This means that their bracket must be zero.

$$\begin{aligned}\left\langle +45\middle|-45\right\rangle &= \frac{1}{2}\left(\left\langle H\right| + e^{-i\phi_V}\left\langle V\right|\right)\left(\left|H\right\rangle + e^{i\phi'_V}\left|V\right\rangle\right)\\ &= \frac{1}{2}\left(\left\langle H\middle|H\right\rangle + e^{i\phi'_V}\left\langle H\middle|V\right\rangle + e^{-i\phi_V}\left\langle V\middle|H\right\rangle + e^{i(\phi'_V-\phi_V)}\left\langle V\middle|V\right\rangle\right)\\ &= \frac{1}{2}\left(1 + e^{i(\phi'_V-\phi_V)}\right) = 0\,.\end{aligned} \tag{3.15}$$

This is satisfied if

$$\phi'_V - \phi_V = \pi. \tag{3.16}$$

It is typical to choose $\phi_V = 0$, which leaves $\phi'_V = \pi$. Our final answer is thus

$$\begin{aligned}\left|+45\right\rangle &= \frac{1}{\sqrt{2}}\left(\left|H\right\rangle + \left|V\right\rangle\right),\\ \left|-45\right\rangle &= \frac{1}{\sqrt{2}}\left(\left|H\right\rangle - \left|V\right\rangle\right).\end{aligned} \tag{3.17}$$

The choice of $\phi_V = 0$ is motivated by the fact that it makes the states in eq. (3.17) consistent with the classical polarization vectors $\boldsymbol{\varepsilon}_{+45}$ and $\boldsymbol{\varepsilon}_{-45}$ (see table 2.1).

Let's recap what we've learned so far from experiment 3 (fig. 3.4). The middle measurement, using a PA$_{45}$, prepares a beam in the state $\left|+45\right\rangle$. A photon in the state $\left|+45\right\rangle$ is then analyzed with a PA$_{\text{HV}}$. This measurements yields 50% probability of finding a horizontally polarized photon and 50% probability of finding a vertically polarized photon. These probabilities are related to the coefficients in eq. (3.9) by

---

[3] Just as the absolute phase of an electromagnetic wave does not affect the measured intensity.

$$P\left(H\,\middle\|\,+45\right\rangle\big) = \left|c_H\right|^2 = \left|\langle H|+45\rangle\right|^2 = \frac{1}{2},$$
$$P\left(V\,\middle\|\,+45\right\rangle\big) = \left|c_V\right|^2 = \left|\langle V|+45\rangle\right|^2 = \frac{1}{2}. \tag{3.18}$$

The notation that we're using for the probabilities makes it clear that the measured probability is a conditional probability (see sec. 1.1.4); $P\big(H\,\|+45\rangle\big)$ is the probability that we measure a horizontally polarized photon, given that the beam is prepared in the state $|+45\rangle$. We'll discuss measurement probabilities in more depth in chapter 5.

Equation (3.18) gives a prescription for finding the magnitudes of the probability amplitudes from measured probabilities. To determine the phases we need more information; here, an important piece of that information comes from the orthogonality of the $|+45\rangle$ and $|-45\rangle$ states.

3.5 COMPLEX PROBABILITY AMPLITUDES

3.5.1 Experiment 4

Experiment 4, shown in fig. 3.5, is the same as experiment 3, except that we replace the middle PA_{45} by a circular polarization analyzer PA_{C}, which splits the beam into left- and right-circular polarization states.[4] Figure 3.5 shows the $|L\rangle$ state being broken into its horizontal and vertical components by a $\mathrm{PA}_{\mathrm{HV}}$, where we find that this state is made of 50% $|H\rangle$ and 50% $|V\rangle$. If we repeat this experiment and instead analyze the $|R\rangle$ state, we will get the same result.

Applying the same arguments we used in analyzing experiment 3, we conclude that the $|L\rangle$ and $|R\rangle$ states must be written as linear combinations of $|H\rangle$ and $|V\rangle$. Since the probabilities of $|H\rangle$ and $|V\rangle$ are equal, we again find that the magnitude of the coefficients in the linear expansion of $|L\rangle$ in the HV-basis are both equal to $1/\sqrt{2}$. So, similar to eqs. (3.13) and (3.14), we write the $|L\rangle$ and $|R\rangle$ states as

$$|L\rangle = \frac{1}{\sqrt{2}}\Big(|H\rangle + e^{i\varphi_V}\,|V\rangle\Big),$$
$$|R\rangle = \frac{1}{\sqrt{2}}\Big(|H\rangle + e^{i\varphi'_V}\,|V\rangle\Big). \tag{3.19}$$

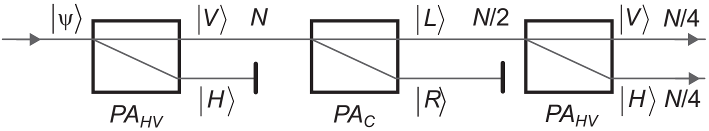


Fig 3.5 Experiment 4.

---

[4] In problem 2.12 you showed that it is possible to construct a PA_{C} out of a $\mathrm{PA}_{\mathrm{HV}}$ and a pair of quarter-wave plates.

Repeating experiment 1 (fig. 3.2) using PA$_C$ measurements instead of PA$_{HV}$ measurements, shows that $|L\rangle$ and $|R\rangle$ are orthogonal, or $\langle L|R\rangle = 0$. Using this fact, and eq. (3.19) yields

$$\varphi'_V - \varphi_V = \pi, \tag{3.20}$$

which is similar to eq. (3.16).

3.5.2 Experiment 5

To this point, everything we did in the search for the phases of the $|+45\rangle$ and $|-45\rangle$ states applies equally as well to the $|L\rangle$ and $|R\rangle$ states. Now, however, we can't choose $\varphi_V = 0$ (making $\varphi'_V = \pi$), because that would make the states $|L\rangle$ and $|R\rangle$ the same as the states $|+45\rangle$ and $|-45\rangle$. We know that they're not the same because of experiment 5, shown in fig. 3.6.

In experiment 5, the last PA$_{45}$ tells us that the probabilities of measuring ±45° polarizations are both 1/2 when a beam is prepared in state $|L\rangle$:

$$\begin{aligned} P(+45 \| L\rangle) &= |\langle +45|L\rangle|^2 = \frac{1}{2}, \\ P(-45 \| L\rangle) &= |\langle -45|L\rangle|^2 = \frac{1}{2}. \end{aligned} \tag{3.21}$$

Using eqs. (3.17) and (3.19), we find that

$$\begin{aligned} \langle +45|L\rangle &= \frac{1}{2}\left(\langle H| + \langle V|\right)\left(|H\rangle + e^{i\varphi_V}|V\rangle\right) \\ &= \frac{1}{2}\left(1 + e^{i\varphi_V}\right), \end{aligned} \tag{3.22}$$

so

$$\begin{aligned} |\langle +45|L\rangle|^2 &= \frac{1}{4}\left(1 + e^{i\varphi_V}\right)\left(1 + e^{-i\varphi_V}\right) \\ &= \frac{1}{4}\left(2 + 2\cos\varphi_V\right). \end{aligned} \tag{3.23}$$

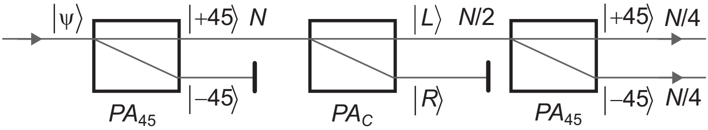


Fig 3.6 Experiment 5.

This must be equal to 1/2 by eq. (3.21), so $\varphi_V = \pm\pi/2$. Make the choice that it's positive, which makes $\varphi'_V = 3\pi/2$ via eq. (3.20). Substituting these values into eq. (3.19) yields

$$\begin{aligned} |L\rangle &= \frac{1}{\sqrt{2}}\left(|H\rangle + i|V\rangle\right), \\ |R\rangle &= \frac{1}{\sqrt{2}}\left(|H\rangle - i|V\rangle\right). \end{aligned} \tag{3.24}$$

We were left with some degree of flexibility in choosing φ_V, but we were forced into having complex probability amplitudes in eq. (3.24). The choice we made makes the polarization states $|L\rangle$ and $|R\rangle$ consistent with the polarization vectors $\boldsymbol{\varepsilon}_L$ and $\boldsymbol{\varepsilon}_R$ (see table 2.1).

3.6 ROW AND COLUMN VECTOR NOTATION

In chapter 1 we said that there are many ways to express a vector. This is also true of quantum mechanical state vectors in Hilbert space. Let $|\psi\rangle = c_H\,|H\rangle + c_V\,|V\rangle$; the coefficients c_H and c_V completely specify the state. Thus, we can write[5]

$$|\psi\rangle \doteq \begin{pmatrix} c_H \\ c_V \end{pmatrix}_{HV} = \begin{pmatrix} \langle H|\psi\rangle \\ \langle V|\psi\rangle \end{pmatrix}_{HV}. \tag{3.25}$$

We've placed the subscript *HV* on the vector to make it clear that the coefficients are expressed in the *HV*-basis. (It's fairly obvious that this is the case here, but it won't always be.) There is a definite ordering implied here—you need to know that the coefficient of $|H\rangle$ comes first, and the coefficient of $|V\rangle$ comes second.

There's nothing special about the *HV*-basis, we could also write

$$|\psi\rangle \doteq \begin{pmatrix} \langle +45|\psi\rangle \\ \langle -45|\psi\rangle \end{pmatrix}_{45}. \tag{3.26}$$

We write kets as column vectors, and bras as row vectors:

$$\langle\psi| = c_H^*\,\langle H| + c_V^*\,\langle V|,$$

---

[5] We're going to continue to use the $\doteq$ symbol to denote "is represented by" for quantum state vectors.

Table 3.2 Bra and ket notation, and column vectors, of important polarization states. The subscript 45 refers to vectors written in the $|+45\rangle$, $|+45\rangle$ basis, while the subscript C refers to the basis of circular polarization states $|L\rangle$ and $|R\rangle$.

Polarization State	Bra, Ket Notation in $\lvert H\rangle, \lvert V\rangle$ Basis	Column Vector in $\lvert H\rangle, \lvert V\rangle$ Basis
$\lvert H\rangle$	$\lvert H\rangle$	$\begin{pmatrix}1\\0\end{pmatrix}_{HV}$
$\lvert V\rangle$	$\lvert V\rangle$	$\begin{pmatrix}0\\1\end{pmatrix}_{HV}$
$\lvert +45\rangle \doteq \begin{pmatrix}1\\0\end{pmatrix}_{45}$	$\lvert +45\rangle = \frac{1}{\sqrt{2}}(\lvert H\rangle + \lvert V\rangle)$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}_{HV}$
$\lvert -45\rangle \doteq \begin{pmatrix}0\\1\end{pmatrix}_{45}$	$\lvert -45\rangle = \frac{1}{\sqrt{2}}(\lvert H\rangle - \lvert V\rangle)$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}_{HV}$
$\lvert L\rangle \doteq \begin{pmatrix}1\\0\end{pmatrix}_{C}$	$\lvert L\rangle = \frac{1}{\sqrt{2}}(\lvert H\rangle + i\lvert V\rangle)$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix}_{HV}$
$\lvert R\rangle \doteq \begin{pmatrix}0\\1\end{pmatrix}_{C}$	$\lvert R\rangle = \frac{1}{\sqrt{2}}(\lvert H\rangle - i\lvert V\rangle)$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}_{HV}$

$$\langle\psi| \doteq \begin{pmatrix}c_H^* & c_V^*\end{pmatrix}_{HV} = \begin{pmatrix}\langle H|\psi\rangle^* & \langle V|\psi\rangle^*\end{pmatrix}_{HV} = \begin{pmatrix}\langle\psi|H\rangle & \langle\psi|V\rangle\end{pmatrix}_{HV}, \tag{3.27}$$

where we've used the fact that $\langle\psi_1|\psi_2\rangle^* = \langle\psi_2|\psi_1\rangle$.

Some specific examples of polarization state vectors we've talked about in this chapter are given in table 3.2. If you compare this to table 2.1, you'll notice that the column vector for a given polarization state is the same as its corresponding classical Jones vector. This is because we have chosen the phase factors in the coefficients to make them the same, as described above. Please keep in mind, however, that even though we're using the same notation, the quantum and classical vectors mean (at this point subtly) different things, as they exist in different vector spaces.

The row and column vector notation for quantum states is useful for doing things like evaluating inner products, as seen in the following example.

EXAMPLE 3.1

Compute $\langle H|+45\rangle$.

Using the vectors from table 3.2, and rewriting $\langle H|$ as a row vector:

$$\begin{aligned}\langle H|+45\rangle &= \begin{pmatrix}1 & 0\end{pmatrix}_{HV}\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}_{HV}\\ &= \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\end{pmatrix}_{HV}\begin{pmatrix}1\\1\end{pmatrix}_{HV}\\ &= \frac{1}{\sqrt{2}}\left[(1)(1)+(0)(1)\right]\\ &= \frac{1}{\sqrt{2}}.\end{aligned} \tag{3.28}$$

3.7 INTERFERENCE

3.7.1 Experiment 6

Look at the (admittedly odd at this point) arrangement for experiment 6, shown in fig. 3.7(a). A beam of photons in the $|+45\rangle$ state is split equally at a PA$_{\mathrm{HV}}$. The $|V\rangle$ beam passes through a half-wave plate with its axis oriented at 45°, and is converted to a beam of photons in state $|H\rangle$—a half-wave plate acts the same way on the polarization states of individual photons as it does on the polarization vectors of classical waves (more on this in the next chapter). This $|H\rangle$ beam is deflected by a PA$_{\mathrm{HV}}$, but otherwise it is unaltered. Finally, the $|H\rangle$ beam is split equally by a PA$_{45}$. Apart from the apparently extraneous half-wave plate and second PA$_{\mathrm{HV}}$, there is nothing unexpected in this experiment. Figure 3.7(b) shows what would happen in a similar arrangement in which the $|V\rangle$ beam is blocked.

What would happen if we perform experiment 6 as shown in fig. 3.7(c), where we don't block either beam? How will the beam split on the final polarization analyzer? A very reasonable thing to do would be to look at figs. 3.7(a) and (b) and say, "When the $|H\rangle$ beam is incident on the last analyzer it splits equally, and the $|V\rangle$ beam behaves the same. Since the beam is split equally in either case, if both beams are incident on the PA$_{45}$ they will both be split equally, and $N/2$ photons will come out in the state $|+45\rangle$ and $N/2$ photons will come out in the state $|-45\rangle$."

Physics is an experimental science, so you don't have to wonder whether your hypothesis is correct for long, you simply do the experiment, as described in more detail in lab 3. The results are shown in fig. 3.8: All N photons come out in the state $|+45\rangle$! Your first reaction will probably be to go back to your original hypothesis and wonder, "If $N/4$ photons come out in the $|-45\rangle$ state when I put the $|H\rangle$ beam into the final analyzer [fig 3.7(a)], and $N/4$ photons come out in that same state when I put in the $|V\rangle$ beam [fig 3.7(b)], how can I get *no* photons in the $|-45\rangle$ state when I use both? How do these photons cancel each other out?"

With some reflection, hopefully you'll recall the discussion of the polarization interferometer in sec. 2.5. Experiment 6 looks similar to the experimental arrangement shown in fig. 2.10, where a classical wave with +45° polarization was incident on a polarization interferometer. With the proper phase adjustment, all of the light emerged

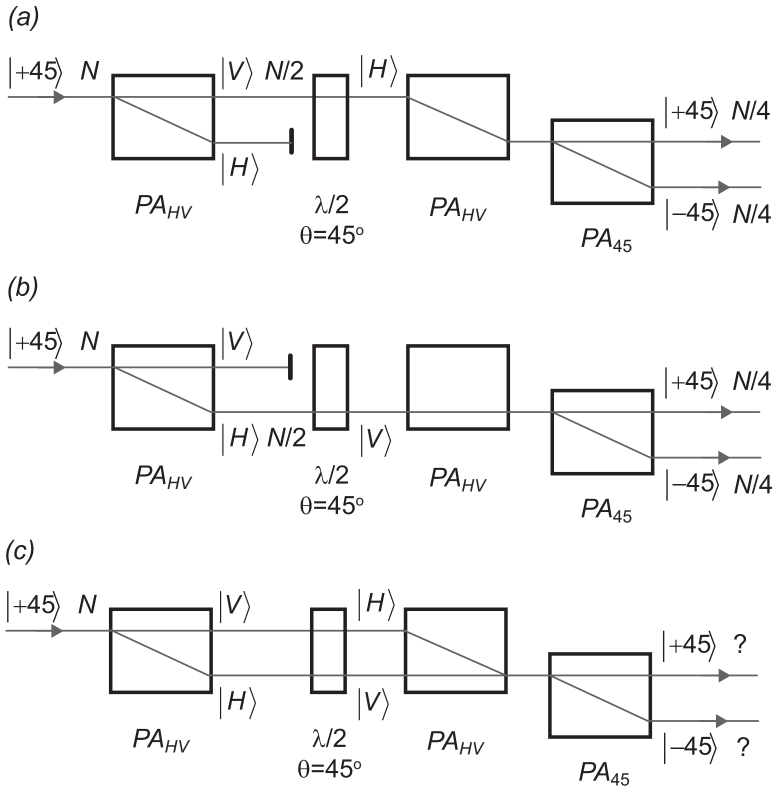


Fig 3.7 (a) and (b) show arrangements for experiment 6 in which one of the beams is blocked, while (c) asks what will happen if neither beam is blocked.

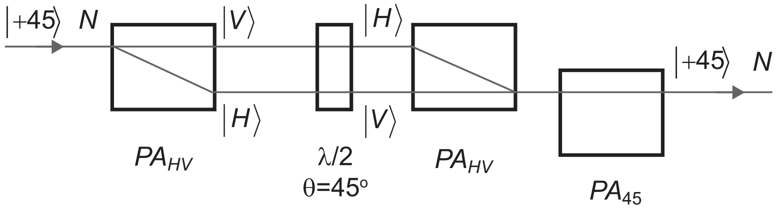


Fig 3.8 The result for experiment 6 when neither beam is blocked.

with +45° polarization. In the classical case we interpreted the results in terms of wave interference.

To verify that interference is occurring in experiment 6, we can vary the path length (phase shift ϕ) between the two beams by simply tilting one of the PA_{HV}'s, just as we

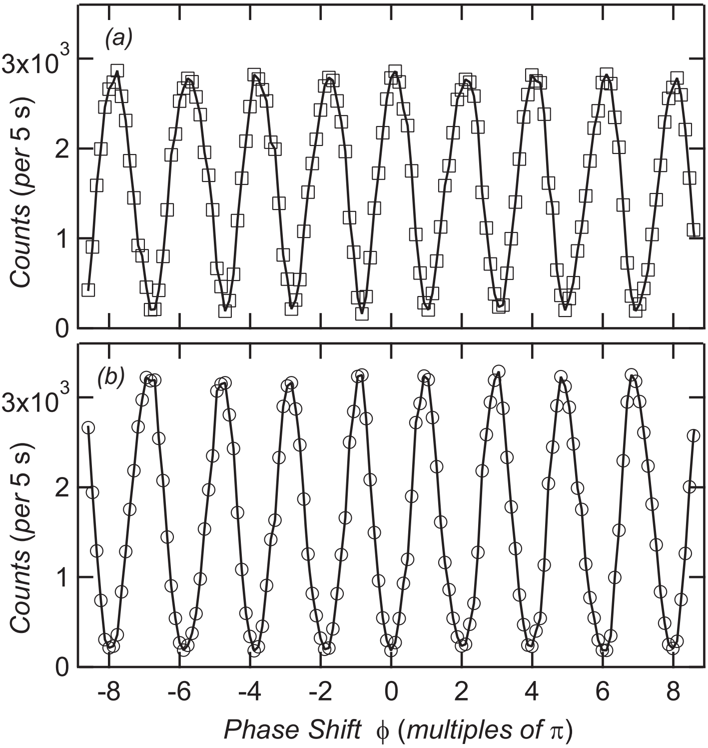


Fig 3.9 Measured data for experiment 6 displaying single photon interference. (a) The count rates for photons exiting in state $|+45\rangle$, and (b) the count rates for photons exiting in state $|-45\rangle$. Using the procedure described in labs 2 and 3, it was verified that only a single photon at a time is present in the interferometer. (These data were acquired by A. Gogo and W.D. Snyder.)

did in the classical case, and we can then count how many photons come out in the $|+45\rangle$ and $|-45\rangle$ states. Data for such an experiment, obtained by students using the procedure described in lab 3, is shown in fig. 3.9. We see that, indeed, when the phase difference between the beams is 0 nearly all of photons exit in the $|+45\rangle$ state. As the phase is varied the output photons oscillate between the $|+45\rangle$ and $|-45\rangle$ states, clearly displaying interference.

3.7.2 Which Way?

Experiment 6 represents a polarization interferometer. The important point is that interference occurs *even though there is only one photon at a time in the interferometer*! This may strike you as odd because, as mentioned at the beginning of this chapter, you may be thinking of photons as “particles” of light, and you’re not used to particles displaying interference. While in some sense photons may behave like particles of

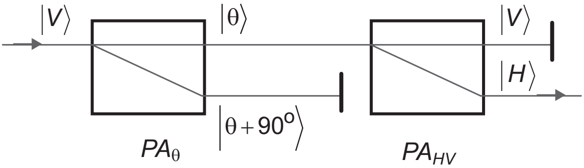


Fig 3.10 The experiment in problem 3.4.

light energy, they are by no means classical particles, as they clearly display wave-like properties as well.

A question that often arises is, "Which path through the interferometer did the photon take?" The interference pattern shown in fig. 3.9 is a function of path length (phase) difference between the two arms of the interferometer, and as such it is a function of both paths. The only way to explain the measured interference pattern is by assuming that *each photon takes both paths* through the interferometer. Anthropomorphizing, if each photon took only one path or the other, there would be no way for it to "know" the path length difference; it would be equally likely to exit as a $|-45\rangle$ or a $|+45\rangle$ photon, independent of the path length difference. This is not what happens.

3.8 PROBLEMS

3.1* $|\theta\rangle$ represents the state of a beam of photons linearly polarized at an angle of θ from the horizontal. Write $|\theta\rangle$ as a linear combination of $|H\rangle$ and $|V\rangle$. [Hint: Use what you know about classical linear polarization.]

3.2 What is the probability that a photon in state $|\theta\rangle$, as described in problem 3.1, will be measured to have vertical polarization?

3.3 What is the probability that a photon in state $|\theta\rangle$, as described in problem 3.1, will be measured to have linear polarization along +45°?

3.4* A beam of photons in state $|V\rangle$ is sent through a series of two polarization analyzers, as illustrated in fig. 3.10. The angle θ is measured with respect to the horizontal, and the state $|\theta\rangle$ is described in problem 3.1.

(a) What fraction of the input photons will survive to the final output?

(b) At what angle θ must the PA_θ be oriented so as to maximize the number of photons that are transmitted by the $\mathrm{PA}_{\mathrm{HV}}$? What fraction of the photons are output for this value of θ?

(c) What fraction of the photons are transmitted if the PA_θ is simply removed from the experiment?

3.5 What is the probability that a photon in state $|\theta\rangle$, as described in problem 3.1, will be measured to have right-circular polarization?

3.6* A general (elliptical) polarization state can be written in terms of 2 parameters, θ and ϕ, as:

$$|e_1\rangle = \cos(\theta)|H\rangle + e^{i\phi}\sin(\theta)|V\rangle. \tag{3.29}$$

Given the state $|e_1\rangle$, find the normalized state $|e_2\rangle$ that is orthogonal to $|e_1\rangle$.

3.7 Using both bras and kets, and row and column vectors, compute $\langle -45|L\rangle$.

3.8 Using both bras and kets, and row and column vectors, compute $\langle +45|e_1\rangle$ [see eq. (3.29)].

3.9 Compute the probability that a photon in the state $|e_1\rangle$ [eq. (3.29)] will be measured to have right-circular polarization.

3.10 Determine the column vectors representing the states $|+45\rangle$ and $|-45\rangle$ using the states $|L\rangle$ and $|R\rangle$ as a basis.

3.11* If the beam entering the interferometer of fig. 3.8 is not in the state $|+45\rangle$, but rather in the state $|V\rangle$, would you expect to see interference? Explain why, or why not.

3.12 If the beam entering the interferometer of fig. 3.8 is not in the state $|+45\rangle$, but rather in the state $|R\rangle$, would you expect to see interference? Explain why, or why not.

Consider the following experimental arrangement, which applies to the following three problems: a beam of photons is sent into a PA_{HV}, which allows us to measure the $|H\rangle$ and $|V\rangle$ components. A measurement of the horizontal/vertical polarization (call it $\wp_{HV}$) that yields a horizontally polarized photon takes on the value $\wp_{HV} = +1$, and a measurement that yields a vertically polarized photon takes on the value $\wp_{HV} = -1$. A series of polarization measurements thus consists of a string of +1's and –1's; we can compute the statistics of these measurements.

3.13* If the beam of photons in state $|H\rangle$:

(a)What is the probability that a given measurement yields +1? –1?

(b) What is the mean value of $\wp_{HV}$?

(c) What is the standard deviation of $\wp_{HV}$?

3.14* If the beam of photons in state $|-45\rangle$:

(a)What is the probability that a given measurement yields +1? –1?

(b) What is the mean value of $\wp_{HV}$?

(c) What is the standard deviation of $\wp_{HV}$?

3.15 If the beam of photons in state $|e_1\rangle$ [eq. (3.29)]:

(a)What is the probability that a given measurement yields +1? –1?

(b) What is the mean value of $\wp_{HV}$?

(c) What is the standard deviation of $\wp_{HV}$?

CHAPTER 4

Operators

Classically and quantum mechanically, polarization can be described in terms of vectors. We've stressed that these vectors are different from each other, but they certainly have many similarities. We've found that objects such as wave plates and polarizers can change the polarization of a wave, and we can describe their operation mathematically using matrices. More generally, the mathematical objects that change one vector into another are called operators, and in quantum mechanics operators change one state into another.

4.1 OPERATORS

Recall example 2.2, in which a beam was incident on a linear polarizer. We found the output polarization vector by multiplying the input polarization vector by the Jones matrix for the polarizer. Quantum mechanically the behavior is analogous. Objects such as polarizers transform one polarization state into another, and we refer to them as operators because they perform a state transformation operation. This looks like

$$\hat{O}\left|\psi_1\right\rangle = c\left|\psi_2\right\rangle, \tag{4.1}$$

where $\hat{O}$ is an operator that changes the state, and the constant c is in general complex. We'll adopt the notation that the caret symbol ^ placed over something denotes an operator. By convention operators get placed next to the vertical bar in the state symbol, so while $\hat{O}\left|\psi\right\rangle$ and $\left\langle\psi\right|\hat{O}$ make sense, $\left|\psi\right\rangle\hat{O}$ and $\hat{O}\left\langle\psi\right|$ don't.

Suppose an input quantum state $\left|\psi_i\right\rangle$ is operated on sequentially by a series of operators $\hat{O}_1, \hat{O}_2, \ldots, \hat{O}_N$. After the first operation the state becomes

$$\left|\psi_1\right\rangle = \hat{O}_1\left|\psi_i\right\rangle. \tag{4.2}$$

After the second operation this is transformed to

$$\left|\psi_2\right\rangle = \hat{O}_2\left|\psi_1\right\rangle = \hat{O}_2\left(\hat{O}_1\left|\psi_i\right\rangle\right) = \hat{O}_2\hat{O}_1\left|\psi_i\right\rangle. \tag{4.3}$$

After *N* operations this is

$$\left|\psi_N\right\rangle = \hat{O}_N\left|\psi_{N-1}\right\rangle = \hat{O}_N...\hat{O}_2\hat{O}_1\left|\psi_i\right\rangle. \tag{4.4}$$

Just as the classical Jones matrices had to be written from right to left to determine the effective Jones matrix, the proper ordering for quantum operators is also from right to left

$$\hat{O}_{eff} = \hat{O}_N...\hat{O}_2\hat{O}_1. \tag{4.5}$$

Also as in classical physics, the order of operations in quantum mechanics matters, in general $\hat{O}_2\hat{O}_1 \neq \hat{O}_1\hat{O}_2$. However, we'll soon see that in quantum mechanics this fact, while not surprising, turns out to be somewhat more profound than it is in classical physics.

While the commutative property does not hold for operators, the distributive property does:

$$\left(\hat{O}_1 + \hat{O}_2\right)\left|\psi\right\rangle = \hat{O}_1\left|\psi\right\rangle + \hat{O}_2\left|\psi\right\rangle, \tag{4.6}$$

and

$$\hat{O}\left(\left|\psi_1\right\rangle + \left|\psi_2\right\rangle\right) = \hat{O}\left|\psi_1\right\rangle + \hat{O}\left|\psi_2\right\rangle. \tag{4.7}$$

It is also customary to refer to powers of operators:

$$\hat{O}^2 = \hat{O}\hat{O}. \tag{4.8}$$

A function of an operator is defined in terms of its power series representation. Thus, for example,

$$e^{\hat{O}} \equiv \sum_{n=0}^{\infty}\frac{1}{n!}\hat{O}^n. \tag{4.9}$$

4.1.1 The Polarization Rotation Operator

Imagine we've got a photon propagating in the *z*-direction, and its wave vector is $\boldsymbol{k} = k\boldsymbol{u}_z$. This photon is linearly polarized in state $\left|H\right\rangle$, and we want to rotate its polarization so that it ends up in state $\left|+45\right\rangle$, as shown in fig. 4.1.

The operator $\hat{R}\left(\theta, \boldsymbol{u}_z\right)$ denotes a right-handed rotation through an angle of θ about the axis $\boldsymbol{u}_z$. At the moment we're interested in polarization rotation, and we'll always be rotating about the propagation direction, so we'll take the $\boldsymbol{u}_z$ to be implied. Thus, we'll use $\hat{R}_p\left(\theta\right)$ to denote the polarization rotation operator, with the subscript *p* making this explicit. From fig. 4.1 we'd expect that

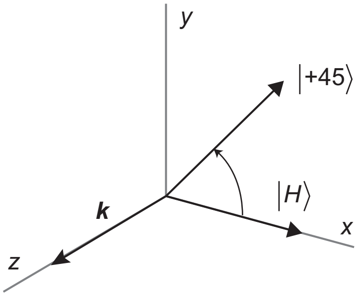


Fig 4.1 Rotation of state $|H\rangle$ into state $|+45\rangle$.

$$\hat{R}_p\left(45^{\mathrm{o}}\right)|H\rangle = |+45\rangle, \tag{4.10}$$

which is correct.

4.2 THE ADJOINT OPERATOR

Given that $\hat{R}_p\left(45^{\mathrm{o}}\right)|H\rangle = |+45\rangle$, is the following true:

$$\langle H|\hat{R}_p\left(45^{\mathrm{o}}\right) \overset{?}{=} \langle +45|? \tag{4.11}$$

To check, imagine that we apply the rotation operator $\hat{R}_p\left(45^{\mathrm{o}}\right)$ twice to the state $|H\rangle$. From fig. 4.1 we see that the first rotation by 45^{o} yields $|+45\rangle$, and we'd expect that the second would yield $|V\rangle$, which is true:

$$\hat{R}_p\left(45^{\mathrm{o}}\right)\hat{R}_p\left(45^{\mathrm{o}}\right)|H\rangle = \hat{R}_p\left(45^{\mathrm{o}}\right)|+45\rangle = |V\rangle. \tag{4.12}$$

From this we see that

$$\begin{aligned}\langle H|\hat{R}_p\left(45^{\mathrm{o}}\right)\hat{R}_p\left(45^{\mathrm{o}}\right)|H\rangle &= \langle H|\left[\hat{R}_p\left(45^{\mathrm{o}}\right)\hat{R}_p\left(45^{\mathrm{o}}\right)|H\rangle\right] \\ &= \langle H|V\rangle \\ &= 0.\end{aligned} \tag{4.13}$$

However, if eq. (4.11) were correct, we would find

$$\begin{aligned}\langle H|\hat{R}_p\left(45^{\mathrm{o}}\right)\hat{R}_p\left(45^{\mathrm{o}}\right)|H\rangle &= \left[\langle H|\hat{R}_p\left(45^{\mathrm{o}}\right)\right]\left[\hat{R}_p\left(45^{\mathrm{o}}\right)|H\rangle\right] \\ &\overset{?}{=} \langle +45|+45\rangle \\ &= 1.\end{aligned} \tag{4.14}$$

This disagrees with eq. (4.13), so it's wrong, and the conjecture we made in eq. (4.11) is as well:

$$\langle H|\hat{R}_p\left(45^\circ\right) \neq \langle +45|. \tag{4.15}$$

In order to get an equivalent result when operating on a bra vector, we need to use the adjoint operator (denoted with a †)

$$\langle H|\hat{R}_p^\dagger\left(45^\circ\right) = \langle +45|. \tag{4.16}$$

We pronounce $\hat{R}_p^\dagger$ as "$\hat{R}_p$-dagger." With this definition we have

$$\begin{aligned} \langle H|\hat{R}_p^\dagger\left(45^\circ\right)\hat{R}_p\left(45^\circ\right)|H\rangle &= \left[\langle H|\hat{R}_p^\dagger\left(45^\circ\right)\right]\left[\hat{R}_p\left(45^\circ\right)|H\rangle\right] \\ &= \langle +45|+45\rangle \\ &= 1. \end{aligned} \tag{4.17}$$

Any operator, regular or adjoint, can operate on either a bra or a ket (i.e., to the right or to the left). However, if we know that $\hat{O}|\psi_1\rangle = |\psi_2\rangle$, then we immediately know that $\langle\psi_1|\hat{O}^\dagger = \langle\psi_2|$, but we don't necessarily know what $\langle\psi_1|\hat{O}$ or $\hat{O}^\dagger|\psi_1\rangle$ are.

The adjoint of a product is

$$\left(\hat{O}_1\hat{O}_2\right)^\dagger = \hat{O}_2^\dagger\hat{O}_1^\dagger. \tag{4.18}$$

Note that the ordering gets reversed.

4.2.1 Unitary Operators

Look again at eq. (4.17); grouping things differently, we find that

$$\begin{aligned} \langle H|\hat{R}_p^\dagger\left(45^\circ\right)\hat{R}_p\left(45^\circ\right)|H\rangle &= \langle H|\left[\hat{R}_p^\dagger\left(45^\circ\right)\hat{R}_p\left(45^\circ\right)|H\rangle\right] \\ &= \langle H|\psi\rangle \\ &= 1. \end{aligned} \tag{4.19}$$

Here we have used the result of eq. (4.17) to set the last line equal to 1. In order to have $\langle H|\psi\rangle = 1$, it must be the case that $|\psi\rangle = |H\rangle$. So

$$\hat{R}_p^\dagger\left(45^\circ\right)\hat{R}_p\left(45^\circ\right)|H\rangle = |H\rangle, \tag{4.20}$$

and

$$\hat{R}_p^\dagger\left(45^\circ\right)\hat{R}_p\left(45^\circ\right) = \hat{1}. \tag{4.21}$$

Here $\hat{1}$ is the identity operator, which has no effect on any state. There's nothing special about 45^{o} here, in general

$$\hat{R}_p^{\dagger}(\theta)\hat{R}_p(\theta) = \hat{1}. \tag{4.22}$$

The rotation operator is a unitary operator. Unitary operators are somewhat special, and they are typically denoted by $\hat{U}$. The definition of a unitary operator is that

$$\hat{U}^{\dagger}\hat{U} = \hat{U}\hat{U}^{\dagger} = \hat{1}. \tag{4.23}$$

Unitary operators preserve the normalization of a state, which means that

$$\hat{U}\left|\psi_1\right\rangle = e^{i\phi}\left|\psi_2\right\rangle. \tag{4.24}$$

A unitary operator changes the state, but the complex constant is always of magnitude 1.

The inverse of $\hat{O}$ is denoted by $\hat{O}^{-1}$, and is defined by

$$\hat{O}^{-1}\hat{O} = \hat{O}\hat{O}^{-1} = \hat{1}. \tag{4.25}$$

For unitary operators $\hat{U}^{-1} = \hat{U}^{\dagger}$, but in general the inverse and the adjoint are not the same.

It should be fairly obvious that if we rotate first by 45^{o}, and then by -45^{o} about the same axis, we have to end up back where we started. This means

$$\hat{R}_p\left(-45^{\mathrm{o}}\right)\hat{R}_p\left(45^{\mathrm{o}}\right)\left|H\right\rangle = \left|H\right\rangle. \tag{4.26}$$

Comparing this to eq. (4.20), we learn that $\hat{R}_p^{\dagger}\left(45^{\mathrm{o}}\right) = \hat{R}_p\left(-45^{\mathrm{o}}\right)$; more generally

$$\hat{R}_p^{\dagger}(\theta) = \hat{R}_p^{-1}(\theta) = \hat{R}_p(-\theta). \tag{4.27}$$

The adjoint of the rotation operator turns out to be equivalent to rotating in the opposite direction.

4.3 THE PROJECTION OPERATOR

A $\mathrm{PA}_{\mathrm{HV}}$ and a beam block can be used to prepare photons in state $\left|H\right\rangle$. The $\mathrm{PA}_{\mathrm{HV}}$ transforms the state $\left|\psi\right\rangle$ into the state $\left|H\right\rangle$ with a certain probability, and this transformation can be described by an operator. With $\left|\psi\right\rangle = c_H\left|H\right\rangle + c_V\left|V\right\rangle$, it is useful to define this operator such that

$$\hat{P}_H\left|\psi\right\rangle = c_H\left|H\right\rangle, \tag{4.28}$$

and the output state is weighted by its probability amplitude. The operator $\hat{P}_H$ is called the projection operator onto $\left|H\right\rangle$, because it projects any state onto $\left|H\right\rangle$.

Looking more carefully at eq. (4.28), we see that

$$\begin{aligned}\hat{P}_H\left|\psi\right\rangle &= c_H\left|H\right\rangle \\ &= \left|H\right\rangle c_H \\ &= \left|H\right\rangle\left\langle H|\psi\right\rangle \\ &= \left(\left|H\right\rangle\left\langle H\right|\right)\left|\psi\right\rangle ,\end{aligned} \tag{4.29}$$

or $\hat{P}_H = \left|H\right\rangle\left\langle H\right|$. In general, the projection operator onto an arbitrary state $\left|\psi\right\rangle$ is given by

$$\hat{P}_\psi = \left|\psi\right\rangle\left\langle\psi\right|. \tag{4.30}$$

This equation does *not* represent an inner product. An inner product of the form $\left\langle\psi_1|\psi_2\right\rangle$ is a complex number, while eq. (4.30) represents an outer product, which is an operator.

In the *HV*-basis we can write an arbitrary state $\left|\psi\right\rangle$ as

$$\begin{aligned}\left|\psi\right\rangle &= c_H\left|H\right\rangle + c_V\left|V\right\rangle \\ &= \left|H\right\rangle c_H + \left|V\right\rangle c_V \\ &= \left|H\right\rangle\left\langle H|\psi\right\rangle + \left|V\right\rangle\left\langle V|\psi\right\rangle \\ &= \left(\left|H\right\rangle\left\langle H\right| + \left|V\right\rangle\left\langle V\right|\right)\left|\psi\right\rangle \\ &= \left(\hat{P}_H + \hat{P}_V\right)\left|\psi\right\rangle .\end{aligned} \tag{4.31}$$

Since $\left|\psi\right\rangle$ is arbitrary, it must be true that

$$\left(\hat{P}_H + \hat{P}_V\right) = \hat{1}. \tag{4.32}$$

The sum of the projection operators onto the basis states is equal to the identity operator. This is true in general, not just for $\hat{P}_H$ and $\hat{P}_V$; if the states $\left|\psi_j\right\rangle$ form an orthonormal basis, then

$$\hat{1} = \sum_j \hat{P}_{\psi j} = \sum_j \left|\psi_j\right\rangle\left\langle\psi_j\right|. \tag{4.33}$$

Indeed, showing that this relationship holds is one way to prove that the states $\left|\psi_j\right\rangle$ form a complete, orthonormal basis.

4.4 THE MATRIX REPRESENTATION OF OPERATORS

When expressing states as column vectors, it is necessary to “order” the basis states, as we learned in sec. 3.6. Equation (4.33) suggests that it is also occasionally useful to label them by integers. With this in mind, let's order and label our horizontal and vertical basis vectors as

$$|H\rangle = |HV_1\rangle, |V\rangle = |HV_2\rangle, \tag{4.34}$$

which is consistent with the ordering of the basis states we've used previously. In this notation a general state is expressed as

$$\begin{aligned} |\psi\rangle &= \psi_1 |HV_1\rangle + \psi_2 |HV_2\rangle \\ &= \sum_j \psi_j |HV_j\rangle, \end{aligned} \tag{4.35}$$

where

$$\psi_j = \langle HV_j|\psi\rangle. \tag{4.36}$$

Using these basis states, the column vector notation for a state is

$$|\psi\rangle \doteq \begin{pmatrix} \langle HV_1|\psi\rangle \\ \langle HV_2|\psi\rangle \end{pmatrix}_{HV} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}_{HV}, \tag{4.37}$$

and eq. (4.33) tells us that the identity operator may be expressed as

$$\hat{1} = \sum_j |HV_j\rangle\langle HV_j|. \tag{4.38}$$

There's no new physics here, not even any new math; it's just new notation.

4.4.1 Matrices

The operator $\hat{O}$ acts on the state $|\psi\rangle$, and produces a new state $|\varphi\rangle$:

$$|\varphi\rangle = \hat{O}|\psi\rangle. \tag{4.39}$$

We know the column vector representation of $|\psi\rangle$ [eq. (4.37)], and we want to determine the column vector representation of $|\varphi\rangle$:

$$|\varphi\rangle \doteq \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}_{HV}. \tag{4.40}$$

We know that

$$\varphi_i = \langle HV_i|\varphi\rangle = \langle HV_i|\hat{O}|\psi\rangle = \langle HV_i|\hat{O}\hat{1}|\psi\rangle. \tag{4.41}$$

If we express the identity using eq. (4.38), then

$$
\begin{aligned}
\varphi_i &= \langle HV_i|\hat{O}\left(\sum_j |HV_j\rangle\langle HV_j|\right)|\psi\rangle \\
&= \sum_j \langle HV_i|\hat{O}|HV_j\rangle\langle HV_j||\psi\rangle \\
&= \sum_j \langle HV_i|\hat{O}|HV_j\rangle\psi_j .
\end{aligned}
\tag{4.42}
$$

We define

$$
O_{ij} \equiv \langle HV_i|\hat{O}|HV_j\rangle, \tag{4.43}
$$

where, in general, the values of O_{ij} are complex. We can then write

$$
\varphi_i = \sum_j O_{ij}\psi_j . \tag{4.44}
$$

Compare this equation to eq. (1.41), and you'll see that it is equivalent to a matrix multiplying a column vector. Specifically, it represents

$$
\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}_{HV} = \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix}_{HV} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}_{HV} . \tag{4.45}
$$

The O_{ij}'s are called the matrix elements of $\hat{O}$, for reasons that should now be obvious.

Equation (4.45) is expressed in the HV-basis, but it is true in any basis, as long as the vectors and the matrix are all expressed in the *same* basis; more on this in the next section. We say that the matrix elements O_{ij}, eq. (4.43), form the matrix representation of the operator $\hat{O}$ in the HV-basis.

EXAMPLE 4.1

Find the matrix representation of the projection operator $\hat{P}_H$ in the HV-basis.

By definition this is

$$
\hat{P}_H \doteq \begin{pmatrix} \langle HV_1|\hat{P}_H|HV_1\rangle & \langle HV_1|\hat{P}_H|HV_2\rangle \\ \langle HV_2|\hat{P}_H|HV_1\rangle & \langle HV_2|\hat{P}_H|HV_2\rangle \end{pmatrix}_{HV} , \tag{4.46}
$$

which we can also write as

$$
\hat{P}_H \doteq \begin{pmatrix} \langle H|\hat{P}_H|H\rangle & \langle H|\hat{P}_H|V\rangle \\ \langle V|\hat{P}_H|H\rangle & \langle V|\hat{P}_H|V\rangle \end{pmatrix}_{HV} . \tag{4.47}
$$

We know that $\hat{P}_H|H\rangle = |H\rangle$ and $\hat{P}_H|V\rangle = 0$, so

$$\hat{P}_H \doteq \begin{pmatrix} \langle H|H\rangle & 0 \\ \langle V|H\rangle & 0 \end{pmatrix}_{HV} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_{HV}. \tag{4.48}$$

As can be seen from this example, the key to writing the matrix representation of an operator is knowing how the operator operates on each of the basis states. These operations essentially define the operator, and knowing them makes determining the matrix elements straightforward. Here's another example.

EXAMPLE 4.2

Find the matrix representation of the polarization rotation operator $\hat{R}_p(\theta)$.

We need to know what this operator does to each of our basis states. Consider rotating a horizontally polarized photon by the angle θ, as shown in fig. 4.2(a). We see that the resulting state is

$$\hat{R}_p(\theta)|HV_1\rangle = \hat{R}_p(\theta)|H\rangle = \cos\theta|H\rangle + \sin\theta|V\rangle. \tag{4.49}$$

Rotation of a vertically polarized photon, shown in fig. 4.2(b), yields

$$\hat{R}_p(\theta)|HV_2\rangle = \hat{R}_p(\theta)|V\rangle = -\sin\theta|H\rangle + \cos\theta|V\rangle. \tag{4.50}$$

The matrix representation of the operator is thus

$$\begin{aligned} \hat{R}_p(\theta) &\doteq \begin{pmatrix} \langle HV_1|\hat{R}_p(\theta)|HV_1\rangle & \langle HV_1|\hat{R}_p(\theta)|HV_2\rangle \\ \langle HV_2|\hat{R}_p(\theta)|HV_1\rangle & \langle HV_2|\hat{R}_p(\theta)|HV_2\rangle \end{pmatrix}_{HV} \\ &= \begin{pmatrix} \langle H|\left[\cos\theta|H\rangle + \sin\theta|V\rangle\right] & \langle H|\left[-\sin\theta|H\rangle + \cos\theta|V\rangle\right] \\ \langle V|\left[\cos\theta|H\rangle + \sin\theta|V\rangle\right] & \langle V|\left[-\sin\theta|H\rangle + \cos\theta|V\rangle\right] \end{pmatrix}_{HV} \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}_{HV}. \end{aligned} \tag{4.51}$$

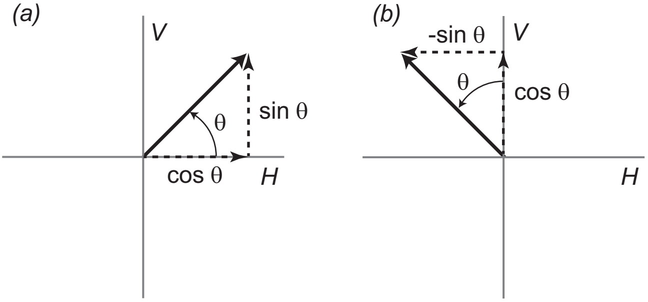


Fig 4.2 Rotating (a) horizontal polarization, and (b) vertical polarization by an angle θ about the $\boldsymbol{u}_z$ axis, which points out of the page.

Once we've determined the matrix representation of an operator, we can use it to determine how the operator will modify a particular state.

EXAMPLE 4.3

Determine $\hat{P}_H\left|+45\right\rangle$.

One way to do this is to use the outer product representation of $\hat{P}_H$:

$$\begin{aligned}\hat{P}_H\left|+45\right\rangle &= \left|H\right\rangle\left\langle H\right|\left[\frac{1}{\sqrt{2}}\left(\left|H\right\rangle+\left|V\right\rangle\right)\right]\\ &= \frac{1}{\sqrt{2}}\left|H\right\rangle,\end{aligned} \tag{4.52}$$

where we've used table 3.2. We could also use the matrix representation of $\hat{P}_H$ [eq. (4.48)]:

$$\begin{aligned}\hat{P}_H\left|+45\right\rangle &\doteq \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}_{HV}\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\end{pmatrix}_{HV}\\ &= \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 0\end{pmatrix}_{HV}\\ &= \frac{1}{\sqrt{2}}\left|H\right\rangle.\end{aligned} \tag{4.53}$$

Whether you use operators and states, or matrices and vectors, to do calculations like this is up to you. In the end, though, it is usually best to express your answer in terms of state vectors.

4.4.2 Correspondence Between Quantum and Classical Matrices

In sec. 4.3 we found that the physical implementation of the projection operator $\hat{P}_H$ was a horizontal polarizer. Furthermore, the matrix for $\hat{P}_H$, eq. (4.48), is the same as the Jones matrix for a horizontal polarizer listed in table 2.2. Each of the optical elements listed in table 2.2 will change the polarization state of an input photon, therefore each of these elements has a corresponding quantum mechanical operator. The matrix representation of the quantum operator, in the *HV*-basis, is exactly the same as the corresponding Jones matrix. This further validates the treatment of the polarization of a single photon presented in chapter 3.

There is no Jones matrix listed in table 2.2 that corresponds to the rotation operator $\hat{R}_p\left(\theta\right)$ of eq. (4.51). Didn't we say in chapter 2 that a half-wave plate rotates linear polarization? Why isn't the half-wave plate the physical implementation of the rotation operator? The answer is that if we apply the rotation operator $\hat{R}_p\left(\theta\right)$ to $\left|\phi\right\rangle$, the state of linear polarization along the angle ϕ, then we must get

$$\hat{R}_p(\theta)|\phi\rangle = |\phi+\theta\rangle, \tag{4.54}$$

independent of the input angle ϕ. In the problems you will verify that $\hat{R}_p(\theta)$, as expressed in eq. (4.51), does indeed do this. However, this is *not* what a half-wave plate does. A half-wave plate at a fixed angle rotates different input polarizations by different amounts. Consider, for example, a half-wave plate whose fast axis is at 45°. A horizontally polarized input photon will be rotated to vertical polarization, while a photon polarized along 45° will not be rotated at all.

There are optical materials, however, that are the physical embodiment of the rotation operator $\hat{R}_p(\theta)$; they rotate all input linear polarizations by the same amount. Such materials are said to exhibit optical activity. Optically active materials include quartz crystals and some sugar solutions (e.g., Karo® syrup) [4.1].

All this may make it seem like there is no difference between the quantum and classical descriptions of polarization; however, this is not the case. One difference is that when speaking of single photons, we can only talk about the probabilities of measurements, not intensities. Furthermore, a key caveat to the discussion so far is that we have been talking about individual photons. As we will learn in chapter 8, the quantum mechanical behavior of *two* photons, whose polarizations are correlated with each other, can be *very* different than we would expect from classical physics. In situations like these quantum mechanics gets especially interesting.

4.5 CHANGING BASES

We have mainly been discussing the *HV*-basis. However, it is occasionally useful to re-express things in a different basis (the basis of $|+45\rangle$ and $|-45\rangle$, for example). To do this, we need to know how to transform the representations of states and operators from one basis to another. There is a formal mathematical way to do this, called a similarity transformation, that is described in complement 4.A. Here we'll describe a less formal way to accomplish this transformation.

Start with the state $|\psi\rangle$, which is known in the *HV*-basis:

$$|\psi\rangle = \psi_H|H\rangle + \psi_V|V\rangle. \tag{4.55}$$

We want to transform $|\psi\rangle$ to the $\pm$45-basis. In other words, we need to find the coefficients in the expansion

$$|\psi\rangle = \psi_{+45}|+45\rangle + \psi_{-45}|-45\rangle. \tag{4.56}$$

To do this we need another piece of information; we need to know the relationships between the sets of basis vectors. In this case, we can find the needed relationships in table 3.2:

$$\begin{aligned} |+45\rangle &= \frac{1}{\sqrt{2}}\left(|H\rangle + |V\rangle\right), \\ |-45\rangle &= \frac{1}{\sqrt{2}}\left(|H\rangle - |V\rangle\right). \end{aligned} \tag{4.57}$$

Now it is useful to go back to first principles. The coefficients in eq. (4.56) are given by $\psi_{+45} = \langle +45 | \psi \rangle$ and $\psi_{-45} = \langle -45 | \psi \rangle$; these are essentially the definitions of the coefficients. Substituting in the known state from eq. (4.55) yields

$$\begin{aligned} \psi_{+45} &= \langle +45 | \left(\psi_H \left| H \right\rangle + \psi_V \left| V \right\rangle \right) = \psi_H \langle +45 | H \rangle + \psi_V \langle +45 | V \rangle , \\ \psi_{-45} &= \langle -45 | \left(\psi_H \left| H \right\rangle + \psi_V \left| V \right\rangle \right) = \psi_H \langle -45 | H \rangle + \psi_V \langle -45 | V \rangle . \end{aligned} \tag{4.58}$$

Switching the kets in eq. (4.57) to bras, and using the inner product, we find

$$\begin{aligned} \langle +45 | H \rangle &= \frac{1}{\sqrt{2}}, \quad \langle +45 | V \rangle = \frac{1}{\sqrt{2}}, \\ \langle -45 | H \rangle &= \frac{1}{\sqrt{2}}, \quad \langle -45 | V \rangle = -\frac{1}{\sqrt{2}}. \end{aligned} \tag{4.59}$$

Substituting these into eq. (4.58) yields the coefficients we want, in terms of the coefficients we know

$$\begin{aligned} \psi_{+45} &= \frac{1}{\sqrt{2}} \left(\psi_H + \psi_V \right), \\ \psi_{-45} &= \frac{1}{\sqrt{2}} \left(\psi_H - \psi_V \right). \end{aligned} \tag{4.60}$$

A different way to attack this problem would be to once again start with the definition $\psi_{+45} = \langle +45 | \psi \rangle$, but to then use the identity operator to write

$$\psi_{+45} = \langle +45 | \hat{1} | \psi \rangle . \tag{4.61}$$

The key now is expressing the identity operator in a useful way. Equation (4.61) already has a $\langle +45 |$ in it, so we need some information from the *HV*-basis. Let's get it by expressing $\hat{1}$ as the sum of the projection operators onto the *HV*-basis [eq. (4.32)]:

$$\begin{aligned} \psi_{+45} &= \langle +45 | \hat{1} | \psi \rangle \\ &= \langle +45 | \left(| H \rangle \langle H | + | V \rangle \langle V | \right) | \psi \rangle \\ &= \langle +45 | H \rangle \langle H | \psi \rangle + \langle 45 | V \rangle \langle V | \psi \rangle . \\ &= \langle +45 | H \rangle \psi_H + \langle 45 | V \rangle \psi_V \\ &= \frac{1}{\sqrt{2}} \left(\psi_H + \psi_V \right). \end{aligned} \tag{4.62}$$

This agrees with eq. (4.60). Once again we've used eq. (4.59), which originally came from the relationship between the basis states, eq. (4.57). ψ_{-45} can be found in a similar manner.

We don't need to know anything *new* to change bases. We don't need to memorize any new formulas, we just need to apply what we've already learned. All we need is the definition of the coefficients (e.g., $\psi_{+45} = \langle +45 | \psi \rangle$, etc.) and the relationship between the basis vectors [e.g., eq. (4.57) and/or eq. (4.59)]. It's also occasionally useful to use the fact that the identity operator $\hat{1}$ can be written as the sum of the projection operators onto the basis vectors.

EXAMPLE 4.4

Write the state $\left|L\right\rangle$ in the ±45-basis.

The representation we're looking for is

$$\left|L\right\rangle = c_{+45}\left|+45\right\rangle + c_{-45}\left|-45\right\rangle, \tag{4.63}$$

where

$$c_{+45} = \left\langle +45\middle|L\right\rangle \quad c_{-45} = \left\langle -45\middle|L\right\rangle. \tag{4.64}$$

We know how the ±45-basis relates to the *HV*-basis [eq. (4.59)], and from table 3.2 we know that

$$\left|L\right\rangle = \frac{1}{\sqrt{2}}\left[\left|H\right\rangle + i\left|V\right\rangle\right], \tag{4.65}$$

or

$$\left\langle H\middle|L\right\rangle = \frac{1}{\sqrt{2}},\ \left\langle V\middle|L\right\rangle = i\frac{1}{\sqrt{2}}. \tag{4.66}$$

Starting from eq. (4.64), c_{+45} can be found by using the identity operator:

$$\begin{aligned} c_{+45} &= \left\langle +45\middle|\hat{1}\middle|L\right\rangle \\ &= \left\langle +45\right|\left(\left|H\right\rangle\left\langle H\right| + \left|V\right\rangle\left\langle V\right|\right)\left|L\right\rangle \\ &= \left\langle +45\middle|H\right\rangle\left\langle H\middle|L\right\rangle + \left\langle 45\middle|V\right\rangle\left\langle V\middle|L\right\rangle \\ &= \left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{2}}\right)\left(i\frac{1}{\sqrt{2}}\right) \\ &= \frac{1}{2}(1+i). \end{aligned} \tag{4.67}$$

In a similar fashion, we find that $c_{-45} = \frac{1}{2}(1-i)$. Equation (4.63) thus becomes

$$\begin{aligned} \left|L\right\rangle &= \frac{1}{2}(1+i)\left|+45\right\rangle + \frac{1}{2}(1-i)\left|-45\right\rangle \\ &= \frac{1}{\sqrt{2}}\left(e^{i\pi/4}\left|+45\right\rangle + e^{-i\pi/4}\left|-45\right\rangle\right) \\ &= e^{i\pi/4}\frac{1}{\sqrt{2}}\left(\left|+45\right\rangle + e^{-i\pi/2}\left|-45\right\rangle\right) \\ &= e^{i\pi/4}\frac{1}{\sqrt{2}}\left(\left|+45\right\rangle - i\left|-45\right\rangle\right). \end{aligned} \tag{4.68}$$

Using the same ideas, we can also change the matrix representation of an operator from one basis set to another.

EXAMPLE 4.5

Transform $\hat{P}_H$ from the HV-basis to the ±45-basis.

$\hat{P}_H$ is given in the HV-basis in eq. (4.48). To write $\hat{P}_H$ in the ±45-basis, we first need to order the basis vectors. Use the ordering $\left|+45\right\rangle = \left|45_1\right\rangle$ and $\left|-45\right\rangle = \left|45_2\right\rangle$, and then

$$\begin{aligned}\hat{P}_H &\doteq \begin{pmatrix} \langle 45_1|\hat{P}_H|45_1\rangle & \langle 45_1|\hat{P}_H|45_2\rangle \\ \langle 45_2|\hat{P}_H|45_1\rangle & \langle 45_2|\hat{P}_H|45_2\rangle \end{pmatrix}_{45} \\ &= \begin{pmatrix} \langle +45|\hat{P}_H|+45\rangle & \langle +45|\hat{P}_H|-45\rangle \\ \langle -45|\hat{P}_H|+45\rangle & \langle -45|\hat{P}_H|-45\rangle \end{pmatrix}_{45}.\end{aligned} \tag{4.69}$$

Using eq. (4.57), we can write

$$\begin{aligned}\langle +45|\hat{P}_H|+45\rangle &= \left[\frac{1}{\sqrt{2}}\left(\langle H| + \langle V|\right)\right]\hat{P}_H\left[\frac{1}{\sqrt{2}}\left(|H\rangle + |V\rangle\right)\right] \\ &= \frac{1}{2}\left(\langle H|\hat{P}_H|H\rangle + \langle H|\hat{P}_H|V\rangle + \langle V|\hat{P}_H|H\rangle + \langle V|\hat{P}_H|V\rangle\right).\end{aligned} \tag{4.70}$$

We've written the matrix element in our new basis in terms of matrix elements of the original basis [eq. (4.48)]. We now use the known matrix elements, and obtain

$$\langle +45|\hat{P}_H|+45\rangle = \frac{1}{2}(1+0+0+0) = \frac{1}{2}. \tag{4.71}$$

The other three matrix elements in eq. (4.69) can be found in a similar manner, yielding

$$\hat{P}_H \doteq \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}_{45}. \tag{4.72}$$

Does this result make sense? In the next example we'll apply it to find out.

EXAMPLE 4.6

Calculate $\hat{P}_H\left|+45\right\rangle$, using eq, (4.72).

$$\begin{aligned}\hat{P}_H|+45\rangle &\doteq \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}_{45}\begin{pmatrix} 1 \\ 0 \end{pmatrix}_{45} \\ &= \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}_{45} \\ &= \frac{1}{2}\left(|+45\rangle + |-45\rangle\right).\end{aligned} \tag{4.73}$$

Using eq. (4.57), this is can be rewritten as

$$\begin{aligned}\hat{P}_H\left|+45\right\rangle &= \frac{1}{2}\left[\frac{1}{\sqrt{2}}\left(\left|H\right\rangle+\left|V\right\rangle\right)+\frac{1}{\sqrt{2}}\left(\left|H\right\rangle-\left|V\right\rangle\right)\right] \\ &= \frac{1}{\sqrt{2}}\left|H\right\rangle,\end{aligned} \tag{4.74}$$

which is in agreement with eq. (4.52).

The technique for changing bases we've discussed in this section has some advantages, and some disadvantages. The advantage is that there are no new formulas to memorize, and we're assured of getting the right answer by applying fundamental relationships. The disadvantage is that, in some sense, we're reinventing the wheel for each problem, and the calculations can be tedious. The technique for changing bases described in complement 4.A is different. It presents a formal method, so we don't need to reinvent the wheel. However, it does require remembering how to construct certain matrices that do the basis transformation, as well as remembering the proper ordering used in applying these matrices.

4.6 HERMITIAN OPERATORS

Recall that if $\hat{O}\left|\psi\right\rangle=\left|\psi'\right\rangle$, then $\left\langle\psi\right|\hat{O}^\dagger=\left\langle\psi'\right|$. Therefore, the matrix elements $O_{ij}^\dagger$ of the adjoint operator $\hat{O}^\dagger$, can be written in terms of the matrix elements of the operator $\hat{O}$ as

$$\begin{aligned}O_{ij}^\dagger &= \left\langle\psi_i\right|\hat{O}^\dagger\left|\psi_j\right\rangle \\ &= \left\langle\psi_i'\middle|\psi_j\right\rangle \\ &= \left\langle\psi_j\middle|\psi_i'\right\rangle^* \\ &= \left\langle\psi_j\right|\hat{O}\left|\psi_i\right\rangle^* \\ &= O_{ji}^*\ .\end{aligned} \tag{4.75}$$

There are certain special operators that are self-adjoint, in other words $\hat{O}^\dagger=\hat{O}$ and $O_{ij}^\dagger=O_{ij}$. Such operators are called Hermitian operators. From eq. (4.75), it can be seen that the matrix elements of a Hermitian operator satisfy $O_{ij}=O_{ji}^*$.

In chapter 1 we talked about the eigenvalues λ and eigenvectors $\boldsymbol{x}$ of the matrix $\bar{\mathbf{M}}$, which are the solutions to the equation $\bar{\mathbf{M}}\boldsymbol{x}=\lambda\boldsymbol{x}$. Since operators can be represented as matrices, it will probably not surprise you that operators have eigenvalues and eigenvectors (or eigenstates). The eigenvalues and eigenstates of $\hat{O}$ are the solutions to the equation

$$\hat{O}\left|\lambda_i\right\rangle=\lambda_i\left|\lambda_i\right\rangle. \tag{4.76}$$

In quantum mechanics it is customary to label the eigenstate and its corresponding eigenvalue by the same symbol, in this case λ_i. In an N-dimensional space an operator has N eigenvalues, so $i = 1, 2, \ldots, N$. One way to find the eigenvalues and eigenstates is to express the operator as a matrix, and then use the techniques described in sec. 1.2.

Hermitian operators have some interesting properties. The eigenvalues corresponding to Hermitian operators are always real, not complex, and the eigenstates are orthogonal to each other; you'll prove these properties in the problems. Furthermore, the eigenstates of a Hermitian operator in a finite-dimensional Hilbert space form a complete set.[1] Since these eigenstates can always be normalized, they form an orthonormal basis that can be used to express any vector in the Hilbert space.

The fact that the eigenstates of a Hermitian operator form an orthonormal basis also means that

$$\hat{1} = \sum_i |\lambda_i\rangle\langle\lambda_i|. \tag{4.77}$$

We can use this to write

$$\begin{aligned}\hat{O} &= \hat{O}\hat{1} \\ &= \sum_i \hat{O}|\lambda_i\rangle\langle\lambda_i| \\ &= \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|.\end{aligned} \tag{4.78}$$

This is frequently a useful way to express a Hermitian operator.

As we'll see in the next chapter, Hermitian operators play a special role in quantum mechanics.

4.7 References

[4.1] E. Hecht, *Optics, 4th ed*. (Addison Wesley, San Francisco, 2002), Sec. 8.10.

[4.2] L. E. Ballentine, *Quantum Mechanics, A Modern Development* (World Scientific, Singapore, 1998), Sec. 1.3.

---

[1] In an infinite-dimensional Hilbert space, the eigenstates may or may not form a complete set [4.2].

4.8 PROBLEMS

4.1* Prove that a unitary operator does not change the magnitude of a state vector.

4.2 Show that $\hat{P}_{\psi}^{2} = \hat{P}_{\psi}$.

4.3 If $\left|\psi\right\rangle = c_H \left|H\right\rangle + c_V \left|V\right\rangle$, what is $\left\langle\psi\right|\hat{P}_H$?

4.4 Using the matrix representation of the polarization rotation operator $\hat{R}_p(\theta)$, verify that it is unitary.

4.5 Using the matrix representation of the polarization rotation operator $\hat{R}_p(\theta)$, verify that $\hat{R}_p^{\dagger}(\theta) = \hat{R}_p(-\theta)$.

4.6* Using the matrix representation of the polarization rotation operator $\hat{R}_p(\theta)$, verify that $\hat{R}_p(\theta)\left|\phi\right\rangle = \left|\phi+\theta\right\rangle$, where $\left|\phi\right\rangle$ is the state of a photon linearly polarized along an angle ϕ from the horizontal.

4.7 Determine the column vectors representing the states $\left|H\right\rangle$ and $\left|V\right\rangle$ using the states $\left|L\right\rangle$ and $\left|R\right\rangle$ as a basis.

4.8 Verify that the vectors you found in problem 4.7 are orthogonal.

4.9 Determine the column vector representing the elliptically polarized state $\left|e\right\rangle = \cos(\theta)\left|H\right\rangle + e^{i\phi}\sin(\theta)\left|V\right\rangle$, using the states $\left|+45\right\rangle$ and $\left|-45\right\rangle$ as a basis.

4.10 Work out the matrix representations of the projection operators $\hat{P}_H = \left|H\right\rangle\left\langle H\right|$ and $\hat{P}_V = \left|V\right\rangle\left\langle V\right|$ using the states $\left|L\right\rangle$ and $\left|R\right\rangle$ as a basis. Check that the relationships $\hat{P}_H^{2} = \hat{P}_H$, $\hat{P}_V^{2} = \hat{P}_V$ and $\hat{P}_H\hat{P}_V = \hat{P}_V\hat{P}_H = 0$ are satisfied using these matrix representations.

4.11 Express $\hat{R}_p(\theta)$ in the ±45-basis.

4.12 Using the result of problem 4.11, verify that $\hat{R}_p(45^{\circ})\left|+45\right\rangle = \left|V\right\rangle$.

4.13 Express $\hat{R}_p(\theta)$ in the circular polarization basis.

4.14* Find the eigenvalues and eigenstates of $\hat{P}_{+45} = \left|+45\right\rangle\left\langle +45\right|$. Physically, do these make sense?

4.15 Find the eigenvalues and eigenstates of $\hat{R}_p(\theta)$.

4.16* Prove that the eigenvalues of a Hermitian operator are real.

4.17 Prove that if $\hat{A} \equiv \left|a_1\right\rangle\left\langle a_2\right|$, then $\hat{A}^{\dagger} = \left|a_2\right\rangle\left\langle a_1\right|$.

4.18* Prove that for a Hermitian operator, eigenstates corresponding to nondegenerate (distinct, not equal) eigenvalues are orthogonal. Hint: Look at the matrix elements $\left\langle\lambda_1\right|\hat{O}\left|\lambda_2\right\rangle$ corresponding to distinct eigenstates. (You won't prove it in this problem, but eigenstates corresponding to degenerate [equal] eigenvalues can be constructed so that they are orthogonal.)

4.19* Compute $e^{\hat{O}}|\lambda\rangle$, assuming $|\lambda\rangle$ is an eigenstate of $\hat{O}$.

4.20* Prove that for a Hermitian operator $\hat{O}$, it is possible to express a function of $\hat{O}$, $f(\hat{O})$, as

$$f(\hat{O}) = \sum_i f(\lambda_i)|\lambda_i\rangle\langle\lambda_i|. \tag{4.79}$$

Here the λ_i's are the eigenvalues of $\hat{O}$, and the $|\lambda_i\rangle$'s are the corresponding eigenstates.

COMPLEMENT 4.A

Similarity Transformations

Here we'll describe a more formal method of changing the representations of vectors and matrices from one basis to another.

4.A.1 Changing the Representation of Vectors

Suppose we know the representation of the state $\left|\psi\right\rangle$ in the basis consisting of the orthonormal states $\left|o_j\right\rangle$:

$$\left|\psi\right\rangle = \sum_{j=1}^{N} \psi_j^o \left|o_j\right\rangle. \tag{4.A.1}$$

The $\left|o_j\right\rangle$'s form what we will refer to as the original basis, or the *o*-basis. We'll use the vector $\boldsymbol{\psi}^o$ to refer to the column vector that represents $\left|\psi\right\rangle$ in the *o*-basis:

$$\boldsymbol{\psi}^o = \begin{pmatrix} \psi_1^o \\ \bullet \\ \bullet \\ \bullet \\ \psi_N^o \end{pmatrix}_o . \tag{4.A.2}$$

We want to know the representation of the state $\left|\psi\right\rangle$ in a new orthonormal basis, consisting of the states $\left|n_i\right\rangle$:

$$\left|\psi\right\rangle = \sum_{i=1}^{N} \psi_i^n \left|n_i\right\rangle. \tag{4.A.3}$$

The ψ_i^n's are the elements of the column vector $\boldsymbol{\psi}^n$.[2]

We find the coefficients ψ_i^n using

$$\begin{aligned}\psi_i^n &= \langle n_i | \psi \rangle \\ &= \langle n_i | \left(\sum_{j=1}^{N} \psi_j^o | o_j \rangle \right) \\ &= \sum_{j=1}^{N} \langle n_i | o_j \rangle \psi_j^o \\ &= \sum_{j=1}^{N} S_{ij} \psi_j^o \,.\end{aligned} \tag{4.A.4}$$

Here we've defined

$$S_{ij} \equiv \langle n_i | o_j \rangle, \tag{4.A.5}$$

which are the matrix elements of $\overline{\mathbf{S}}$. Comparing eq. (4.A.4) to eq. (1.41), we see that eq. (4.A.4) is equivalent to the multiplication of a matrix and a column vector. Specifically, eq. (4.A.4) is equivalent to

$$\boldsymbol{\psi}^n = \overline{\mathbf{S}} \boldsymbol{\psi}^o. \tag{4.A.6}$$

Thus, to change from $\boldsymbol{\psi}^o$, which is the representation of $|\psi\rangle$ in the *o*-basis, to the representation of $|\psi\rangle$ in the *n*-basis, $\boldsymbol{\psi}^n$, we use eq. (4.A.6). In sec. 4.5 we said that in order to change bases it was necessary to have information about the relationships between the sets of basis vectors; in this method of changing bases, this information is contained in $\overline{\mathbf{S}}$.

The inverse transformation, from the *n*-basis to the *o*-basis, is $\boldsymbol{\psi}^o = \overline{\mathbf{S}}^{-1} \boldsymbol{\psi}^n$. However, the matrix $\overline{\mathbf{S}}$ is unitary, as you'll prove in the problems, so $\overline{\mathbf{S}}^{-1} = \overline{\mathbf{S}}^{\dagger}$. Thus, the inverse transformation is usually written as

$$\boldsymbol{\psi}^o = \overline{\mathbf{S}}^{\dagger} \boldsymbol{\psi}^n. \tag{4.A.7}$$

The adjoint matrix is most easily determined by remembering that its matrix elements are related to those of the original matrix by $S_{ij}^{\dagger} = S_{ji}^{*}$ [eq. (4.75)].

---

[2] We need to refer to the vectors $\boldsymbol{\psi}^o$ and $\boldsymbol{\psi}^n$ separately, because they are basis-dependent representations of the basis-independent state $|\psi\rangle$. Because the superscripts specify the basis, we can use = instead of $\doteq$ for these vectors.

EXAMPLE 4.A.1

Write the state $|L\rangle$ in the ±45-basis.

From table 3.2, we know $|L\rangle$ as a column vector in the HV-basis:

$$\mathbf{L}^{HV} \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}_{HV}. \tag{4.A.8}$$

The matrix that takes us from the HV-basis to the ±45-basis has elements

$$S_{ij} \equiv \langle 45_i | HV_j \rangle, \tag{4.A.9}$$

using the notation in secs. 4.4 and 4.5. Thus,

$$\begin{aligned}\bar{\mathbf{S}} &\doteq \begin{pmatrix}\langle 45_1|HV_1\rangle & \langle 45_1|HV_2\rangle\\ \langle 45_2|HV_1\rangle & \langle 45_2|HV_2\rangle\end{pmatrix}_{HV\to\pm45}\\ &= \begin{pmatrix}\langle +45|H\rangle & \langle +45|V\rangle\\ \langle -45|H\rangle & \langle -45|V\rangle\end{pmatrix}_{HV\to\pm45}.\end{aligned} \tag{4.A.10}$$

We've written the subscript $HV \to \pm45$ on this matrix because it's not really written in either basis, it's used to take us from one basis to another. The matrix elements can be obtained from eq. (4.59), yielding

$$\bar{\mathbf{S}} \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}_{HV\to\pm45}. \tag{4.A.11}$$

Using this, we find

$$\begin{aligned}\mathbf{L}^{\pm45} &= \bar{\mathbf{S}}\mathbf{L}^{HV}\\ &\doteq \left[\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}_{HV\to\pm45}\right]\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}_{HV}\\ &= \frac{1}{2}\begin{pmatrix}(1+i)\\ (1-i)\end{pmatrix}_{\pm45},\end{aligned} \tag{4.A.12}$$

or

$$|L\rangle = \frac{1}{2}(1+i)|+45\rangle + \frac{1}{2}(1-i)|-45\rangle. \tag{4.A.13}$$

This agrees with eq. (4.68).

4.A.2 Changing the Representation of Matrices

Consider $\bar{\mathbf{M}}^o$, which is expressed in the original basis. It performs the transformation

$$\varphi^o = \bar{\mathbf{M}}^o \psi^o. \tag{4.A.14}$$

Transforming φ^o to φ^n, we see that:

$$\begin{aligned}\varphi^n &= \bar{\mathbf{S}}\varphi^o \\ &= \bar{\mathbf{S}}\left(\bar{\mathbf{M}}^o\psi^o\right) \\ &= \bar{\mathbf{S}}\bar{\mathbf{M}}^o\left(\bar{\mathbf{S}}^\dagger\psi^n\right) \\ &= \left(\bar{\mathbf{S}}\bar{\mathbf{M}}^o\bar{\mathbf{S}}^\dagger\right)\psi^n \\ &= \bar{\mathbf{M}}^n\psi^n,\end{aligned} \tag{4.A.15}$$

where we've used eqs. (4.A.6) and (4.A.7). Evidently, matrices are transformed from one basis to another using

$$\bar{\mathbf{M}}^n = \bar{\mathbf{S}}\bar{\mathbf{M}}^o\bar{\mathbf{S}}^\dagger. \tag{4.A.16}$$

This represents a similarity transformation.[3]

EXAMPLE 4.A.2

Transform the matrix representation of $\hat{P}_H$ from the HV-basis to the ±45-basis.

The HV-basis representation of $\hat{P}_H$ is given in eq. (4.48). Using $\bar{\mathbf{S}}$ from eq. (4.A.11), and the fact that $S^\dagger_{ij} = S^*_{ji}$, we find

$$\begin{aligned}\bar{\mathbf{P}}_H^{\pm 45} &= \bar{\mathbf{S}}\bar{\mathbf{P}}_H^{HV}\bar{\mathbf{S}}^\dagger \\ &\doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}_{HV\to\pm 45}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}_{HV}\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}_{HV\to\pm 45} \\ &= \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}_{HV\to\pm 45}\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}_{\pm 45}.\end{aligned} \tag{4.A.17}$$

This is in agreement with eq. (4.72).

---

[3] Since $\bar{\mathbf{S}}$ is unitary, it is more properly a unitary similarity transformation.

4.A.3 Problems

4.A.1* Prove that $\overline{\mathbf{S}}$ is unitary.

4.A.2 Transform the column vectors representing the states $\left|H\right\rangle$ and $\left|V\right\rangle$ from the *HV*-basis to the circular polarization basis.

4.A.3 Transform the column vector representing the elliptical polarization state $\left|e\right\rangle = \cos\left(\theta\right)\left|H\right\rangle + e^{i\phi}\sin\left(\theta\right)\left|V\right\rangle$ from the *HV*-basis to the ±45-basis.

4.A.4 Transform the matrix representing $\hat{R}_p\left(\theta\right)$ from the *HV*-basis to the ±45-basis.

CHAPTER 5

Measurement

In classical physics there is usually little confusion about how we make the correspondence between the mathematical formalism we use to describe physical systems, and the measurements we make on those systems. For example, Newton's Second Law says that $\boldsymbol{F} = m\boldsymbol{a}$. Here, $\boldsymbol{F}$ is the mathematical representation of the force; it can be measured, at least in principle, using something like a spring scale. However, at this point it's probably not obvious to you how to relate the results of a physical measurement to the quantum state—things are not as clear-cut in quantum mechanics as they are in classical mechanics. The purpose of this chapter is to describe how we make correspondences between the mathematical formalism that is quantum mechanics, and the results of specific measurements.

5.1 MEASURING POLARIZATION

Suppose we have a photon prepared in state $|\psi\rangle$, and we'd like to learn something about its polarization. To do this we might employ a measurement apparatus like that shown in fig. 5.1, where the photon passes through a horizontal/vertical polarization analyzer PA_{HV}, and then through one of two detectors. The detectors pictured in fig. 5.1 are of a special type, which we'll refer to as nondestructive detectors. We can think of them as being made of a transparent material, like glass, that transmits a photon without changing its polarization, yet it still produces a signal when a photon passes through. A nondestructive measurement of a photon is extremely difficult, but in principle possible to perform.[1]

Note that the PA_{HV} by itself is not sufficient to perform a useful measurement. We, being macroscopic classical objects, want to obtain information from a single photon, which is a microscopic quantum-mechanical object. We desire to know whether the photon emerged from the V-port or the H-port of the PA_{HV}. That information must be "amplified" from

---

[1] Much more common is a destructive photon measurement, in which the photon is absorbed by the detector.

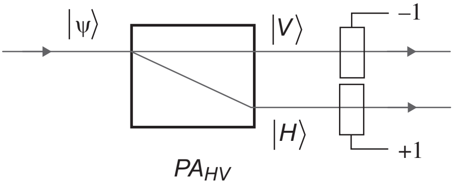


Fig 5.1 Performing a polarization measurement using nondestructive photon detectors.

microscopic to macroscopic. This amplification is irreversible; we can't perfectly regenerate the input state by running things backward.[2] We use the term "meter" to describe an apparatus that forms such a bridge from the microscopic to the macroscopic. In fig. 5.1, the meters that are performing the actual measurement are the two detectors.

The measurement described in fig. 5.1 has two possible outcomes: The photon is observed to be either horizontally polarized, or vertically polarized. In order for us to compute the statistics of the results of many such measurements, it is convenient to assign a numerical value to each of the possible measurement outcomes. We'll assign a value of +1 to a measurement that yields a horizontally polarized photon, and −1 to a measurement that yields a vertically polarized photon.

If we perform many measurements, when the input state is $|H\rangle$, the result is a series of +1s, which average to +1. Likewise, if the input state is $|V\rangle$, the result is a series of −1s, and the average is −1. A general input state is a linear combination of horizontal and vertical, $|\psi\rangle = c_H|H\rangle + c_V|V\rangle$. The measurement results are random for this state, and the probabilities of obtaining +1 or –1 are

$$\begin{aligned} P(+1\,\|\,\psi\rangle) &= P(H\,\|\,\psi\rangle) = |\langle H|\psi\rangle|^2 = |c_H|^2, \\ P(-1\,\|\,\psi\rangle) &= P(V\,\|\,\psi\rangle) = |\langle V|\psi\rangle|^2 = |c_V|^2. \end{aligned} \tag{5.1}$$

The notation here emphasizes that the probabilities are conditioned on the input state (see sec. 1.1.4).

Since we know the probability of obtaining a particular measurement, we can obtain the average of the measurements by weighting the possible values by their corresponding probabilities and summing, as in eq. (1.16). Using eq. (5.1), we find

$$\begin{aligned} \text{average polarization } &= (+1)P(+1\,\|\,\psi\rangle) + (-1)P(-1\,\|\,\psi\rangle) \\ &= |c_H|^2 - |c_V|^2. \end{aligned} \tag{5.2}$$

Remember that this average was computed for a measurement performed with a PA_{HV}. Using an apparatus that measures circular polarization (a PA_C), for example, would yield a different answer.

---

[2] Many physicists believe that irreversibility plays an important role in what we refer to as a quantum measurement [5.1].

5.1.1 The Polarization Operator

Let's define what we will refer to as the HV-polarization operator, $\hat{\wp}_{HV}$. One way to define an operator is by its action on a complete set of basis states. In this way $\hat{\wp}_{HV}$ can be defined by

$$\hat{\wp}_{HV}\left|H\right\rangle = (+1)\left|H\right\rangle,\ \hat{\wp}_{HV}\left|V\right\rangle = (-1)\left|V\right\rangle. \tag{5.3}$$

Each of the basis states is an eigenstate of $\hat{\wp}_{HV}$; $\left|H\right\rangle$ corresponds to eigenvalue +1, and $\left|V\right\rangle$ corresponds to eigenvalue –1. In a 2-D Hilbert space, operators have two eigenvalues and eigenstates, so these are all of the eigenvalues and eigenstates of $\hat{\wp}_{HV}$. Another way to express $\hat{\wp}_{HV}$ is as [eq. (4.78)]:

$$\hat{\wp}_{HV} = (+1)\left|H\right\rangle\left\langle H\right| + (-1)\left|V\right\rangle\left\langle V\right|. \tag{5.4}$$

The operator $\hat{\wp}_{HV}$ is Hermitian: $\hat{\wp}_{HV}^{\dagger} = \hat{\wp}_{HV}$. The facts that: (i) its eigenvalues are real, (ii) its eigenstates are orthogonal, and (iii) its eigenstates form a basis, are all consistent with $\hat{\wp}_{HV}$ being Hermitian.

5.2 THE POSTULATES OF QUANTUM MECHANICS

Not all textbooks agree on the number of postulates that there are in quantum mechanics, or on their ordering. Some books don't enumerate postulates at all; they simply lump them all together and refer to them as the statistical interpretation of quantum mechanics. However, I want to be clear that the connections we make between mathematics and experiments are based on postulates. Fortunately, these postulates have served us well for many years, and are all the stronger for it.

5.2.1 Postulate I—States

> *Postulate I—At any instant in time, the state of a physical system is specified by a ket vector,* $\left|\psi\right\rangle$.

It is common to say that the system is "in" state $\left|\psi\right\rangle$. This is essentially an assumption we've been making all along, and we've just codified it as a postulate. Different people think about the state in different ways. I find it useful to think about quantum states in terms of information. The state contains all of the knowable information about a quantum system. If we know the state of a system, we know the probabilities for all possible outcomes, of all possible measurements, that we could perform on that system. Notice that we said that we'd know "the probabilities for all possible outcomes," not the outcomes themselves. In quantum mechanics there are very few cases where the outcome is known beforehand—all measurements are probabilistic. The best that quantum mechanics can do is predict the probabilities of measurement results, and this information is contained in the state.

We need to add one caveat to Postulate I: not *all* quantum states can be described by ket vectors $|\psi\rangle$. So far we've specialized to the case of pure states, and all pure states are described by ket vectors. However, in later chapters we'll learn that quantum mechanics also allows for mixed states, which cannot be represented by kets. At that point we will need to generalize our notion of the quantum state, but the fact that there exists a state that contains all information about a quantum system remains valid.

5.2.2 Postulate II—Observables

Postulate II—Physically measurable quantities are represented by observables (e.g., O). Every observable has a corresponding Hermitian operator ($\hat{O}$).

As an example we've used before, polarization is an observable—it is possible to measure it. The *HV*-polarization observable is $\wp_{HV}$, and its corresponding Hermitian operator is $\hat{\wp}_{HV}$. Other observables include energy, position, momentum, etc., and they all have their own corresponding Hermitian operators.

A more mathematical definition of an observable in quantum mechanics is a Hermitian operator whose eigenstates form a complete basis set [5.2]. Since the eigenstates of a finite-dimensional Hermitian operator always form a complete basis, by this definition all finite-dimensional Hermitian operators correspond to observables. Given this definition, some physicists don't bother to distinguish between the observable and the operator. However, I believe it is useful to make a distinction between the physically observed quantity, and its corresponding Hermitian operator.

5.2.3 Postulate III—Eigenvalues and Eigenstates

This postulate has two parts. The first is:

Postulate III(a)—If an observable O, with corresponding Hermitian operator $\hat{O}$, is measured, the possible results of the measurement are given by the eigenvalues of $\hat{O}$.

When we measure polarization using the apparatus in fig. 5.1, the possible outcomes are +1 for horizontal, and –1 for vertical, which are the eigenvalues of $\hat{\wp}_{HV}$. We never obtain a measurement value of 0, even though the *average* polarization may be 0 (see example 5.2, below). Quantum mechanics gets its name from the fact that it allows for only discrete measurement values of certain quantities, in situations where classical mechanics allows continuous values. In these situations the corresponding Hermitian operator has a discrete eigenvalue spectrum, and these discrete eigenvalues represent the allowed outcomes of measurements.

It is this postulate that requires us to associate a Hermitian operator with an observable in postulate II. A measurement returns a real number, not a complex number.[3]

---

[3] You might imagine determining a complex value by measuring *two* observables, corresponding to the real and imaginary parts. A single observable cannot take on complex values.

In order to postulate that measurements are given by eigenvalues, while still guaranteeing that they are real, we must associate observables with Hermitian operators, because only these operators are guaranteed to have real eigenvalues.

> *Postulate III(b)—For a measurement of O, with corresponding Hermitian operator $\hat{O}$, on a system prepared in state $|\psi\rangle$, the probability of obtaining the eigenvalue λ as a measurement result is $P(\lambda\,\|\,\psi\rangle) = |\langle\lambda|\psi\rangle|^2$, where $|\lambda\rangle$ is the eigenstate corresponding to λ. After a measurement returning the value λ, the system is left in state $|\lambda\rangle$.*

We have already discussed the first part of this postulate, regarding the probability of a measurement result, in detail. I'd like to stress, however, that the probability refers to *measuring the value* λ, not *being in state* $|\lambda\rangle$ prior to the measurement—as stated in the postulate, prior to measurement the system was in state $|\psi\rangle$. The probability postulate is called Born's rule, after Max Born.

The second part of postulate III(b), about the state of the system after the measurement, is new. It says that the act of measurement may *change* the state of the system; before the measurement the system is in state $|\psi\rangle$, and after the measurement the system is in state $|\lambda\rangle$. This is sometimes referred to as the von Neumann projection postulate, after John von Neumann, who is credited with introducing it.

Suppose that the +1 detector in fig. 5.1 fires, indicating that the photon is horizontally polarized. Postulate III says that when leaving the measurement apparatus, the photon is in state $|H\rangle$. To verify this, we must perform a second polarization measurement, as shown in fig. 5.2. If the photon leaves the first measurement apparatus in state $|H\rangle$, the second measurement will yield +1 with probability $P(+1\,\|H\rangle) = P(H\,\|H\rangle) = |\langle H|H\rangle|^2 = 1$. This is indeed what we would find.

This result shouldn't surprise you for polarization measurements, but postulate III applies to *any* quantum measurement, with occasionally surprising results. To summarize: After a measurement returning the result λ, the system is left in the

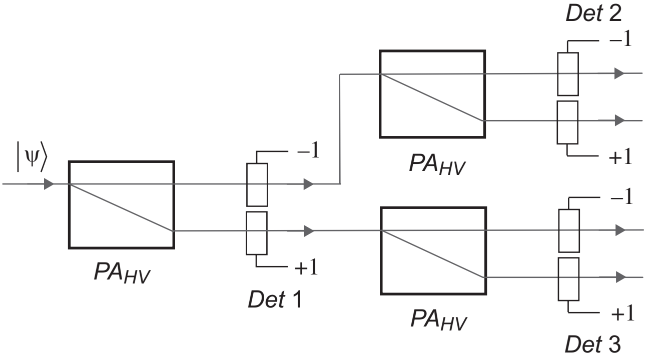


Fig 5.2 Repeated measurements. If detector 1 (Det 1) measures +1, then detector 3 (Det 3) will measure +1 with 100% probability, and detector 2 (Det 2) will not register anything.

corresponding eigenstate $\left|\lambda\right\rangle$. If a subsequent measurement is performed, before the state of the system has a chance to change, the result λ will again be obtained, with 100% probability.

EXAMPLE 5.1

A measurement of $\wp_{HV}$ is performed for a beam of photons prepared in the right-circular polarization state $\left|R\right\rangle$. What are the possible outcomes of this measurement? What are their probabilities? For each outcome, what state is the system left in?

The possible results are the eigenvalues: +1 for horizontal, and –1 for vertical.

For a measurement yielding +1, the probability is $P\left(+1\middle\|R\right\rangle)=\left|\left\langle H|R\right\rangle\right|^2=1/2$. After such a measurement, the beam of photons is left in state $\left|H\right\rangle$.

For a measurement yielding –1, the probability is $P\left(-1\middle\|R\right\rangle)=\left|\left\langle V|R\right\rangle\right|^2=1/2$. After such a measurement, the beam of photons is left in state $\left|V\right\rangle$.

We once again need to make a caveat to the wording we've used in a postulate. Postulate III(b) assumes that each eigenvalue has a *unique* corresponding eigenstate. If this is the case, then the operator is said to possess a nondegenerate eigenvalue spectrum. However, sometimes there is more than one eigenstate corresponding to a particular eigenvalue; this eigenvalue is said to be degenerate. The degeneracy of the eigenvalue equals the number of eigenstates that correspond to the eigenvalue (e.g., if an eigenvalue has three eigenstates, it is said to be three-fold degenerate). For degenerate eigenvalues, each of the corresponding eigenstates contributes to the probability of the measurement. We will generalize postulate III(b) to account for this fact when the need arises.

5.3 EXPECTATION VALUES

Let's examine some further ways in which the polarization operator $\hat{\wp}_{HV}$ is useful by computing $\left\langle\psi\right|\hat{\wp}_{HV}\left|\psi\right\rangle$ for the general state $\left|\psi\right\rangle=c_H\left|H\right\rangle+c_V\left|V\right\rangle$. Using eq. (5.4), we find

$$\begin{aligned}\left\langle\psi\right|\hat{\wp}_{HV}\left|\psi\right\rangle&=\left\langle\psi\right|\left[(+1)\left|H\right\rangle\left\langle H\right|+(-1)\left|V\right\rangle\left\langle V\right|\right]\left|\psi\right\rangle\\&=(+1)\left\langle\psi|H\right\rangle\left\langle H|\psi\right\rangle+(-1)\left\langle\psi|V\right\rangle\left\langle V|\psi\right\rangle\\&=(+1)\left|\left\langle H|\psi\right\rangle\right|^2+(-1)\left|\left\langle V|\psi\right\rangle\right|^2\\&=\left|c_H\right|^2-\left|c_V\right|^2.\end{aligned}\tag{5.5}$$

Note that this is equal to the average polarization we found in eq. (5.2), and since we've done this calculation for an arbitrary state, $\left\langle\psi\right|\hat{\wp}_{HV}\left|\psi\right\rangle$ will always yield the average polarization.

We call the quantity $\langle \hat{O} \rangle \equiv \langle \psi | \hat{O} | \psi \rangle$ the expectation value of O. The expectation value is equal to the average we'd obtain from measurements of O, to within the experimental and statistical errors, assuming the system is prepared in state $|\psi\rangle$.

Applying the procedures described in chapter 1 to $\hat{O}$, we find that the variance ΔO^2 is given by

$$\begin{aligned}
\Delta O^2 &= \left\langle \left(\delta \hat{O}\right)^2 \right\rangle \\
&= \left\langle \left(\hat{O} - \langle \hat{O} \rangle\right)^2 \right\rangle \\
&= \left\langle \left(\hat{O}^2 - 2\hat{O}\langle \hat{O} \rangle + \langle \hat{O} \rangle^2\right) \right\rangle \\
&= \langle \hat{O}^2 \rangle - 2\langle \hat{O} \rangle \langle \hat{O} \rangle + \langle \hat{O} \rangle^2 \\
&= \langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2 .
\end{aligned} \tag{5.6}$$

The standard deviation ΔO is once again given by the square root of the variance.

EXAMPLE 5.2

Compute the expectation value and the standard deviation of $\wp_{HV}$, for a beam of photons prepared in the state $|R\rangle$.

The expectation value is given by

$$\begin{aligned}
\langle \hat{\wp}_{HV} \rangle &= \langle R | \hat{\wp}_{HV} | R \rangle \\
&= \langle R | \left[(+1)|H\rangle\langle H| + (-1)|V\rangle\langle V| \right] | R \rangle \\
&= (+1)\left|\langle H | R \rangle\right|^2 + (-1)\left|\langle V | R \rangle\right|^2 \\
&= \left|\frac{1}{\sqrt{2}}\right|^2 - \left|\frac{-i}{\sqrt{2}}\right|^2 \\
&= 0.
\end{aligned} \tag{5.7}$$

Physically, this can be interpreted by the fact that a right-circularly polarized beam will split equally at a PA_{HV}. Half the time we will measure a vertically polarized photon (–1), and half the time we will measure a horizontally polarized photon (+1). These measurements will average to 0 (to within the statistical errors).

To find variance, we must first determine $\hat{\wp}_{HV}{}^2$:

$$\begin{aligned}
\hat{\wp}_{HV}{}^2 &= \left[(+1)|H\rangle\langle H| + (-1)|V\rangle\langle V| \right]\left[(+1)|H\rangle\langle H| + (-1)|V\rangle\langle V| \right] \\
&= (+1)^2 |H\rangle\langle H| + (-1)^2 |V\rangle\langle V| \\
&= |H\rangle\langle H| + |V\rangle\langle V| \\
&= \hat{1}.
\end{aligned} \tag{5.8}$$

The last line follows from eq. (4.33). The variance $\Delta \wp_{HV}{}^2$ is then

$$\begin{aligned}\Delta \wp_{HV}^{\,2} &= \left\langle \hat{\wp}_{HV}^{\,2} \right\rangle - \left\langle \hat{\wp}_{HV} \right\rangle^2 \\ &= \left\langle R \middle| \hat{1} \middle| R \right\rangle - \left\langle \hat{\wp}_{HV} \right\rangle^2 \\ &= 1 - 0 \\ &= 1\,.\end{aligned} \tag{5.9}$$

The standard deviation is $\Delta \wp_{HV} = \sqrt{\Delta \wp_{HV}^{\,2}} = 1$, which is reasonable for a value that alternates randomly between +1 and –1.

The term expectation value can be misleading, as seen from the previous example. There we found that the expectation value of $\wp_{HV}$ was 0, so you might expect that 0 would be the most likely result of a measurement. However, the only allowed measurement values are ±1, so you would *never* expect to measure 0. The expectation value is the expected value of the *average* of a series of measurements, not the expected value of any *individual* measurement.

5.4 OPERATORS AND MEASUREMENTS

We should clarify something that can be confusing. Measurements correspond to Hermitian operators; operators change one state into another; after a measurement the state of a system is changed. Given these facts, it is tempting to believe that the output state resulting from a measurement can be obtained mathematically by applying the corresponding Hermitian operator to the input state. In other words, assume we want to make a measurement of observable O, for a system prepared in the input state $\left|\psi_i\right\rangle$. You are likely to *want* to say that after the measurement the system is left in state $\hat{O}\left|\psi_i\right\rangle$. However, after a measurement the system is generally *not* left in state $\hat{O}\left|\psi_i\right\rangle$. I'll illustrate this with an example.

EXAMPLE 5.3

Is $\hat{\wp}_{HV}\left|R\right\rangle$ the output state that would be obtained from a measurement of $\wp_{HV}$, for a beam of photons prepared in state $\left|R\right\rangle$?

We can answer this using the matrix representation of $\hat{\wp}_{HV}$. In general, a Hermitian operator expressed in the basis of its eigenstates is diagonal, with its eigenvalues making up the diagonal elements, and in the problems you will verify that

$$\hat{\wp}_{HV} \doteq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{HV}. \tag{5.10}$$

Using this,

$$\begin{aligned}\hat{\wp}_{HV}\left|R\right\rangle &\doteq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{HV} \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} \\ &\doteq \left|L\right\rangle .\end{aligned} \tag{5.11}$$

Clearly this cannot be the output state of a measurement of $\wp_{HV}$ because $\left|L\right\rangle$ is not an eigenstate of $\hat{\wp}_{HV}$, and we know that the output state must be an eigenstate. The possible outputs for this measurement were discussed in example 5.1. The answer to our question is "no."

In hindsight, it should be obvious that simply applying an operator to the input state cannot always yield the output state. This is because applying an operator to a state is a deterministic process; you always obtain the same result. Measurements are probabilistic; in general there are many possible output states for a given measurement, and each is obtained with a different probability. There is no way to predict exactly what the state of a system will be after a measurement. The take-home message is this: Applying an operator and performing a measurement are *not* the same thing.

5.5 COMMUTATION AND INDETERMINACY RELATIONS

In sec. 4.1 we said that operators do not in general commute, $\hat{A}\hat{B} \neq \hat{B}\hat{A}$. The commutator of $\hat{A}$ and $\hat{B}$ is defined as

$$\left[\hat{A},\hat{B}\right] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}, \tag{5.12}$$

and is itself an operator. If $\left[\hat{A},\hat{B}\right] = 0$ the operators commute, and the order of the operations doesn't matter.

Now consider observables A and B, with corresponding Hermitian operators $\hat{A}$ and $\hat{B}$. We make a series of measurements of A, on an ensemble of many copies of a system, all prepared in the state $\left|\psi\right\rangle$. The uncertainty in such a measurement is expressed as the standard deviation ΔA. Thus, the uncertainty is the square root of the variance

$$\begin{aligned}\Delta A^2 &= \left\langle \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)^2 \right\rangle \\ &= \left\langle \psi \right| \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)^\dagger \left(\hat{A} - \left\langle \hat{A} \right\rangle\right) \left|\psi\right\rangle,\end{aligned} \tag{5.13}$$

where the second line follows because $\hat{A}$ is Hermitian. If we define the new state

$$\left|a\right\rangle \equiv \left(\hat{A} - \left\langle \hat{A} \right\rangle\right)\left|\psi\right\rangle, \tag{5.14}$$

then $\Delta A^2 = \left\langle a \middle| a \right\rangle$. After making these measurements of A, we change our apparatus to perform a series of measurements of B, with a new ensemble of many copies of the system, all prepared in the same state $\left|\psi\right\rangle$. Defining

$$\left|b\right\rangle \equiv \left(\hat{B} - \left\langle \hat{B} \right\rangle\right)\left|\psi\right\rangle, \tag{5.15}$$

the variance of these measurements is $\Delta B^2 = \left\langle b \middle| b \right\rangle$.

The Schwartz inequality gives a relationship between the inner products,

$$\left\langle a \middle| a \right\rangle \left\langle b \middle| b \right\rangle \geq \left|\left\langle a \middle| b \right\rangle\right|^2, \tag{5.16}$$

which you'll prove in the problems. This means that

$$\begin{aligned} \Delta A^2 \Delta B^2 &\geq \left| \langle a | b \rangle \right|^2, \\ \Delta A \Delta B &\geq \left| \langle a | b \rangle \right|. \end{aligned} \tag{5.17}$$

Since $\langle a | b \rangle$ is in general complex,

$$\begin{aligned} \left| \langle a | b \rangle \right|^2 &= \left[\mathrm{Re}\left(\langle a | b \rangle \right) \right]^2 + \left[\mathrm{Im}\left(\langle a | b \rangle \right) \right]^2 \\ &\geq \left[\mathrm{Im}\left(\langle a | b \rangle \right) \right]^2 \\ &= \left[\frac{1}{2i} \left(\langle a | b \rangle - \langle b | a \rangle \right) \right]^2 . \end{aligned} \tag{5.18}$$

Furthermore,

$$\begin{aligned} \langle a | b \rangle &= \langle \psi | \left(\hat{A} - \langle \hat{A} \rangle \right)^{\dagger} \left(\hat{B} - \langle \hat{B} \rangle \right) | \psi \rangle \\ &= \left\langle \left(\hat{A} - \langle \hat{A} \rangle \right) \left(\hat{B} - \langle \hat{B} \rangle \right) \right\rangle \\ &= \left\langle \hat{A}\hat{B} - \hat{A} \langle \hat{B} \rangle - \hat{B} \langle \hat{A} \rangle + \langle \hat{A} \rangle \langle \hat{B} \rangle \right\rangle \\ &= \langle \hat{A}\hat{B} \rangle - \langle \hat{A} \rangle \langle \hat{B} \rangle . \end{aligned} \tag{5.19}$$

The equation for $\langle b | a \rangle$ looks the same, with the $\hat{A}$'s and $\hat{B}$'s exchanged. Combining eqs. (5.17), (5.18), and (5.19) yields

$$\begin{aligned} \Delta A \Delta B &\geq \left| \frac{1}{2i} \left(\langle a | b \rangle - \langle b | a \rangle \right) \right| \\ &= \frac{1}{2} \left| \left(\langle \hat{A}\hat{B} \rangle - \langle \hat{A} \rangle \langle \hat{B} \rangle - \langle \hat{B}\hat{A} \rangle + \langle \hat{B} \rangle \langle \hat{A} \rangle \right) \right| \\ &= \frac{1}{2} \left| \langle \hat{A}\hat{B} - \hat{B}\hat{A} \rangle \right| . \end{aligned} \tag{5.20}$$

Finally, we recognize the commutator in the last line, so

$$\Delta A \Delta B \geq \frac{1}{2} \left| \left\langle \left[\hat{A}, \hat{B} \right] \right\rangle \right| . \tag{5.21}$$

This is an extremely important property of quantum measurements. It says that there is a lower bound on the product of the uncertainties of the measurements of two observables, which is set by the expectation value of the commutator. If this expectation value is nonzero, the uncertainty in one of the measurements must increase if the uncertainty in the other decreases.

Equation (5.21) represents the indeterminacy relation between $\hat{A}$ and $\hat{B}$. Some people refer to this as an uncertainty relation, or an uncertainty principle, but I prefer the term indeterminacy relation (my reasoning for this is given below).[4]

EXAMPLE 5.4

A beam of photons is prepared in the state $|R\rangle$. Measurements of $\wp_{HV}$ and $\wp_{45}$ [$\hat{\wp}_{45} = (+1)|+45\rangle\langle+45| + (-1)|-45\rangle\langle-45|$] are performed separately on this beam. Determine the uncertainties in these measurements, and show that they are consistent with the corresponding indeterminacy relation.

The indeterminacy relation we need to verify is

$$\Delta\wp_{HV}\Delta\wp_{45} \geq \frac{1}{2}\left|\left\langle\left[\hat{\wp}_{HV}, \hat{\wp}_{45}\right]\right\rangle\right|. \tag{5.22}$$

In example 5.2 we showed that $\Delta\wp_{HV} = 1$. To solve the rest of the problem, start with the matrix representation of $\hat{\wp}_{45}$. In the problems you will verify that this is given by

$$\hat{\wp}_{45} \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{HV}. \tag{5.23}$$

We find $\Delta\wp_{45}$ by finding $\langle\hat{\wp}_{45}\rangle$ and $\langle\hat{\wp}_{45}{}^2\rangle$:

$$\begin{aligned} \langle\hat{\wp}_{45}\rangle &= \langle R|\hat{\wp}_{45}|R\rangle \\ &\doteq \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \end{pmatrix}_{HV}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{HV}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}_{HV} \\ &= \frac{1}{2}\begin{pmatrix} 1 & i \end{pmatrix}_{HV}\begin{pmatrix} -i \\ 1 \end{pmatrix}_{HV} \\ &= \frac{1}{2}(-i+i) \\ &= 0. \end{aligned} \tag{5.24}$$

$$\begin{aligned} \hat{\wp}_{45}{}^2 &\doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{HV}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{HV} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}_{HV} \\ &\doteq \hat{1}. \end{aligned} \tag{5.25}$$

---

[4] The original Heisenberg indeterminacy relation (which we'll discuss later) applies to two specific observables, position and momentum.

Therefore,

$$\begin{aligned}\Delta \wp_{45}{}^{2} &= \left\langle \hat{\wp}_{45}{}^{2}\right\rangle - \left\langle \hat{\wp}_{45}\right\rangle^{2} \\ &= \left\langle R\middle|\hat{1}\middle|R\right\rangle - \left\langle \hat{\wp}_{45}\right\rangle^{2} \\ &= 1-0 \\ &= 1,\end{aligned} \tag{5.26}$$

and the uncertainty is $\Delta \wp_{45} = \sqrt{\Delta \wp_{45}{}^{2}} = 1$.

The matrix representation of $\hat{\wp}_{HV}$ is given in eq. (5.10). The matrix representation of the commutator is then

$$\begin{aligned}\left[\hat{\wp}_{HV}, \hat{\wp}_{45}\right] &= \hat{\wp}_{HV}\hat{\wp}_{45} - \hat{\wp}_{45}\hat{\wp}_{HV} \\ &\doteq \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}_{HV}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}_{HV} - \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}_{HV}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}_{HV} \\ &= \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}_{HV} - \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}_{HV} \\ &= \begin{pmatrix}0 & 2\\ -2 & 0\end{pmatrix}_{HV}.\end{aligned} \tag{5.27}$$

The expectation value of the commutator is

$$\begin{aligned}\left\langle\left[\hat{\wp}_{HV}, \hat{\wp}_{45}\right]\right\rangle &\doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1 & i\end{pmatrix}_{HV}\begin{pmatrix}0 & 2\\ -2 & 0\end{pmatrix}_{HV}\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -i\end{pmatrix} \\ &= \begin{pmatrix}1 & i\end{pmatrix}_{HV}\begin{pmatrix}-i\\ -1\end{pmatrix} \\ &= (-i-i) \\ &= -2i.\end{aligned} \tag{5.28}$$

Substituting our values into eq. (5.22) yields

$$\begin{aligned}(1)(1) &\geq \frac{1}{2}\left|-2i\right|, \\ 1 &\geq 1,\end{aligned} \tag{5.29}$$

which shows that the indeterminacy relation is satisfied.

The state $\left|R\right\rangle$ is said to be a minimum uncertainty state for $\wp_{HV}$ and $\wp_{45}$ measurements, because the uncertainty product is equal to the minimum value allowed by the indeterminacy relation.

The indeterminacy relation is often presented as an explanation of why a particular measurement apparatus cannot measure two noncommuting observables with arbitrary precision. These explanations are usually unsatisfactory, because they always leave you wondering whether or not you could increase the precision with a different apparatus. Note, however, that eq. (5.21) does not refer to any particular apparatus. The state of the system $\left|\psi\right\rangle$ determines the minimum value of the uncertainty product, not a particular measurement scheme.

This is one reason why I prefer the term indeterminacy relation over the term uncertainty relation. Uncertainty relation might imply that the lower bound is due to uncertainties in the measurements. Indeterminacy relation makes it more clear that properties corresponding to noncommuting observables are inherently not well defined in quantum systems prior to a measurement.

We need to be clear that the indeterminacy relation of eq. (5.21) applies to measurements of A and B performed on *separate* sub-ensembles. We prepare a system many times in state $\left|\psi\right\rangle$ and make measurements of A using one set of particles. We then change our apparatus to make measurements of B, and then make these measurements on the same system, prepared in the same state, but with a different set of particles. We never try to make measurements of both A and B on the *same* particle.

Suppose we try to make both $\wp_{HV}$ and $\wp_{45}$ measurements on the same photons for a system prepared in state $\left|\psi\right\rangle$. If we make the $\wp_{HV}$ measurement first, then by postulate III(b) the state of the system after the measurement will be $\left|H\right\rangle$ or $\left|V\right\rangle$, not $\left|\psi\right\rangle$. The state of the system has changed before the $\wp_{45}$ measurement, so it no longer makes sense to use the state $\left|\psi\right\rangle$ when we are computing quantities such as $\Delta\wp_{45}$. Clearly eq. (5.21) cannot correspond to measurements performed in this way.

In order to make meaningful measurements of two different, noncommuting observables on the same ensemble, you must sacrifice some measurement precision. For this reason, it can be shown that the uncertainty product for such measurements is always *larger* than that for separate measurements [5.3].

5.6 COMPLEMENTARITY

Look back at experiment 6 in sec. 3.7, which describes single-photon interference. This is an experiment that you can perform, as described in lab 3. Here we want to explain the presence or absence of a single-photon interference pattern in terms of information, rather than in terms of wave interference.

The first PA_{HV} in combination with a beam block [figs. 3.7(a) and (b)] effectively performs a measurement of polarization, changing the state of the system. This measurement yields information—we know which path the photon took through the interferometer. Since we know which path it took, it clearly took only one path, so the relative phase between the two arms makes no difference, and we see no interference. If the beam block is removed [as in fig. 3.8] no measurement is performed; we can obtain no information about the path of the photon, so it takes both paths and interferes with itself.

Lab 3 describes a variation of this experiment, which involves examining the behavior of the interferometer with different wave-plate settings, and is often referred to as a

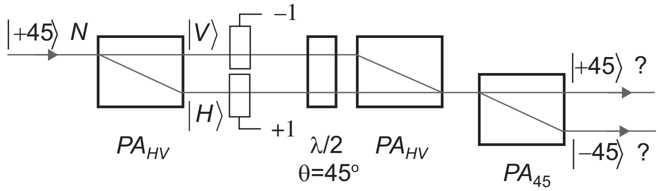


Fig 5.3 A single photon interference experiment with a nondestructive polarization measurement performed inside the interferometer.

quantum eraser. The idea in this variation is that for certain wave-plate settings, we know which path the photon takes through the interferometer, so no interference is observed. For other wave-plate settings the which-path information is not available (it's "erased"), allowing the interference pattern to be recovered.

Now let's examine a different experiment that has been proposed, but to my knowledge has not yet been performed. A diagram of this experiment is shown in fig. 5.3. The experimental arrangement is the same as that in experiment 6 (fig. 3.8), except that a nondestructive measurement takes place inside the interferometer. In experiment 6 we said that all N photons that entered the interferometer leave through the $|+45\rangle$ output port. Does the internal polarization measurement have any effect on this result?

It does. With the internal polarization measurement apparatus in place, half of the photons will leave from the $|+45\rangle$ port, and the other half will leave from the $|-45\rangle$ port. No interference will be observed if the phase of the interferometer is varied. This is because even though neither arm is blocked, the internal apparatus performs a measurement of the polarization. For example, if the nondestructive measurement says that the photon is horizontally polarized, after the measurement the state of the photon is $|H\rangle$, and we know which arm of the interferometer the photon travels through. Since the photon takes only one path, there is no interference.[5]

If this result unsettles you, you are not alone. You probably want to believe that whether or not we measure the polarization of the photon inside the interferometer, the experimental results should not be affected. However, that's not what quantum mechanics predicts. Someday, someone will do the experiment outlined in fig. 5.3, to test whether or not the quantum prediction holds. I'm willing to bet that the quantum prediction will be verified.[6]

The experiment in fig. 5.3 would be a very stringent test of the principle of complementarity, a concept first put forward by Niels Bohr. The idea is that quantum objects have mutually exclusive, complementary properties. For example, we can conceive of light being made of waves, or being made of particles, but not both. Yet, photons seem to exhibit both wave-like and particle-like behaviors.

---

[5] It is in principle possible to retrieve the interference pattern, even with the internal measurement apparatus in place, by erasing the which-path information obtained from the measurement. See ref. [5.4] for an example of how to do this.

[6] Another explanation for the lack of an interference pattern with the internal measurement apparatus in place is that the nondestructive measurement alters the phase of photons passing through it, even though it does not affect their polarization. This measurement induced phase change is enough to destroy the interference.

A key component of complementarity is that we can either observe the wave-like (unlocalized), or the particle-like (localized), nature of light, but not both simultaneously. If we remove the internal measurement apparatus from fig. 5.3, we have no knowledge about which path the photon took, and we see wave-like interference. With the apparatus in place, a measurement localizes the photon to a particular arm, and we observe particle-like behavior (no interference).

This need not be an either-or proposition. If we obtain partial information about the path of the photon, the interference is partially destroyed. For example, if 10% of the time we know the photon's path (maybe the nondestructive measurement succeeds only 10% of the time), then the visibility of the interference pattern, averaged over a large ensemble, might be 90%, whereas if we know the path 50% of the time, the visibility would be reduced to 50%.[7]

One last comment. In lab 3 you are able to see an interference pattern while simultaneously verifying that only a single photon at a time is inside the interferometer. The interference pattern illustrates the wave-like nature of light, while the fact that the photon leaves the interferometer from either one port or the other, but not both, illustrates the particle-like nature of light. Does this violate complementarity? No. The measurement that displays the particle-like behavior occurs *outside* the interferometer. *Inside* the interferometer we have no way of knowing which way the photon went, so it takes both paths and interferes with itself. The wave-like nature of the photon manifests itself inside the interferometer, and the particle-like nature outside. This is consistent with Bohr's principle of complementarity.

5.7 References

[5.1] O.R. Frisch, "Take a photon . . .," Contemp. Phys. **7**, 45 (1965).

[5.2] C. Cohen-Tannoudji, B. Diu, and Franck Laloë, *Quantum Mechanics* (Wiley, New York, 1977), Sec. II.D.2.

[5.3] M. G. Raymer, "Uncertainty principle for joint measurement of noncommuting variables," Am. J. Phys. **62**, 986 (2004).

[5.4] C. Ferrari and B. Braunecker, "Entanglement, which-way measurements, and quantum erasure," Am. J. Phys. **78**, 792 (2010).

5.8 PROBLEMS

5.1* Prove that $\hat{\wp}_{HV}$ is Hermitian.

5.2* Determine the matrix representation of $\hat{\wp}_{HV}$ in the *HV*-basis.

5.3 What is the matrix representation of the polarization operator $\hat{\wp}_{HV}$, using the states $|L\rangle$ and $|R\rangle$ as a basis?

5.4 Prove that eq. (5.3) and eq. (5.4) are equivalent. In other words, prove that for a general state $|\psi\rangle = c_H|H\rangle + c_V|V\rangle$, both of these equations yield the same value for $\hat{\wp}_{HV}|\psi\rangle$.

---

[7] See complement 2.A for a discussion of visibility. Lower visibility means, in some sense, less interference.

5.5 A measurement of $\wp_{HV}$ is performed for a beam of photons prepared in the state $\left|\psi\right\rangle = \left(1/\sqrt{3}\right)\left|H\right\rangle + \sqrt{2/3}\,e^{i\pi/3}\left|V\right\rangle$. What are the possible outcomes of this measurement? What are their probabilities? For each outcome, what state is the system left in?

5.6* Does the projection operator $\hat{P}_{+45} = \left|+45\right\rangle\left\langle+45\right|$ correspond to an observable? If not, why not? If it does: How might you implement this measurement? What are the possible outcomes of measurements of P_{+45}? For each outcome, what state is the system left in?

5.7* Does the rotation operator $\hat{R}_p\left(\theta\right)$ correspond to an observable? If not, why not? If it does: How might you implement this measurement? What are the possible outcomes of measurements of $R_p\left(\theta\right)$? For each outcome, what state is the system left in?

5.8 Calculate the expectation value and standard deviation of $\wp_{HV}$ for a beam of photons prepared in the state $\left|\psi\right\rangle = \left(1/\sqrt{3}\right)\left|H\right\rangle + \sqrt{2/3}\,e^{i\pi/3}\left|V\right\rangle$.

5.9 Calculate the expectation value and standard deviation of $\wp_{HV}$ for a beam of photons prepared in an elliptical polarization state $\left|e\right\rangle = \cos\theta\left|H\right\rangle + \sin\theta e^{i\phi}\left|V\right\rangle$.

5.10 Calculate the expectation value and standard deviation of $\hat{P}_{+45} = \left|+45\right\rangle\left\langle+45\right|$ for a beam of photons prepared in a linear polarization state $\left|\theta\right\rangle = \cos\theta\left|H\right\rangle + \sin\theta\left|V\right\rangle$.

5.11 Determine the matrix representation of $\hat{\wp}_{45}$ in the *HV*-basis.

5.12* In analogy with $\hat{\wp}_{HV}$ and $\hat{\wp}_{45}$, define $\hat{\wp}_C$, which is the operator corresponding to measurement of circular polarization. Define this operator such that measurements of left-circular polarization yield positive values. Find the matrix representation of $\hat{\wp}_C$ in the *HV*-basis.

5.13 Show that $\left[\hat{\wp}_{HV}, \hat{\wp}_{45}\right] = 2i\hat{\wp}_C$, where $\hat{\wp}_C$ is defined in problem 5.12.

5.14 Show that $\left[\hat{\wp}_{HV}, \hat{\wp}_C\right] = -2i\hat{\wp}_{45}$, where $\hat{\wp}_C$ is defined in problem 5.12.

5.15 Calculate $\left[\hat{P}_{+45}, \hat{P}_V\right]$. Express your answer in the *HV*-basis.

5.16 A measurement of $\wp_{45}$ is performed for a beam of photons prepared in the state $\left|\psi\right\rangle = \left(1/\sqrt{3}\right)\left|H\right\rangle + \sqrt{2/3}\,e^{i\pi/3}\left|V\right\rangle$. What are the possible outcomes of this measurement? What are their probabilities? For each outcome, what state is the system left in?

5.17* A measurement of $\wp_{HV}$ is performed on a photon prepared in state $\left|L\right\rangle$. A measurement of $\wp_C$ (problem 5.12) is then performed on a second photon prepared in the same state. What is the probability that the first measurement returns +1, and the second returns −1?

5.18* A measurement of $\wp_{HV}$ is performed on a photon prepared in state $\left|L\right\rangle$. A measurement of $\wp_C$ (problem 5.12) is then performed on the same photon. What is the probability that the first measurement returns +1, and the second returns −1?

5.19 Verify that measurements of $\wp_{HV}$ and $\wp_{45}$ satisfy the appropriate indeterminacy relation for a beam of photons prepared in the state $\left|\psi\right\rangle = \left(1/\sqrt{3}\right)\left|H\right\rangle + \sqrt{2/3}\, e^{i\pi/3}\left|V\right\rangle$.

5.20 Verify that measurements of $\wp_{HV}$ and $\wp_C$ (problem 5.12) satisfy the appropriate indeterminacy relation for a beam of photons prepared in the elliptical polarization state $\left|e\right\rangle = \cos\theta\left|H\right\rangle + \sin\theta e^{i\phi}\left|V\right\rangle$.

5.21 Verify that measurements of $\wp_{45}$ and $\wp_C$ (problem 5.12) satisfy the appropriate indeterminacy relation for a beam of photons prepared in the elliptical polarization state $\left|e\right\rangle = \cos\theta\left|H\right\rangle + \sin\theta e^{i\phi}\left|V\right\rangle$.

5.22 Verify that measurements of $\wp_{45}$ and $\wp_C$ (problem 5.12) satisfy the appropriate indeterminacy relation for a beam of photons prepared in an eigenstate of $\hat{\wp}_{45}$.

5.23* Prove the Schwartz inequality: $\left\langle a|a\right\rangle\left\langle b|b\right\rangle \geq \left|\left\langle a|b\right\rangle\right|^2$. Hint: Let

$$\left|\psi\right\rangle = \left|a\right\rangle - \left(\frac{\left\langle b|a\right\rangle}{\left\langle b|b\right\rangle}\right)\left|b\right\rangle, \tag{5.30}$$

and use $\left\langle\psi|\psi\right\rangle \geq 0$.

5.24 An operator is said to be anti-Hermitian if $\hat{O}^\dagger = -\hat{O}$. If $\hat{A}$ and $\hat{B}$ are Hermitian, prove that $\left[\hat{A}, \hat{B}\right]$ is anti-Hermitian.

5.25 If $\hat{A}$ and $\hat{B}$ are Hermitian, and $\left[\hat{A}, \hat{B}\right] = i\hat{C}$, prove that $\hat{C}$ is Hermitian. Given this definition of $\hat{C}$, you will often see the indeterminacy principle written as $\Delta A \Delta B \geq \frac{1}{2}\left|\left\langle\hat{C}\right\rangle\right|$.

COMPLEMENT 5.A

"Measuring" a Quantum State

The state $|\psi\rangle$ of a quantum mechanical system determines the probabilities of the outcomes of measurements that can be performed on that system. In this sense it contains *all* of the knowable information about the system; if you know the state, you know the probability of *any* measurement you could perform on that system. It is thus natural to ask: Is it possible to measure (or otherwise determine) the state of a system?

The answer is yes, and here we'll describe a process for reconstructing the state of a very simple system. You'll be able to perform this state measurement in lab 4. We'll assume that the state is pure, so that it can be represented by a ket vector $|\psi\rangle$.[8]

5.A.1 Reconstructing a Polarization State

Imagine that we have a beam of photons prepared in a particular polarization state $|\psi\rangle$. To measure the state we must determine the probability amplitudes of the representation of the state in some basis. If we represent the state as $|\psi\rangle = c_H |H\rangle + c_V |V\rangle$, we need to determine c_H and c_V to determine the state.

Start by passing the beam through a PA_{HV} and splitting it into its horizontal and vertical components, as shown in fig. 5.A.1. As described in eq. (5.1), the probabilities of horizontal and vertical measurements are

$$
\begin{aligned}
P(H \| \psi\rangle) &= |\langle H|\psi\rangle|^2 = |c_H|^2, \\
P(V \| \psi\rangle) &= |\langle V|\psi\rangle|^2 = |c_V|^2.
\end{aligned}
\tag{5.A.1}
$$

Many measurements, with the system prepared in the same state, are needed to get an accurate determination of these probabilities.

---

[8] The general process of determining mixed states of more complicated systems is often referred to as quantum state tomography. The word tomography is used because the original state measurements used a mathematical reconstruction algorithm similar to that used in tomographic imaging. See also refs. [5.A.1] and [5.A.2].

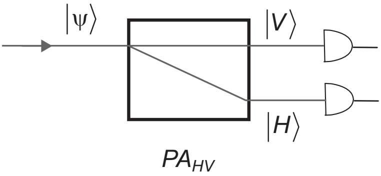


Fig 5.A.1 Performing a (destructive) measurement of polarization.

The measurements yield probabilities, but what we really need are probability *amplitudes*. Probability amplitudes are complex numbers; the measurements determine their magnitudes, but we still need to determine their phases. With this in mind, let's rewrite the state in a manner that explicitly indicates the magnitudes and phases of the probability amplitudes:

$$\begin{aligned} |\psi\rangle &= c_H |H\rangle + c_V |V\rangle \\ &= |c_H| e^{i\phi_H} |H\rangle + |c_V| e^{i\phi_V} |V\rangle \\ &= e^{i\phi_H} \left(a|H\rangle + b e^{i(\phi_V - \phi_H)} |V\rangle \right). \end{aligned} \tag{5.A.2}$$

Here we've defined $a \equiv |c_H|$ and

$$b \equiv |c_V| = \sqrt{1-a^2}, \tag{5.A.3}$$

which holds because the probabilities must be normalized. The overall phase factor in eq. (5.A.2) does not change the state, so we can safely set $\phi_H = 0$. Define $\phi \equiv \phi_V$, and we can then rewrite eq. (5.A.2) as

$$|\psi\rangle = a|H\rangle + b e^{i\phi} |V\rangle. \tag{5.A.4}$$

Equation (5.A.4) indicates that we must determine three real-valued quantities, a, b and ϕ, to fully determine the state. The measurements shown in fig. 5.A.1 determine a and b, because a is given by

$$a = \sqrt{P\left(H \| \psi\rangle\right)}, \tag{5.A.5}$$

and b is easily obtained from a using eq. (5.A.3). There is no way to determine ϕ from these measurements, however, so we need to perform other measurements as well.

Next we perform a similar series of measurements, but replace the $\mathrm{PA}_{\mathrm{HV}}$ in fig. 5.A.1 with a PA_{45}. The state preparation procedure must remain the same; this ensures that the state $|\psi\rangle$ is the same for all measurements. The probability of obtaining a +45° polarized photon is given by

$$
\begin{aligned}
P\left(+45\middle\|\psi\right\rangle) &= \left|\langle +45|\psi\rangle\right|^2 \\
&= \left|\langle +45|\left(a|H\rangle + b\,e^{i\phi}|V\rangle\right)\right|^2 \\
&= \left|\frac{a}{\sqrt{2}} + \frac{b}{\sqrt{2}}\,e^{i\phi}\right|^2 \\
&= \frac{1}{2}\left[a^2 + 2ab\cos\phi + b^2\right] \\
&= \frac{1}{2}\left[1 + 2ab\cos\phi\right],
\end{aligned}
\tag{5.A.6}
$$

where we've used eq. (5.A.3). Since a and b are known from the first set of measurements, this equation can be inverted to obtain $\cos\phi$ from $P\left(+45\|\psi\rangle\right)$, which is determined by the second set of measurements.

However, the inverse cosine is not unique. Suppose, for example, that we determine $\cos\phi = 0.5$, then $\phi = \pi/3$ or $\phi = -\pi/3$. With the information that we have from our two sets of measurements, we cannot in general choose between these possibilities. We need to perform a third set of measurements, on yet more members of the ensemble.

Now we modify our apparatus to perform measurements of circular polarization, using a PA_C in place of the PA_HV in fig. 5.A.1. Once again, it is necessary to use the same state preparation procedure, in order to produce the same state. The probability of measuring a left-circularly polarized photon is

$$
\begin{aligned}
P\left(L\|\psi\rangle\right) &= \left|\langle L|\psi\rangle\right|^2 \\
&= \left|\langle L|\left(a|H\rangle + b\,e^{i\phi}|V\rangle\right)\right|^2 \\
&= \left|\frac{a}{\sqrt{2}} - i\frac{b}{\sqrt{2}}\,e^{i\phi}\right|^2 \\
&= \frac{1}{2}\left[a^2 - ab\left(ie^{i\phi} - ie^{-i\phi}\right) + b^2\right] \\
&= \frac{1}{2}\left[1 + 2ab\sin\phi\right].
\end{aligned}
\tag{5.A.7}
$$

This can be inverted to obtain $\sin\phi$. Knowing both $\sin\phi$ and $\cos\phi$ uniquely determines ϕ. Above we were trying to choose between $\phi = \pi/3$ and $\phi = -\pi/3$; simply knowing whether $\sin\phi$ is positive or negative is enough to make this determination.

EXAMPLE 5.A.1

Polarization measurements are performed for three different device settings, as described above, for many photons all prepared in the same state. If the following probabilities are obtained: $P\left(H\|\psi\rangle\right) = 0.50$, $P\left(+45\|\psi\rangle\right) = 0.93$, and $P\left(L\|\psi\rangle\right) = 0.25$, what is the polarization state of the system?

Express the state using eq. (5.A.4). First, we determine *a* and *b*:

$$a=\sqrt{P\left(H\middle\|\psi\rangle\right)}=\sqrt{0.50}=\frac{1}{\sqrt{2}}, \tag{5.A.8}$$

$$b=\sqrt{1-a^2}=\frac{1}{\sqrt{2}}. \tag{5.A.9}$$

Substituting these into eq. (5.A.6), we find

$$P\left(+45\middle\|\psi\rangle\right)=\frac{1}{2}\left[1+\cos\phi\right]. \tag{5.A.10}$$

Solving for $\cos\phi$ yields

$$\begin{aligned}\cos\phi &= 2P\left(+45\middle\|\psi\rangle\right)-1 \\ &= 2\left(0.93\right)-1 \\ &= 0.86\,.\end{aligned} \tag{5.A.11}$$

The two solutions for ϕ are then

$$\begin{aligned}\phi &= \cos^{-1}\left(0.86\right) \\ &= \pm 0.54 \text{ rad} \\ &= \pm 0.17\,\pi \\ &\cong \pm\pi/6\,.\end{aligned} \tag{5.A.12}$$

Examining eq. (5.A.7) shows that if $\phi>0, P\left(L\middle\|\psi\rangle\right)>0.5$, and if $\phi<0, P\left(L\middle\|\psi\rangle\right)<0.5$ Our data then indicate that we should choose the negative solution, which is $\phi\cong-\pi/6$. The final solution is thus

$$|\psi\rangle\cong\frac{1}{\sqrt{2}}\left(|H\rangle+e^{-i\pi/6}|V\rangle\right). \tag{5.A.13}$$

5.A.2 Discussion

Note that it was necessary to perform more than one set of measurements, corresponding to more than one measurement apparatus (observable), in order to determine the state. This is a general property of all state measurement procedures. You can view this as a manifestation of complementarity. Quantum systems have complementary aspects, so it is necessary to perform measurements of multiple observables in order to illuminate the properties of a state which reflects these different aspects.

The state that we determine describes the state of all of the copies of the system, which must all be identically prepared. It is impossible to measure the state of a single member of the ensemble. You cannot, even in principle, determine the polarization state of a single photon. In order to determine a state you must measure probabilities, which requires multiple trials. Once the first measurement is performed, the state of the system is changed (remember postulate III(b)?). You can't use this photon for the next trial, because it is no longer in the same state. It is necessary to repeat the state preparation procedure for the next measurement.

In this light, we can view the state as being a consequence of the procedure used to prepare the system. Any individual member of the system that has undergone the proper preparation will be in that state.

5.A.3 References

[5.A.1] M.G. Raymer, "Measuring the quantum mechanical wave function," Contemp. Phys. **38**, 343 (1997).

[5.A.2] U. Leonhard, *Measuring the Quantum State of Light* (Cambridge University Press, Cambridge, 1997).

5.A.4 PROBLEMS

5.A.1 Polarization measurements are performed for three different device settings, as described above, for photons all prepared in the same state. The following probabilities are obtained: $P\left(H\|\psi\rangle\right)=0.50$, $P\left(+45\|\psi\rangle\right)=0.75$, and $P\left(L\|\psi\rangle\right)=0.93$. What is the polarization state of the system?

5.A.2 Polarization measurements are performed for three different device settings, as described above, for photons all prepared in the same state. The following probabilities are obtained: $P\left(H\|\psi\rangle\right)=0.67$, $P\left(+45\|\psi\rangle\right)=0.50$, and $P\left(L\|\psi\rangle\right)=0.97$. What is the polarization state of the system?

5.A.3 You perform polarization measurements using a PA_{HV} on 1000 photons prepared in an unknown state. 997 of the photons are measured to have horizontal polarization. What can you say about the polarization state?

5.A.4 Suppose you are performing a polarization state measurement, and you know that the state is linearly polarized. How would this change the measurement procedure described above? How many different device settings would you need to uniquely determine the state?

CHAPTER 6

Spin-1/2

So far we've been studying polarization, and using photon polarization as an example of a two-dimensional (2-D) quantum system. Polarization was a nice place to start, because there are strong analogies between quantum and classical polarization. Now we'll study a new system that has no classical counterpart—spin-1/2 particles. This system is also two-dimensional, and it has many analogs in polarization. However, you'll see that many of the classical correspondences we were able to make before no longer apply.

6.1 THE STERN-GERLACH EXPERIMENT

6.1.1 Force on a Magnetic Dipole

A magnetic dipole with dipole moment $\boldsymbol{\mu}$, in a magnetic field $\boldsymbol{\mathcal{B}}$, has potential energy

$$V = -\boldsymbol{\mu}\bullet\boldsymbol{\mathcal{B}}, \tag{6.1}$$

and the force on the dipole is

$$\mathbf{F} = -\nabla V = \nabla\left(\boldsymbol{\mu}\bullet\boldsymbol{\mathcal{B}}\right). \tag{6.2}$$

If the dipole is small, and the magnetic field is uniform, the gradient is 0, so $\mathbf{F} = 0$. If, however, the field is nonuniform, there will be a net force. Assuming the gradient points in the z-direction, and the dipole is small compared to the length scale over which the field changes, the force is

$$\mathbf{F} = \mu_z \frac{\partial}{\partial z}\mathcal{B}_z. \tag{6.3}$$

One way to create a nonuniform magnetic field would be to arrange some permanent magnets into a geometry like that shown in fig. 6.1. Otto Stern and Walther Gerlach used a device like this in 1922 to perform an important experiment.

Consider a beam of atoms emitted from an oven, as shown in fig. 6.2. There are slits placed so that the atoms in the beam are roughly collimated, and the beam passes

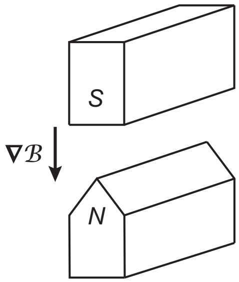


Fig 6.1 A Stern-Gerlach magnet. The magnetic field $\mathcal{B}$ between the north and south poles points upward, but the gradient of the field $\nabla\mathcal{B}$ points downward.

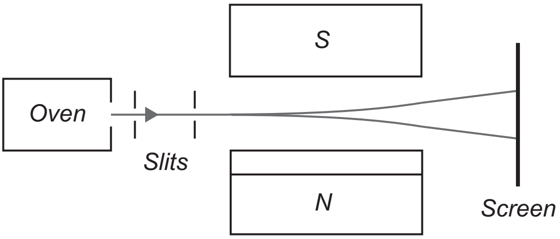


Fig 6.2 The Stern-Gerlach experiment.

through the Stern-Gerlach magnet shown in fig. 6.1. A screen is placed beyond the magnet, and the distribution of atoms striking the screen is viewed to determine how the magnetic field deflects them. What would you predict will happen?.

In fig. 6.2 the gradient of $\mathcal{B}$ is pointing down ($\partial\mathcal{B}_z / \partial z$ is negative). If the z-component of the magnetic dipole moment of an atom is pointing down as well (μ_z negative), then eq. (6.3) predicts an upward force on the atom, so it will be deflected up. If μ_z is positive the atom will be deflected down. The larger the magnitude of μ_z, the larger the force, hence the larger the deflection. By knowing a few things like the mass of the atoms and the geometry, you can calibrate the device so that by measuring the amount of deflection you measure μ_z.

The oven is hot, and shielded from magnetic fields, so we'd expect that the directions of the dipoles would be random. If the dipole is straight up, there will be a maximum value for the z-component of the dipole moment $\mu_z = \mu$, producing a maximum downward deflection; if the dipole is straight down, there will be a maximum upward deflection. Most of the atoms will have deflections somewhere between these two extremes. So, we'd expect to see a distribution of atoms like that shown in fig. 6.3(a), centered on no deflection, and extending upward and downward the same amount. However, this is not what Stern and Gerlach observed in their experiment. Instead, they

Fig 6.3 Expected (a) and observed (b) distributions of atoms in the Stern-Gerlach experiment.

saw a two-peaked distribution like that shown in fig. 6.3(b), with one peak above the center and one below, and with no atoms going straight through.

The history of the Stern-Gerlach experiment is interesting, complete with wrong interpretations and experimental issues related to cigar smoke [6.1]. The present-day interpretation of the results is that μ_z takes on two distinct values, one positive and one negative, not the continuous distribution of values from $-\mu$ to μ as classical physics predicts. The z-component of the magnetic dipole moment is quantized.

6.1.2 Spin

Stern and Gerlach used silver atoms in their experiment, and we now know that the magnetic dipole moment of silver comes predominately from a single unpaired electron in its outer orbital shell (the magnetic moments of paired electrons point in opposite directions and cancel out, while the nuclear contribution to the magnetic moment is small). This magnetic moment comes from a property of particles called spin. The magnetic dipole moment of a particle is related to its spin **S** by

$$\boldsymbol{\mu} = \gamma \mathbf{S}, \tag{6.4}$$

where γ is a property of the particle called the gyromagnetic ratio.

Spin was first postulated by George Uhlenbeck and Samuel Goudsmit in 1925 in order to explain some spectroscopic observations [6.2]. The word spin comes from the fact that it is natural to associate a magnetic moment with a charged spinning particle (see problem 6.2). Uhlenbeck and Goudsmit originally believed that the particle was actually spinning, but Uhlenbeck soon realized that the magnitude of the spin of an electron could not be accounted for in this way—it would need to be spinning too fast. Now we know that spin is an intrinsic angular momentum associated with a particle (like mass or charge), and is not due to the particle actually spinning. This angular momentum leads to a dipole moment via eq. (6.4). Spin is a purely quantum mechanical property, with no classical analog. In non-relativistic quantum mechanics we need to add spin to the theory in an ad hoc manner; however, it arises naturally in Dirac's relativistic theory of the electron.

Since the magnetic dipole moment of silver atoms is due to a single electron, we could repeat the Stern-Gerlach experiment of fig. 6.2 using electrons instead of silver atoms, and get essentially the same result as shown in fig. 6.3. Because electrons have negative charge, they have a negative gyromagnetic ratio ($\gamma_e = -1.76 \times 10^{11}\ \mathrm{s}^{-1}\mathrm{T}^{-1}$). Therefore, by eq. (6.4), the magnetic dipole moment and the spin of an electron point in opposite directions. In the Stern-Gerlach experiment (fig. 6.2) with electrons, a downward deflected electron has a positive μ_z, so it follows that the z-component of spin S_z is negative for downward deflected electrons. Similarly, for upward deflected electrons S_z is positive. The two allowed values of μ_z mean that there only two allowed values of S_z for an electron, and they are

$$S_z = \pm\frac{1}{2}\hbar. \tag{6.5}$$

The constant $\hbar$ is pronounced "h-bar," and is equal to Planck's constant h, divided by 2π; $\hbar = 1.055 \times 10^{34}\ \mathrm{J \cdot s}$, and it has units of angular momentum. Electrons which have positive S_z are said to be "spin-up," while electrons with negative S_z are said to be "spin-down." As we'll see in more detail in chapter 7, the factor of 1/2 in front of the $\hbar$ in eq. (6.5) comes from the fact that electrons are spin-1/2 particles. All spin-1/2 particles have two allowed values for the z-component of spin, given by eq. (6.5).

6.2 SPIN STATES

In chapter 3 (and lab 3) we discussed a series of experiments exploring the polarization states of individual photons. We used these experiments to develop the notion of quantum states. Now we're going to talk about analogous experiments exploring spin states.[1] There are some similarities with the polarization experiments, but also some important differences. You won't actually do these experiments, but experiments like these have been performed [6.4].

We'll describe these experiments using spin analyzers, SAs, which are analogous to the polarization analyzers we've discussed previously. A spin analyzer consists of Stern-Gerlach magnets, oriented as shown in fig. 6.4, which split a beam of electrons into spin-up and spin-down components.

6.2.1 Experiment 1

Consider the experimental arrangement of fig. 6.5. A beam of electrons prepared in an unknown spin state $|\psi\rangle$ enters a spin analyzer oriented in the z-direction, SA_z. Electrons with spin-down ($S_z = -\hbar/2$) are deflected downward and blocked, while N electrons with spin-up ($S_z = \hbar/2$) are passed to a second SA_z. All N of these electrons emerge from the second SA_z with spin-up. If we were to block the spin-up electrons instead, we would find that all of the electrons would emerge from the second SA_z with spin-down.

---

[1] These experiments are similar to those described in ref. [6.3].

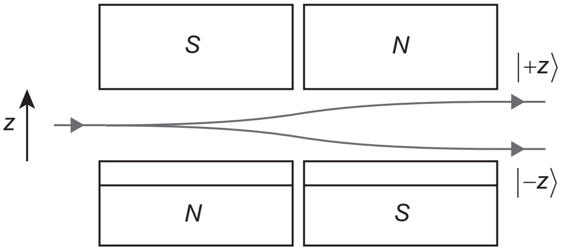


Fig 6.4 A spin analyzer oriented along the *z*-direction, SA$_z$. The first pair of magnets deflects spin-up electrons upward, and spin-down electrons downward. The second set of magnets deflects the beams so that they emerge parallel to the incoming beam.

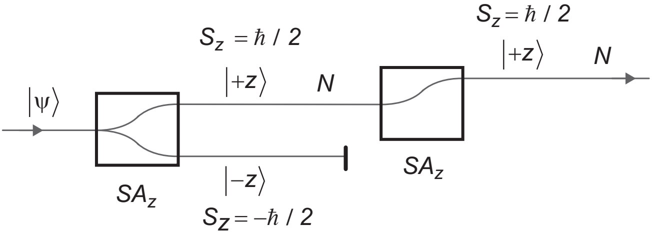


Fig 6.5 Experiment 1.

In fig. 6.5 the first SA$_z$, in combination with the beam block, serves as a state preparation device—the electrons are prepared in a state corresponding to spin-up along the *z*-direction. We will refer to this state as $|+z\rangle$; electrons with spin-down along the *z*-direction are in state $|-z\rangle$. These states are normalized so $\langle +z|+z\rangle=1$ and $\langle -z|-z\rangle=1$. Electrons which enter the second SA$_z$ in state $|+z\rangle$ emerge in the same state with 100% probability. There is no probability of emerging in state $|-z\rangle$, so $|+z\rangle$ and $|-z\rangle$ must be orthogonal:

$$\langle +z|-z\rangle=0. \tag{6.6}$$

Think carefully about what eq. (6.6) says, because it may surprise you. It says that the quantum mechanical states corresponding to spin pointing up, and spin pointing down, are orthogonal. However, the 3-D, real-space vectors corresponding to up and down are NOT orthogonal: $(+\boldsymbol{u}_z)\cdot(-\boldsymbol{u}_z)=-1\neq 0$! When we were examining polarization we were lucky: If the Hilbert-space state vectors were orthogonal, then the corresponding classical polarization vectors were orthogonal as well. Clearly this is not the case for spin-1/2 particles; our classical intuition breaks down. Remember that Hilbert-space vectors do not really "point" in a particular direction—they are abstract quantities.

States that are orthogonal in Hilbert space need not correspond to 3-D–space vectors that are orthogonal.

6.2.2 The $\hat{S}_z$ Operator

If we detect the electrons emerging from an SA$_z$, then we've performed a measurement of the z-component of spin. The z-component of spin is an observable, S_z, and it has a corresponding Hermitian operator $\hat{S}_z$. Since the result of a measurement must be an eigenvalue, Eq. (6.5) says that the eigenvalues of $\hat{S}_z$ are $\pm\hbar/2$, and experiment 1 indicates that the corresponding eigenstates are $|\pm z\rangle$:

$$\hat{S}_z|+z\rangle = \frac{\hbar}{2}|+z\rangle, \quad \hat{S}_z|-z\rangle = -\frac{\hbar}{2}|-z\rangle. \tag{6.7}$$

The eigenstates form an orthonormal basis, and we can express them as row and column vectors

$$|+z\rangle \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix}_z, \quad |-z\rangle \doteq \begin{pmatrix} 0 \\ 1 \end{pmatrix}_z. \tag{6.8}$$

In this basis the matrix representation of $\hat{S}_z$ is

$$\hat{S}_z \doteq \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_z = \frac{\hbar}{2}\bar{\boldsymbol{\sigma}}_z. \tag{6.9}$$

Here $\bar{\boldsymbol{\sigma}}_z$ is one of the Pauli matrices (or Pauli spin matrices); we'll come across the others soon.

6.3 MORE SPIN STATES

6.3.1 Experiment 2

Now consider experiment 2, shown in fig. 6.6. The SA$_z$ and the beam block prepare electrons in the state $|+z\rangle$. These electrons then pass through a spin analyzer oriented along the x-direction SA$_x$. Half of the electrons emerge in state $|+x\rangle$ with $S_x = \hbar/2$, while the other half emerge in state $|-x\rangle$ with $S_x = -\hbar/2$. The fact that there are positive

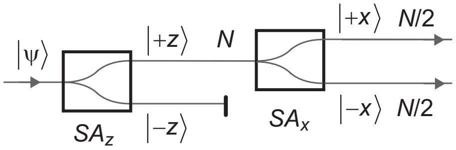


Fig 6.6 Experiment 2.

and negative spin components along the x-direction should not be surprising. There's no preferred direction in space, and since there's nothing special about the z-direction, we expect the x-direction to look the same.

Looking at fig. 6.6, what's probably running through your mind is that Newtonian physics teaches us that the motion of objects along orthogonal directions are independent of each other. Since the x- and z-directions are orthogonal, Newton would argue that the SA$_x$ splits the beam in two components because these components were present in the original state $|\psi\rangle$. The splitting of the beam on the SA$_x$ should have nothing to do with the presence of the SA$_z$. However, experiment 1 told us that there's something funny about the orthogonality of spin states. Our classical notion of orthogonality isn't always correct, so we need to check if the x- and z-directions really are independent of each other.

6.3.2 Experiment 3

We can test the independence of the x- and z-directions using experiment 3, shown in fig 6.7. If the x- and z-directions are independent, then the initially prepared $|+z\rangle$ electrons should propagate through the SA$_x$ with the z-component of their spins unaffected. None of the electrons should emerge from the final SA$_z$ in state $|-z\rangle$. However, this is not what happens! Experiment 3 indicates that the x- and z-components of spin are *not* independent of (orthogonal to) each other!

The SA$_x$ performs a measurement; after this measurement, the electrons leave the SA$_x$ in state $|+x\rangle$. The probabilities of the S_z measurements performed by the second SA$_z$, given the input state $|+x\rangle$, are

$$P\left(S_z = \hbar/2 \,\middle|\, |+x\rangle\right) = \left|\langle +z|+x\rangle\right|^2 = \frac{1}{2},$$

$$P\left(S_z = -\hbar/2 \,\middle|\, |+x\rangle\right) = \left|\langle -z|+x\rangle\right|^2 = \frac{1}{2}. \tag{6.10}$$

We can write $|+x\rangle$ as a linear combination of $|+z\rangle$ and $|-z\rangle$, and eq. (6.10) indicates that the magnitudes of the probability amplitudes are the same. Repeating experiment 3 using the $|-x\rangle$ state yields the same results. Following the same procedure described

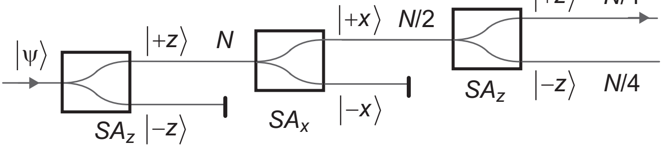


Fig 6.7 Experiment 3.

in sec. 3.4 for polarization states, we are free to make the probability amplitudes real, and find

$$\left|+x\right\rangle = \frac{1}{\sqrt{2}}\left(\left|+z\right\rangle + \left|-z\right\rangle\right), \quad \left|-x\right\rangle = \frac{1}{\sqrt{2}}\left(\left|+z\right\rangle - \left|-z\right\rangle\right). \tag{6.11}$$

This says that spin-sideways is a superposition of spin-up and spin-down, which is definitely nonclassical.

The states of eq. (6.11) are eigenstates of the operator $\hat{S}_x$, with eigenvalues $\pm\hbar/2$. We can also express $\hat{S}_x$ in the z-basis, as shown in the following example.

EXAMPLE 6.1

Find the matrix representation of $\hat{S}_x$ in the z-basis.

In the z-basis we know that

$$\hat{S}_x \doteq \begin{pmatrix} \left\langle +z\right|\hat{S}_x\left|+z\right\rangle & \left\langle +z\right|\hat{S}_x\left|-z\right\rangle \\ \left\langle -z\right|\hat{S}_x\left|+z\right\rangle & \left\langle -z\right|\hat{S}_x\left|-z\right\rangle \end{pmatrix}_z . \tag{6.12}$$

There are two ways to proceed. We know how $\hat{S}_x$ acts on the x-states (they're eigenstates), so we can rewrite the z-states in terms of $\left|+x\right\rangle$ and $\left|-x\right\rangle$ to calculate the matrix elements. You'll take this approach in the problems.

Another approach is to use eq. (4.78) to write $\hat{S}_x$ as

$$\hat{S}_x = \left(\frac{\hbar}{2}\right)\left|+x\right\rangle\left\langle +x\right| + \left(-\frac{\hbar}{2}\right)\left|-x\right\rangle\left\langle -x\right|. \tag{6.13}$$

Then

$$\begin{aligned} \hat{S}_x\left|+z\right\rangle &= \frac{\hbar}{2}\left(\left|+x\right\rangle\left\langle +x\middle|+z\right\rangle - \left|-x\right\rangle\left\langle -x\middle|+z\right\rangle\right) \\ &= \frac{\hbar}{2}\frac{1}{\sqrt{2}}\left(\left|+x\right\rangle - \left|-x\right\rangle\right) \\ &= \frac{\hbar}{2}\left|-z\right\rangle, \end{aligned} \tag{6.14}$$

where we've used eq. (6.11) to calculate the inner products. Similarly,

$$\hat{S}_x\left|-z\right\rangle = \frac{\hbar}{2}\left|+z\right\rangle. \tag{6.15}$$

Substituting these back into eq. (6.12) yields

$$\hat{S}_x \doteq \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_z = \frac{\hbar}{2}\overline{\boldsymbol{\sigma}}_x, \tag{6.16}$$

where $\overline{\boldsymbol{\sigma}}_x$ is another Pauli matrix.

6.3.3 Experiments 4 and 5

If we repeat experiment 3, replacing the SA$_x$ with an SA$_y$, we get experiment 4, which is shown in fig. 6.8. The results are essentially the same as found in experiment 3, indicating that the y- and z-components of spin are not independent, and that

$$P\left(S_z = \hbar/2 \| +y\rangle\right) = \left|\langle +z | +y\rangle\right|^2 = \frac{1}{2},$$

$$P\left(S_z = -\hbar/2 \| +y\rangle\right) = \left|\langle -z | +y\rangle\right|^2 = \frac{1}{2}. \tag{6.17}$$

The same results are obtained using the $|-y\rangle$ state.

As with the x-states, we can once again write the y-states as linear combinations of $|+z\rangle$ and $|-z\rangle$, with the magnitudes of the probability amplitudes being equal. However, just as we found for polarization states, we can no longer choose the probability amplitudes to be real, because that would make the x-states and the y-states the same. We know that these states are not the same by experiment 5, depicted in fig. 6.9 (e.g., $|+y\rangle$ splits into $|+x\rangle$ and $|-x\rangle$ components). Because we defined the x-states using real probability amplitudes, we are forced to use complex probability amplitudes to describe the y-states. In analogy with the calculation in sec. 3.5, the y-states are found to be

$$|+y\rangle = \frac{1}{\sqrt{2}}\left(|+z\rangle + i|-z\rangle\right), \quad |-y\rangle = \frac{1}{\sqrt{2}}\left(|+z\rangle - i|-z\rangle\right). \tag{6.18}$$

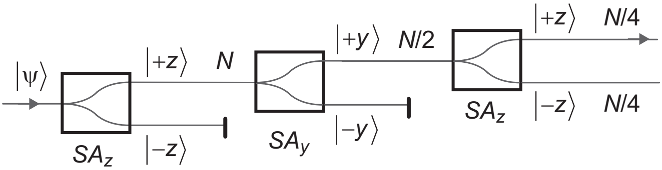


Fig 6.8 Experiment 4.

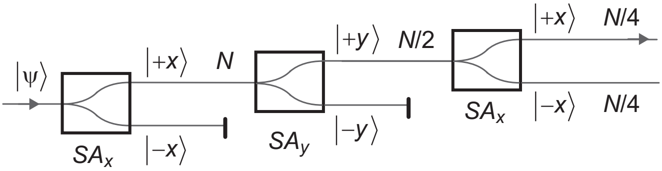


Fig 6.9 Experiment 5.

These states are eigenstates of the operator $\hat{S}_y$, which you'll show in problem 6.8 to be

$$\hat{S}_y \doteq \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}_z = \frac{\hbar}{2}\overline{\boldsymbol{\sigma}}_y \tag{6.19}$$

Here $\overline{\boldsymbol{\sigma}}_y$ is the third and final Pauli matrix.

Just as we did for photons, we can calculate expectation values for measurements performed on spin-1/2 particles, as in the following example.

EXAMPLE 6.2

Calculate the expectation value of S_y, for electrons prepared in the state $\left|+x\right\rangle$.

$$\begin{aligned} \left\langle \hat{S}_y \right\rangle &= \left\langle +x \right| \hat{S}_y \left| +x \right\rangle \\ &\doteq \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix}_z \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}_z \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}_z \\ &= \frac{\hbar}{4}\begin{pmatrix} 1 & 1 \end{pmatrix}_z \begin{pmatrix} -i \\ i \end{pmatrix}_z \\ &= \frac{\hbar}{4}\left(-i+i\right) = 0\,. \end{aligned} \tag{6.20}$$

6.4 COMMUTATION RELATIONS

Do the operators corresponding to the components of spin commute? Let's see.

EXAMPLE 6.3

Find the commutator of $\hat{S}_x$ and $\hat{S}_y$.

$$\begin{aligned} \left[\hat{S}_x, \hat{S}_y\right] &= \hat{S}_x\hat{S}_y - \hat{S}_y\hat{S}_x \\ &\doteq \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_z \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}_z - \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}_z \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_z \\ &= \frac{\hbar^2}{4}\left\{\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}_z - \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}_z\right\} \\ &= \frac{\hbar^2}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}_z \\ &= i\frac{\hbar^2}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_z \\ &= i\hbar\left(\frac{\hbar}{2}\overline{\boldsymbol{\sigma}}_z\right). \end{aligned} \tag{6.21}$$

This means

$$\left[\hat{S}_x, \hat{S}_y\right] = i\hbar\hat{S}_z. \tag{6.22}$$

Equation (6.22) remains true for cyclic permutations of the subscripts (the cyclic permutations of *x, y, z* are *y, z, x* and *z, x, y*), so

$$\left[\hat{S}_y, \hat{S}_z\right] = i\hbar\hat{S}_x, \left[\hat{S}_z, \hat{S}_x\right] = i\hbar\hat{S}_y. \tag{6.23}$$

Flipping the order of the operators in the commutator introduces a minus sign, so

$$\left[\hat{S}_y, \hat{S}_x\right] = -i\hbar\hat{S}_z. \tag{6.24}$$

Equation (6.24) also holds for cyclic permutation of the subscripts.

Since the operators corresponding to the components of spin do not commute, measurements of the spin components must obey an indeterminacy relation [eq. (5.21)]. For instance, eq. (6.22) implies that

$$\Delta S_x \Delta S_y \geq \frac{\hbar}{2}\left|\left\langle \hat{S}_z \right\rangle\right|. \tag{6.25}$$

In chapter 7 we'll explore some further ramifications of the fact that the spin component operators do not commute.

6.5 PARTICLE INTERFERENCE

6.5.1 Experiment 6

Now, look at the experimental arrangement in fig. 6.10(a). A beam of electrons in state $|+x\rangle$ is split into $|+z\rangle$ and $|-z\rangle$ states by an SA_z, and the $|-z\rangle$ electrons are blocked. The $|+z\rangle$ electrons then pass through an SA_{-z}, which is simply an SA_z with the poles of the magnets reversed, in order to deflect electrons in the opposite directions. The SA_{-z} deflects the $|+z\rangle$ electrons, but does not change their state. Last, an SA_x splits the $|+z\rangle$ electrons equally into $|+x\rangle$ and $|-x\rangle$ beams. After what we've already discussed, there shouldn't be anything here that surprises you. Figure 6.10(b) shows the same basic experimental arrangement, only with the $|+z\rangle$ electrons blocked and the $|-z\rangle$ electrons passed.

Now consider what would happen if we perform experiment 6, shown in fig 6.10(c). How many electrons will come out in the $|+x\rangle$ and $|-x\rangle$ states? The first answer that will probably come to mind is that the $|+z\rangle$ electrons will split equally, and the $|-z\rangle$ electrons will split equally, so *N*/2 electrons will emerge in each of the output beams. That's certainly one possibility.

However, hopefully you're beginning to appreciate some of the oddities of quantum mechanics by now, and will reconsider your first inclinations. The SA_z and the SA_{-z} simply split the original beam apart and recombine it. No measurement is performed, and no information is gained. Therefore, after these spin analyzers

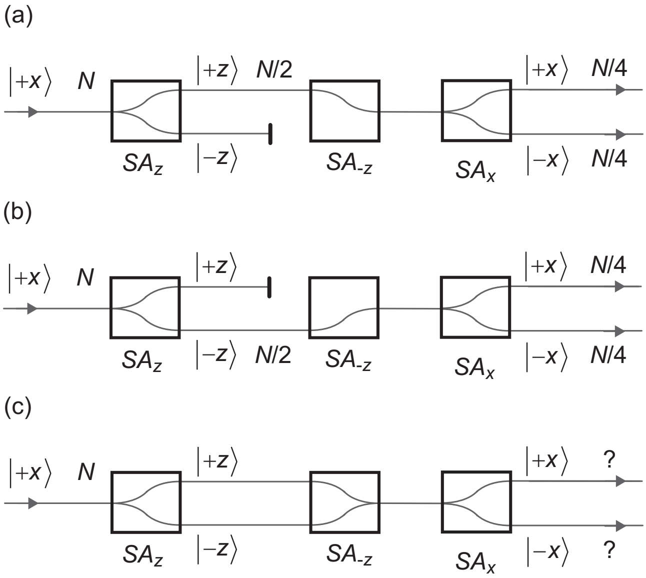


Fig 6.10 (a) and (b) show arrangements for experiment 6 in which one of the beams is blocked, while (c) asks what will happen if neither beam is blocked.

the state of the beam should be unchanged, and the electrons should all be in state $\left|+x\right\rangle$. If this reasoning is correct, then all *N* electrons will emerge from the SA$_x$ in state $\left|+x\right\rangle$.

The only way to decide between these two plausible arguments is to actually do the experiment. The result is shown in fig. 6.11. Evidently, the argument that the combination of the SA$_z$ and the SA$_{-z}$ has no effect on the electrons is the correct one, as all of the electrons exit in state $\left|+x\right\rangle$ (see problem 6.23).

But what happened to the electrons that were coming out in state $\left|-x\right\rangle$ when one of the beams was blocked? How can we get *N*/4 electrons in state $\left|-x\right\rangle$ when the $\left|+z\right\rangle$ beam is passed, and another *N*/4 electrons in state $\left|-x\right\rangle$ when the $\left|-z\right\rangle$ beam is passed, but get *no* electrons in state $\left|-x\right\rangle$ when *both* beams are passed? The answer is interference! The $\left|+z\right\rangle$ and $\left|-z\right\rangle$ beams interfere with each other to produce electrons in state $\left|+x\right\rangle$. This is more obvious if you recall that $\left|+x\right\rangle = \left(1/\sqrt{2}\right)\left(\left|+z\right\rangle + \left|-z\right\rangle\right)$.[2] Interference is a property of waves, and confirmation that electrons have wave-like properties comes from the observation that electrons can exhibit a two-slit diffraction pattern [6.5].

---

[2] It is also possible to insert a phase shifter inside the interferometer formed by the SA$_z$ and the SA$_{-z}$. This would shift the relative phase of the $\left|+z\right\rangle$ and $\left|-z\right\rangle$ states. A π phase shift between them would produce the state $\left|-x\right\rangle = \left(1/\sqrt{2}\right)\left(\left|+z\right\rangle - \left|-z\right\rangle\right)$, so all *N* electrons would leave the SA$_x$ in state $\left|-x\right\rangle$.

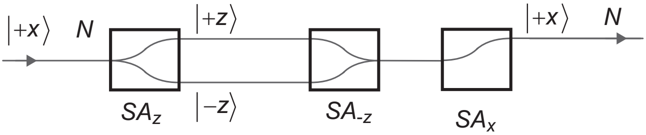


Fig 6.11 The result for experiment 6 when neither beam is blocked.

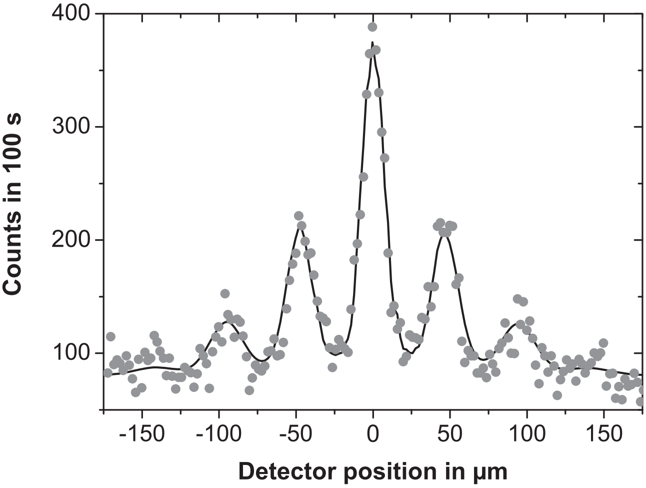


Fig 6.12 The diffraction pattern produced when a beam of C_{60} molecules passes through a diffraction grating having a slit width of 50 nm, and a period of 100 nm. The points represent experimentally measured values, while the solid curve is the theoretically predicted pattern. Reprinted with permission from O. Nairz, M. Arndt, and A. Zeilinger, Am. J. Phys. **71**, 319 (2003). Copyright 2003, American Association of Physics Teachers.

The fact that particles such as electrons can exhibit wave-like interference is surprising to macroscopic human beings, who obey the laws of classical physics. However, the idea that particles exhibit wave-like properties is probably one that you have encountered before. The particle wavelength λ is related to its momentum p by the de Broglie relation $\lambda = h/p$. The frequency f of a particle is related to its energy E by Planck's famous formula $E = hf$. For macroscopic particles the momentum is large enough, and $h = 6.63\times10^{-34}$ $\text{J}\cdot\text{s}$ small enough, to make the wavelength too small to be observable.

The boundary between microscopic and macroscopic is an interesting place, and physicists are constantly pushing it toward larger objects. Figure 6.12 clearly shows interference of C_{60} molecules (carbon atoms arranged in a soccer-ball structure, commonly known as fullerenes or "buckyballs"). These molecules are certainly large compared to electrons, or even a single atom, yet they still show wave-like properties. In this experiment the velocity of the molecules was 136 m/s, which corresponds to a wavelength of $\lambda = 4\times 10^{-3}$ nm. Note that the diameter of the C_{60} molecule itself is approximately 1 nm.

6.6 References

[6.1] B. Friedrich and D. Herschbach, "Stern and Gerlach: how a bad cigar helped reorient atomic physics," *Physics Today* **56**, 53 (2003).

[6.2] S.A. Goudsmit, "The discovery of the electron spin," in *Foundations of Modern EPR*, G.R. Eaton, S.S. Eaton, and K.M. Salikhov, eds. (World Scientific, Singapore, 1998), p. 1.

[6.3] J.S. Townsend, *A Modern Approach to Quantum Mechanics* (McGraw-Hill, New York, 1992), Ch. 1.

[6.4] J. Summhammer *et al.*, "Direct observation of Fermion spin superposition by neutron interferometry," Phys. Rev. A **27**, 2523 (1983).

[6.5] A. Tonomura, *et al.*, "Demonstration of single-electron buildup of an interference pattern," Am. J. Phys. **57**, 117 (1989).

6.7 PROBLEMS

6.1 Silver atoms pass through a Stern-Gerlach magnet with a field gradient of 10 T/cm. If the magnet is 3.5 cm long, and the atoms are moving at 500 m/s, by how far will the atoms be deflected?

6.2* Classically, a uniform, spinning sphere with charge q and mass m has a gyromagnetic ratio (ratio of magnetic dipole moment to angular momentum) of $\gamma = q/2m$. In the relativistic quantum theory of the electron, this formula is modified by the g factor: $\gamma = gq/2m$. What is the g factor for the electron?

6.3 Write $\hat{S}_z$ in terms of the projection operators onto the states $|+z\rangle$ and $|-z\rangle$.

6.4 Write the states $|+z\rangle$ and $|-z\rangle$, using the states $|+x\rangle$ and $|-x\rangle$ as a basis.

6.5 Using your result from problem 6.4, find matrix representation of $\hat{S}_x$ in the z-basis.

6.6 Find the eigenstates and eigenvalues of the matrix representation of $\hat{S}_x$ in the z-basis.

6.7 Write the states $|+z\rangle$ and $|-z\rangle$, using the states $|+y\rangle$ and $|-y\rangle$ as a basis.

6.8* Find the matrix representation of $\hat{S}_y$ in the z-basis.

6.9 Show that

$$\hat{S}_y|+z\rangle = i\frac{\hbar}{2}|-z\rangle, \quad \hat{S}_y|-z\rangle = -i\frac{\hbar}{2}|+z\rangle. \tag{6.26}$$

6.10 Find the eigenstates and eigenvalues of the matrix representation of $\hat{S}_y$ in the z-basis.

6.11 Calculate $P\left(S_y = -\hbar/2 \,\middle\|\, +x\rangle\right)$.

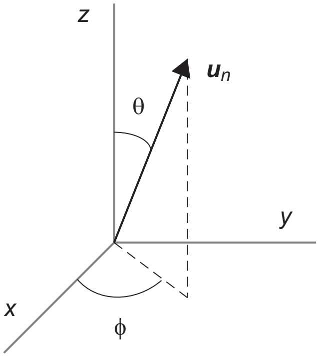


Fig 6.13 An arbitrary unit vector in 3-D space.

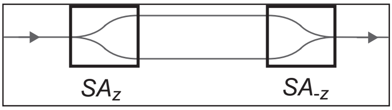


Fig 6.14 The experimental arrangement for problem 6.23.

6.12 Calculate $P\left(S_x = -\hbar/2 \middle\| -y\right\rangle\big)$.

Some of the following problems refer to the state $|+n\rangle = \cos(\theta/2)|+z\rangle + e^{i\phi}\sin(\theta/2)|-z\rangle$, which is a state corresponding to spin-up along the direction $\boldsymbol{u}_n$ in fig. 6.13.

6.13* Find the state $|-n\rangle$, which corresponds to spin down along $\boldsymbol{u}_n$. (Hint: this would be spin-up along $-\boldsymbol{u}_n$.)

6.14 Prove that $|-n\rangle$ is orthogonal to $|+n\rangle$.

6.15 Find the matrix representation of $\hat{S}_n$ in the z-basis.

6.16 Calculate $P\left(S_n = \hbar/2 \| -z\rangle\right)$.

6.17 Calculate $P\left(S_n = -\hbar/2 \| +x\rangle\right)$.

6.18 Calculate the mean and standard deviation of measurements of S_z, for a beam of electrons prepared in state $|+n\rangle$.

6.19* Calculate the mean and standard deviation of measurements of S_x, for a beam of electrons prepared in state $|-n\rangle$.

6.20* A measurement of S_z is performed on an electron prepared in state $|\psi\rangle = \frac{1}{2}|+z\rangle + i\frac{\sqrt{3}}{2}|-z\rangle$. A measurement of S_x is subsequently performed on the same electron. What is the probability that the first measurement returns $-\hbar/2$, and the second returns $\hbar/2$?

6.21 Calculate $\left[\hat{S}_x, \hat{S}_z\right]$.

6.22 Verify that measurements of S_x and S_z satisfy the appropriate indeterminacy relation for a beam of electrons prepared in the state $\left|\psi\right\rangle = \frac{1}{2}\left|+z\right\rangle + i\frac{\sqrt{3}}{2}\left|-z\right\rangle$.

6.23* Show that the operator corresponding to the experimental setup shown in fig. 6.14 is equal to the identity operator. Thus, the spin state of the particles is not changed on propagation through this combination of spin analyzers. (Hint: you might want to look back at sec. 2.5.)

CHAPTER 7

Angular Momentum and Rotation

We now know that spin is an intrinsic angular momentum associated with particles, and we've discussed some of the properties of spin-1/2 particles. In this chapter we'll talk about other types of spin, and about angular momentum in general. We'll also see how rotation is intimately linked with angular momentum. Before we get to that, however, we need to discuss an important property of observables whose corresponding operators commute.

7.1 COMMUTING OBSERVABLES

Consider the Hermitian operators $\hat{A}$ and $\hat{B}$, which have corresponding observables A and B, and assume that they commute: $\left[\hat{A},\hat{B}\right]=0$. It is shown in complement 7.A that there exists a complete set of states $\left|\lambda\right\rangle$ that are simultaneously eigenstates of both $\hat{A}$ and $\hat{B}$. In other words,

$$\hat{A}\left|\lambda\right\rangle = a_\lambda\left|\lambda\right\rangle \text{ and } \hat{B}\left|\lambda\right\rangle = b_\lambda\left|\lambda\right\rangle, \tag{7.1}$$

where the a_λ's are the eigenvalues of $\hat{A}$ corresponding to the states $\left|\lambda\right\rangle$, and the b_λ's are the eigenvalues of $\hat{B}$ corresponding to these same states. It is convenient to label eigenstates by their corresponding eigenvalues; in this case there are two sets of eigenvalues, and we label the states by both. Making this notational change, we let $\left|\lambda\right\rangle = \left|a,b\right\rangle$, and then

$$\hat{A}\left|a,b\right\rangle = a\left|a,b\right\rangle \text{ and } \hat{B}\left|a,b\right\rangle = b\left|a,b\right\rangle. \tag{7.2}$$

The inverse of the above situation is also true: If $\hat{A}$ and $\hat{B}$ do not commute, then they do not have a complete set of simultaneous eigenstates (although they may have some simultaneous eigenstates). This is proved in complement 7.A as well. We also show there that the observables A and B are compatible if $\left[\hat{A},\hat{B}\right]=0$. For compatible observables, the indeterminacy principle allows the system to have well-defined values of both A and B simultaneously. A measurement of B will not change a previously measured value for *A*; if *A* is measured again after the *B* measurement, its value will be the same.

Observables that do not commute are incompatible. Measurements performed on one of the observables can change measurements performed on the other. For example, in eq. (5.27) we showed that $\left[\hat{\wp}_{HV}, \hat{\wp}_{45}\right] \neq 0$, so the observables $\wp_{HV}$ and $\wp_{45}$ are incompatible. To see how this manifests itself, consider a measurement like that shown in fig. 3.4. A measurement of $\wp_{HV}$ is performed on a photon in an arbitrary polarization state, yielding the result *V*. Next a measurement of $\wp_{45}$ is performed, yielding the result +45. If these observables were compatible, a second $\wp_{HV}$ measurement would yield *V* with 100% certainty. However, since they are incompatible, the $\wp_{45}$ measurement alters the polarization state of the photon, so that the final measurement does not yield *V* 100% of the time.

7.2 ANGULAR MOMENTUM OPERATORS

7.2.1 Total Angular Momentum

We can combine the spin component operators to create a "spin vector" operator:

$$\hat{\boldsymbol{S}} = \hat{S}_x \boldsymbol{u_x} + \hat{S}_y \boldsymbol{u}_y + \hat{S}_z \boldsymbol{u}_z . \tag{7.3}$$

In addition to angular momentum associated with their spin, particles can have angular momentum associated with their motion, which in quantum mechanics we refer to as orbital angular momentum. The vector operator corresponding to orbital angular momentum is $\hat{\boldsymbol{L}}$. Unlike spin, which is purely quantum mechanical, orbital angular momentum has a classical counterpart.

The total angular momentum is

$$\begin{aligned} \hat{\boldsymbol{J}} &= \hat{\boldsymbol{S}} + \hat{\boldsymbol{L}} \\ &= \hat{J}_x \boldsymbol{u_x} + \hat{J}_y \boldsymbol{u}_y + \hat{J}_z \boldsymbol{u}_z . \end{aligned} \tag{7.4}$$

The components of $\hat{\boldsymbol{J}}$ (and $\hat{\boldsymbol{L}}$) satisfy the same commutation relations as the components of $\hat{\boldsymbol{S}}$:

$$\left[\hat{J}_x, \hat{J}_y\right] = i\hbar \hat{J}_z, \left[\hat{J}_y, \hat{J}_z\right] = i\hbar \hat{J}_x, \left[\hat{J}_z, \hat{J}_x\right] = i\hbar \hat{J}_y. \tag{7.5}$$

In chapter 6 we demonstrated these relations for the special case of spin-1/2 particles, but they are true for any angular momentum operators.

The states $|+z\rangle$ and $|-z\rangle$ describe only the spin contribution to the total angular momentum, so they have no orbital angular momentum. If $\hat{\boldsymbol{L}} = 0$, then $\hat{\boldsymbol{J}} = \hat{\boldsymbol{S}}$, and

$$\hat{J}_z |+z\rangle = \frac{\hbar}{2}|+z\rangle, \ \hat{J}_z |-z\rangle = -\frac{\hbar}{2}|-z\rangle. \tag{7.6}$$

A consequence of this is that for spin-1/2 the matrix representations of the components of $\hat{\boldsymbol{J}}$ in the *z*-basis are then the same as the corresponding spin operators given in chapter 6: $\hat{J}_i \doteq (\hbar/2)\overline{\boldsymbol{\sigma}}_i$ $(i = x, y, z)$.

7.2.2 The $\hat{J}^2$ Operator

You might think that $\hat{\boldsymbol{J}}$ would play an important role in quantum mechanics, as it does in classical physics. However, except as a notational convenience, we refer to $\hat{\boldsymbol{J}}$ infrequently, because it does not correspond to an observable. The operators that make up the components of $\hat{\boldsymbol{J}}$ do not commute [eq. (7.5)], and therefore they are incompatible observables. We cannot simultaneously measure values for all three components of angular momentum for a single particle. Indeed, experiments 3, 4 and 5 in chapter 6 show us that we can only know one component at a time; measuring any other component disturbs the state of the particle. Furthermore, there are strong arguments to suggest that it's not just that we can't simultaneously *measure* the angular momentum components, but that simultaneous values *don't even exist*. It doesn't make sense to talk about a particle having a definite value for more than one component of its angular momentum at a time. If this is the case, then clearly $\hat{\boldsymbol{J}}$ does not correspond to an observable. The same goes for both $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{L}}$.

The problem with $\hat{\boldsymbol{J}}$ is that its components are incompatible. To learn something about the total angular momentum, we need to find some other operator that contains this information. This new operator needs to be compatible with (at least one of) the components of the angular momentum; this operator is $\hat{J}^2$:

$$\hat{J}^2 = \hat{\boldsymbol{J}} \bullet \hat{\boldsymbol{J}} = \hat{J}_x^{\,2} + \hat{J}_y^{\,2} + \hat{J}_z^{\,2}. \tag{7.7}$$

In the problems you'll prove that $\hat{J}^2$ is Hermitian, and since we can measure J^2, it is an observable. $\hat{J}^2$ is also compatible with each of the components of $\hat{\boldsymbol{J}}$. To show this, let's find the commutator of $\hat{J}^2$ and $\hat{J}_z$. We'll find the following relation, which you'll prove in the problems, useful:

$$\left[\hat{A}\hat{B},\hat{C}\right] = \hat{A}\left[\hat{B},\hat{C}\right] + \left[\hat{A},\hat{C}\right]\hat{B}. \tag{7.8}$$

Using this, the commutator of $\hat{J}^2$ and $\hat{J}_z$ is

$$\begin{aligned}
\left[\hat{J}^2,\hat{J}_z\right] &= \left[\hat{J}_x^{\,2}+\hat{J}_y^{\,2}+\hat{J}_z^{\,2},\hat{J}_z\right] \\
&= \left[\hat{J}_x^{\,2},\hat{J}_z\right]+\left[\hat{J}_y^{\,2},\hat{J}_z\right]+\underset{=0}{\cancel{\left[\hat{J}_z^{\,2},\hat{J}_z\right]}} \\
&= \hat{J}_x\left[\hat{J}_x,\hat{J}_z\right]+\left[\hat{J}_x,\hat{J}_z\right]\hat{J}_x+\hat{J}_y\left[\hat{J}_y,\hat{J}_z\right]+\left[\hat{J}_y,\hat{J}_z\right]\hat{J}_y \\
&= -i\hbar\hat{J}_x\hat{J}_y - i\hbar\hat{J}_y\hat{J}_x + i\hbar\hat{J}_y\hat{J}_x + i\hbar\hat{J}_x\hat{J}_y \\
&= 0\,.
\end{aligned} \tag{7.9}$$

The other components of $\hat{\boldsymbol{J}}$ commute with $\hat{J}^2$ as well. Because they are compatible, we can simultaneously measure J^2 and one component of $\boldsymbol{J}$. Remember, however, that we cannot simultaneously determine more than one component of $\boldsymbol{J}$ at a time.

For a spin-1/2 particle we can calculate the matrix representation of $\hat{J}^2$:

$$\begin{aligned}\hat{J}^2 &= \hat{J}_x^{\,2} + \hat{J}_y^{\,2} + \hat{J}_z^{\,2} \\ &\doteq \frac{\hbar^2}{4}\left\{\overline{\boldsymbol{\sigma}}_x^{\,2} + \overline{\boldsymbol{\sigma}}_y^{\,2} + \overline{\boldsymbol{\sigma}}_z^{\,2}\right\} \\ &= \frac{\hbar^2}{4}\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right\} \\ &= \frac{\hbar^2}{4}\left\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\} \\ &= \frac{3}{4}\hbar^2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.\end{aligned} \tag{7.10}$$

From this equation, it is straightforward to see that both $|+z\rangle$ and $|-z\rangle$ are eigenstates of $\hat{J}^2$, with eigenvalues $3\hbar^2/4$.

From here on, we will assume that all vectors and matrices corresponding to spin-1/2 particles are written in the $|+z\rangle, |-z\rangle$ basis, unless otherwise indicated.

7.3 EIGENVALUES AND EIGENSTATES

We shouldn't be too surprised that $|+z\rangle$ and $|-z\rangle$ are eigenstates of both $\hat{J}^2$ and $\hat{J}_z$, because these are compatible observables, and hence they are guaranteed to have simultaneous eigenstates. These are not the only eigenstates of $\hat{J}^2$ and $\hat{J}_z$, however. In complement 7.B we find all of the allowed eigenvalues and eigenstates. There we find that

$$\hat{J}^2\left|j,m_j\right\rangle = j(j+1)\hbar^2\left|j,m_j\right\rangle, \tag{7.11}$$

$$\hat{J}_z\left|j,m_j\right\rangle = m_j\hbar\left|j,m_j\right\rangle. \tag{7.12}$$

The allowed values for j are

$$j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots, \tag{7.13}$$

and for a given value of j, the allowed values for m_j are

$$m_j = -j, -j+1, -j+2, \ldots, j-2, j-1, j. \tag{7.14}$$

Thus, for a given value of j, there are $2j+1$ allowed values for m_j. For example:

$$j = 0,\ m_j = 0, \tag{7.15}$$

$$j = \frac{1}{2},\ m_j = -\frac{1}{2}, \frac{1}{2}, \tag{7.16}$$

$$j = 1,\ m_j = -1, 0, 1. \tag{7.17}$$

We refer to j as the total angular momentum quantum number [the corresponding eigenvalue of $\hat{J}^2$ is $j(j+1)\hbar^2$]. Likewise, m_j is the quantum number corresponding to the z-component of angular momentum (with corresponding eigenvalue $m_j\hbar$).

The square root of J^2 gives us the magnitude of the total angular momentum: $J=\sqrt{j(j+1)}\,\hbar$. For a given J, there are $2j+1$ allowed values for J_z. J_z has a maximum value (for $m_j = j$) of $J_{z\,\max} = j\hbar$. Notice that J is *always* greater than $J_{z\,\max}$ (except for the case of 0 angular momentum, $j = 0$), which means that the angular momentum can never be perfectly aligned with the z-axis (the same is true of any other axis). This is a consequence of the indeterminacy principle. The components of angular momentum satisfy indeterminacy relations, an example of which is

$$\Delta J_x \Delta J_y \geq \frac{\hbar}{2}\left|\left\langle \hat{J}_z \right\rangle\right|. \tag{7.18}$$

If the expectation value of the z-component of angular momentum is nonzero, there is uncertainty in the x- and y-components. This means that the x- and y-components of angular momentum cannot be 0, so the angular momentum can never be perfectly aligned with the z-axis.

7.3.1 Raising and Lowering Operators

When dealing with angular momentum, it is often convenient to use the raising operator $\hat{J}_+$ and the lowering operator $\hat{J}_-$, which are defined by

$$\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y. \tag{7.19}$$

These operators are described in detail in complement 7.B, but for our purposes here we are mainly interested in their action on the angular momentum states:

$$\hat{J}_+\left|j,m_j\right\rangle = \hbar\left[j(j+1)-m_j(m_j+1)\right]^{1/2}\left|j,m_j+1\right\rangle, \tag{7.20}$$

$$\hat{J}_-\left|j,m_j\right\rangle = \hbar\left[j(j+1)-m_j(m_j-1)\right]^{1/2}\left|j,m_j-1\right\rangle. \tag{7.21}$$

As you can see, the raising operator raises the quantum number corresponding to the z-component of angular momentum by 1, while the lowering operator lowers it by 1.

What does $\hat{J}_+$ do to a state whose z-component of angular momentum is already maximized? A state with maximum z-component of angular momentum has $m_j = j$, so eq. (7.20) yields

$$\hat{J}_+\left|j,j\right\rangle = \hbar\left[j(j+1)-j(j+1)\right]^{1/2}\left|j,j\right\rangle = 0. \tag{7.22}$$

The raising operator cannot raise the z-component of angular momentum higher than its maximum value. You can also show that the lowering operator will not lower the z-component of angular momentum beyond its minimum value.

7.3.2 Spin

For a free particle $\hat{\boldsymbol{J}} = \hat{\boldsymbol{S}}$, so $\hat{J}^2 = \hat{S}^2$. Equations (7.11) and (7.12) thus describe eigenvalues and eigenstates of $\hat{S}^2$ and $\hat{S}_z$. For spin we use the spin quantum numbers s and m_s, so

$$\hat{S}^2 \left| s, m_s \right\rangle = s\left(s+1\right)\hbar^2 \left| s, m_s \right\rangle, \tag{7.23}$$

$$\hat{S}_z \left| s, m_s \right\rangle = m_s \hbar \left| s, m_s \right\rangle, \tag{7.24}$$

with

$$s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots, \tag{7.25}$$

$$m_s = -s, -s+1, -s+2, \ldots, s-2, s-1, s. \tag{7.26}$$

In the previous section we found that $\left|+z\right\rangle$ and $\left|-z\right\rangle$ are eigenstates of $\hat{J}^2$, with eigenvalue $3\hbar^2/4$. It thus follows that

$$\hat{S}^2 \left|+z\right\rangle = \frac{3}{4}\hbar^2 \left|+z\right\rangle = \frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2 \left|+z\right\rangle, \tag{7.27}$$

$$\hat{S}^2 \left|-z\right\rangle = \frac{3}{4}\hbar^2 \left|-z\right\rangle = \frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2 \left|-z\right\rangle. \tag{7.28}$$

We already know that

$$\hat{S}_z \left|+z\right\rangle = \frac{1}{2}\hbar \left|+z\right\rangle, \ \hat{S}_z \left|-z\right\rangle = -\frac{1}{2}\hbar \left|-z\right\rangle. \tag{7.29}$$

Thus, we learn that $s = 1/2$ for a spin-1/2 particle (in general we refer to particles as being spin-s). It follows that

$$\left|+z\right\rangle = \left|\frac{1}{2}, \frac{1}{2}\right\rangle, \left|-z\right\rangle = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle. \tag{7.30}$$

Here the states are the same, they're just expressed in different notation.

Particles with integer spin ($s = 1, 2, \ldots$) are bosons, while those with half-integer spin ($s = 1/2, 3/2, \ldots$) are fermions. These different types of particles have surprisingly different behaviors, which we'll talk more about in sec. 13.5.

EXAMPLE 7.1

What are the allowed eigenstates of $\hat{S}^2$ and $\hat{S}_z$ for a spin-2 particle?

A spin-2 particle has $s = 2$, so

$$m_s = -2, -1, 0, 1, 2. \tag{7.31}$$

The eigenstates are thus

$$\left|2,-2\right\rangle, \left|2,-1\right\rangle, \left|2,0\right\rangle, \left|2,1\right\rangle, \left|2,2\right\rangle. \tag{7.32}$$

EXAMPLE 7.2

Express the eigenstates of $\hat{S}_x$ for a spin-1/2 particle using the $\left|s,m_s\right\rangle$ state notation.

In sec. 6.3 we found that the eigenstates of $\hat{S}_x$ are

$$\left|+x\right\rangle=\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle+\left|-z\right\rangle\right),\ \left|-x\right\rangle=\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle-\left|-z\right\rangle\right). \tag{7.33}$$

These states can be written as

$$\begin{aligned}\left|+x\right\rangle&=\left|\frac{1}{2},\frac{1}{2}\right\rangle_x\\&=\frac{1}{\sqrt{2}}\left(\left|\frac{1}{2},\frac{1}{2}\right\rangle+\left|\frac{1}{2},-\frac{1}{2}\right\rangle\right),\end{aligned} \tag{7.34}$$

$$\begin{aligned}\left|-x\right\rangle&=\left|\frac{1}{2},-\frac{1}{2}\right\rangle_x\\&=\frac{1}{\sqrt{2}}\left(\left|\frac{1}{2},\frac{1}{2}\right\rangle-\left|\frac{1}{2},-\frac{1}{2}\right\rangle\right).\end{aligned} \tag{7.35}$$

The subscripts x on the states indicate eigenstates of $\hat{S}_x$; if there is no subscript, we assume that we're talking about eigenstates of $\hat{S}_z$. Note that the eigenstates of $\hat{S}_x$ are also eigenstates of $\hat{S}^2$.

7.4 SPIN-1

There are three eigenstates for spin-1 particles: $\left|1,1\right\rangle$, $\left|1,0\right\rangle$ and $\left|1,-1\right\rangle$. Given this, what would happen if a beam of spin-1 particles passes through an SA$_z$? Since there are three allowed values for the z-component of spin, in general the beam will be split in three, as shown in fig. 7.1. The same result will be obtained if the beam passes through spin analyzers oriented in other directions (e.g., an SA$_x$ or an SA$_y$).

Since $\left|1,1\right\rangle$, $\left|1,0\right\rangle$ and $\left|1,-1\right\rangle$ are eigenstates of $\hat{S}_z$, in this basis the matrix representation of $\hat{S}_z$ is diagonal. You should be able to verify that

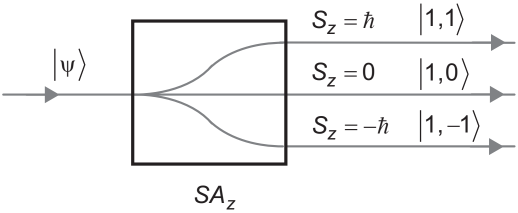


Fig 7.1 A beam of spin-1 particles is incident on a spin analyzer oriented along the z-direction, SA$_z$.

$$\hat{S}_z \doteq \begin{pmatrix} \langle 1,1|\hat{S}_z|1,1\rangle & \langle 1,1|\hat{S}_z|1,0\rangle & \langle 1,1|\hat{S}_z|1,-1\rangle \\ \langle 1,0|\hat{S}_z|1,1\rangle & \langle 1,0|\hat{S}_z|1,0\rangle & \langle 1,0|\hat{S}_z|1,-1\rangle \\ \langle 1,-1|\hat{S}_z|1,1\rangle & \langle 1,-1|\hat{S}_z|1,0\rangle & \langle 1,-1|\hat{S}_z|1,-1\rangle \end{pmatrix} \tag{7.36}$$

$$= \hbar \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

You'll show in the problems that

$$\hat{S}_x \doteq \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \hat{S}_y \doteq \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \tag{7.37}$$

In order to obtain the spin eigenstates along the x- and y-directions, we can diagonalize these matrices.

7.5 ROTATION

In chapter 4 we discussed the polarization rotation operator. Now we're going to look at rotation more generally, and find an operator that will rotate things besides just polarization.

7.5.1 Infinitesimal Rotations

Imagine that we have a spin-1/2 particle in state $|+z\rangle$, and we rotate the spin by an infinitesimal angle $d\theta$ about the x-axis, as shown in fig. 7.2. Thus,

$$\hat{R}(d\theta, \boldsymbol{u}_x)|+z\rangle = |\psi\rangle, \tag{7.38}$$

where $|\psi\rangle$ is the rotated spin state.

We know that $|\psi\rangle$ can be written as a linear combination of $|+z\rangle$ and $|-z\rangle$, because they form a basis. Since $d\theta$ is infinitesimal, $|\psi\rangle$ will be only slightly different from the initial state $|+z\rangle$, so the $|-z\rangle$ contribution will be small. To first order in $d\theta$ we'd expect

$$\hat{R}(d\theta, \boldsymbol{u}_x)|+z\rangle = |+z\rangle - i d\theta\, c|-z\rangle, \tag{7.39}$$

where c is a constant.[1] It's not obvious at this point that the factor of $-i$ is needed, but if we allow c to be a complex number, there is no loss in generality in using this expression. Equation (7.39) implies that

---

[1] For the moment we're working with equations that are correct to first-order in $d\theta$, and eq. (7.39) is normalized to this order.

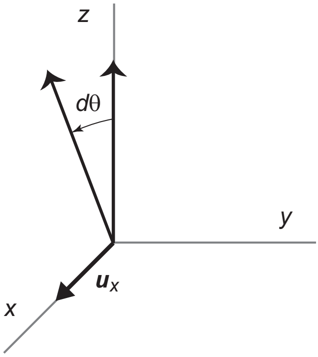


Fig 7.2 Rotation by the angle $d\theta$ about the axis $\boldsymbol{u}_x$.

$$\hat{R}\left(d\theta,\boldsymbol{u}_x\right)=\hat{1}-id\theta\,\hat{G}_x, \tag{7.40}$$

where $\hat{G}_x$ is a dimensionless operator that "generates" rotation about $\boldsymbol{u}_x$.

In classical physics the generator of rotation is angular momentum, so it's reasonable to use the angular momentum operator as the generator of rotation in quantum mechanics. The generator of rotation about $\boldsymbol{u}_x$ is thus $\hat{J}_x$. Using a factor of $\hbar$ to make the units work, eq. (7.40) becomes

$$\hat{R}\left(d\theta,\boldsymbol{u}_x\right)=\hat{1}-\frac{i}{\hbar}\hat{J}_x d\theta. \tag{7.41}$$

7.5.2 The Rotation Operator

Rotation through an angle $\theta+d\theta$ is equivalent to sequential rotations of θ and $d\theta$:

$$\hat{R}\left(\theta+d\theta,\boldsymbol{u}_x\right)=\hat{R}\left(d\theta,\boldsymbol{u}_x\right)\hat{R}\left(\theta,\boldsymbol{u}_x\right). \tag{7.42}$$

Using eq. (7.41), this is

$$\begin{aligned}\hat{R}\left(\theta+d\theta,\boldsymbol{u}_x\right)&=\left(\hat{1}-\frac{i}{\hbar}\hat{J}_x d\theta\right)\hat{R}\left(\theta,\boldsymbol{u}_x\right)\\&=\hat{R}\left(\theta,\boldsymbol{u}_x\right)-\frac{i}{\hbar}\hat{J}_x\hat{R}\left(\theta,\boldsymbol{u}_x\right)d\theta\,.\end{aligned} \tag{7.43}$$

Rearranging, we find

$$\frac{\hat{R}\left(\theta+d\theta,\boldsymbol{u}_x\right)-\hat{R}\left(\theta,\boldsymbol{u}_x\right)}{d\theta}=-\frac{i}{\hbar}\hat{J}_x\hat{R}\left(\theta,\boldsymbol{u}_x\right). \tag{7.44}$$

If we define the derivative of an operator in the same way as the derivative of a function, that is,

$$\lim_{\Delta\theta\to 0}\frac{\hat{R}(\theta+\Delta\theta,\boldsymbol{u}_x)-\hat{R}(\theta,\boldsymbol{u}_x)}{\Delta\theta}=\frac{d}{d\theta}\hat{R}(\theta,\boldsymbol{u}_x), \tag{7.45}$$

then eq. (7.44) becomes

$$\frac{d}{d\theta}\hat{R}(\theta,\boldsymbol{u}_x)=-\frac{i}{\hbar}\hat{J}_x\hat{R}(\theta,\boldsymbol{u}_x). \tag{7.46}$$

This is a straightforward differential equation to solve, and the solution is

$$\hat{R}(\theta,\boldsymbol{u}_x)=e^{-i\theta\hat{J}_x/\hbar}. \tag{7.47}$$

The generator of rotation about an arbitrary axis $\boldsymbol{u}_n$ is $\hat{J}_n=\hat{\boldsymbol{J}}\cdot\boldsymbol{u}_n$, so

$$\hat{R}(\theta,\boldsymbol{u}_n)=e^{-i\theta\hat{J}_n/\hbar}\ . \tag{7.48}$$

The properties of the polarization rotation operator that we noted in sec. 4.2 still hold for $\hat{R}(\theta,\boldsymbol{u}_n)$: the rotation operator is unitary, and $\hat{R}^{\dagger}(\theta,\boldsymbol{u}_n)=\hat{R}(-\theta,\boldsymbol{u}_n)$.[2]

EXAMPLE 7.3

Compute $\hat{R}(\theta,\boldsymbol{u}_z)\left|+z\right\rangle$.

$$\begin{aligned}\hat{R}(\theta,\boldsymbol{u}_z)\left|+z\right\rangle &= e^{-i\theta\hat{J}_z/\hbar}\left|+z\right\rangle\\ &=\sum_{n=0}^{\infty}\frac{1}{n!}\left(-i\frac{\theta\hat{J}_z}{\hbar}\right)^n\left|+z\right\rangle\\ &=\sum_{n=0}^{\infty}\frac{1}{n!}\left(-i\frac{\theta(\hbar/2)}{\hbar}\right)^n\left|+z\right\rangle\\ &=\left[\sum_{n=0}^{\infty}\frac{1}{n!}\left(-i\frac{\theta}{2}\right)^n\right]\left|+z\right\rangle\\ &=e^{-i\theta/2}\left|+z\right\rangle .\end{aligned} \tag{7.49}$$

The state $\left|+z\right\rangle$ does not change on rotation about $\boldsymbol{u}_z$, it merely picks up an angle-dependent phase. Physically, we wouldn't expect the state to change; if you rotate something that's pointing up along $\boldsymbol{u}_z$ about the z-axis, it will still be pointing up along $\boldsymbol{u}_z$.

You've probably noticed that we could have performed the computation in eq. (7.49) much more easily by simply replacing $\hat{J}_z$ by its eigenvalue:

$$\begin{aligned}\hat{R}(\theta,\boldsymbol{u}_z)\left|+z\right\rangle &= e^{-i\theta\hat{J}_z/\hbar}\left|+z\right\rangle\\ &=e^{-i\theta(\hbar/2)/\hbar}\left|+z\right\rangle\\ &=e^{-i\theta/2}\left|+z\right\rangle .\end{aligned} \tag{7.50}$$

---

[2] The factor of $-i$ in eq. (7.39) is necessary to make $\hat{R}(\theta,\boldsymbol{u}_z)$ unitary.

This does indeed work, but it *only* works because $\left|+z\right\rangle$ is an eigenstate of $\hat{J}_z$. Applying $\hat{R}(\theta, \boldsymbol{u}_z)$ to any state that is *not* an eigenstate of $\hat{J}_z$ is not so easy. For this, it's usually easiest to rewrite the state as a linear combination of eigenstates of $\hat{J}_z$, as shown in the following example.

EXAMPLE 7.4

Compute $\hat{R}(\pi/2, \boldsymbol{u}_z)\left|+x\right\rangle$.

$$\begin{aligned} \hat{R}(\pi/2, \boldsymbol{u}_z)\left|+x\right\rangle &= e^{-i(\pi/2)\hat{J}_z/\hbar}\left|+x\right\rangle \\ &= e^{-i(\pi/2)\hat{J}_z/\hbar}\left[\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle + \left|-z\right\rangle\right)\right] \\ &= \frac{1}{\sqrt{2}}\left(e^{-i(\pi/2)(\hbar/2)/\hbar}\left|+z\right\rangle + e^{-i(\pi/2)(-\hbar/2)/\hbar}\left|-z\right\rangle\right) \\ &= \frac{1}{\sqrt{2}}e^{-i\pi/4}\left(\left|+z\right\rangle + e^{i\pi/2}\left|-z\right\rangle\right) \\ &= e^{-i\pi/4}\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle + i\left|-z\right\rangle\right) \\ &= e^{-i\pi/4}\left|+y\right\rangle. \end{aligned} \tag{7.51}$$

Spin up along the *x*-axis, rotated by 90° about the *z*-axis, yields spin up along the *y*-axis, which makes sense physically, if you think about the geometry.

7.6 SPIN OF A PHOTON

The polarization rotation operator $\hat{R}_p(\theta)$(chapter 4) rotates the polarization of a photon about its propagation direction, which is assumed to be $\boldsymbol{u}_z$. There's nothing special about this operator, it's just a special case of the rotation operator described in the last section. In other words, when applied to a photon polarization state

$$\begin{aligned} \hat{R}(\theta, \boldsymbol{u}_z) &= \hat{R}_p(\theta) \\ &\doteq \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}_{HV}. \end{aligned} \tag{7.52}$$

Notice that

$$\begin{aligned} \hat{R}(\theta, \boldsymbol{u}_z)\left|L\right\rangle &\doteq \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}_{HV}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}_{HV} \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix} \cos\theta - i\sin\theta \\ \sin\theta + i\cos\theta \end{pmatrix}_{HV} \end{aligned}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} \cos\theta - i\sin\theta \\ i(\cos\theta - i\sin\theta) \end{pmatrix}_{HV} \tag{7.53}$$
$$= e^{-i\theta} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}_{HV}$$
$$\doteq e^{-i\theta} |L\rangle .$$

In other words, $|L\rangle$ is an eigenstate of $\hat{R}(\theta, \boldsymbol{u}_z)$, with eigenvalue $e^{-i\theta}$. The other polarization eigenstate is $|R\rangle$, with eigenvalue $e^{i\theta}$. Using eqs. (7.48) and (7.53), we find

$$\hat{R}(\theta, \boldsymbol{u}_z)|L\rangle = e^{-i\theta\hat{J}_z/\hbar}|L\rangle \tag{7.54}$$
$$= e^{-i\theta}|L\rangle .$$

In order for this to be true, it must be the case that

$$\hat{J}_z|L\rangle = \hbar|L\rangle. \tag{7.55}$$

Thus, $|L\rangle$ is an eigenstate of $\hat{J}_z$, with eigenvalue $\hbar$.

While it is possible for photons to have orbital angular momentum (see ref. [7.1]), ordinarily they do not, in which case $\hat{J}_z = \hat{S}_z$. Equation (7.55) then means that $|L\rangle$ is an eigenstate of $\hat{S}_z$, with eigenvalue $\hbar$ ($m_s = 1$); it is also true that $|R\rangle$ is an eigenstate of $\hat{S}_z$, with eigenvalue $-\hbar$ ($m_s = -1$).[3] Photons have $s = 1$, so they are spin-1 particles.

We've always treated photon polarization as a 2-D system, but spin-1 is a 3-D system. If photons have a spin-1, what happened to the third eigenstate? Where is $m_s = 0$? It turns out that the $m_s = 0$ state is not allowed because the photon mass is 0. The electromagnetic field is transverse, so the field is perpendicular to the propagation direction, and the spin eigenstates can only be parallel or antiparallel to the propagation direction.[4]

7.7 References

[7.1] L. Allen, M. J. Padgett, and M. Babiker, "The Orbital Angular Momentum of Light," in *Progress in Optics*, E. Wolf, editor (Elsevier, Amsterdam, 1999), vol. 39, p. 291.

[7.2] J.J. Sakurai, *Advanced Quantum Mechanics* (Benjamin/Cummings, Menlo Park, CA, 1967) sec. 2–3.

---

[3] A clarification about signs: remember that the definition of left-circular polarization is that you point your *left thumb toward the source*, and your fingers curl in the direction that the polarization rotates. Thus, if you point your *right thumb along the propagation direction*, $\boldsymbol{u}_z$, your fingers will curl along the polarization rotation direction of a left-circularly polarized photon. This is why left-circularly polarized photons have positive helicity (spin parallel to motion, $m_s = 1$), while right-circularly polarized photons have negative helicity.

[4] For more details, see ref. [7.2].

7.8 PROBLEMS

7.1* Show that

$$\left[\hat{A}\hat{B},\hat{C}\right]=\hat{A}\left[\hat{B},\hat{C}\right]+\left[\hat{A},\hat{C}\right]\hat{B}. \tag{7.56}$$

7.2* Show that $\hat{J}^2$ is Hermitian.

7.3 Compute $\left[\hat{J}^2,\hat{J}_x\right]$ and $\left[\hat{J}^2,\hat{J}_y\right]$.

7.4 Show that for a spin-1/2 particle, the operator $\hbar\left|+z\right\rangle\left\langle -z\right|$ is equivalent to the operator $\hat{J}_+$.

7.5 Write the operators $\hat{J}_x$ and $\hat{J}_y$ in terms of the operators $\hat{J}_+$ and $\hat{J}_-$.

7.6 Find the matrix representations of $\hat{J}_+$ and $\hat{J}_-$ for a spin-1 particle, in the basis of the eigenstates of $\hat{J}_z$.

7.7* Find the matrix representations of $\hat{S}_x$ and $\hat{S}_y$ for a spin-1 particle, in the basis of the eigenstates of $\hat{S}_z$. You might find the results of problems 7.5 and 7.6 useful.

7.8* Find the eigenstates of $\hat{S}_x$ for a spin-1 particle, in the basis of the eigenstates of $\hat{S}_z$.

7.9* Find the eigenstates of $\hat{S}_y$ for a spin-1 particle, in the basis of the eigenstates of $\hat{S}_z$.

7.10 A spin-1 particle is measured to have $S_y=\hbar$. What is the probability that a subsequent measurement will yield $S_z=0$? $S_z=\hbar$? $S_z=-\hbar$?

7.11 A spin-1 particle is measured to have $S_y=0$. What is the probability that a subsequent measurement will yield $S_x=-\hbar$?

7.12* A beam of spin-1 particles is prepared in the state $\left|\psi\right\rangle=(1/3)\left(2\left|1,1\right\rangle-i\left|1,0\right\rangle+2\left|1,-1\right\rangle\right)$. What is the probability that measurements performed on the particles will yield $S_y=-\hbar$?

7.13* A beam of spin-1 particles is prepared in the state $\left|\psi\right\rangle=(1/3)\left(2\left|1,1\right\rangle-i\left|1,0\right\rangle+2\left|1,-1\right\rangle\right)$. Calculate the expectation values of S^2, S_y and S_z for these particles.

7.14 Find the matrix representation of the operator $\hat{R}(\theta,\boldsymbol{u}_z)$ for a spin-1/2 particle, in the $\left|+z\right\rangle$, $\left|-z\right\rangle$ basis.

7.15 Find the matrix representation of the operator $\hat{R}(\theta,\boldsymbol{u}_y)$ for a spin-1/2 particle, in the $\left|+z\right\rangle$, $\left|-z\right\rangle$ basis.

7.16* Starting from the state $\left|+z\right\rangle$, apply rotation operators to generate the state $\left|+n\right\rangle$, which is spin up along $\boldsymbol{u}_n$, as shown in fig. 7.3. Show that $\left|+n\right\rangle=\cos(\theta/2)\left|+z\right\rangle+e^{i\phi}\sin(\theta/2)\left|-z\right\rangle$.

7.17 Calculate $\hat{R}(\pi,\boldsymbol{u}_x)\left|+z\right\rangle$.

7.18 Calculate $\hat{R}(\theta,\boldsymbol{u}_z)\left|1,1\right\rangle$, $\hat{R}(\theta,\boldsymbol{u}_z)\left|1,0\right\rangle$ and $\hat{R}(\theta,\boldsymbol{u}_z)\left|1,-1\right\rangle$.

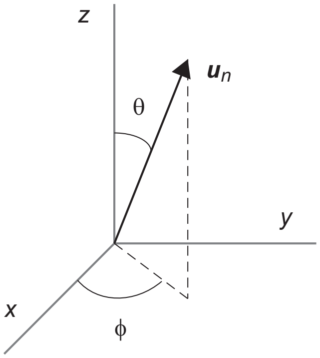


Fig 7.3 The direction of $\boldsymbol{u}_n$.

7.19 Calculate $\hat{R}\left(\pi/2, \boldsymbol{u}_x\right)|1,1\rangle$.

7.20 Calculate $\hat{R}\left(\pi, \boldsymbol{u}_x\right)|1,1\rangle$.

COMPLEMENT 7.A

Compatible Observables

In chapter 7 there were several assertions about compatible observables that were made without proof; the purpose of this complement is to prove them.

7.A.1 Commuting Operators Have a Simultaneous Eigenbasis

In sec. 7.1 we said that if the operators $\hat{A}$ and $\hat{B}$, which correspond to observables A and B, commute, they have a complete set of simultaneous eigenstates $|a,b\rangle$ (an eigenbasis). To prove this, start with the eigenstates $|a\rangle$ and $|b\rangle$:

$$\hat{A}|a\rangle = a|a\rangle, \tag{7.A.1}$$

$$\hat{B}|b\rangle = b|b\rangle. \tag{7.A.2}$$

Since $\hat{A}$ and $\hat{B}$ correspond to observables, we know that their eigenvalues are real, and that their eigenstates (both $|a\rangle$ and $|b\rangle$) form a complete, orthonormal set. Since the states $|b\rangle$ form a complete set, I can expand any of the states $|a\rangle$ as a linear combination of them:

$$|a\rangle = \sum_b c_b |b\rangle. \tag{7.A.3}$$

Here we have assumed that the eigenvalues b are nondegenerate (there is a unique eigenstate corresponding to each eigenvalue).[5]

From eq. (7.A.1) we know that

$$\left(\hat{A} - a\right)|a\rangle = 0. \tag{7.A.4}$$

Using this, and eq. (7.A.3), we find

---

[5] For a proof that accounts for degeneracy, see ref. [7.A.1].

$$\begin{aligned}\left(\hat{A}-a\right)|a\rangle &= \left(\hat{A}-a\right)\sum_b c_b\,|b\rangle \\ &= \sum_b \left(\hat{A}-a\right)c_b\,|b\rangle \\ &= 0.\end{aligned} \tag{7.A.5}$$

If we operate $\hat{B}$ on a single term in this expansion, and use the fact that $\hat{A}$ and $\hat{B}$ commute, we find that

$$\begin{aligned}\hat{B}\left(\hat{A}-a\right)c_b\,|b\rangle &= \left(\hat{A}-a\right)c_b\hat{B}|b\rangle \\ &= b\left(\hat{A}-a\right)c_b\,|b\rangle,\end{aligned} \tag{7.A.6}$$

where we've used eq. (7.A.2). Equation (7.A.6) tells us that $\left(\hat{A}-a\right)c_b\,|b\rangle$ is an eigenstate of $\hat{B}$ with eigenvalue b. Since the eigenstates of Hermitian operators are orthogonal, the states $\left(\hat{A}-a\right)c_b\,|b\rangle$ are orthogonal, and hence all of the terms in the expansion of eq. (7.A.5) are orthogonal to each other. The only way that orthogonal terms can sum to 0 is if each of the terms is individually 0. This means

$$\begin{aligned}\left(\hat{A}-a\right)c_b\,|b\rangle &= 0, \\ \hat{A}c_b\,|b\rangle &= ac_b\,|b\rangle, \\ \hat{A}|b\rangle &= a|b\rangle,\end{aligned} \tag{7.A.7}$$

and the states $|b\rangle$ are eigenstates of both $\hat{A}$ (with eigenvalue a) and $\hat{B}$ (with eigenvalue b). We've already said that the $|b\rangle$ states are complete, so $\hat{A}$ and $\hat{B}$ have a complete set of simultaneous eigenstates. We can relabel the $|b\rangle$ states as $|a,b\rangle$, to explicitly indicate the eigenvalues of both operators.

7.A.2 Non-Commuting Operators Don't Have a Simultaneous Eigenbasis

Now we would like to prove the inverse of the above: If $\hat{A}$ and $\hat{B}$ do not commute, then they do not have a complete set of simultaneous eigenstates. We'll do this by first proving that if $\hat{A}$ and $\hat{B}$ have a complete set of simultaneous eigenstates, then they must commute.

For notational simplicity let's assume that the eigenstates of $\hat{A}$ and $\hat{B}$ are $|\lambda\rangle$, and

$$\hat{A}|\lambda\rangle = a_\lambda\,|\lambda\rangle \text{ and } \hat{B}|\lambda\rangle = b_\lambda\,|\lambda\rangle. \tag{7.A.8}$$

We can expand the Hermitian operators $\hat{A}$ and $\hat{B}$ as

$$\hat{A} = \sum_\lambda a_\lambda\,|\lambda\rangle\langle\lambda| \text{ and } \hat{B} = \sum_{\lambda'} b_{\lambda'}\,|\lambda'\rangle\langle\lambda'|. \tag{7.A.9}$$

This means

$$\begin{aligned}\hat{A}\hat{B} &= \sum_{\lambda} a_{\lambda}\left|\lambda\right\rangle\left\langle\lambda\right| \sum_{\lambda'} b_{\lambda'}\left|\lambda'\right\rangle\left\langle\lambda'\right| \\ &= \sum_{\lambda,\lambda'} a_{\lambda}b_{\lambda'}\left|\lambda\right\rangle\left\langle\lambda|\lambda'\right\rangle\left\langle\lambda'\right|\end{aligned} \tag{7.A.10}$$

The eigenstates of Hermitian operators are orthogonal, so $\left\langle\lambda|\lambda'\right\rangle = 0$ unless $\lambda = \lambda'$, which means $\left\langle\lambda|\lambda'\right\rangle = \delta_{\lambda,\lambda'}$. We can substitute this into eq. (7.A.10), and do the sum over λ', to obtain

$$\begin{aligned}\hat{A}\hat{B} &= \sum_{\lambda,\lambda'} a_{\lambda}b_{\lambda'}\left|\lambda\right\rangle\left\langle\lambda'\right|\delta_{\lambda,\lambda'} \\ &= \sum_{\lambda} a_{\lambda}b_{\lambda}\left|\lambda\right\rangle\left\langle\lambda\right| .\end{aligned} \tag{7.A.11}$$

Likewise, you can show that

$$\hat{B}\hat{A} = \sum_{\lambda} a_{\lambda}b_{\lambda}\left|\lambda\right\rangle\left\langle\lambda\right|, \tag{7.A.12}$$

so $\hat{A}\hat{B} = \hat{B}\hat{A}$, and $\left[\hat{A},\hat{B}\right] = 0$.

Thus, if $\hat{A}$ and $\hat{B}$ have a complete set of simultaneous eigenstates, they commute. From this it follows logically that if $\hat{A}$ and $\hat{B}$ do not commute, they cannot have a complete set of simultaneous eigenstates. They may have *some* simultaneous eigenstates, but these states cannot form a complete set.

7.A.3 Measurements of Compatible Observables

One of the key assertions we made in chapter 7 was that observables whose corresponding operators commute are compatible. By this we mean that it is possible to obtain values for both observables simultaneously. For instance, suppose we measure A and get the value a, and then measure B and get the value b. If these observables are compatible, then the second measurement shouldn't change the first; if I perform another measurement of A, I should once again obtain the value a.

To see that this is the case, we note that compatible observables have corresponding Hermitian operators that commute, so they have simultaneous eigenstates $\left|a,b\right\rangle$, where

$$\hat{A}\left|a,b\right\rangle = a\left|a,b\right\rangle \text{ and } \hat{B}\left|a,b\right\rangle = b\left|a,b\right\rangle. \tag{7.A.13}$$

For simplicity, we'll assume that the eigenvalues are non-degenerate (the eigenstates are uniquely specified by the two eigenvalues a and b).[6] The states $\left|a,b\right\rangle$ are a complete set, so an arbitrary state can be written as a linear combination of them as

---

[6] A generalization for degenerate eigenvalues can be found in ref. [7.A.2].

$$\left|\psi\right\rangle = \sum_{a',b'} c_{a',b'} \left|a',b'\right\rangle. \tag{7.A.14}$$

This state is normalized if

$$\sum_{a',b'} \left|c_{a',b'}\right|^2 = 1. \tag{7.A.15}$$

Suppose now that a system is prepared in the state $\left|\psi\right\rangle$, and that we make a measurement of A which yields the value a. By postulate III(b) (sec. 5.2), after the measurement the system is left in an eigenstate corresponding to the eigenvalue a. However, this does not uniquely specify the state, since we have no information about b. In general the system will be left in a state with contributions from all possible values of b. Thus, after the measurement system is left in state $\left|\psi'\right\rangle$, with a fixed value of a, but a superposition of many b values:

$$\left|\psi'\right\rangle = \frac{1}{\sqrt{\sum_{b'} \left|c_{a,b'}\right|^2}} \sum_{b'} c_{a,b'} \left|a,b'\right\rangle. \tag{7.A.16}$$

The square root factor in front of the sum normalizes $\left|\psi'\right\rangle$. This generalizes postulate III(b).

Now a measurement of B is made, but prior to this measurement system is in state $\left|\psi'\right\rangle$, not $\left|\psi\right\rangle$. This measurement returns the result b, and after the measurement the state $\left|\psi'\right\rangle$ is collapsed to the state $\left|\psi''\right\rangle$, which is an eigenstate corresponding to the eigenvalue b:

$$\left|\psi''\right\rangle = \frac{1}{\sqrt{\left|c_{a,b}\right|^2}} c_{a,b} \left|a,b\right\rangle. \tag{7.A.17}$$

Once again, the square root factor normalizes the state.

Finally, we repeat the measurement of A, now with the system being in-state $\left|\psi''\right\rangle$. But $\left|\psi''\right\rangle$ is an eigenstate of $\hat{A}$, with eigenvalue a:

$$\begin{aligned}
\hat{A}\left|\psi''\right\rangle &= \frac{1}{\sqrt{\left|c_{a,b}\right|^2}} c_{a,b} \hat{A} \left|a,b\right\rangle \\
&= \frac{1}{\sqrt{\left|c_{a,b}\right|^2}} c_{a,b} a \left|a,b\right\rangle \\
&= a\left|\psi''\right\rangle,
\end{aligned} \tag{7.A.18}$$

so the measurement must yield the value a. Thus, A and B are compatible observables, because the B measurement in the middle does not change the measured value of A.

It is also possible to show that, for compatible observables, the measurement order does not affect the probability of measurements. The joint probability of measuring a

and b, for a system prepared in an arbitrary state $|\psi\rangle$, $P\left(a,b\big\||\psi\rangle\right)$, is the same, whether the measurement of A is performed before the measurement of B, or vice versa.[7]

The above arguments are all based on the fact that the states $|a,b\rangle$ are a complete set of simultaneous eigenstates of $\hat{A}$ and $\hat{B}$, which exist only if $\hat{A}$ and $\hat{B}$ commute. If $\hat{A}$ and $\hat{B}$ do not commute the argument breaks down, and A and B are not compatible observables.

7.A.4 References

[7.A.1] L. E. Ballentine, *Quantum Mechanics, A Modern Development* (World Scientific, Singapore, 1998), pg. 24.

[7.A.2] C. Cohen-Tannoudji, B. Diu, and Franck Laloë, *Quantum Mechanics* (Wiley, New York, 1977), Sec. III.C.6.

---

[7] For a proof of this, see ref. [7.A.2].

COMPLEMENT 7.B

Eigenvalues and Eigenstates of Angular Momentum

In this complement we'll derive the eigenvalues and simultaneous eigenstates of the angular momentum operators $\hat{J}^2$ and $\hat{J}_z$.

From sec. 7.2 we know that the commutation relations between the angular momentum components are

$$\left[\hat{J}_x, \hat{J}_y\right] = i\hbar\hat{J}_z,\ \left[\hat{J}_y, \hat{J}_z\right] = i\hbar\hat{J}_x,\ \left[\hat{J}_z, \hat{J}_x\right] = i\hbar\hat{J}_y, \tag{7.B.1}$$

and that

$$\hat{J}^2 = \hat{\boldsymbol{J}}\bullet\hat{\boldsymbol{J}} = \hat{J}_x{}^2 + \hat{J}_y{}^2 + \hat{J}_z{}^2. \tag{7.B.2}$$

Furthermore, all of the operators $\hat{J}^2$, $\hat{J}_x$, $\hat{J}_y$ and $\hat{J}_z$ are Hermitian, and correspond to observables. These pieces of information are enough to uniquely determine the eigenvalues and eigenstates of angular momentum.

Since the angular momentum component operators do not commute with each other, they do not correspond to compatible observables, and we can only determine one of them at a time; it is traditional to choose J_z. $\hat{J}^2$ and $\hat{J}_z$ correspond to observables, so their eigenvalues are real. We showed in sec. 7.2 that $\hat{J}^2$ and $\hat{J}_z$ commute, so they are compatible, and have a complete set of simultaneous eigenstates. For the moment we'll refer to these eigenstates as $\left|\alpha, m\right\rangle$, where

$$\hat{J}^2\left|\alpha, m\right\rangle = \alpha\hbar^2\left|\alpha, m\right\rangle, \tag{7.B.3}$$

$$\hat{J}_z\left|\alpha, m\right\rangle = m\hbar\left|\alpha, m\right\rangle. \tag{7.B.4}$$

We've pulled out the factors of $\hbar$ and $\hbar^2$ because $\hbar$ has units of angular momentum; the constants α and m are thus real, and dimensionless.

7.B.1 Limits on the Eigenvalues

Using eqs. (7.B.2)–(7.B.4), the expectation value of J^2 for state $\left|\alpha, m\right\rangle$ is

$$\begin{aligned}\left\langle\alpha, m\right|\hat{J}^2\left|\alpha, m\right\rangle &= \alpha\hbar^2 \\ &= \left\langle\alpha, m\right|\hat{J}_x^{\,2}\left|\alpha, m\right\rangle + \left\langle\alpha, m\right|\hat{J}_y^{\,2}\left|\alpha, m\right\rangle + \left\langle\alpha, m\right|\hat{J}_z^{\,2}\left|\alpha, m\right\rangle \\ &= \left\langle\alpha, m\right|\hat{J}_x^{\,2}\left|\alpha, m\right\rangle + \left\langle\alpha, m\right|\hat{J}_y^{\,2}\left|\alpha, m\right\rangle + m^2\hbar^2\ .\end{aligned} \tag{7.B.5}$$

Looking at the first term in this sum

$$\begin{aligned}\left\langle\alpha, m\right|\hat{J}_x^{\,2}\left|\alpha, m\right\rangle &= \left\langle\alpha, m\right|\hat{J}_x^{\dagger}\hat{J}_x\left|\alpha, m\right\rangle \\ &= \left(\left\langle\alpha, m\right|\hat{J}_x^{\dagger}\right)\left(\hat{J}_x\left|\alpha, m\right\rangle\right) \\ &\geq 0\ ,\end{aligned} \tag{7.B.6}$$

where we've used the facts that $\hat{J}_x$ is Hermitian, and that the inner product of a vector with itself is a positive real number. Similarly,

$$\left\langle\alpha, m\right|\hat{J}_y^{\,2}\left|\alpha, m\right\rangle \geq 0. \tag{7.B.7}$$

Combining eqs. (7.B.5)–(7.B.7), we see that

$$\alpha \geq m^2. \tag{7.B.8}$$

Since m is real, this implies that α is positive. Furthermore, for a fixed value of α, m is bounded; it has maximum and minimum values such that

$$-\sqrt{\alpha} \leq m_{\min} \leq m \leq m_{\max} \leq \sqrt{\alpha}\,. \tag{7.B.9}$$

7.B.2 The Raising and Lowering Operators

Consider the operators $\hat{J}_+$ and $\hat{J}_-$, defined by

$$\hat{J}_{\pm} = \hat{J}_x \pm i\hat{J}_y. \tag{7.B.10}$$

They both commute with $\hat{J}^2$, since

$$\left[\hat{J}^2, \hat{J}_{\pm}\right] = \underbrace{\cancel{\left[\hat{J}^2, \hat{J}_x\right]}}_{=0} \pm i\,\underbrace{\cancel{\left[\hat{J}^2, \hat{J}_y\right]}}_{=0} = 0. \tag{7.B.11}$$

Their commutators with $\hat{J}_z$ are

$$\begin{aligned}\left[\hat{J}_z,\hat{J}_\pm\right] &= \left[\hat{J}_z,\hat{J}_x\right] \pm i\left[\hat{J}_z,\hat{J}_y\right] \\ &= i\hbar\hat{J}_y \pm i\left(-i\hbar\hat{J}_x\right) \\ &= \hbar\left(\pm\hat{J}_x + i\hat{J}_y\right) \\ &= \pm\hbar\hat{J}_\pm \; .\end{aligned} \qquad (7.B.12)$$

This means that

$$\begin{aligned}\left[\hat{J}_z,\hat{J}_\pm\right] &= \hat{J}_z\hat{J}_\pm - \hat{J}_\pm\hat{J}_z \\ &= \pm\hbar\hat{J}_\pm \; ,\end{aligned} \qquad (7.B.13)$$

and so

$$\begin{aligned}\hat{J}_z\hat{J}_\pm &= \hat{J}_\pm\hat{J}_z \pm \hbar\hat{J}_\pm \\ &= \hat{J}_\pm\left(\hat{J}_z \pm \hbar\right).\end{aligned} \qquad (7.B.14)$$

This is a relationship that we'll find useful shortly. Another useful relationship is

$$\begin{aligned}\hat{J}_+\hat{J}_- &= \left(\hat{J}_x + i\hat{J}_y\right)\left(\hat{J}_x - i\hat{J}_y\right) \\ &= \hat{J}_x^2 + \hat{J}_y^2 + i\left(\hat{J}_y\hat{J}_x - \hat{J}_x\hat{J}_y\right) \\ &= \hat{J}^2 - \hat{J}_z^2 + i\left[\hat{J}_y,\hat{J}_x\right] \\ &= \hat{J}^2 - \hat{J}_z^2 + \hbar\hat{J}_z \; .\end{aligned} \qquad (7.B.15)$$

Similarly, you can show that

$$\hat{J}_-\hat{J}_+ = \hat{J}^2 - \hat{J}_z^2 - \hbar\hat{J}_z. \qquad (7.B.16)$$

What happens when we apply $\hat{J}_+$ to one of our eigenstates? We know that we obtain a new state:

$$\hat{J}_+\left|\alpha,m\right\rangle = \left|\psi\right\rangle, \qquad (7.B.17)$$

but what are its properties? Let's learn something about $\left|\psi\right\rangle$ by applying $\hat{J}_z$ to it:

$$\begin{aligned}\hat{J}_z\left|\psi\right\rangle &= \hat{J}_z\hat{J}_+\left|\alpha,m\right\rangle \\ &= \hat{J}_+\left(\hat{J}_z + \hbar\right)\left|\alpha,m\right\rangle \\ &= \hat{J}_+\left(m\hbar + \hbar\right)\left|\alpha,m\right\rangle \\ &= \left(m+1\right)\hbar\hat{J}_+\left|\alpha,m\right\rangle \\ &= \left(m+1\right)\hbar\left|\psi\right\rangle ,\end{aligned} \qquad (7.B.18)$$

where we've used eqs. (7.B.4), (7.B.14), and (7.B.17). Equation (7.B.18) says that $\left|\psi\right\rangle$ is an eigenstate of $\hat{J}_z$, with eigenvalue $(m+1)\hbar$. In other words

$$\left|\psi\right\rangle = \hat{J}_+\left|\alpha,m\right\rangle = c_+\hbar\left|\alpha,m+1\right\rangle, \tag{7.B.19}$$

where c_+ is a constant that you'll determine in problem 7.B.3. Because $\hat{J}_+$ raises the value of m by 1, it's referred to as the raising operator. In problem 7.B.4 you'll show that

$$\hat{J}_-\left|\alpha,m\right\rangle = c_-\hbar\left|\alpha,m-1\right\rangle, \tag{7.B.20}$$

so $\hat{J}_-$ is called the lowering operator.

7.B.3 Allowed Eigenvalues

By eq. (7.B.9) we know that there is a maximum value of m, and it is customary to set $m_{\max} = j$. Equation (7.B.19) suggests that applying $\hat{J}_+$ to $\left|\alpha,j\right\rangle$ would raise j to j+1, but that can't happen because we know that j is the maximum allowed value. In order to ensure that we don't have a contradiction, it is necessary that

$$\hat{J}_+\left|\alpha,j\right\rangle = 0. \tag{7.B.21}$$

Similarly, $\hat{J}_-$ cannot be allowed to lower m below $m_{\min}$, so

$$\hat{J}_-\left|\alpha,m_{\min}\right\rangle = 0. \tag{7.B.22}$$

Using eqs. (7.B.3), (7.B.4), (7.B.16), and (7.B.21), we find

$$\begin{aligned} \hat{J}_-\hat{J}_+\left|\alpha,j\right\rangle &= \left(\hat{J}^2-\hat{J}_z^2-\hbar\hat{J}_z\right)\left|\alpha,j\right\rangle, \\ \hat{J}_-\left(\hat{J}_+\left|\alpha,j\right\rangle\right) &= \hbar^2\left(\alpha-j^2-j\right)\left|\alpha,j\right\rangle, \\ 0 &= \hbar^2\left[\alpha-j\left(j+1\right)\right]\left|\alpha,j\right\rangle. \end{aligned} \tag{7.B.23}$$

This means that

$$\alpha = j\left(j+1\right). \tag{7.B.24}$$

In a similar fashion, eqs. (7.B.3), (7.B.4), (7.B.15), and (7.B.22) show that

$$\begin{aligned} \hat{J}_+\hat{J}_-\left|\alpha,m_{\min}\right\rangle &= \left(\hat{J}^2-\hat{J}_z^2+\hbar\hat{J}_z\right)\left|\alpha,m_{\min}\right\rangle, \\ \hat{J}_+\left(\hat{J}_-\left|\alpha,m_{\min}\right\rangle\right) &= \hbar^2\left(\alpha-m_{\min}^2+m_{\min}\right)\left|\alpha,m_{\min}\right\rangle, \\ 0 &= \hbar^2\left[\alpha+m_{\min}\left(-m_{\min}+1\right)\right]\left|\alpha,m_{\min}\right\rangle, \end{aligned} \tag{7.B.25}$$

so

$$\alpha = -m_{\min}\left(-m_{\min}+1\right). \tag{7.B.26}$$

Comparing eqs. (7.B.24) and (7.B.26) indicates that $m_{\min} = -j$.

In light of the fact that $\alpha = j\left(j+1\right)$, we can label the eigenstates by j, rather than α. Also, since the value of m is associated with j, it is often written as m_j. Thus, eqs. (7.B.3) and (7.B.4) become

$$\hat{J}^2\left|j,m_j\right\rangle = j\left(j+1\right)\hbar^2\left|j,m_j\right\rangle, \tag{7.B.27}$$

$$\hat{J}_z\left|j,m_j\right\rangle = m_j\hbar\left|j,m_j\right\rangle, \tag{7.B.28}$$

where $-j \le m_j \le j$.

Assume that we fix the value of j, and start in the state $\left|j,-j\right\rangle$, which has the lowest allowed value of m_j. We generate the next allowed value for m_j by applying $\hat{J}_+$ to this state. From eq. (7.B.19):

$$\hat{J}_+\left|j,-j\right\rangle = c_+\hbar\left|j,-j+1\right\rangle, \tag{7.B.29}$$

so $-j+1$ is an allowed value for m_j. We continue applying $\hat{J}_+$ to the new states that we generate, obtaining more allowed values for m_j. Eventually we must reach the state $\left|j,j\right\rangle$, at which point $\hat{J}_+\left|j,j\right\rangle = 0$, and we've generated all the allowed states for a given value of j. This process indicates that the allowed values for m_j are

$$m_j = -j,-j+1,-j+2,\ldots,j-2,j-1,j. \tag{7.B.30}$$

Notice that the difference between j and $-j$ must be an integer, so

$$\begin{aligned} j-\left(-j\right) &= \text{int}, \\ 2j &= \text{int}, \\ j &= \frac{\text{int}}{2}. \end{aligned} \tag{7.B.31}$$

The allowed values for j are thus

$$j = 0,\frac{1}{2},1,\frac{3}{2},2,\ldots \tag{7.B.32}$$

For a given value of j, there are $2j+1$ allowed values for m_j. Some examples are:

$$j = 0,\ m_j = 0, \tag{7.B.33}$$

$$j = \frac{1}{2},\ m_j = -\frac{1}{2},\frac{1}{2}, \tag{7.B.34}$$

$$j = 1,\ m_j = -1, 0, 1, \tag{7.B.35}$$

$$j = \frac{3}{2},\ m_j = -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}. \tag{7.B.36}$$

7.B.4 Problems

7.B.1 Show that $\hat{J}_+^\dagger = \hat{J}_-$ and $\hat{J}_-^\dagger = \hat{J}_+$.

7.B.2 Show that $\left[\hat{J}_+, \hat{J}_-\right] = 2\hbar\hat{J}_z$.

7.B.3* Show that

$$\hat{J}_+ \left| j, m_j \right\rangle = \hbar\left[j(j+1) - m_j(m_j+1)\right]^{1/2} \left| j, m_j + 1\right\rangle \tag{7.B.37}$$

Do this by calculating the matrix elements $\left\langle j, m_j \middle| \hat{J}_- \hat{J}_+ \middle| j, m_j \right\rangle$. You will find eqs. (7.B.16), and (7.B.19), and problem 7.B.1 useful.

7.B.4* Verify eq. (7.B.20). Then, using steps similar to problem 7.B.3, show that $c_- = \hbar\left[j(j+1) - m_j(m_j - 1)\right]^{1/2}$.

CHAPTER 8

Two-Particle Systems and Entanglement

Up to now we've been talking about the quantum mechanics of individual particles. In this chapter we'll discuss the quantum mechanics of systems consisting of two particles. There are certain behaviors that two-particle systems exhibit that have no one-particle analog. In particular, we'll see that the states of two particles can become "entangled" with each other. Entanglement is a purely quantum mechanical effect. Experiments using entangled states can be used to prove that any theory based on the usual assumptions of classical physics *cannot* explain certain observed phenomena, while quantum mechanics can. You can perform such an experiment in lab 5.

8.1 PAIRS OF PHOTONS

8.1.1 Two-Photon States and Operators

Consider the process used in the laboratory experiments, spontaneous parametric downconversion, which is described in lab 1, and pictured in fig. 8.1. In this process a single photon from a pump laser is split into two photons, called the signal and the idler, in a crystal. The signal and the idler emerge from the crystal at essentially the same time, and their frequencies and momenta are correlated. For more details about the properties of downconverted photon pairs, see lab 1. Here we are interested in the polarizations of the downconverted photons. For the type-I downconversion used in the labs, the polarization of the two downconverted photons are the same, and orthogonal to that of the pump.

In order to specify the state of this system we need to specify the states of both of the downconverted photons. For many experiments we don't worry about this detail, because the idler photon is used merely to herald the presence of the signal photon. Now, however, we are going to start thinking about the complete polarization state of the two-photon system.

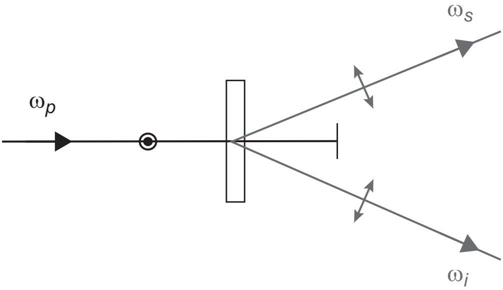


Fig 8.1 Spontaneous parametric downconversion. One pump photon at angular frequency ω_p is converted into signal and idler photons at angular frequencies ω_s and ω_i inside of a crystal. The polarizations of the signal and idler photons are orthogonal to that of the pump.

In order to describe the polarization state of the two-photon system shown in fig. 8.1, the polarization of each photon must be specified. This polarization state is

$$|H,H\rangle = |H\rangle_s \otimes |H\rangle_i . \tag{8.1}$$

The symbol ⊗ denotes the direct product, which combines state vectors in different Hilbert spaces (one for each particle) to create a new vector that specifies the state of the two-particle system in an enlarged Hilbert space. We will usually use the notation $|H,H\rangle$, but there are several different notations that specify this same state; these include

$$\begin{aligned} |H,H\rangle &= |H\rangle_s |H\rangle_i \\ &= |HH\rangle . \end{aligned} \tag{8.2}$$

We will find the notation $|H\rangle_s |H\rangle_i$ to be frequently useful, because it makes very clear which particle is in which state.

If we place a half-wave plate in the signal beam of fig. 8.1, and orient it so that the polarization of the signal beam is rotated to +45°, the state of the two-photon system would then be $|+45,H\rangle$. This state can be rewritten as

$$\begin{aligned} |+45,H\rangle &= |+45\rangle_s \otimes |H\rangle_i \\ &= \frac{1}{\sqrt{2}}\left(|H\rangle_s + |V\rangle_s\right) \otimes |H\rangle_i \\ &= \frac{1}{\sqrt{2}}\left(|H\rangle_s \otimes |H\rangle_i + |V\rangle_s \otimes |H\rangle_i\right) \\ &= \frac{1}{\sqrt{2}}\left(|H,H\rangle + |V,H\rangle\right) . \end{aligned} \tag{8.3}$$

If the polarization state of the idler beam is then modified using a quarter-wave plate to change it to right-circular, the state becomes $\left|+45,R\right\rangle$:

$$\begin{aligned}\left|+45,R\right\rangle &= \left|+45\right\rangle_s \otimes \left|R\right\rangle_i \\ &= \frac{1}{\sqrt{2}}\left(\left|H\right\rangle_s + \left|V\right\rangle_s\right) \otimes \frac{1}{\sqrt{2}}\left(\left|H\right\rangle_i - i\left|V\right\rangle_i\right) \\ &= \frac{1}{2}\left(\left|H\right\rangle_s\left|H\right\rangle_i - i\left|H\right\rangle_s\left|V\right\rangle_i + \left|V\right\rangle_s\left|H\right\rangle_i - i\left|V\right\rangle_s\left|V\right\rangle_i\right) \\ &= \frac{1}{2}\left(\left|H,H\right\rangle - i\left|H,V\right\rangle + \left|V,H\right\rangle - i\left|V,V\right\rangle\right).\end{aligned} \tag{8.4}$$

The four states appearing in the last line of eq. (8.4) are the most commonly used basis states for the 4-dimensional space describing the polarizations of two photons.

We can also compute inner products of two-particle states. For example,

$$\begin{aligned}\left\langle V,+45\middle|R,H\right\rangle &= \left({}_s\left\langle V\right|{}_i\left\langle +45\right|\right)\left(\left|R\right\rangle_s\left|H\right\rangle_i\right) \\ &= {}_s\left\langle V\middle|R\right\rangle_s\,{}_i\left\langle +45\middle|H\right\rangle_i \\ &= \left(\frac{-i}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right) \\ &= \frac{-i}{2}.\end{aligned} \tag{8.5}$$

When computing inner products, the state vectors corresponding to each particle are combined. Thus, in eq. (8.5) the answer is the product of two inner products, one for the signal and one for the idler.

Operators can act in the subspace of one particle, in the subspace of the other, or in the entire space of both particles. For example, consider the *HV*-polarization operator $\hat{\wp}_{HV}$ defined in sec. 5.1; $\left|H\right\rangle$ is an eigenstate of this operator with eigenvalue +1, and $\left|V\right\rangle$ is an eigenstate with eigenvalue –1. We can define $\hat{\wp}_{HV}^{s}$ which acts only on the state of the signal photon, and $\hat{\wp}_{HV}^{i}$ which acts only on the state of the idler photon. Operators corresponding to different particles always commute, for example,

$$\left[\hat{\wp}_{HV}^{s}, \hat{\wp}_{HV}^{i}\right] = 0. \tag{8.6}$$

We can also define operators such as $\hat{\wp}_{HV}^{si} \equiv \hat{\wp}_{HV}^{s}\hat{\wp}_{HV}^{i}$, which acts on the state of both photons, as shown in example 8.1.

EXAMPLE 8.1

Calculate the action of the operators $\hat{\wp}_{HV}^{s}$, $\hat{\wp}_{HV}^{i}$, and $\hat{\wp}_{HV}^{si}$ on the state $\left|V,+45\right\rangle$.

$$\begin{aligned}\hat{\wp}_{HV}^{s}\left|V,+45\right\rangle &= \left(\hat{\wp}_{HV}^{s}\left|V\right\rangle_s\right) \otimes \left|+45\right\rangle_i \\ &= (-1)\left|V\right\rangle_s \otimes \left|+45\right\rangle_i \\ &= -\left|V,+45\right\rangle,\end{aligned} \tag{8.7}$$

$$\begin{aligned}\hat{\wp}^{i}_{HV}\left|V,+45\right\rangle &= \left|V\right\rangle_s \otimes \left(\hat{\wp}^{i}_{HV}\left|+45\right\rangle_i\right)\\ &= \left|V\right\rangle_s \otimes \left[\hat{\wp}^{i}_{HV}\frac{1}{\sqrt{2}}\left(\left|H\right\rangle_i + \left|V\right\rangle_i\right)\right]\\ &= \left|V\right\rangle_s \otimes \left[\frac{1}{\sqrt{2}}\left(\left|H\right\rangle_i - \left|V\right\rangle_i\right)\right]\\ &= \left|V\right\rangle_s \otimes \left|-45\right\rangle_i\\ &= \left|V,-45\right\rangle,\end{aligned} \tag{8.8}$$

$$\begin{aligned}\hat{\wp}^{si}_{HV}\left|V,+45\right\rangle &= \hat{\wp}^{s}_{HV}\hat{\wp}^{i}_{HV}\left|V,+45\right\rangle\\ &= \left(\hat{\wp}^{s}_{HV}\left|V\right\rangle_s\right)\otimes\left(\hat{\wp}^{i}_{HV}\left|+45\right\rangle_i\right)\\ &= \left(-1\right)\left|V\right\rangle_s \otimes \left|-45\right\rangle_i\\ &= -\left|V,-45\right\rangle.\end{aligned} \tag{8.9}$$

8.1.2 Probabilities

Postulate III(b) (sec. 5.2) tells us how to determine the probability of a measurement result. Now let's formulate the probability postulate in a slightly different way.

We want to make measurements of observable O. The eigenvalues and eigenstates of the corresponding Hermitian operator $\hat{O}$ are given by

$$\hat{O}\left|\lambda_n\right\rangle = \lambda_n\left|\lambda_n\right\rangle. \tag{8.10}$$

We'll assume that the eigenvalues are nondegenerate. Postulate III(b) tells us that for a system prepared in state $\left|\psi\right\rangle$, the probability of obtaining λ_n as a measurement result is

$$\begin{aligned}P\left(\lambda_n \,\middle|\, \left|\psi\right\rangle\right) &= \left|\left\langle\lambda_n\middle|\psi\right\rangle\right|^2\\ &= \left\langle\psi\middle|\lambda_n\right\rangle\left\langle\lambda_n\middle|\psi\right\rangle\\ &= \left\langle\psi\right|\left(\left|\lambda_n\right\rangle\left\langle\lambda_n\right|\right)\left|\psi\right\rangle\\ &= \left\langle\psi\right|\hat{P}_{\lambda_n}\left|\psi\right\rangle\\ &= \left\langle\hat{P}_{\lambda_n}\right\rangle,\end{aligned} \tag{8.11}$$

where $\hat{P}_{\lambda_n}$ is the projection operator onto the state $\left|\lambda_n\right\rangle$. Thus, the probability of obtaining a particular eigenvalue as a measurement result can be written as the expectation value of the projection operator onto its corresponding eigenstate. This is another way to formulate Born's rule.

To this point we have been explicitly indicating that probabilities are conditioned upon the state preparation. This notation becomes increasingly cumbersome, however, when talking about probabilities of two or more quantities. So, we're going to change this notation. We will not explicitly indicate that measurement probabilities are conditioned on the state preparation, unless it's absolutely necessary to avoid confusion. The state preparation will be assumed, and we'll use the notation

$$P(\lambda_n) \equiv P\big(\lambda_n \,\big\|\, |\psi\rangle\big). \tag{8.12}$$

Let's look at some examples.

EXAMPLE 8.2

Calculate the probability that the signal photon will be measured to have vertical polarization, and the idler photon will be measured to have horizontal polarization, on a system prepared in the state $|R,+45\rangle$.

We are interested in the joint probability that both measurements yield particular results: $P(V_s,H_i)$. The projection operator onto the state corresponding to this measurement is $\hat{P}_{V_s,H_i} = |V,H\rangle\langle V,H|$, so the probability is given by

$$\begin{aligned} P(V_s,H_i) &= \langle \hat{P}_{V_s,H_i} \rangle \\ &= \langle R,+45|V,H\rangle\langle V,H|R,+45\rangle \\ &= \left|\langle R,+45|V,H\rangle\right|^2 \\ &= \left|{}_s\langle R|V\rangle_s \, {}_i\langle +45|H\rangle_i\right|^2 \\ &= \left|\frac{i}{\sqrt{2}}\frac{1}{\sqrt{2}}\right|^2 \\ &= \frac{1}{4}. \end{aligned} \tag{8.13}$$

EXAMPLE 8.3

Calculate the probability that the idler photon will be measured to have horizontal polarization, on a system prepared in the state $|R,+45\rangle$.

We are interested in the probability $P(H_i)$. The projection operator onto the eigenstate corresponding to this measurement is $\hat{P}_{H_i} = |H\rangle_{i\,i}\langle H|$, so the probability is given by

$$\begin{aligned}
P(H_i) &= \langle \hat{P}_{H_i} \rangle \\
&= \langle R,+45 | \left(|H\rangle_{i\ i}\langle H| \right) | R,+45 \rangle \\
&= \left({}_s\langle R|\ {}_i\langle +45| \right) \left(|H\rangle_{i\ i}\langle H| \right) \left(|R\rangle_s |+45\rangle_i \right) \\
&= {}_s\langle R|R\rangle_{s\ i}\langle +45|H\rangle_{i\ i}\langle H|+45\rangle_i \\
&= (1)\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right) \\
&= \frac{1}{2}.
\end{aligned} \tag{8.14}$$

This probability is independent of any measurement performed on the signal beam.

As we learned in sec 1.1, the conditional probability that a measurement of an observable A yields a, given that a measurement of B yields b, is written as $P(a|b)$. How do we determine $P(a|b)$? We start with eq. (1.27), which says

$$P(a,b) = P(a|b)P(b). \tag{8.15}$$

Rearranging this, we obtain

$$P(a|b) = \frac{P(a,b)}{P(b)}, \tag{8.16}$$

which is Bayes' formula.

EXAMPLE 8.4

Calculate the probability that the signal photon will be measured to have vertical polarization, given that the idler photon is measured to be horizontally polarized, for a system prepared in the state $|R,+45\rangle$.

We are interested in the probability $P(V_s \mid H_i)$. Using eq. (8.16), this is given by

$$P(V_s \mid H_i) = \frac{P(V_s, H_i)}{P(H_i)}. \tag{8.17}$$

The probabilities on the right-hand side of this equation were determined in examples 8.2 and 8.3, and they yield

$$P(V_s \mid H_i) = \frac{1/4}{1/2} = \frac{1}{2}. \tag{8.18}$$

8.2 ENTANGLED STATES

Now that we have described the basic mathematics of two-particle systems, we can discuss a very important class of two-particle states, entangled states. Indeed, entanglement is one of, if not *the* most important feature of quantum mechanics that sets it apart from classical physics.

Suppose that we use the downconversion source shown in fig. 8.2, instead of the simpler one of fig. 8.1. As described in lab 1, for the downconversion process to be efficient, the propagation directions and polarizations of the pump, signal, and idler photons must be properly oriented with respect to the crystal axes. For the source in fig. 8.2 there are two crystals sandwiched together, with their orientations rotated by 90° with respect to each other. One crystal converts vertically polarized pump photons into horizontally polarized signal and idler photons, while the other converts horizontally polarized pump photons into vertically polarized signal and idler photons. If the pump is polarized at 45°, each of these processes is equally likely. If the crystals are thin enough, observers detecting the signal and idler photons have no information about which crystal a given photon was produced in. If care is taken, an observer performing a measurement on the signal (or idler) photon would have absolutely no way of distinguishing between a horizontally or vertically polarized photon, short of actually performing a polarization measurement. If the photons are indistinguishable in this way, the polarization state is a superposition of the two possible states generated by the downconversion process:

$$\left|\phi^{+}\right\rangle=\frac{1}{\sqrt{2}}\left(\left|H,H\right\rangle+\left|V,V\right\rangle\right). \tag{8.19}$$

In sec. 3.7, when discussing single-photon interference, we said that in order to explain the data it was necessary for the photon to take both paths through the interferometer, not one or the other. Likewise, we will soon see that in order to explain experiments performed using the source of fig. 8.2, it is necessary to interpret the state of eq. (8.19) as meaning that the photons are in *both* states $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$ at the same time, not as meaning that they are in one state or the other. In other words, the photons are in a superposition of these states, not a mixture (a term we will define more fully below).

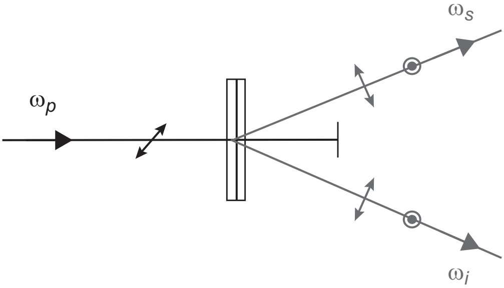


Fig 8.2 A source that uses two crystals, whose axes are rotated 90° with respect to each other, to produce polarization entangled photons. The pump is polarized at +45°, and produces signal and idler photons in the polarization entangled state of eq. (8.19).

Above we said that if we knew the state of particle A to be $\left|\psi_A\right\rangle_A$, and the state of particle B to be $\left|\psi_B\right\rangle_B$, then the state of the total system would be

$$\left|\psi_A,\psi_B\right\rangle=\left|\psi_A\right\rangle_A\otimes\left|\psi_B\right\rangle_B. \tag{8.20}$$

Each pair of basis states of the individual particles can be combined to create an allowed state of the total system. Such states are called product states, and these product states form a basis for the combined system. These states are also pure states, which means they are fully described by a state (e.g., ket) vector.

However, it is not possible to write every allowed state of the combined system as a product state. The state of eq. (8.19) is an example of this—it cannot be factorized into separate states corresponding to the two particles, so it is not a product state (in problem 8.5 you'll prove this). It doesn't make sense to talk about the signal or idler photons as "being" in states by themselves.[1] States of the combined system which cannot be written as product states are known as entangled states. Entangled states are inherently states of the combined system.

Let's examine the probabilities of various measurements performed on photons prepared in the state $\left|\phi^+\right\rangle$ of eq. (8.19).

EXAMPLE 8.5

For a two-photon system prepared in the state $\left|\phi^+\right\rangle$ of eq. (8.19), determine the probabilities of obtaining: (a) the signal photon is measured to be horizontally polarized, (b) the signal photon is measured to be horizontally polarized, given that the idler photon is found to have horizontal polarization.

(a) The probability that the signal photon is measured to be horizontally polarized is

$$\begin{aligned}
P(H_s)&=\left\langle \hat{P}_{H_S}\right\rangle\\
&=\frac{1}{\sqrt{2}}\left(\langle H,H|+\langle V,V|\right)\left(|H\rangle_s\,{}_s\langle H|\right)\frac{1}{\sqrt{2}}\left(|H,H\rangle+|V,V\rangle\right)\\
&=\frac{1}{2}\left[{}_s\langle H|H\rangle_s\,{}_i\langle H|+{}_s\langle V|H\rangle_s\,{}_i\langle V|\right]\left[{}_s\langle H|H\rangle_s|H\rangle_i+{}_s\langle H|V\rangle_s|V\rangle_i\right]\\
&=\frac{1}{2}\,{}_i\langle H|H\rangle_i\\
&=\frac{1}{2}.
\end{aligned} \tag{8.21}$$

(b) The probability that the signal photon is measured to be horizontally polarized, given that the idler photon is found to have horizontal polarization is $P(H_s\,|\,H_i)$. By eq. (8.16) we can write this as

---

[1] At least not pure states; the individual particles are not in states described by a ket vector.

$$P\left(H_s \mid H_i\right)=\frac{P\left(H_s,H_i\right)}{P\left(H_i\right)}. \tag{8.22}$$

The calculation of $P\left(H_i\right)$ proceeds in the same manner as the calculation of $P\left(H_s\right)$ above, except that the projection operator projects onto the idler, as opposed to the signal. You should be able to verify that the result is

$$P\left(H_i\right)=\frac{1}{2}. \tag{8.23}$$

The probability appearing in the numerator of eq. (8.22) is

$$\begin{aligned} P\left(H_s,H_i\right)&=\left\langle \hat{P}_{H_s,H_i}\right\rangle \\ &=\frac{1}{\sqrt{2}}\left(\langle H,H|+\langle V,V|\right)\left(|H,H\rangle\langle H,H|\right)\frac{1}{\sqrt{2}}\left(|H,H\rangle+|V,V\rangle\right) \\ &=\frac{1}{2}\left[\left(\langle H,H|+\langle V,V|\right)|H,H\rangle\right]\left[\langle H,H|\left(|H,H\rangle+|V,V\rangle\right)\right] \\ &=\frac{1}{2}\left[\langle H,H|H,H\rangle+\langle V,V|H,H\rangle\right]\left[\langle H,H|H,H\rangle+\langle H,H|V,V\rangle\right] \\ &=\frac{1}{2}. \end{aligned} \tag{8.24}$$

Substituting eqs. (8.23) and (8.24) into eq. (8.22) yields

$$P\left(H_s \mid H_i\right)=\frac{1/2}{1/2}=1. \tag{8.25}$$

Part (a) of example 8.5 tells us that for photons in the entangled state of eq. (8.16), measurements of $\wp_{HV}^{s}$ yield perfectly random results: the signal photon is measured to have horizontal polarization half of the time (and consequently vertical polarization the other half). The results for the idler are the same: It will be measured to be horizontally polarized half of the time, and vertically polarized the other half.

However, part (b) of example 8.5 tells us that if the idler is found to be horizontally polarized, the signal will be measured to be horizontally polarized with 100% certainty. The ordering of the measurements doesn't matter: If the signal is found to be horizontal, so will the idler. You can perform a similar calculation to show that if the idler is measured to be vertically polarized, the signal will be vertically polarized as well. Simply by looking at the form of the state $\left|\phi^{+}\right\rangle$ you could probably guess that it would yield these probabilities, however, it's wise to explicitly calculate them to validate your intuition.

We learn the following: Despite the fact that the measured individual polarizations are purely random, the results of the measurements are perfectly correlated. Neither photon has a well-defined polarization, yet if the polarization of one photon is known from a measurement, the polarization of the other is determined, at least in the *HV*-basis. Example 8.6 illustrates that this fact is true for measurements in other bases as well.

EXAMPLE 8.6

For a two-photon system prepared in the state $\left|\phi^+\right\rangle$ of eq. (8.19), determine the probability that the signal photon is measured to be polarized along +45°, given that the idler photon is found to be polarized along this same direction.

We want to find $P\left(+45_s \mid +45_i\right)$, which is given by

$$P\left(+45_s \mid +45_i\right)=\frac{P\left(+45_s,+45_i\right)}{P\left(+45_i\right)}. \tag{8.26}$$

The joint probability in the numerator is

$$\begin{aligned}
P\left(+45_s,+45_i\right)&=\left\langle \hat{P}_{+45_s,+45_i}\right\rangle\\
&=\frac{1}{\sqrt{2}}\left(\left\langle H,H\right|+\left\langle V,V\right|\right)\left(\left|+45,+45\right\rangle\left\langle +45,+45\right|\right)\frac{1}{\sqrt{2}}\left(\left|H,H\right\rangle+\left|V,V\right\rangle\right)\\
&=\frac{1}{2}\left|\left\langle H,H\middle|+45,+45\right\rangle+\left\langle V,V\middle|+45,+45\right\rangle\right|^2\\
&=\frac{1}{2}\left|{}_s\left\langle H\middle|+45\right\rangle_s\,{}_i\left\langle H\middle|+45\right\rangle_i+{}_s\left\langle V\middle|+45\right\rangle_s\,{}_i\left\langle V\middle|+45\right\rangle_i\right|^2\\
&=\frac{1}{2}\left|\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)+\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\right|^2\\
&=\frac{1}{2}.
\end{aligned} \tag{8.27}$$

The probability in the denominator is

$$\begin{aligned}
P\left(+45_i\right)&=\left\langle \hat{P}_{+45_i}\right\rangle\\
&=\frac{1}{\sqrt{2}}\left(\left\langle H,H\right|+\left\langle V,V\right|\right)\left(\left|+45\right\rangle_i\,{}_i\left\langle +45\right|\right)\frac{1}{\sqrt{2}}\left(\left|H,H\right\rangle+\left|V,V\right\rangle\right)\\
&=\frac{1}{2}\left[{}_s\left\langle H\right|\,{}_i\left\langle H\middle|+45\right\rangle_i+{}_s\left\langle V\right|\,{}_i\left\langle V\middle|+45\right\rangle_i\right]\left[{}_i\left\langle +45\middle|H\right\rangle_i\left|H\right\rangle_s+{}_i\left\langle +45\middle|V\right\rangle_i\left|V\right\rangle_s\right]\\
&=\frac{1}{2}\left[\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)+0+0+\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\right]\\
&=\frac{1}{2}.
\end{aligned} \tag{8.28}$$

Therefore

$$P\left(+45_s \mid +45_i\right)=\frac{1/2}{1/2}=1. \tag{8.29}$$

In the problems you'll generalize example 8.6 to show that for pairs of photons prepared in state $\left|\phi^{+}\right\rangle$, the polarizations of the two photons are perfectly correlated, for measurements in *any* elliptical polarization basis. Classical physics cannot explain such strong polarization correlations, as we'll soon see.

8.3 MIXED STATES

In the previous section we stressed that the state $\left|\phi^{+}\right\rangle$ of eq. (8.19) must be interpreted as being in both states $|H,H\rangle$ and $|V,V\rangle$ at the same time. How do we know this? Why can't it mean that the system is randomly in either state $|H,H\rangle$ or state $|V,V\rangle$? The answer is that if the system is randomly in one state or the other, the measured probabilities are different than if the system is in both states at the same time.

Before showing that this is the case, we need to formalize our description of the "either-or" state. The entangled state $\left|\phi^{+}\right\rangle$ is a pure state; it is a superposition of two states that is expressed as a state vector. The state which describes a system as being either in state $|H,H\rangle$ or state $|V,V\rangle$ is called a mixed state, or a mixture. It is not pure, because it cannot be described by a single state vector—the system is randomly prepared in one of two possible states.

This state preparation randomness represents a lack of knowledge in a classical sense—like a coin flip, there's nothing intrinsically quantum mechanical about it. For example, to produce a state which is a mixture of $|H,H\rangle$ and $|V,V\rangle$, I could randomly orient the polarization of the pump beam to produce photons in either one of these states.

The best way to deal with mixed states in quantum mechanics is to use the density operator $\hat{\rho}$ (also referred to as the density matrix or the state operator). The density operator is a generalized quantum state. It can describe both pure and mixed states, whereas the state vectors we have been using so far can describe only pure states. A description of the density operator is given in complement 8.A. In the discussion here we won't use the density operator, but instead use the fact that measured probabilities are conditioned on the state that the system is prepared in.

The probability that we will measure $A=a$ for a system prepared in state $|\psi_1\rangle$ is $P\left(a\middle\|\psi_1\rangle\right)$, and the probability that we will obtain $A=a$ for a system prepared in state $|\psi_2\rangle$ is $P\left(a\middle\|\psi_2\rangle\right)$. The total probability of measuring $A=a$ is

$$P(a)=P\left(a\middle\|\psi_1\rangle\right)P\left(|\psi_1\rangle\right)+P\left(a\middle\|\psi_2\rangle\right)P\left(|\psi_2\rangle\right). \tag{8.30}$$

This is simply the classical rule that probabilities for different possibilities add to give the total probability. If there are many different states that the system could be prepared in, we'd modify eq. (8.30) by including a term in the sum for each possible state. For further discussion, see ref. [8.1].

We can use eq. (8.30) to calculate the probabilities of measurements for systems prepared in mixed states. For example, you'll show in the problems that for a mixed state containing 50% $|H,H\rangle$ and 50% $|V,V\rangle$, $P(H_s \mid H_i)=1$. Measurements in the horizontal/vertical basis of each photon are perfectly correlated for this mixed state, just as we found they were for the entangled state $\left|\phi^+\right\rangle$ in example 8.5. Measurements in this basis cannot distinguish photons being in the entangled state or the mixed state. However, the next example shows that there *are* measurements that can distinguish these possibilities.

EXAMPLE 8.7

For a two-photon system prepared in an equal mixture of states $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$, determine the probability that the signal photon is measured to be polarized along +45°, given that the idler photon is found to be polarized along this same direction.

We want to find $P\left(+45_s \mid +45_i\right)$. We can determine it the same way we did in example 8.6, by using eq. (8.26); we just need to calculate the probabilities using the mixed state rather than the entangled state.

An equal mixture means $P\left(\left|H,H\right\rangle\right)=1/2$ and $P\left(\left|V,V\right\rangle\right)=1/2$. Applying eq. (8.30) to the numerator of eq. (8.26), we find

$$\begin{aligned}
P\left(+45_s,+45_i\right) &= P\left(+45_s,+45_i \,\middle\|\, \left|H,H\right\rangle\right)P\left(\left|H,H\right\rangle\right) \\
&\quad + P\left(+45_s,+45_i \,\middle\|\, \left|V,V\right\rangle\right)P\left(\left|V,V\right\rangle\right) \\
&= \left\langle H,H\right|\hat{P}_{+45_s,+45_i}\left|H,H\right\rangle\frac{1}{2}+\left\langle V,V\right|\hat{P}_{+45_s,+45_i}\left|V,V\right\rangle\frac{1}{2} \\
&= \frac{1}{2}\left\langle H,H\middle|+45,+45\right\rangle\left\langle +45,+45\middle|H,H\right\rangle \\
&\quad +\frac{1}{2}\left\langle V,V\middle|+45,+45\right\rangle\left\langle +45,+45\middle|V,V\right\rangle \\
&= \frac{1}{2}\left[\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)+\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\right] \\
&= \frac{1}{4}.
\end{aligned} \tag{8.31}$$

Applying eq. (8.30) to the denominator of eq. (8.26), we find

$$\begin{aligned}
P\left(+45_i\right) &= P\left(+45_i \,\middle\|\, \left|H,H\right\rangle\right)P\left(\left|H,H\right\rangle\right)+P\left(+45_i \,\middle\|\, \left|V,V\right\rangle\right)P\left(\left|V,V\right\rangle\right) \\
&= \left\langle H,H\right|\hat{P}_{+45_i}\left|H,H\right\rangle\frac{1}{2}+\left\langle V,V\right|\hat{P}_{+45_i}\left|V,V\right\rangle\frac{1}{2} \\
&= \frac{1}{2}\,{}_i\left\langle H\middle|+45\right\rangle_i\,{}_i\left\langle +45\middle|H\right\rangle_i\,{}_s\left\langle H\middle|H\right\rangle_s+\frac{1}{2}\,{}_i\left\langle V\middle|+45\right\rangle_i\,{}_i\left\langle +45\middle|V\right\rangle_i\,{}_s\left\langle V\middle|V\right\rangle_s \\
&= \frac{1}{2}\left[\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)+\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\right] \\
&= \frac{1}{2}.
\end{aligned} \tag{8.32}$$

Substituting back into eq. (8.26) yields the final result

$$P\left(+45_s \mid +45_i\right) = \frac{1/4}{1/2} = \frac{1}{2}. \tag{8.33}$$

Let's summarize what we've discovered. For the entangled state $\left|\phi^+\right\rangle$ of eq. (8.19), the polarization correlations between the signal and idler are perfect in both the horizontal/vertical basis [$P\left(H_s \mid H_i\right) = 1$, example 8.5] and the ±45° basis [$P\left(+45_s \mid +45_i\right) = 1$, example 8.6]. For the state that is an equal mixture of $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$ the polarization correlations are perfect in the horizontal/vertical basis [$P\left(H_s \mid H_i\right) = 1$, problem 8.10], but *not* in the ±45° basis [$P\left(+45_s \mid +45_i\right) = 1/2$, example 8.7]. Clearly it is possible to perform measurements that distinguish between these two possible states, and you can perform them for yourself in lab 5.[2] You can determine which state, entangled or mixed, best explains the data.

8.4 TESTING LOCAL REALISM

We're going to describe an experiment which demonstrates that the natural world violates local realism; this is an experiment you can perform in lab 5. Furthermore, we'll find that quantum mechanics perfectly describes the experimental results. We haven't yet defined what we mean by locality or reality, but we'll discuss these concepts after we describe the experiment. In complement 8.B we will give mathematical definitions of these terms.

8.4.1 Alice and Bob's Experiment

A downconversion source produces signal photons that are sent to Alice, and idler photons that are sent to Bob, as shown in fig. 8.3. The source is similar to the polarization-entangled source shown in fig. 8.2, but it has some differences, which will be discussed below. Alice's photons pass through a polarization analyzer oriented along θ_A (a PA_{θ_A}). She finds her photon to be polarized parallel to θ_A with probability $P\left(\theta_A\right)$, and perpendicular to θ_A with probability $P\left(\theta_A^{\perp}\right)$. Bob performs similar measurements on his photons, to determine whether they are polarized parallel or perpendicular to θ_B. Alice is free to rotate her PA how she chooses, and she performs measurements with it in two different orientations, θ_{A1} and θ_{A2}, randomly choosing between them. Bob also randomly switches his PA between two different orientations, θ_{B1} and θ_{B2}.

---

[2] Note that we must perform measurements in two different bases to distinguish between the states. It is possible to create a classical mixed state that will have perfect polarization correlations between two beams in any chosen basis, but those classical correlations will not be maintained in all bases, as they are for the entangled state. See problems 8.12 and 8.13.

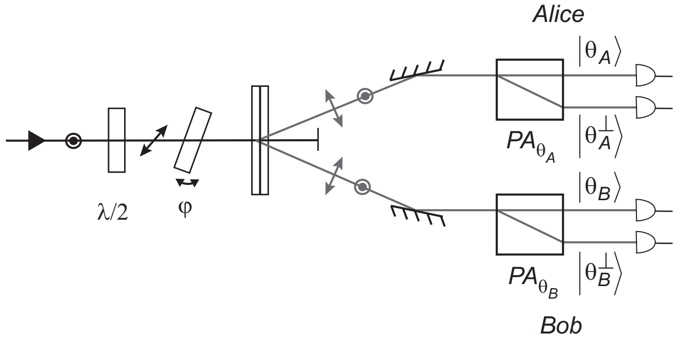


Fig 8.3 An experiment performed by Alice and Bob. A polarization-entangled source directs signal photons to Alice, and idler photons to Bob. Alice and Bob have linear polarization analyzers (PA_{θ_A} and PA_{θ_B}), which they use to measure the polarization of their photons. The half-wave plate, $\lambda/2$, and the birefringent phase plate, φ, are used to control the polarization of the pump.

Alice and Bob are a great distance apart from each other, and they perform their measurements independently, and without communicating with each other. The source is halfway between Alice and Bob, so that they receive their photons at the same time. As they perform their polarization measurements they note the precise arrival time of the photons (using atomic clocks), so that they can go back, after all the measurements are recorded, and compare results corresponding to each photon pair. So, for every photon that she detects, Alice records its arrival time, the orientation of her PA (θ_{A1} or θ_{A2}), and the measured polarization (θ_{A1}, $\theta_{A2}^{\perp}$, etc.). Bob records the same information about his measurements.

After all the data have been recorded, Alice and Bob get together to analyze their data, and determine quantities such as $P\left(\theta_{A1},\theta_{B2}\right)$, which is the joint probability that Alice finds her photon be polarized along θ_{A1}, and Bob finds his photon to be polarized along θ_{B2}. During this comparison Alice and Bob make four observations about their data.[3] The first three are:

1. On those trials where Alice sets her PA to θ_{A1} and Bob sets his to θ_{B1}, they measure photons polarized along these directions 9% of the time [$P\left(\theta_{A1},\theta_{B1}\right)=0.09$].
2. On those trials where Alice sets her PA to θ_{A1} and Bob sets his to θ_{B2}, if Alice measures her photon to be polarized along θ_{A1}, Bob *always* measures his photon to be polarized along θ_{B2} [$P\left(\theta_{B2}\,|\,\theta_{A1}\right)=1$].
3. On those trials where Alice sets her PA to θ_{A2} and Bob sets his to θ_{B1}, if Bob measures his photon to be polarized along θ_{B1}, Alice *always* measures her photon be polarized along θ_{A2} [$P\left(\theta_{A2}\,|\,\theta_{B1}\right)=1$].

---

[3] We'll assume that the number of measurements that Alice and Bob have performed is large, large enough that the statistical errors in the probabilities cannot account for any deviations from the observations described here.

There are four possibilities for the settings of the two PAs, and these observations apply to three of those settings. Let's use these three observations to infer something about the data for the fourth setting. Consider only those trials described by observation 1, where Alice measures photons to be polarized along θ_{A1}, and Bob measures photons to be polarized along θ_{B1}; this occurs 9% of the time. What happens if Bob changes his mind and orients his PA to θ_{B2}? Since Alice measures her photon to be polarized along θ_{A1}, observation 2 tells us that Bob *must* measure his photon to be polarized along θ_{B2}. What if, instead, Bob leaves his PA at θ_{B1}, but Alice changes her mind and orients hers to θ_{A2}? Since this is a trial where Bob will measure his photon be polarized along θ_{B1}, observation 3 says that Alice *must* measure her photon be polarized along θ_{A2}.

If Alice changes her PA to θ_{A2} she will measure her photon to be polarized along that direction. If Bob changes his PA to θ_{B2} he will measure his photon be polarized along that direction. It follows that if they *both* switch their PAs, Alice will measure her photon to be polarized along θ_{A2} and Bob will measure his photon to be polarized along θ_{B2}. Observations 2 and 3 force us to conclude that every time photons would be measured to be polarized along θ_{A1} and θ_{B1}, if the PAs are both switched, the photons must be measured to be polarized along θ_{A2} and θ_{B2}. Measurements of photons polarized along θ_{A2} and θ_{B2} must occur at least as often as measurements of photons polarized along θ_{A1} and θ_{B1}:

$$P\left(\theta_{A2},\theta_{B2}\right) \geq P\left(\theta_{A1},\theta_{B1}\right). \tag{8.34}$$

Since observation 1 tells us that $P\left(\theta_{A1},\theta_{B1}\right) = 0.09$, we conclude that when Alice and Bob examine their data they must find $P\left(\theta_{A2},\theta_{B2}\right) \geq 0.09$.

We now have a prediction about what Alice and Bob should see for measurements in which their PAs are oriented along θ_{A2} and θ_{B2}. What do they in fact observe? Their fourth observation is:

4. On those trials where Alice sets her PA to θ_{A2} and Bob sets his to θ_{B2}, they *never* observe photons to be simultaneously polarized along these two directions [$P\left(\theta_{A2},\theta_{B2}\right) = 0$].

What's wrong? The measured data do not agree with our prediction, so there must be an error somewhere. You can go back and double check, but I assure you that there is no flaw in the logic that leads us to the prediction in eq. (8.34). Furthermore, in complement 8.B we will prove a generalized version of eq. (8.34), under some very reasonable assumptions.

Ah, assumptions! There must be some assumption that we are making about this experiment that is violated. Clearly if we're making a false assumption, even one that we're not explicitly aware of, we could arrive at the wrong prediction. It turns out that there are two assumptions we've been making that are implicit in classical probability theory.

8.4.2 Locality and Reality

First, we've been making the locality assumption. We've been assuming that measurements performed in one location do not influence measurements performed in another. If the photons in one beam can somehow signal the photons in the other that they have been measured to have a particular polarization, it would be possible in the scenario described above to have $P\left(\theta_{A2},\theta_{B2}\right)=0$, despite any reasoned arguments to the contrary.

Another way that nonlocality could influence results would be if the source was somehow affected by the measurements. You can imagine that if the source somehow "knew" what the settings of the PAs were, it could produce photons with polarizations consistent with the measurement results, but not consistent with our inference.

One way to try to ensure the validity of the locality assumption is to place the observers far enough apart so that the measurements are space-like separated. In other words, the measurement apparatuses are far enough apart so that a signal would need to travel at faster than the speed of light in order for it to be possible for one measurement to influence the other. In the experiment just described, this is done by having Alice and Bob very far apart, and having their measurements synchronized to high precision with an atomic clock. In order to assure that the measurements cannot influence the source, it is necessary for Alice and Bob to randomly change the settings of their PAs on a timescale that is short compared to the time it takes the photons travel from the source to their detectors.

If Alice and Bob work very hard to do these things, they will find that their observations are not affected. They would still infer that $P\left(\theta_{A2},\theta_{B2}\right)\geq 0.09$, and measure that $P\left(\theta_{A2},\theta_{B2}\right)=0$. They have eliminated the possibility that this difference is due to communication at the speed of light, but they cannot rule out nonlocal influences which are due to faster-than-light (possibly instantaneous) communication.

While the locality assumption is subtle, the reality assumption is even more so. The reality assumption is that physically measurable quantities have definite values before (and whether or not) they are actually measured. In our example one way for the reality assumption to be satisfied would be if our source emitted photon pairs with polarizations that are determined when they leave the source. The polarizations may be random, and they may be correlated with each other, but once they leave the source the polarizations are fixed. Classical physics is consistent with reality defined in this way.

Note that the reality assumption need *not* mean that the results of measurements are predetermined at the source. For example, a source could emit a photon polarized along +45°. If this photon strikes a $\mathrm{PA}_{\mathrm{HV}}$ it will randomly split between the horizontal and vertical output ports. The measurement result is random, even though the polarization of the photon before the measurement corresponds to an "element of reality," to use the phrase Einstein, Podolsky, and Rosen (EPR) used in the famous paper describing their belief that quantum mechanics was incomplete [8.2].

In summary: Our predictions are not in agreement with the experimental results, and there is no flaw in our logic. The only other possibility is that an assumption we've used in making the predictions must be violated. Our only assumptions are locality and reality, so we must abandon one or the other (or both).[4] The experiment described here offers no clues as to which assumption we must abandon. However, there are other experiments that have been performed in which even certain types of nonlocal communication between detectors is not enough to explain the data [8.3].

8.4.3 The Quantum Mechanical Explanation

Let's see how quantum mechanics can explain our observations, even though a local-realistic model cannot. First, we need to describe how the source of fig. 8.3 is different from that of fig. 8.2. Notice that in fig. 8.3 there is a half-wave plate and a birefringent phase shifter in front of the downconversion crystals. By rotating the half-wave plate, we can change the relative magnitudes of the horizontal and vertical components of the polarization of the pump beam. If the pump beam has a larger horizontal component, the state of the signal and idler photons will have a larger $\left|V,V\right\rangle$ contribution; similarly, with a larger vertical component to the pump beam, the state of the emerging photons will have a larger $\left|H,H\right\rangle$ contribution. The phase shifter in fig. 8.3 is a birefringent plate; tilting it varies the relative phase shift between the horizontal and vertical polarizations of the pump beam. This relative phase shift between the components of the pump polarizations gets mapped into the relative phases of the $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$ contribution to the signal state. The state of the signal photons leaving the source is thus

$$\left|\psi\right\rangle = \sqrt{a}\left|H,H\right\rangle + \sqrt{1-a}\,e^{i\varphi}\left|V,V\right\rangle. \tag{8.35}$$

The parameter a is adjusted by rotating the half-wave plate, while φ is adjusted by tilting the birefringent plate. The source of fig. 8.3 is capable of producing a state consisting of an arbitrary linear combination of $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$.

For Alice and Bob's experiment, the source is adjusted to produce photons in the state

$$\left|\psi_1\right\rangle = \sqrt{0.2}\left|H,H\right\rangle + \sqrt{0.8}\left|V,V\right\rangle. \tag{8.36}$$

In the problems you will verify that with photons in this state, and with proper settings of the polarization analyzers, Alice and Bob will indeed obtain measurement results described by observations 1–4. The measured data are consistent with the quantum mechanical predictions, but not the predictions of local realism.

Quantum mechanics can explain the results of this experiment by violating either the locality assumption, or the reality assumption, or both. In sec. 8.2 we've already stressed that entangled states effectively violate the reality postulate. There we said that for entangled states, the state of the two-particle system is well defined, but it is not

---

[4] In truth, locality and reality are the only *fundamental* assumptions that have been made. In all real experiments there are other assumptions, due to the technical implementation of the experiment. For example, most experiments require the fair sampling assumption, which says that the photons that are actually detected are a fair sample of those that are emitted by the source. In other words, there is no bias in the detectors that is somehow skewing measured statistics.

possible to associate a definite pure state (an element of reality) with each individual particle. However, particles in classical mixed states are consistent with the reality postulate. In such a state, particles are in one state or the other, we just don't know which one. Particles in classical mixed states do not violate tests of local realism.[5]

Another possibility is that nature (and its quantum mechanical description) could violate locality. Indeed, quantum mechanics does. You might think of this in terms of the state collapse associated with the projection postulate. Note that the projection postulate doesn't say anything about the speed at which the state collapse spreads outward from the location of the measurement. Performing a measurement on one particle instantaneously collapses the state of a two-particle system throughout all of space, which effectively changes the state of the other particle. Before Alice's measurement Bob's particle is not in a definite pure state, but after her measurement it is.

Note, however, that this instantaneous collapse of the state does *not* violate causality; Alice and Bob cannot use this state collapse to transmit any information. For example, Alice's measurement of a certain polarization allows Alice to instantaneously determine the measurement result that Bob will obtain, but what good is that? There is no way for her to transmit that information to Bob faster than the speed of light. As far as Bob is concerned, his measurement is random, since he can't know what Alice measured. Furthermore, Alice cannot use the state collapsed to send a signal to Bob. There is no measurement (or other operation) that Alice can perform locally that will change the result of any measurement that Bob might perform. We'll talk more about this in sec. 17.3.

The ideas behind the particular test of local realism that we have just described were originally due to Lucian Hardy. Other descriptions of this test can be found in refs. [8.4] and [8.5]. Note that the observations we've described are idealized, and the inequality of eq. (8.34) is not applicable to real experiments. In complement 8.B we'll derive an experimentally testable inequality that is consistent with local realism, but is violated by quantum mechanics and by nature.

8.4.4 Other Tests of Local Realism

The first person to prove that it is possible to test local realism experimentally was John Bell. In a famous 1964 paper, Bell derived the first "Bell inequality" that must be satisfied by any local-realistic theory, but is violated by quantum mechanics [8.6]. Since that time, others have derived similar inequalities, which are frequently referred to generically as Bell inequalities. In particular, these newer inequalities are actually accessible to experimental test, whereas the original Bell inequality was for an idealized system. In addition to Hardy's test of local realism, you'll also perform one of these tests in lab 5. For a further discussion of Bell inequalities, and experimental tests of those inequalities, I refer you to [8.7], and the references therein.

The first suggestion of a test of local realism that did not involve an inequality was given by Daniel Greenberger, Michael Horne, and Anton Zeilinger (GHZ). For a discussion of the GHZ test refer to [8.8].

---

[5] It is possible for quantum states that have some mixed character (that are not 100% pure) to violate local realism. Indeed, no real experimentally produced state is 100% pure.

8.4.5 One More Thought

In this chapter we've discussed the differences that one would observe between measurements performed using entangled states and measurements performed using mixed states. However, the proof that nature allows entangled states can only come from experiments. The first experiments that showed entanglement actually exists were those performed to test local realism.

We now know that entanglement has a number of interesting and important uses, including quantum teleportation and quantum computing. We'll discuss some of these in chapter 17.

8.5 References

[8.1] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics with Applications, 7th ed.* (Pearson Prentice Hall, Upper Saddle River, NJ, 2004) p. 48.

[8.2] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?," Phys. Rev. **47**, 777 (1935).

[8.3] S. Gröblacher et al., "An experimental test of non-local realism," Nature **446**, 871 (2007).

[8.4] N. D. Mermin, "Quantum mysteries refined," Am. J. Phys. **62**, 880 (1994).

[8.5] P. G. Kwiat and L. Hardy, "The mystery of the quantum cakes," Am. J. Phys. **68**, 33 (2000).

[8.6] J. S. Bell, "On the Einstein–Podolsky–Rosen paradox," Physics (Long Island City, N.Y.) **1**, 195 (1964).

[8.7] D. Dehlinger and M. W. Mitchell, "Entangled photon apparatus for the undergraduate laboratory," Am. J. Phys. **70**, 898 (2002); D. Dehlinger and M. W. Mitchell, "Entangled photons, nonlocality, and Bell inequalities in the undergraduate laboratory," Am. J. Phys. **70**, 903 (2002).

[8.8] N. D. Mermin, "Quantum mysteries revisited," Am. J. Phys. **58**, 731 (1990).

8.6 PROBLEMS

8.1 Calculate the expectation values of $\wp^{s}_{HV}$, $\wp^{i}_{HV}$, and $\wp^{si}_{HV}$ for photons prepared in the state $\left|V,L\right\rangle$.

8.2 Calculate the expectation values of $\wp^{s}_{HV}$, $\wp^{i}_{HV}$, and $\wp^{si}_{HV}$ for photons prepared in the state $\left|\psi\right\rangle = \sqrt{1/3}\left|H,+45\right\rangle + \sqrt{2/3}\left|H,-45\right\rangle$.

8.3 Calculate the probability that both the signal and the idler photons will be measured to have +45° polarization, for photons prepared in the state $\left|\psi\right\rangle = \sqrt{1/3}\left|H,+45\right\rangle + \sqrt{2/3}\left|H,-45\right\rangle$.

8.4 Calculate the probability that the signal and idler photons will be measured to have orthogonal polarizations, for photons prepared in the state $\left|R,+45\right\rangle$. Both photons are measured in the horizontal/vertical basis.

8.5* Prove that the entangled state $\left|\phi\right\rangle = a\left|H,H\right\rangle + b\left|V,V\right\rangle$, where a and b are arbitrary, nonzero constants, cannot be factorized into a product state (i.e. $\left|\phi\right\rangle \neq \left|\psi_s\right\rangle_s \left|\psi_i\right\rangle_i$). (Hint: Start by assuming that $\left|\phi\right\rangle$ can be written as a product state.)

8.6* The state $\left|\phi^{+}\right\rangle$ produced in the downconversion source of fig. 8.2 is one of the four Bell states, which are given by

$$\left|\phi^{\pm}\right\rangle=\frac{1}{\sqrt{2}}\left(|H,H\rangle\pm|V,V\rangle\right). \tag{8.37}$$

$$\left|\psi^{\pm}\right\rangle=\frac{1}{\sqrt{2}}\left(|H,V\rangle\pm|V,H\rangle\right). \tag{8.38}$$

Prove that the Bell states form a basis in the Hilbert space describing the polarizations of two photons.

8.7 For a two-photon system prepared in the state $|\psi\rangle=\left(1/\sqrt{2}\right)\left(|+45,+45\rangle+|-45,-45\rangle\right)$, determine the probabilities of obtaining the following measurements:

(a) the signal photon is polarized at +45°;

(b) the signal photon is polarized at –45°;

(c) the signal photon is polarized at +45°, given that the idler photon is polarized at +45°.

8.8 For a two-photon system prepared in the state $|\psi\rangle=\left(1/\sqrt{2}\right)\left(|+45,+45\rangle+|-45,-45\rangle\right)$, determine the probabilities of obtaining the following measurements:

(a) the signal photon is horizontally polarized;

(b) the signal photon is vertically polarized;

(c) the signal photon is horizontally polarized, given that the idler photon is horizontally polarized.

8.9* The probability of measuring a signal photon to have an elliptical polarization corresponding to the state $\left|e_{s}\right\rangle=\cos\left(\theta_{s}\right)|H\rangle+e^{i\varphi_{s}}\sin\left(\theta_{s}\right)|V\rangle$ is $P\left(e_{s}\right)$. The probability of measuring an idler photon to have an elliptical polarization corresponding to the state $\left|e_{i}\right\rangle=\cos\left(\theta_{i}\right)|H\rangle+e^{i\varphi_{i}}\sin\left(\theta_{i}\right)|V\rangle$ is $P\left(e_{i}\right)$. A suitable parametric downconversion source is used to create light in the state $\left|\phi^{+}\right\rangle$ of eq. (8.19).

(a) Calculate the joint measurement probability $P\left(e_{s},e_{i}\right)$ for photons prepared in state $\left|\phi^{+}\right\rangle$.

(b) Calculate the probability $P\left(e_{i}\right)$.

(c) Calculate the conditional probability $P\left(e_{s}\mid e_{i}\right)$.

(d) What is the probability $P\left(e_{s}\mid e_{i}\right)$ for the special case of $\theta_{s}=\theta_{i}$ and $\varphi_{s}=-\varphi_{i}$ (i.e., the polarizations of the two beams are complex conjugates of each other)? (You should find that the polarizations are perfectly correlated.)

(e) What does this say about the special case of linear polarizations ($\varphi_{s}=\varphi_{i}=0$)? Do you find anything interesting about this result?

8.10* Show that the probabilities calculated in example 8.5 are unchanged if the state $\left|\phi^{+}\right\rangle$ is replaced by a mixed state containing 50% $\left|H,H\right\rangle$ and 50% $\left|V,V\right\rangle$.

8.11 Here I'd like you to use your intuition to answer these questions, without resorting to calculations. Of course, you are free to perform calculations to confirm your intuition. When you're done, compare the results you get here to those of example 8.7.

A source produces signal and idler photons that both pass through linear polarizers oriented at +45°. If this source produces photons in the state $\left|H,H\right\rangle$:

(a) What is the probability of detecting a signal photon?
(b) What is the probability of detecting an idler photon?
(c) What is the probability of detecting both signal and idler photons?
(d) What is the probability of detecting the signal photon, given that an idler photon is detected?

If this source produces photons in the state $\left|V,V\right\rangle$:

(e) What is the probability of detecting a signal photon?
(f) What is the probability of detecting an idler photon?
(g) What is the probability of detecting both signal and idler photons?
(h) What is the probability of detecting the signal photon, given that an idler photon is detected?

If this source randomly produces photons in a mixture of the states $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$:

(i) What is the probability of detecting a signal photon?
(j) What is the probability of detecting an idler photon?
(k) What is the probability of detecting both signal and idler photons?
(l) What is the probability of detecting the signal photon, given that an idler photon is detected?

8.12 For a system prepared in an equal mixture of the states $\left|+45,+45\right\rangle$ and $\left|-45,-45\right\rangle$, determine the probabilities of obtaining the following measurements:

(a) the signal photon is polarized at +45°,
(b) the signal photon is polarized at –45°,
(c) the signal photon is polarized at +45°, given that the idler photon is polarized at +45°. Compare your results to those of problem 8.7.

8.13 For a system prepared in an equal mixture of the states $\left|+45,+45\right\rangle$ and $\left|-45,-45\right\rangle$, determine the probabilities of obtaining the following measurements:

(a) the signal photon is horizontally polarized;
(b) the signal photon is vertically polarized;
(c) the signal photon is horizontally polarized, given that the idler photon is horizontally polarized. Compare your results to those of problem 8.8.

8.14* A suitable parametric downconversion source is used to create light in the state $\left|\psi_1\right\rangle$ of eq. (8.36), where signal and idler photons travel to Alice and Bob, respectively.

(a) Calculate the joint probability $P(\theta_A, \theta_B)$ that Alice measures her photon to be linearly polarized along the angle θ_A, and Bob measures his photon to be linearly polarized along the angle θ_B.

(b) For the angles $\theta_{A1} = 19^\circ, \theta_{A2} = -35^\circ, \theta_{B1} = -19^\circ$, and $\theta_{B2} = 35^\circ$, calculate the four probabilities $P(\theta_{A1}, \theta_{B1})$, $P(\theta_{B2} \mid \theta_{A1})$, $P(\theta_{A2} \mid \theta_{B1})$ and $P(\theta_{A2}, \theta_{B2})$.

(c) Compare your results to see if they are consistent with observations 1–4 in sec. 8.4. Are your results consistent with local realism?

8.15* A suitable parametric downconversion source is used to create light in the state $|\psi_1\rangle$ of eq. (8.36), where signal and idler photons travel to Alice and Bob, respectively.

(a) Calculate the probability $P(\theta_A)$ that Alice measures her photon to be linearly polarized along the angle θ_A, and the probability $P(\theta_B)$ that Bob measures his photon to be linearly polarized along the angle θ_B.

(b) Assume that Alice and Bob do not communicate with each other, or with the source. Can Alice use her *measurement* to send a signal to Bob? Can she perform a measurement that will affect the probabilities of any measurement that Bob might perform?

COMPLEMENT 8.A

The Density Operator

In this complement we'll describe the density operator. The density operator is a way to represent a general state in quantum mechanics, as it can be used to represent both pure and mixed states.

8.A.1 Definition and Properties

The density operator $\hat{\rho}$ corresponding to an arbitrary pure state $|\psi\rangle$ is

$$\hat{\rho} = |\psi\rangle\langle\psi|. \tag{8.A.1}$$

For a mixture of states $|\psi_j\rangle$, each prepared with probability $p_j = P(|\psi_j\rangle)$, the density operator is given by

$$\hat{\rho} = \sum_j p_j |\psi_j\rangle\langle\psi_j|. \tag{8.A.2}$$

It can be readily seen from eq. (8.A.2) that the density operator is Hermitian: $\hat{\rho}^\dagger = \hat{\rho}$.

Since p_j is the probability of preparing the system in the state $|\psi_j\rangle$, it must be a real number. It must also satisfy

$$0 \leq p_j \leq 1. \tag{8.A.3}$$

Probabilities must be normalized, so it is also necessary that

$$\sum_j p_j = 1. \tag{8.A.4}$$

We will assume that the states $|\psi_j\rangle$ are normalized, but otherwise make no more assumptions about them. They do not need to form a basis; they don't even need to be orthogonal. They are simply a set of states that the system can be prepared in with some probability.

It is important to note that the mixed state density operator in eq. (8.A.2) is *not* the density operator corresponding to the pure state

$$|\psi\rangle = \sum_j \sqrt{p_j}\, |\psi_j\rangle. \tag{8.A.5}$$

The density operator corresponding to this pure state is given by eq. (8.A.1):

$$\begin{aligned}\hat{\rho} &= |\psi\rangle\langle\psi| \\ &= \left(\sum_j \sqrt{p_j}\,|\psi_j\rangle\right)\left(\sum_k \sqrt{p_k}\,\langle\psi_k|\right) \\ &= \sum_{j,k} \sqrt{p_j p_k}\,|\psi_j\rangle\langle\psi_k| .\end{aligned} \tag{8.A.6}$$

8.A.2 The Trace

The trace of a matrix is equal to the sum of its diagonal elements. It is therefore natural to define the trace of an operator as being equal to the sum of the diagonal elements of its matrix representation. For example, the trace of $\hat{A}$ is

$$\begin{aligned}\mathrm{Tr}\left(\hat{A}\right) &= \sum_n A_{nn} \\ &= \sum_n \langle\phi_n|\hat{A}|\phi_n\rangle ,\end{aligned} \tag{8.A.7}$$

where the states $|\phi_n\rangle$ form an orthonormal basis. The trace is independent of the chosen basis, and is linear:

$$\mathrm{Tr}\left(a\hat{A}+b\hat{B}\right) = a\mathrm{Tr}\left(\hat{A}\right)+b\mathrm{Tr}\left(\hat{B}\right). \tag{8.A.8}$$

Using eq. (8.A.7), we can see that

$$\begin{aligned}\mathrm{Tr}\left(|a\rangle\langle b|\right) &= \sum_n \langle\phi_n|a\rangle\langle b|\phi_n\rangle \\ &= \sum_n \langle b|\phi_n\rangle\langle\phi_n|a\rangle \\ &= \langle b|\left(\sum_n |\phi_n\rangle\langle\phi_n|\right)|a\rangle \\ &= \langle b|a\rangle ,\end{aligned} \tag{8.A.9}$$

where we've used the completeness of the basis states.

The trace of the density operator is given by

$$\begin{aligned}\operatorname{Tr}(\hat{\rho}) &= \operatorname{Tr}\left(\sum_j p_j \left|\psi_j\right\rangle\left\langle\psi_j\right|\right)\\ &= \sum_j p_j \operatorname{Tr}\left(\left|\psi_j\right\rangle\left\langle\psi_j\right|\right)\\ &= \sum_j p_j \left\langle\psi_j\middle|\psi_j\right\rangle\\ &= \sum_j p_j\\ &= 1,\end{aligned} \tag{8.A.10}$$

where we've used eqs. (8.A.2), (8.A.8), (8.A.9), and (8.A.4). $\operatorname{Tr}(\hat{\rho}) = 1$ is the normalization condition for a density operator. Like any other operator in a discrete basis, the density operator can be represented as a matrix, referred to as the density matrix. The elements of the density matrix are

$$\rho_{mn} = \left\langle\phi_m\middle|\hat{\rho}\middle|\phi_n\right\rangle. \tag{8.A.11}$$

In terms of these matrix elements, the normalization condition for the density matrix is

$$\operatorname{Tr}(\hat{\rho}) = \sum_n \rho_{nn} = 1. \tag{8.A.12}$$

For a pure state the density operator is given by $\hat{\rho} = |\psi\rangle\langle\psi|$, so clearly $\hat{\rho}^2 = \hat{\rho}$. Thus, for pure states $\operatorname{Tr}(\hat{\rho}^2)=\operatorname{Tr}(\hat{\rho}) = 1$. The density operator for a mixed state must be normalized, so $\operatorname{Tr}(\hat{\rho}) = 1$. However, it can be shown that for mixed states $\hat{\rho}^2 \neq \hat{\rho}$ and $\operatorname{Tr}(\hat{\rho}^2) < 1$. For this reason $\operatorname{Tr}(\hat{\rho}^2)$ can be taken as a measure of the "purity" of a state. The closer that $\operatorname{Tr}(\hat{\rho}^2)$ is to 1, the more pure it is; only pure states achieve $\operatorname{Tr}(\hat{\rho}^2) = 1$.

8.A.3 Expectation Values and Probabilities

Suppose that we wish to determine the expectation value of observable A, for a system prepared in a mixed state. When the system is prepared in state $\left|\psi_j\right\rangle$, the expectation value is

$$\left\langle\hat{A}\right\rangle_j = \left\langle\psi_j\middle|\hat{A}\middle|\psi_j\right\rangle \tag{8.A.13}$$

When calculating the expectation value for the mixed state, we must weight $\left\langle \hat{A} \right\rangle_j$ by the probability p_j that the system is prepared in state $\left|\psi_j\right\rangle$, and then sum over all the possible state preparations. In other words

$$\begin{aligned}
\left\langle \hat{A} \right\rangle &= \sum_j \left\langle \hat{A} \right\rangle_j p_j \\
&= \sum_j p_j \left\langle \psi_j \middle| \hat{A} \middle| \psi_j \right\rangle \\
&= \sum_j p_j \mathrm{Tr}\left(\hat{A} \left|\psi_j\right\rangle \left\langle \psi_j \right| \right) \\
&= \mathrm{Tr}\left(\hat{A} \sum_j p_j \left|\psi_j\right\rangle \left\langle \psi_j \right| \right) \\
&= \mathrm{Tr}\left(\hat{A}\hat{\rho} \right).
\end{aligned} \tag{8.A.14}$$

Since the probability of obtaining a particular eigenvalue as a measurement result can be written as the expectation value of the projection operator onto its corresponding eigenstate, this expression can also be used to calculate probabilities of measurements.

EXAMPLE 8.A.1

Compare the density operators that correspond to the following two states: (a) a superposition that consists of equal parts $|H,H\rangle$ and $|V,V\rangle$ (assuming a relative phase of zero), and (b) a mixture that consists of equal parts $|H,H\rangle$ and $|V,V\rangle$.

(a) The state vector corresponding to this pure state is

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|H,H\rangle + |V,V\rangle\right). \tag{8.A.15}$$

Its corresponding density operator is

$$\begin{aligned}
\hat{\rho} &= |\psi\rangle\langle\psi| \\
&= \frac{1}{\sqrt{2}}\left(|H,H\rangle + |V,V\rangle\right)\frac{1}{\sqrt{2}}\left(\langle H,H| + \langle V,V|\right) \\
&= \frac{1}{2}|H,H\rangle\langle H,H| + \frac{1}{2}|H,H\rangle\langle V,V| + \frac{1}{2}|V,V\rangle\langle H,H| + \frac{1}{2}|V,V\rangle\langle V,V|.
\end{aligned} \tag{8.A.16}$$

(b) The density operator corresponding to this mixed state is

$$\hat{\rho} = \frac{1}{2}\left|H,H\right\rangle\left\langle H,H\right| + \frac{1}{2}\left|V,V\right\rangle\left\langle V,V\right| \tag{8.A.17}$$

Clearly these are different; the density operator corresponding to the pure state has two more terms. These extra terms, which intermingle the $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$ contributions, contain information about the entanglement between the states.

EXAMPLE 8.A.2

For a two-photon system prepared in an equal mixture of states $\left|H,H\right\rangle$ and $\left|V,V\right\rangle$, determine the probability that the signal photon is measured to be polarized along +45°, given that the idler photon is found to be polarized along this same direction.

The density operator corresponding to this state is given by eq. (8.A.17). We want to find $P\left(+45_s \mid +45_i\right)$, which is

$$P\left(+45_s \mid +45_i\right) = \frac{P\left(+45_s, +45_i\right)}{P\left(+45_i\right)}. \tag{8.A.18}$$

Using eq. (8.A.14) we find the numerator of this expression to be

$$\begin{aligned}
P\left(+45_s, +45_i\right) &= \left\langle \hat{P}_{+45_s,+45_i} \right\rangle \\
&= \mathrm{Tr}\left[\hat{P}_{+45_s,+45_i}\hat{\rho}\right] \\
&= \mathrm{Tr}\left[\left(\left|+45,+45\right\rangle\left\langle +45,+45\right|\right)\left(\frac{1}{2}\left|H,H\right\rangle\left\langle H,H\right|\right)\right] \\
&\quad +\mathrm{Tr}\left[\left(\left|+45,+45\right\rangle\left\langle +45,+45\right|\right)\left(\frac{1}{2}\left|V,V\right\rangle\left\langle V,V\right|\right)\right] \\
&= \mathrm{Tr}\left[\left(\frac{1}{2}\right)\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\left|+45,+45\right\rangle\left\langle H,H\right|\right] \\
&\quad +\mathrm{Tr}\left[\left(\frac{1}{2}\right)\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\left|+45,+45\right\rangle\left\langle V,V\right|\right] \\
&= \frac{1}{4}\left(\left\langle H,H \middle| +45,+45\right\rangle + \left\langle V,V \middle| +45,+45\right\rangle\right) \\
&= \frac{1}{4}\left[\left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}\right)\right] \\
&= \frac{1}{4}.
\end{aligned} \tag{8.A.19}$$

Applying eq. (8.A.14) to the denominator of eq. (8.A.18) we find

$$
\begin{aligned}
P(+45_i) &= \left\langle \hat{P}_{+45_i} \right\rangle \\
&= \mathrm{Tr}\left[\hat{P}_{+45_i} \hat{\rho} \right] \\
&= \mathrm{Tr}\left[\left(|+45\rangle_{i\ i}\langle +45| \right) \left(\frac{1}{2} |H,H\rangle\langle H,H| \right) \right] \\
&\quad +\mathrm{Tr}\left[\left(|+45\rangle_{i\ i}\langle +45| \right) \left(\frac{1}{2} |V,V\rangle\langle V,V| \right) \right] \\
&= \frac{1}{2}\, {}_i\langle H|+45\rangle_{i\ i}\langle +45|H\rangle_{i\ s}\langle H|H\rangle_s \\
&\quad +\frac{1}{2}\, {}_i\langle V|+45\rangle_{i\ i}\langle +45|V\rangle_{i\ s}\langle V|V\rangle_s \\
&= \frac{1}{2}\left[\left(\frac{1}{\sqrt{2}} \right)\left(\frac{1}{\sqrt{2}} \right) + \left(\frac{1}{\sqrt{2}} \right)\left(\frac{1}{\sqrt{2}} \right) \right] \\
&= \frac{1}{2}.
\end{aligned}
\tag{8.A.20}
$$

Substituting back into eq. (8.A.18) yields the final result

$$
P(+45_s \mid +45_i) = \frac{1/4}{1/2} = \frac{1}{2}. \tag{8.A.21}
$$

This agrees with the results of example 8.7.

8.A.4 Schrödinger's Cat

As an example of the usefulness of the density matrix, consider Schrödinger's cat paradox. Schrödinger imagined placing a cat in a box with some radioactive material. After one hour there is 50% probability that one of the radioactive atoms decays, and this decay triggers a mechanism which kills the cat. Before the box is opened, and a measurement is performed to determine whether the cat is alive or dead, the state of the atom-cat system should thus be

$$
|\psi\rangle = \frac{1}{\sqrt{2}}\left(|no\ decay, alive\rangle + |decay, dead\rangle \right). \tag{8.A.22}
$$

which is an entangled state. Of course, from our discussions above this would imply that the cat is simultaneously alive and dead, which is preposterous. The alternative is that the system is described by the mixed-state density operator

$$
\hat{\rho}_{mix} = \frac{1}{2}|no\ decay, alive\rangle\langle no\ decay, alive| + \frac{1}{2}|decay, dead\rangle\langle decay, dead|. \tag{8.A.23}
$$

This means that the cat is either alive or dead, which is a much more reasonable proposition.

Even if, immediately after the atomic decay, the system is created in the entangled state $|\psi\rangle$, it will very rapidly evolve into the mixed state $\hat{\rho}_{mix}$, so we never actually see the atom-cat system in the entangled state. The reason for this is known as decoherence. In order for the system to remain in the entangled state, the phase relationship between the states $|no\ decay, alive\rangle$ and $|decay, dead\rangle$ must remain stable—these states must be coherent with each other (similar to the coherence described in complement 2.A). To see this, rewrite eq. (8.A.22) as

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|no\ decay, alive\rangle + e^{i\phi}|decay, dead\rangle\right). \tag{8.A.24}$$

which makes this phase relationship more apparent. The density operator corresponding to this state is

$$\begin{aligned} \hat{\rho}_{entangle} &= |\psi\rangle\langle\psi| \\ &= \frac{1}{2}|no\ decay, alive\rangle\langle no\ decay, alive| + \frac{1}{2}e^{-i\phi}|no\ decay, alive\rangle\langle decay, dead| \\ &\quad + \frac{1}{2}e^{i\phi}|decay, dead\rangle\langle no\ decay, alive| + \frac{1}{2}|decay, dead\rangle\langle decay, dead|. \end{aligned} \tag{8.A.25}$$

If ϕ fluctuates, the exponentials in this expression average to zero, and the density operator reduces to that of eq. (8.A.23).

With care microscopic systems can be placed in entangled states, but the cat is a macroscopic system containing of the order 10^{23} atoms, all subject to thermal excitations which cause phase fluctuations. For all intents and purposes the coherence of the atom-cat system decays instantaneously, leaving it in the mixed state $\hat{\rho}_{mix}$. That said, and despite the experimental difficulty in entangling macroscopic objects, entanglement involving $\sim 10^{16}$ atoms in two diamond crystals at room temperature has recently been observed [8.A.1]. This macroscopic entanglement only persists on picosecond timescales, however.

8.A.5 References

[8.A.1] K. C. Lee et al., "Entangling macroscopic diamonds at room temperature," Science **334**, 1253 (2011).

8.A.6 PROBLEMS

8.A.1 For a system prepared in an equal mixture of the states $|+45,+45\rangle$ and $|-45,-45\rangle$, use the density operator to determine the probabilities of obtaining the following measurements:
(a) the signal photon is horizontally polarized;
(b) the signal photon is vertically polarized;
(c) the signal photon is horizontally polarized, given that the idler photon is horizontally polarized. Compare your results to those of problem 8.13.

8.A.2 Write down the density matrix that corresponds to the mixed state that is one third $|\phi^+\rangle = (1/\sqrt{2})(|H,H\rangle + |V,V\rangle)$, and two thirds $|V,V\rangle$. Verify that $\text{Tr}(\hat{\rho})$ and $\text{Tr}(\hat{\rho}^2)$ yield what you would expect for this mixture.

COMPLEMENT 8.B

The Bell-Clauser-Horne Inequality

In this complement we'll prove an experimentally testable inequality that generalizes eq. (8.34), and that must be satisfied by all local, realistic theories. This is the inequality that will be tested in lab 5. Throughout this discussion we will be assuming that Alice and Bob are performing an experiment like that depicted in fig. 8.3.

8.B.1 Reality

We are interested in probabilities such as $P(\theta_A,\theta_B)$, which is the joint probability that Alice measures her photon be polarized along θ_A, and Bob measures his photon to be polarized along θ_B. In quantum mechanics such probabilities depend on the state $|\psi\rangle$ of the signal and idler photons pairs. In classical physics we will assume that probabilities of this type depend on a variable we will call λ, which describes the source. If we know λ, we know what polarizations the signal and idler photons have, and hence we know what the joint probability of the measurements will be. In this discussion we'll be assuming that λ is a discrete random variable.[6] The probability that Alice measures θ_A, and Bob measures θ_B, and the source is described by λ is $P(\theta_A,\theta_B,\lambda)$. The probability that Alice measures θ_A, and Bob measures θ_B, given that the source is described by λ is $P(\theta_A,\theta_B\,|\,\lambda)$.

If we know λ, we know what the joint probability of measurements will be. However, for any given measurement we don't know λ, because we don't have direct access to it; for this reason λ is often referred to as a "hidden variable." Therefore, we must sum over the possible values of λ, which occur with probability $P(\lambda)$. This means

$$\begin{aligned} P(\theta_A,\theta_B) &= \sum_{\lambda} P(\theta_A,\theta_B,\lambda) \\ &= \sum_{\lambda} P(\theta_A,\theta_B\,|\,\lambda)P(\lambda). \end{aligned} \tag{8.B.1}$$

---

[6] It is straightforward to generalize the arguments presented here to allow λ to be a continuous random variable.

This can also be written as:

$$\begin{aligned} P(\theta_A,\theta_B) &= \sum_\lambda P(\theta_A,\theta_B,\lambda) \\ &= \sum_\lambda P(\theta_A \mid \theta_B,\lambda)P(\theta_B,\lambda) \\ &= \sum_\lambda P(\theta_A \mid \theta_B,\lambda)P(\theta_B \mid \lambda)P(\lambda). \end{aligned} \tag{8.B.2}$$

In writing eq. (8.B.2) we have used the reality assumption. This assumption has two parts. The first is that it is possible to describe the source using the hidden variable λ (i.e., that λ exists). The second is that all of the probability distributions we'll be talking about have the properties of classical probability distributions; they are real, take on values between 0 and 1, and are normalized.

So you see, it is possible to define reality mathematically! The reality assumption is consistent with classical physics.

8.B.2 Locality

The locality assumption states that the results of measurements obtained by Alice are independent of those obtained by Bob, and vice versa. This assumption can be enforced in an experiment by ensuring that Alice's and Bob's measurements are space-like separated. Locality manifests itself mathematically as follows. Consider the probability $P(\theta_A \mid \theta_B,\lambda)$ that appears in the last line of eq. (8.B.2), which is the probability that Alice measures her photon be polarized along θ_A, given that Bob measures his photon to be polarized along θ_B, and that the source is described by λ. Under the locality assumption, the probability that Alice measures a certain polarization cannot depend on what Bob measures, so we must have

$$P(\theta_A \mid \theta_B,\lambda) = P(\theta_A \mid \lambda). \tag{8.B.3}$$

In other words, a measurement cannot be conditioned on something that it is independent of. The correlations between Alice's and Bob's measurements are maintained because both measurements depend on λ. Rewriting eq. (8.B.2) using eq. (8.B.3) yields

$$P(\theta_A,\theta_B) = \sum_\lambda P(\theta_A \mid \lambda)P(\theta_B \mid \lambda)P(\lambda). \tag{8.B.4}$$

In sec. 8.4 we also stated that Alice and Bob want to ensure that their measurements cannot affect the source. This assumption means that λ is independent of θ_A and θ_B.

8.B.3 The Inequality

The source produces photons described by λ, and Alice sets her polarization analyzer along the angle θ_A. There are two possible outcomes for Alice's measurement: the photon is parallel to θ_A, or the photon is perpendicular to θ_A. Since all probability distributions need to be normalized, the sum of the probabilities of these outcomes must be 1:

$$P(\theta_A\,|\,\lambda)+P\left(\theta_A^{\perp}\,|\,\lambda\right)=1. \tag{8.B.5}$$

This same condition applies to Bob's measurements.

The experimenters each perform measurements at two different settings of their polarization analyzers: Alice at θ_{A1} and θ_{A2}, Bob at θ_{B1} and θ_{B2}. We can use eq. (8.B.4) to express the probability $P(\theta_{A1},\theta_{B1})$ that Alice measures her photon be polarized along θ_{A1} and Bob measures his to be polarized along θ_{B1}. We can also apply the normalization condition of eq. (8.B.5) to insert some factors of 1, and thus expand this expression as

$$\begin{aligned}
P(\theta_{A1},\theta_{B1}) &= \sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P(\lambda)\\
&= \sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P_B(\theta_{B1}\,|\,\lambda)\left[P\left(\theta_{A2}\,|\,\lambda\right)+P\left(\theta_{A2}^{\perp}\,|\,\lambda\right)\right]P(\lambda)\\
&= \sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}\,|\,\lambda\right)P(\lambda)\\
&\quad+\sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}^{\perp}\,|\,\lambda\right)P(\lambda)\\
&= \sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}\,|\,\lambda\right)\left[P\left(\theta_{B2}\,|\,\lambda\right)+P\left(\theta_{B2}^{\perp}\,|\,\lambda\right)\right]P(\lambda)\\
&\quad+\sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}^{\perp}\,|\,\lambda\right)P(\lambda)\\
&= \sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}\,|\,\lambda\right)P\left(\theta_{B2}\,|\,\lambda\right)P(\lambda)\\
&\quad+\sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}\,|\,\lambda\right)P\left(\theta_{B2}^{\perp}\,|\,\lambda\right)P(\lambda)\\
&\quad+\sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}^{\perp}\,|\,\lambda\right)P(\lambda).
\end{aligned} \tag{8.B.6}$$

Let's look at the last three sums in this expression separately, keeping in mind that each of the probabilities is real, positive, and less than 1. The first of these says

$$\begin{aligned}
\sum_{\lambda} P(\theta_{A1}\,|\,\lambda)P(\theta_{B1}\,|\,\lambda)P\left(\theta_{A2}\,|\,\lambda\right)P\left(\theta_{B2}\,|\,\lambda\right)P(\lambda) &\le \sum_{\lambda} P\left(\theta_{A2}\,|\,\lambda\right)P\left(\theta_{B2}\,|\,\lambda\right)P(\lambda)\\
&= P\left(\theta_{A2},\theta_{B2}\right),
\end{aligned} \tag{8.B.7}$$

where once again we've used eq. (8.B.4). The penultimate sum in eq. (8.B.6) states

$$\sum_{\lambda} P(\theta_{A1}|\lambda)P(\theta_{B1}|\lambda)P(\theta_{A2}|\lambda)P(\theta_{B2}^{\perp}|\lambda)P(\lambda) \le \sum_{\lambda} P(\theta_{A1}|\lambda)P(\theta_{B2}^{\perp}|\lambda)P(\lambda)$$
$$= P(\theta_{A1},\theta_{B2}^{\perp}). \tag{8.B.8}$$

The last sum in eq. (8.B.6) says that

$$\begin{aligned}\sum_{\lambda} P(\theta_{A1}|\lambda)P(\theta_{B1}|\lambda)P(\theta_{A2}^{\perp}|\lambda)P(\lambda) &\le \sum_{\lambda} P(\theta_{B1}|\lambda)P(\theta_{A2}^{\perp}|\lambda)P(\lambda)\\ &= P(\theta_{A2}^{\perp},\theta_{B1}).\end{aligned} \tag{8.B.9}$$

Combining eqs. (8.B.6)–(8.B.9) yields the inequality

$$P(\theta_{A1},\theta_{B1}) \le P(\theta_{A2},\theta_{B2}) + P(\theta_{A1},\theta_{B2}^{\perp}) + P(\theta_{A2}^{\perp},\theta_{B1}). \tag{8.B.10}$$

This is a form of the Bell-Clauser-Horne inequality, which we have applied to the experimental configuration of fig. 8.3, and which must be satisfied by any local, realistic theory [8.B.1].

8.B.4 Discussion

Now we need to show that the inequality of eq. (8.B.10) applies to the discussion of sec. 8.4. Let's begin by examining observation 2, which says that

$$P(\theta_{B2}|\theta_{A1}) = 1. \tag{8.B.11}$$

For the polarizer settings applicable to observation 2, the only possibilities for Bob's measurements are θ_{B2} and $\theta_{B2}^{\perp}$, so

$$\begin{aligned}P(\theta_{B2}|\theta_{A1}) + P(\theta_{B2}^{\perp}|\theta_{A1}) &= 1\\ P(\theta_{B2}^{\perp}|\theta_{A1}) &= 1 - P(\theta_{B2}|\theta_{A1})\\ &= 1 - 1 = 0,\end{aligned} \tag{8.B.12}$$

where we've used eq. (8.B.11). We can use this result to see that

$$P(\theta_{A1},\theta_{B2}^{\perp}) = P(\theta_{B2}^{\perp}|\theta_{A1})P(\theta_{A1}) = 0. \tag{8.B.13}$$

Thus, eqs. (8.B.11) and (8.B.13) are equivalent, which makes intuitive sense. If Bob always measures photon polarized along θ_{B2} when Alice measures hers to be polarized along θ_{A1}, then it must be true that Bob and Alice never measure their photons to be polarized along θ_{A1} and $\theta_{B2}^{\perp}$. Similar reasoning applied to observation 3 [$P(\theta_{A2}|\theta_{B1}) = 1$] tells us that Alice and Bob find

$$P\left(\theta_{A2}^{\perp},\theta_{B1}\right)=0. \tag{8.B.14}$$

Inserting eqs. (8.B.13) and (8.B.14) into eq. (8.B.10) yields

$$P(\theta_{A1},\theta_{B1})\leq P\left(\theta_{A2},\theta_{B2}\right), \tag{8.B.15}$$

which is the same as eq. (8.34) in sec. 8.4. Thus, the Bell-Clause-Horne inequality of eq. (8.B.10) applies to the discussion of Alice and Bob's experiment in that section.

It's important to understand why we said in sec 8.4 that eq. (8.B.15) [which is the same as eq. (8.34)] is not applicable to real experiments, necessitating the derivation of eq. (8.B.10). The only assumptions that go into eq. (8.B.10) are the locality and reality assumptions. Equation (8.B.15) requires two more assumptions: $P\left(\theta_{A2}^{\perp},\theta_{B1}\right)=0$ and $P\left(\theta_{A1},\theta_{B2}^{\perp}\right)=0$. These assumptions were satisfied in our discussion of Alice and Bob's experiment, so there we could get away with using eq. (8.B.15).

However, in real experiments we can *never* say that a measured probability is equal to 0; the best we can do for a probability is set an upper bound. For example, if you perform 1 million measurements and never observe a particular event, you can't be certain that the probability is 0. The best you can determine is that the probability of your event is less than 1/1,000,000. Furthermore, real experiments have background noise. This means that you will almost always see something, even if you expect to see nothing. In a well performed experiment $P\left(\theta_{A2}^{\perp},\theta_{B1}\right)$ and $P\left(\theta_{A1},\theta_{B2}^{\perp}\right)$ will be small, but they'll never be 0. Since this is the case, the full inequality of eq. (8.B.10) must be used in any experimental test.

8.B.5 References

[8.B.1] N. D. Mermin, "Quantum mysteries refined," Am. J. Phys. **62**, 880 (1994).

8.B.6 PROBLEMS

8.B.1* A suitable parametric downconversion source is used to create light in the state

$$\left|\psi_1\right\rangle=\sqrt{0.2}\left|H,H\right\rangle+\sqrt{0.8}\left|V,V\right\rangle, \tag{8.B.16}$$

where signal and idler photons travel to Alice and Bob, respectively.

(a) Calculate the joint probability $P(\theta_A,\theta_B)$ that Alice measures her photon to be linearly polarized along the angle θ_A, and Bob measures his photon to be linearly polarized along the angle θ_B.

(b) Given the angles $\theta_{A1} = 19^{\text{o}}$, $\theta_{A2} = -35^{\text{o}}$, $\theta_{B1} = -19^{\text{o}}$, and $\theta_{B2} = 35^{\text{o}}$, calculate the four probabilities $P\left(\theta_{A1},\theta_{B1}\right)$, $P\left(\theta_{A2},\theta_{B2}\right)$ $P\left(\theta_{A1},\theta_{B2}^{\perp}\right)$ and $P\left(\theta_{A2}^{\perp},\theta_{B1}\right)$.

(c) Check whether these probabilities are consistent with the Bell-Clauser-Horne inequality. Are your results consistent with local realism?

8.B.2 A suitable parametric downconversion source creates photon pairs in a mixed state that is 20% $\left|H,H\right\rangle$ and 80% $\left|V,V\right\rangle$. The signal and idler photons travel to Alice and Bob, respectively.

(a) Calculate the joint probability $P\left(\theta_A,\theta_B\right)$ that Alice measures her photon to be linearly polarized along the angle θ_A, and Bob measures his photon to be linearly polarized along the angle θ_B.

(b) Given the angles $\theta_{A1} = 19^{\text{o}}$, $\theta_{A2} = -35^{\text{o}}$, $\theta_{B1} = -19^{\text{o}}$, and $\theta_{B2} = 35^{\text{o}}$, calculate the four probabilities $P\left(\theta_{A1},\theta_{B1}\right)$, $P\left(\theta_{A2},\theta_{B2}\right)$, $P\left(\theta_{A1},\theta_{B2}^{\perp}\right)$, and $P\left(\theta_{A2}^{\perp},\theta_{B1}\right)$.

(c) Check whether these probabilities are consistent with the Bell-Clauser-Horne inequality. Are your results consistent with local realism?

Two Spin-1/2 Particles

In this complement we'll discuss a system of two spin-1/2 particles (such as the proton and the electron in a hydrogen atom).

8.C.1 States and Operators

The basis states of a system consisting of two spin-1/2 particles can be given by the direct product. For example, the state corresponding to particle one having spin-up along the z-direction, and particle two having spin-down along that same direction is

$$\left|+z\right\rangle_1 \otimes \left|-z\right\rangle_2 = \left|+z\right\rangle_1 \left|-z\right\rangle_2 = \left|+z,-z\right\rangle. \tag{8.C.1}$$

A common set of basis states for this 4-dimensional system consists of the states

$$\left|+z,+z\right\rangle, \left|+z,-z\right\rangle, \left|-z,+z\right\rangle, \left|-z,-z\right\rangle. \tag{8.C.2}$$

We will denote the spin operator corresponding to particle 1 as $\hat{\boldsymbol{S}}^{(1)}$, which is written in terms of its component operators as

$$\hat{\boldsymbol{S}}^{(1)} = \hat{S}_x^{(1)}\boldsymbol{u}_x + \hat{S}_y^{(1)}\boldsymbol{u}_y + \hat{S}_z^{(1)}\boldsymbol{u}_z. \tag{8.C.3}$$

These are the same operators we discussed in chapters 6 and 7, so they satisfy the same commutation relations, for example

$$\left[\hat{S}_x^{(1)}, \hat{S}_y^{(1)}\right] = i\hbar\hat{S}_z^{(1)}, \tag{8.C.4}$$

and so on. Similarly, we will denote the spin operator corresponding to particle 2 as $\hat{\boldsymbol{S}}^{(2)}$. All operators corresponding to different particles commute with each other, for example

$$\left[\hat{S}_x^{(1)}, \hat{S}_x^{(2)}\right] = 0, \left[\hat{S}_x^{(1)}, \hat{S}_y^{(2)}\right] = 0, \left[\hat{S}_x^{(1)}, \hat{S}_z^{(2)}\right] = 0. \tag{8.C.5}$$

As with photons, spin operators corresponding to a given particle only operate on the part of the state corresponding to that particle:

$$\hat{S}_z^{(2)}\left|+z,-z\right\rangle = -\frac{\hbar}{2}\left|+z,-z\right\rangle, \tag{8.C.6}$$

$$\left(\hat{S}^{(2)}\right)^2\left|+z,-z\right\rangle = \frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2\left|+z,-z\right\rangle = \frac{3}{4}\hbar^2\left|+z,-z\right\rangle \tag{8.C.7}$$

Equation (8.C.7) follows from eq. (7.27). The operator corresponding to the total spin of the two particles is

$$\begin{aligned} \hat{\boldsymbol{S}} &= \hat{\boldsymbol{S}}^{(1)} + \hat{\boldsymbol{S}}^{(2)} \\ &= \left(\hat{S}_x^{(1)} + \hat{S}_x^{(2)}\right)\boldsymbol{u}_x + \left(\hat{S}_y^{(1)} + \hat{S}_y^{(2)}\right)\boldsymbol{u}_y + \left(\hat{S}_z^{(1)} + \hat{S}_z^{(2)}\right)\boldsymbol{u}_z \\ &= \hat{S}_x\boldsymbol{u}_x + \hat{S}_y\boldsymbol{u}_y + \hat{S}_z\boldsymbol{u}_z\ . \end{aligned} \tag{8.C.8}$$

8.C.2 Eigenstates of Total Spin

Remember that the square of the spin operator, $\hat{S}^2$, is important, as it gives us information about the magnitude of the angular momentum. For a two particle system

$$\begin{aligned} \hat{S}^2 &= \left(\hat{\boldsymbol{S}}^{(1)} + \hat{\boldsymbol{S}}^{(2)}\right)\bullet\left(\hat{\boldsymbol{S}}^{(1)} + \hat{\boldsymbol{S}}^{(2)}\right) \\ &= \left(\hat{S}^{(1)}\right)^2 + \left(\hat{S}^{(2)}\right)^2 + 2\hat{\boldsymbol{S}}^{(1)}\bullet\hat{\boldsymbol{S}}^{(2)}, \end{aligned} \tag{8.C.9}$$

where we've used the fact that $\hat{\boldsymbol{S}}^{(1)}$ and $\hat{\boldsymbol{S}}^{(2)}$ commute. From sec. 7.3 we know that the eigenstates and eigenvalues of $\hat{S}^2$ and $\hat{S}_z$ must be

$$\hat{S}^2\left|s,m_s\right\rangle = s\left(s+1\right)\hbar^2\left|s,m_s\right\rangle \tag{8.C.10}$$

$$\hat{S}_z\left|s,m_s\right\rangle = m_s\hbar\left|s,m_s\right\rangle, \tag{8.C.11}$$

where

$$s = 0,\frac{1}{2},1,\frac{3}{2},2,\ldots, \tag{8.C.12}$$

$$m_s = -s,-s+1,-s+2,\ldots,s-2,s-1,s\,. \tag{8.C.13}$$

A remaining question is, "How do the states of total spin $\left|s,m_s\right\rangle$ relate to the states of eq. (8.C.2)?" Since the states of eq. (8.C.2) form a basis, it must be possible to write the states $\left|s,m_s\right\rangle$ as linear combinations of them. The brute force approach to answering this question is to write $\hat{S}^2$ as a 4×4 matrix using the basis states of eq. (8.C.2), and to then diagonalize the matrix to find the eigenstates. These eigenstates must be equivalent to the states $\left|s,m_s\right\rangle$.

We'll approach this problem in a different manner. First note that all of the states in eq. (8.C.2) are eigenstates of both $\left(\hat{S}^{(1)}\right)^2$ and $\left(\hat{S}^{(2)}\right)^2$, with eigenvalue $3\hbar^2/4$ [eq. (8.C.7)]. So, any linear combination of these states will also be eigenstates of these operators. Equation (8.C.9) then indicates that the eigenstates of $\hat{\mathbf{S}}^{(1)} \bullet \hat{\mathbf{S}}^{(2)}$ are the eigenstates of $\hat{S}^2$; we will look for these eigenstates.

It is useful to use the matrix representations of the spin component operators to determine their action on the spin states. For example

$$\hat{S}_x^{(1)}\left|+z\right\rangle_1 \doteq \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \doteq \frac{\hbar}{2}\left|-z\right\rangle_1 . \tag{8.C.14}$$

You should be able to verify that

$$\hat{S}_x^{(1)}\left|-z\right\rangle_1 = \frac{\hbar}{2}\left|+z\right\rangle_1,\ \hat{S}_y^{(1)}\left|+z\right\rangle_1 = i\frac{\hbar}{2}\left|-z\right\rangle_1,\ \hat{S}_y^{(1)}\left|-z\right\rangle_1 = -i\frac{\hbar}{2}\left|+z\right\rangle_1 . \tag{8.C.15}$$

These same relationships hold for particle 2 as well. Equations (8.C.14) and (8.C.16) will be used frequently in what follows.

In terms of component operators, $\hat{\mathbf{S}}^{(1)} \bullet \hat{\mathbf{S}}^{(2)}$ can be written as

$$\hat{\mathbf{S}}^{(1)} \bullet \hat{\mathbf{S}}^{(2)} = \hat{S}_x^{(1)}\hat{S}_x^{(2)} + \hat{S}_y^{(1)}\hat{S}_y^{(2)} + \hat{S}_z^{(1)}\hat{S}_z^{(2)} . \tag{8.C.16}$$

The action of this operator on $\left|+z,+z\right\rangle$ is

$$\begin{aligned} \hat{\mathbf{S}}^{(1)} \bullet \hat{\mathbf{S}}^{(2)}\left|+z,+z\right\rangle &= \hat{S}_x^{(1)}\hat{S}_x^{(2)}\left|+z,+z\right\rangle + \hat{S}_y^{(1)}\hat{S}_y^{(2)}\left|+z,+z\right\rangle + \hat{S}_z^{(1)}\hat{S}_z^{(2)}\left|+z,+z\right\rangle \\ &= \frac{1}{4}\hbar^2\left|-z,-z\right\rangle - \frac{1}{4}\hbar^2\left|-z,-z\right\rangle + \frac{1}{4}\hbar^2\left|+z,+z\right\rangle \\ &= \frac{1}{4}\hbar^2\left|+z,+z\right\rangle , \end{aligned} \tag{8.C.17}$$

so $\left|+z,+z\right\rangle$ is an eigenstate of $\hat{\mathbf{S}}^{(1)} \bullet \hat{\mathbf{S}}^{(2)}$, and hence an eigenstate of $\hat{S}^2$. To determine its corresponding eigenvalue, we compute

$$\begin{aligned} \hat{S}^2\left|+z,+z\right\rangle &= \left(\hat{S}^{(1)}\right)^2\left|+z,+z\right\rangle + \left(\hat{S}^{(2)}\right)^2\left|+z,+z\right\rangle + 2\hat{\mathbf{S}}^{(1)} \bullet \hat{\mathbf{S}}^{(2)}\left|+z,+z\right\rangle \\ &= \left(\frac{3}{4}\hbar^2 + \frac{3}{4}\hbar^2 + \frac{2}{4}\hbar^2\right)\left|+z,+z\right\rangle \\ &= 2\hbar^2\left|+z,+z\right\rangle \\ &= 1(1+1)\hbar^2\left|+z,+z\right\rangle . \end{aligned} \tag{8.C.18}$$

So $\left|+z,+z\right\rangle$ is an eigenstate of $\hat{S}^2$ with eigenvalue $2\hbar^2$, which corresponds to the quantum number $s=1$. We know $\left|+z,+z\right\rangle$ must also be an eigenstate of $\hat{S}_z$, and we can find its eigenvalue using

$$
\begin{aligned}
\hat{S}_z\left|+z,+z\right\rangle &= \left(\hat{S}_z^{(1)}+\hat{S}_z^{(2)}\right)\left|+z,+z\right\rangle \\
&= \left(\frac{\hbar}{2}+\frac{\hbar}{2}\right)\left|+z,+z\right\rangle \\
&= \hbar\left|+z,+z\right\rangle .
\end{aligned}
\tag{8.C.19}
$$

The eigenvalue is thus $\hbar$, which corresponds to $m_s = 1$. Putting all this together means that

$$
\left|1,1\right\rangle = \left|+z,+z\right\rangle. \tag{8.C.20}
$$

While not everything about spin states makes intuitive sense, this seems reasonable. It says that two spin-1/2 particles with their *z*-components pointing up have a total spin of 1, with a maximally upward *z*-component. Given this, it would also make sense for the maximally downward pointing state to be

$$
\left|1,-1\right\rangle = \left|-z,-z\right\rangle . \tag{8.C.21}
$$

You will verify that this is indeed the case in the problems.

One way to find the state $\left|1,0\right\rangle$ is to apply the lowering operator to the state $\left|1,1\right\rangle$. The two-particle lowering operator is

$$
\hat{S}_- = \hat{S}_x - i\hat{S}_y = \left(\hat{S}_x^{(1)}+\hat{S}_x^{(2)}\right) - i\left(\hat{S}_y^{(1)}+\hat{S}_y^{(2)}\right). \tag{8.C.22}
$$

Applying this operator to the state $\left|1,1\right\rangle = \left|+z,+z\right\rangle$ yields

$$
\begin{aligned}
\hat{S}_-\left|1,1\right\rangle &= \hat{S}_-\left|+z,+z\right\rangle \\
&= \hat{S}_x^{(1)}\left|+z,+z\right\rangle + \hat{S}_x^{(2)}\left|+z,+z\right\rangle - i\hat{S}_y^{(1)}\left|+z,+z\right\rangle - i\hat{S}_y^{(2)}\left|+z,+z\right\rangle \\
&= \frac{\hbar}{2}\left(\left|-z,+z\right\rangle + \left|+z,-z\right\rangle - i(i)\left|-z,+z\right\rangle - i(i)\left|+z,-z\right\rangle\right).
\end{aligned}
\tag{8.C.23}
$$

Using eq. (7.21) yields

$$
\begin{aligned}
\hat{S}_-\left|1,1\right\rangle &= \hbar\left[1(1+1)-1(1-1)\right]^{1/2}\left|1,0\right\rangle \\
&= \hbar\sqrt{2}\left|1,0\right\rangle .
\end{aligned}
\tag{8.C.241}
$$

Combining these last two equations, we find that

$$
\begin{aligned}
\left|1,0\right\rangle &= \frac{1}{2\sqrt{2}}\left(\left|-z,+z\right\rangle + \left|+z,-z\right\rangle + \left|-z,+z\right\rangle + \left|+z,-z\right\rangle\right) \\
&= \frac{1}{\sqrt{2}}\left(\left|+z,-z\right\rangle + \left|-z,+z\right\rangle\right).
\end{aligned}
\tag{8.C.25}
$$

We've found all three eigenstates corresponding to the quantum number $s = 1$, so the fourth and final eigenstate must have a different value for s. There's only one possible value for s that has only a single eigenstate, and that's $s = 0$, which has the corresponding eigenstate $|0,0\rangle$. In problems 8.C.3 and 8.C.4 you'll show that

$$|0,0\rangle = \frac{1}{\sqrt{2}}\left(|+z,-z\rangle - |-z,+z\rangle\right). \tag{8.C.26}$$

8.C.3 PROBLEMS

8.C.1* Verify that $|1,-1\rangle = |-z,-z\rangle$ (i.e., verify that this state corresponds to the total spin quantum numbers $s = 1$, $m_s = -1$).

8.C.2* Verify that $|1,0\rangle = \frac{1}{\sqrt{2}}\left(|+z,-z\rangle + |-z,+z\rangle\right)$ (i.e., verify that this state corresponds to the total spin quantum numbers $s = 1$, $m_s = 0$).

8.C.3 Find the state $|0,0\rangle$ by noting that it must be orthogonal to all three of the $s = 1$ states.

8.C.4* Verify that $|0,0\rangle = \frac{1}{\sqrt{2}}\left(|+z,-z\rangle - |-z,+z\rangle\right)$ (i.e., verify that this state corresponds to the total spin quantum numbers $s = 0$, $m_s = 0$).

8.C.5 Calculate the expectation values of $\hat{S}^2$ and $\hat{S}_z$ for two spin-1/2 particles in the state $|+x,-x\rangle$.

CHAPTER 9

Time Evolution and the Schrödinger Equation

Thus far we've described a number of different quantum systems. We've discussed quantum states, operators that change states, and measurements that project systems into certain states. However, we haven't described how quantum systems evolve in time. In classical physics we can use Newton's laws, or Lagrange's or Hamilton's equations, to describe how systems evolve. In quantum mechanics it is the Schrödinger equation that describes temporal evolution.

9.1 THE TIME-EVOLUTION OPERATOR

In the Schrödinger picture of quantum mechanics states evolve in time $|\psi\rangle \rightarrow |\psi(t)\rangle$, and observables are time independent.[1] We know that operators change one state into another, so we can use an operator to change the state at some initial time $|\psi(t_0)\rangle$ into a state at a later time $|\psi(t)\rangle$:

$$|\psi(t)\rangle = \hat{U}(t,t_0)|\psi(t_0)\rangle. \tag{9.1}$$

We will generally take $t_0 = 0$, and define $\hat{U}(t) \equiv \hat{U}(t,t_0 = 0)$, so

$$|\psi(t)\rangle = \hat{U}(t)|\psi(0)\rangle. \tag{9.2}$$

We'll assume that our states are normalized at $t = 0$, and we'd like them to remain that way as they evolve in time. This means

---

[1] The alternative to this is the Heisenberg picture, in which states don't change in time, and operators are time dependent. These two pictures are equivalent. Here we will concentrate on the Schrödinger picture; for more information on the Heisenberg picture, see chap. 16.

$$\langle\psi(t)|\psi(t)\rangle = \langle\psi(0)|\hat{U}^\dagger(t)\hat{U}(t)|\psi(0)\rangle = \langle\psi(0)|\psi(0)\rangle = 1, \tag{9.3}$$

or

$$\hat{U}^\dagger(t)\hat{U}(t) = \hat{1}. \tag{9.4}$$

Thus, $\hat{U}(t)$ is unitary, and is referred to as the unitary time-evolution operator.

Think about a system evolving for an infinitesimally small amount of time dt. The state will be nearly the same as the initial state, so $\hat{U}(t)$ will be nearly $\hat{1}$, and we'd expect the difference from $\hat{1}$ to be linear in dt:

$$\hat{U}(dt) = \hat{1} - i\hat{G}_t dt. \tag{9.5}$$

In this equation $\hat{G}_t$ is the generator of temporal evolution, which is analogous to the generator of rotation we discussed in sec. 7.5. Substituting eq. (9.5) into eq. (9.4), and keeping terms to first order in dt, yields

$$\begin{aligned}\hat{U}^\dagger(dt)\hat{U}(dt) &= \left(\hat{1} + i\hat{G}_t^\dagger dt\right)\left(\hat{1} - i\hat{G}_t dt\right) \\ &\cong \hat{1} + i\left(\hat{G}_t^\dagger - \hat{G}_t\right)dt \\ &= \hat{1}.\end{aligned} \tag{9.6}$$

So $\hat{G}_t^\dagger = \hat{G}_t$, and $\hat{G}_t$ must be Hermitian [the i in eq. (9.5) is necessary to ensure that this is the case].

9.2 THE SCHRÖDINGER EQUATION

9.2.1 The Hamiltonian

In classical mechanics the generator of temporal evolution is the Hamiltonian, so we'll take $\hat{G}_t \propto \hat{H}$, where $\hat{H}$ is the Hamiltonian operator. In classical physics the Hamiltonian tells us the total energy, so $\hat{H}$ is the energy operator; it is Hermitian and corresponds to an observable. From eq. (9.5) we see that $\hat{G}_t$ has units of s^{-1}, or frequency. We can take a hint from Planck's relation between energy and frequency,

$$E = hf = \hbar\omega, \tag{9.7}$$

to write $\hat{G}_t = \hat{H}/\hbar$. Equation (9.5) then becomes

$$\hat{U}(dt) = \hat{1} - \frac{i}{\hbar}\hat{H}dt. \tag{9.8}$$

This equation looks remarkably similar to eq. (7.41) for rotation. We can follow the same procedure as in sec. 7.5, and use eq. (9.8) to write a differential equation for $\hat{U}(t)$:

$$\frac{d}{dt}\hat{U}(t) = -\frac{i}{\hbar}\hat{H}\hat{U}(t). \tag{9.9}$$

This is the temporal equivalent to eq. (7.46). Equation (9.9) is the Schrödinger equation for $\hat{U}(t)$. Applying it to $|\psi(0)\rangle$ yields

$$\frac{d}{dt}\hat{U}(t)|\psi(0)\rangle = -\frac{i}{\hbar}\hat{H}\hat{U}(t)|\psi(0)\rangle. \tag{9.10}$$

Using eq. (9.2), we find

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}|\psi(t)\rangle, \tag{9.11}$$

which is the Schrödinger equation for the time evolution of a state. If you've encountered the Schrödinger equation before, it probably didn't look much like this. In chapter 11 we'll reformulate it to make it look more familiar.

9.2.2 Solutions to the Schrödinger Equation

Let's solve eq. (9.9). We'll make the assumption that the Hamiltonian $\hat{H}$ is independent of time, which is true for a large number of important problems. With this assumption, and the initial condition $\hat{U}(0) = \hat{1}$, the solution to eq. (9.9) is

$$\hat{U}(t) = e^{-i\hat{H}t/\hbar}. \tag{9.12}$$

Substituting this into eq. (9.2) yields

$$|\psi(t)\rangle = e^{-i\hat{H}t/\hbar}|\psi(0)\rangle. \tag{9.13}$$

The Hamiltonian is the energy operator, so its eigenstates are the energy eigenstates, and its eigenvalues are the allowed energies E_n:

$$\hat{H}|E_n\rangle = E_n|E_n\rangle. \tag{9.14}$$

If the initial state is an energy eigenstate, $|\psi(0)\rangle = |E_n\rangle$, then eq. (9.13) tells us that the state at a later time will be

$$\begin{aligned}|\psi(t)\rangle &= e^{-i\hat{H}t/\hbar}|E_n\rangle \\ &= e^{-iE_nt/\hbar}|E_n\rangle \\ &= e^{-i\omega_nt}|E_n\rangle.\end{aligned} \tag{9.15}$$

Here we've used eqs. (9.7) and (9.14). Remember, we can only substitute E_n in place of $\hat{H}$ in the argument of the exponential because $|E_n\rangle$ is an eigenstate of $\hat{H}$.[2]

Note that the state in eq. (9.15) does not change in time! The overall phase changes in time, but the state remains $|E_n\rangle$. Recall that our solution is applicable to the case of

---

[2] Refer to sec. 7.5 for more discussion of this.

a time-independent Hamiltonian, but otherwise this result is quite general. If the Hamiltonian is time independent, and the system starts in an energy eigenstate, it will remain in that state for all time.

In order for the state to change in time, we need to start in a state that is *not* an energy eigenstate. For example, the initial state could be a linear combination of two states with different energies:

$$|\psi(0)\rangle = c_1|E_1\rangle + c_2|E_2\rangle. \tag{9.16}$$

At later times the state will become

$$\begin{aligned}
|\psi(t)\rangle &= e^{-i\hat{H}t/\hbar}|\psi(0)\rangle \\
&= c_1 e^{-i\hat{H}t/\hbar}|E_1\rangle + c_2 e^{-i\hat{H}t/\hbar}|E_2\rangle \\
&= c_1 e^{-iE_1t/\hbar}|E_1\rangle + c_2 e^{-iE_2t/\hbar}|E_2\rangle \\
&= c_1 e^{-i\omega_1 t}|E_1\rangle + c_2 e^{-i\omega_2 t}|E_2\rangle \\
&= e^{-i\omega_1 t}\left(c_1|E_1\rangle + c_2 e^{-i(\omega_2-\omega_1)t}|E_2\rangle\right).
\end{aligned} \tag{9.17}$$

The overall phase factor in the front does not change the state, but the relative phase factor does. We can see that this state oscillates in time at frequency $\omega_2 - \omega_1$, which is the frequency difference of the individual states.

9.3 EXPECTATION VALUES

We can also describe the time dependence of expectation values. Let's start with the expectation value of the Hamiltonian. Once again, we will assume that $\hat{H}$ itself is time independent, so any time dependence to its expectation value comes from the time dependence of the state. The expectation value of H is given by

$$\begin{aligned}
\langle H\rangle(t) &= \langle\psi(t)|\hat{H}|\psi(t)\rangle \\
&= \langle\psi(0)|e^{i\hat{H}t/\hbar}\hat{H}e^{-i\hat{H}t/\hbar}|\psi(0)\rangle \\
&= \langle\psi(0)|e^{i\hat{H}t/\hbar}e^{-i\hat{H}t/\hbar}\hat{H}|\psi(0)\rangle \\
&= \langle\psi(0)|\hat{H}|\psi(0)\rangle \\
&= \langle H\rangle(t=0),
\end{aligned} \tag{9.18}$$

where we've used the fact that any function of $\hat{H}$ commutes with $\hat{H}$.

We have not assumed anything about the state in eq. (9.18). In particular, we have not assumed that the initial state is an eigenstate of $\hat{H}$, so the state itself will in general change in time. Despite this fact, the expectation value of the Hamiltonian is constant.

Why? Remember that the Hamiltonian is the energy operator, so eq. (9.18) says that the expectation value of the energy does not change in time. If the Hamiltonian has no time dependence, energy is conserved. In this light, eq. (9.18) makes a great deal of sense.

What about the expectation values of other observables? If A is a time-independent observable, we can write the rate of change of its expectation value as

$$\begin{aligned}\frac{d}{dt}\langle A\rangle &= \frac{d}{dt}\langle\psi(t)|\hat{A}|\psi(t)\rangle \\ &= \left(\frac{d}{dt}\langle\psi(t)|\right)\hat{A}|\psi(t)\rangle + \langle\psi(t)|\hat{A}\left(\frac{d}{dt}|\psi(t)\rangle\right) \\ &= \frac{i}{\hbar}\langle\psi(t)|\hat{H}\hat{A}|\psi(t)\rangle - \frac{i}{\hbar}\langle\psi(t)|\hat{A}\hat{H}|\psi(t)\rangle \\ &= \frac{i}{\hbar}\langle\psi(t)|\left[\hat{H},\hat{A}\right]|\psi(t)\rangle .\end{aligned} \tag{9.19}$$

Here we have used the Schrödinger equation [eq. (9.11)]. If we know the commutator of $\hat{H}$ and $\hat{A}$, we can write down a differential equation for $\langle A\rangle$. More importantly, for our purposes here, eq. (9.19) says that if $\hat{A}$ is independent of time, and commutes with the Hamiltonian, the expectation value of A will be constant in time. In this case $\langle A\rangle$ is a constant of the motion; it is a conserved quantity. This is true even if the state is changing in time.

9.4 SPIN-1/2 PARTICLE IN A MAGNETIC FIELD

We've been discussing time dependence in general terms. Now let's discuss a specific example, that of a spin-1/2 particle placed in a magnetic field.

The energy of magnetic dipole (with dipole moment $\boldsymbol{\mu}$) in a magnetic field $\boldsymbol{\mathcal{B}}$ is given by eq. (6.1). Since the Hamiltonian is the energy, we have

$$H = -\boldsymbol{\mu}\cdot\boldsymbol{\mathcal{B}}. \tag{9.20}$$

Assume that the magnetic field points in the z-direction: $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_z$. Recall that $\boldsymbol{\mu}$ is related to the spin of a particle by $\boldsymbol{\mu} = \gamma\boldsymbol{S}$, where γ is the gyromagnetic ratio. The Hamiltonian is then

$$H = -\gamma S_z\mathcal{B}. \tag{9.21}$$

We'll treat the field classically, so the Hamiltonian operator is

$$\begin{aligned}\hat{H} &= -\gamma\hat{S}_z\mathcal{B} \\ &= -\Omega\hat{S}_z,\end{aligned} \tag{9.22}$$

where $\Omega = \gamma\mathcal{B}$ has units of s^{-1} and is called the Larmor frequency.

Time evolution is most easily dealt with using a basis consisting of the eigenstates of $\hat{H}$, so we need to find these eigenstates. We often say that we need to diagonalize $\hat{H}$, because the matrix for $\hat{H}$ is diagonal in a basis consisting of its eigenstates. In this case, however, eq. (9.22) tells us that $\hat{H} \propto \hat{S}_z$, so the eigenstates of $\hat{H}$ are the same as the eigenstates of $\hat{S}_z$. We already know what these states are: $\left|+z\right\rangle$ and $\left|-z\right\rangle$.

Assume that at $t = 0$ we start in state $\left|+z\right\rangle$ (spin parallel to $\boldsymbol{\mathcal{B}}$). Equation (9.13) then tells us that

$$\begin{aligned}\left|\psi(t)\right\rangle &= e^{-i\hat{H}t/\hbar}\left|+z\right\rangle \\ &= e^{i\Omega\hat{S}_z t/\hbar}\left|+z\right\rangle \\ &= e^{i\Omega(\hbar/2)t/\hbar}\left|+z\right\rangle \\ &= e^{i\Omega t/2}\left|+z\right\rangle .\end{aligned} \tag{9.23}$$

The state acquires a time-dependent phase shift, but does not change in time.

Suppose, instead, that we start in the state $\left|+x\right\rangle$ at $t = 0$ (spin perpendicular to $\boldsymbol{\mathcal{B}}$). To find the time evolution of this state, we should express $\left|+x\right\rangle$ in terms of $\left|+z\right\rangle$ and $\left|-z\right\rangle$, because they are the eigenstates of $\hat{H}$. Doing this, we find that the state at future times is

$$\begin{aligned}\left|\psi(t)\right\rangle &= e^{-i\hat{H}t/\hbar}\left|+x\right\rangle \\ &= e^{-i\hat{H}t/\hbar}\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle + \left|-z\right\rangle\right) \\ &= \frac{1}{\sqrt{2}}\left(e^{i\Omega\hat{S}_z t/\hbar}\left|+z\right\rangle + e^{i\Omega\hat{S}_z t/\hbar}\left|-z\right\rangle\right) \\ &= \frac{1}{\sqrt{2}}\left(e^{i\Omega(\hbar/2)t/\hbar}\left|+z\right\rangle + e^{i\Omega(-\hbar/2)t/\hbar}\left|-z\right\rangle\right) \\ &= e^{i\Omega t/2}\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle + e^{-i\Omega t}\left|-z\right\rangle\right).\end{aligned} \tag{9.24}$$

At $t = \pi/2\Omega$ the state is

$$\begin{aligned}\left|\psi(t = \pi/2\Omega)\right\rangle &= e^{i\pi/4}\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle + e^{-i\pi/2}\left|-z\right\rangle\right) \\ &= e^{i\pi/4}\left|-y\right\rangle .\end{aligned} \tag{9.25}$$

At twice this time, $t = \pi/\Omega$, the state is

$$\begin{aligned}\left|\psi(t = \pi/\Omega)\right\rangle &= e^{i\pi/2}\frac{1}{\sqrt{2}}\left(\left|+z\right\rangle + e^{-i\pi}\left|-z\right\rangle\right) \\ &= e^{i\pi/2}\left|-x\right\rangle .\end{aligned} \tag{9.26}$$

Clearly the state is changing in time, and the spin appears to be precessing about the z-axis. We can confirm this by looking at the time dependence of the expectation value of $\boldsymbol{S}$, which is given by

$$\langle \boldsymbol{S}\rangle(t)=\langle S_x\rangle(t)\boldsymbol{u}_x+\langle S_y\rangle(t)\boldsymbol{u}_y+\langle S_z\rangle(t)\boldsymbol{u}_z. \tag{9.27}$$

The expectation value for the spin along the x-direction is:

$$\begin{aligned}\langle S_x\rangle(t)&=\langle\psi(t)|\hat{S}_x|\psi(t)\rangle\\&=\left[e^{-i\Omega t/2}\frac{1}{\sqrt{2}}\left(\langle +z|+e^{i\Omega t}\langle -z|\right)\right]\hat{S}_x\left[e^{i\Omega t/2}\frac{1}{\sqrt{2}}\left(|+z\rangle+e^{-i\Omega t}|-z\rangle\right)\right]\\&=\frac{1}{2}\left[\left(\langle +z|+e^{i\Omega t}\langle -z|\right)\right]\left[\left(\frac{\hbar}{2}|-z\rangle+e^{-i\Omega t}\frac{\hbar}{2}|+z\rangle\right)\right]\\&=\frac{\hbar}{4}\left(e^{-i\Omega t}+e^{i\Omega t}\right)\\&=\frac{\hbar}{2}\cos\Omega t\,,\end{aligned} \tag{9.28}$$

where we've used eqs. (6.14) and (6.15). This expectation value oscillates at the Larmor frequency. The expectation value for the spin along the y-direction is

$$\begin{aligned}\langle S_y\rangle(t)&=\langle\psi(t)|\hat{S}_y|\psi(t)\rangle\\&=\left[e^{-i\Omega t/2}\frac{1}{\sqrt{2}}\left(\langle +z|+e^{i\Omega t}\langle -z|\right)\right]\hat{S}_y\left[e^{i\Omega t/2}\frac{1}{\sqrt{2}}\left(|+z\rangle+e^{-i\Omega t}|-z\rangle\right)\right]\\&=\frac{1}{2}\left[\left(\langle +z|+e^{i\Omega t}\langle -z|\right)\right]\left[\left(i\frac{\hbar}{2}|-z\rangle-ie^{-i\Omega t}\frac{\hbar}{2}|+z\rangle\right)\right]\\&=\frac{\hbar}{4}\left(-ie^{-i\Omega t}+ie^{i\Omega t}\right)\\&=-\frac{\hbar}{2}\sin\Omega t,\end{aligned} \tag{9.29}$$

where we've used eq. (6.26). This expectation value also oscillates at the Larmor frequency, but is out of phase with $\langle S_x\rangle(t)$. Finally, the expectation value for the spin along the z-direction is

$$\begin{aligned}\langle S_z\rangle(t)&=\langle\psi(t)|\hat{S}_z|\psi(t)\rangle\\&=\left[e^{-i\Omega t/2}\frac{1}{\sqrt{2}}\left(\langle +z|+e^{i\Omega t}\langle -z|\right)\right]\hat{S}_z\left[e^{i\Omega t/2}\frac{1}{\sqrt{2}}\left(|+z\rangle+e^{-i\Omega t}|-z\rangle\right)\right]\\&=\frac{1}{2}\left[\left(\langle +z|+e^{i\Omega t}\langle -z|\right)\right]\left[\left(\frac{\hbar}{2}|+z\rangle-e^{-i\Omega t}\frac{\hbar}{2}|-z\rangle\right)\right]\\&=\frac{\hbar}{4}(1-1)\\&=0,\end{aligned} \tag{9.30}$$

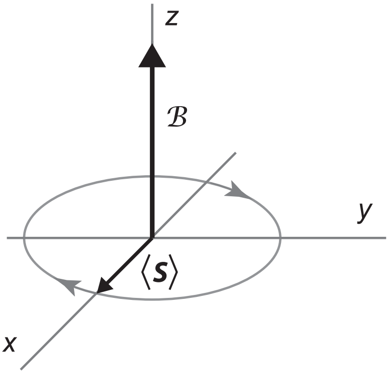


Fig 9.1 The time evolution of $\langle \mathbf{S} \rangle$ for a spin-1/2 particle (e.g., a proton) in a magnetic field. The particle's spin is whose spin is perpendicular to the field at $t = 0$.

which is time independent.

Figure 9.1 depicts $\langle \mathbf{S} \rangle (t)$, and you can see that the expectation value of the spin does indeed precess around the direction of the applied field. This behavior would describe, for example, the spin of a hydrogen nucleus (a proton) in a static magnetic field.

Normally when we describe the behavior of magnetic dipoles in magnetic fields, we find that the field applies a torque that causes the dipole to align with the field. Since the magnetic dipole moment of a proton is parallel to its spin, one might therefore expect the spin to align with the field, not precess about it. In order to understand spin precession, think about a gyroscope. A spinning gyroscope has a large angular momentum, and the torque exerted by the gravitational field causes this angular momentum to precess about the direction of the field, rather than causing it to align with the field. Indeed, if the spin is not initially perpendicular to the magnetic field, $\langle \mathbf{S} \rangle$ will sweep out a cone centered on the magnetic field (see problem 9.6), just as the angular momentum of a gyroscope sweeps out a cone centered on the gravitational field. This classical analogy will be explored more fully in problem 9.11.

9.5 NEUTRINO OSCILLATIONS

In the standard model of particle physics, the fundamental particles are divided into the quarks and the leptons. Quarks and leptons each have three generations, ordered by mass, with lighter masses in the first generation and higher masses in the third. Each generation consists of two particles of different flavors. For the quarks, the generations are (up, down), (charm, strange) and (bottom, top); each of these also has a corresponding antiquark.[3] Most of the ordinary matter in the universe is made of protons and neutrons, which are combinations of the two lightest quarks, the up and the down. Each of the three generations of leptons consists of a charged particle and an associated uncharged neutrino (there are also the antiparticles for each of these). In order of

---

[3] Frequently the quarks are referred to simply by the first letter of their name.

increasing mass, the charged leptons are the electron e^-, the muon μ^-, and the tau τ^-. The neutrinos are believed to have very small masses, and are simply named by their corresponding charged particle: electron ν_e, muon ν_μ, and tau ν_τ neutrinos.

Neutrinos interact with ordinary matter only via the weak force, which is, well, very weak. One way neutrinos are produced is by nuclear reactions, for example, fusion in the Sun. The flux of these solar neutrinos at the Earth's surface is approximately 10^{15} $\text{m}^{-2}\text{s}^{-1}$, and they're passing through your body 24-7, with no obvious ill effects.

Solar neutrinos are interesting because they tell us about the Sun. Detailed models have been developed describing the nuclear burning occurring in the Sun. These models predict the neutrino flux arising from various nuclear reactions, and astrophysicists are fairly confident in them. Combining these models with an understanding of the ways in which neutrinos can be detected yields predictions for the number of observed neutrinos in detectors here on Earth. The first experiment designed to detect solar neutrinos began in the early 1960s. Since neutrinos interact so weakly, the detectors need to be very large, and even then the detection rates are extremely low.

By the 1970s it was apparent that there was a discrepancy between the solar models and the experiments. Only about one-third of the predicted number of neutrinos were actually detected, and this became known as the "solar neutrino problem." This deficit of solar neutrinos was confirmed in other experiments extending into the 1990s. The solar neutrino problem was finally solved in 2002, with data from the Sudbury Neutrino Observatory (SNO) collaboration [9.1].

Part of the problem was that the standard model of particle physics assumed that neutrinos were massless. If neutrinos are massless they must propagate at the speed of light, and all flavors of neutrinos must then have the same speed. We now believe that neutrinos have very small but finite masses. Furthermore, the flavor eigenstates ($|\nu_e\rangle$, $|\nu_\mu\rangle$ and $|\nu_\tau\rangle$) are *not* the same as the mass eigenstates ($|\nu_1\rangle$, $|\nu_2\rangle$ and $|\nu_3\rangle$); the flavor eigenstates are linear combinations of mass eigenstates. Since the different mass eigenstates have different masses, they can propagate at different speeds. Neutrino oscillations arise from the fact that as a linear combination of mass eigenstates propagates, the phase difference between the mass states changes, resulting in a different linear combination.

Imagine the flavor eigenstate $|\nu_e\rangle$, for example, being made up of a linear combination of $|\nu_1\rangle$ and $|\nu_2\rangle$. It is the mass eigenstates that are the energy eigenstates (more on this below), and which propagate at particular speeds. A neutrino that starts as $|\nu_e\rangle$ at the Sun may end up as $|\nu_\mu\rangle$ on Earth because the original linear combination of $|\nu_1\rangle$ and $|\nu_2\rangle$ has changed on propagation. If your detector is sensitive to ν_e's, but not ν_μ's, you would perceive this as a deficit of ν_e's unless you account for the oscillations. If you think neutrinos are massless, you won't account for the oscillation because there shouldn't be any. SNO was able to confirm the observation of neutrino oscillations, and solve the solar neutrino problem, because it was sensitive to all three neutrino flavors.

Let's simplify the discussion by assuming only two mass eigenstates:

$$|\nu_1\rangle = \cos\theta|\nu_e\rangle + \sin\theta|\nu_x\rangle, |\nu_2\rangle = -\sin\theta|\nu_e\rangle + \cos\theta|\nu_x\rangle. \tag{9.31}$$

Here $\left|\nu_x\right\rangle$ corresponds to $\left|\nu_\mu\right\rangle$ or $\left|\nu_\tau\right\rangle$, or some linear combination of the two. The angle θ is called the mixing angle, and is a convenient parameter to describe the unitary transformation between the flavor eigenstates and the mass eigenstates.

Neutrinos are highly relativistic particles, so we must use the relativistic expression for the energy

$$E^2 = p^2c^2 + m^2c^4, \tag{9.32}$$

where p is the relativistic momentum, and m is the rest mass. Neutrinos have very little mass, so most of their energy comes from their momentum, and to second order in the mass we can write

$$\begin{aligned} E &= pc\left(1+\frac{m^2c^2}{p^2}\right)^{1/2} \\ &\cong pc\left(1+\frac{m^2c^2}{2p^2}\right). \end{aligned} \tag{9.33}$$

States with different masses will have different energies, even if they have the same momentum. In terms of the zeroth order energy $E_0 = pc$, this can be rewritten as

$$E \cong E_0\left(1+\frac{m^2c^4}{2E_0{}^2}\right). \tag{9.34}$$

Assume that an electron neutrino is created with momentum p in a fusion reaction in the Sun. This neutrino is a linear combination of states with masses m_1 and m_2. As it propagates toward Earth at essentially the speed of light, the mass states get out of phase, and the neutrino undergoes oscillations. You will show in problem 9.12 that after propagating a distance $L \cong ct$, the probability that the neutrino will survive, and be detected as an electron neutrino is

$$P(\nu_e) = 1 - \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2c^4L}{4E_0\hbar c}\right), \tag{9.35}$$

where $\Delta m^2 \equiv \left|m_2^2 - m_1^2\right|$ is the difference of the squares of the masses (*not* the square of the difference of the masses).

A plot of neutrino survival probability is shown in fig. 9.2 for data from the Kamioka Liquid-scintillator Anti-Neutrino Detector (KamLAND) [9.2]. KamLAND measures electron antineutrinos $\overline{\nu}_e$ produced by nuclear reactors at various distances from the detector.[4] Figure 9.2 clearly shows that these antineutrinos undergo oscillations as they propagate.

---

[4] It was electron antineutrinos $\overline{\nu}_e$ that were produced and measured in this experiment, so the figure plots $P(\overline{\nu}_e)$.

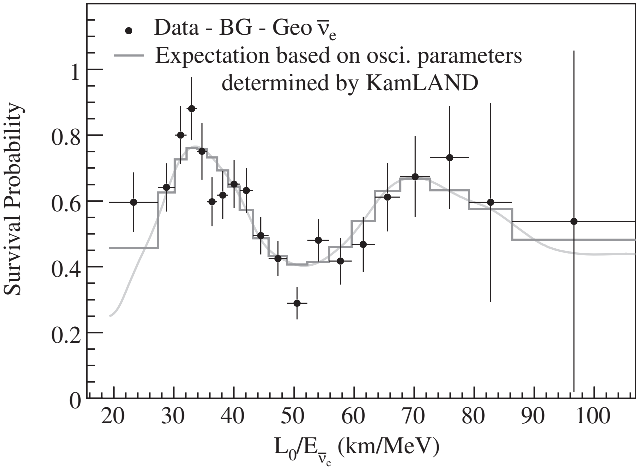


Fig 9.2 Neutrino oscillations observed by KamLAND. Electron antineutrinos $\overline{\nu}_e$ produced in nuclear reactors at various distances from the detector are measured, and the probability that they remain $\overline{\nu}_e$'s is plotted. The horizontal axis is L/E_0 in the notation of eq. (9.35). Reprinted with permission from S. Abe et al., Phys. Rev. Lett. **100**, 221803 (2008). Copyright 2008 by the American Physical Society.

Note that the probability in eq. (9.35) is a function of the mixing angle θ, and the difference of the squares of the masses Δm^2, but not the individual masses themselves. Recent experiments, such as those at SNO and KamLAND, have measured values for θ and Δm^2. At the moment there are some constraints on the neutrino masses, but no measurements of the masses themselves. There are experiments underway that are hoping to alleviate this shortcoming.

For more discussion of neutrinos and neutrino oscillations, see ref. [9.3].

9.6 References

[9.1] Q. R. Ahmad et al. (SNO collaboration), "Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury neutrino observatory," Phys. Rev. Lett. **89**, 011301 (2002); Q. R. Ahmad et al. (SNO collaboration), "Measurement of day and night neutrino energy spectra at SNO and constraints on neutrino mixing parameters," Phys. Rev. Lett. **89**, 011302 (2002).

[9.2] S. Abe et al., "Precision measurement of neutrino oscillation parameters with KamLAND," Phys. Rev. Lett. **100**, 221803 (2008).

[9.3] C. Waltham, "Teaching neutrino oscillations," Am. J. Phys. **72**, 742 (2004).

9.7 PROBLEMS

9.1 If we do not assume that A is time independent, how does the differential equation for $\langle A \rangle$ [eq. (9.19)] change?

9.2* The Hamiltonian of a system is given by $\hat{H} = \beta\hat{Q}$, where β is a constant, and $\hat{Q}$ corresponds to an observable. $\hat{Q}$ is known to have 25 eigenvalues: q_i, $i = 1, 2, \ldots, 25$, with 25 corresponding eigenstates $|q_i\rangle$. A series of measurements of a time-independent observable, Y, are performed on this system.

(a) At time $t = 0$, the state of the system is $|\psi(0)\rangle = |q_{12}\rangle$. In general, will $\langle Y\rangle$ depend on time? Why or why not? (b) At time $t = 0$, the state of the system is instead $|\psi(0)\rangle = \frac{1}{\sqrt{3}}\left(|q_{12}\rangle + |q_{14}\rangle + |q_{16}\rangle\right)$. In general, will $\langle Y\rangle$ depend on time? Why or why not? (c) Are there any situations in which $\langle Y\rangle$ is time independent, regardless of the input state? Explain.

Problems 9.3–9.6 concern the following system: a spin-1/2 particle is placed in a magnetic field pointing in the z-direction $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_z$.

9.3 At $t = 0$ the particle is in state $|+x\rangle$. Calculate the probabilities that the spin will be measured to be pointing up along the x-, y-, and z-directions at time t.

9.4 At $t = 0$ the particle is in state $|+x\rangle$. Solve the differential equation for $\langle S_x\rangle(t)$. Compare your answer to eq. (9.28).

9.5 Calculate the commutator of the projection operator $\hat{P}_{+z}$ and $\hat{H}$. What does this say about your answer to problem 9.3?

9.6* At $t = 0$ the particle is in state $|+n\rangle = \cos(\theta/2)|+z\rangle + e^{i\phi}\sin(\theta/2)|-z\rangle$, which is a state corresponding to spin up along the direction $\boldsymbol{u}_n$ (see problem 6.13). Calculate $\langle \boldsymbol{S}\rangle(t)$, and describe its motion.

9.7 A spin-1/2 particle is placed in a magnetic field pointing in the x-direction $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_x$. At $t = 0$ the particle is in state $|+z\rangle$. Calculate the expectation value of $\boldsymbol{S}$ at time t.

9.8 A spin-1 particle is placed in a magnetic field pointing in the z-direction $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_z$. At $t = 0$ the particle is in state $|1,-1\rangle_x = (1/2)\left(|1,1\rangle - \sqrt{2}|1,0\rangle + |1,-1\rangle\right)$. Calculate the expectation value of $\boldsymbol{S}$ at time t.

9.9 A system of two spin-1/2 particles (complement 8.C) is placed in a magnetic field pointing in the z-direction $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_z$. At $t = 0$ the system is in state $|1,0\rangle$. Calculate the probability $P(+z_1, -z_2, t)$.

9.10 A system of two spin-1/2 particles (complement 8.C) is placed in a magnetic field pointing in the z-direction $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_z$. At $t = 0$ the system is in state $|-z, +z\rangle$. Calculate the probability $P(+z_1, -z_2, t)$.

9.11 A classical spinning sphere with uniformly distributed charge q and mass m has a magnetic dipole moment $\boldsymbol{\mu} = \frac{q}{2m}\boldsymbol{L}$, where $\boldsymbol{L}$ is its angular momentum. At $t = 0$ the angular momentum is $\boldsymbol{L} = L\boldsymbol{u}_x$, and the sphere is placed in a magnetic field pointing in the z-direction $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_z$. Describe the motion of $\boldsymbol{\mu}(t)$, assuming that L is extremely large.

9.12* A neutrino is created in state $|\nu_e\rangle$, with momentum p. Show that after propagating a distance $L \cong ct$, the probability that it will be detected as an electron neutrino is given by eq. (9.35).

9.13* Using the data in fig. 9.2, and assuming that antineutrinos behave in the same manner as neutrinos (although there is some recent evidence to suggest that maybe they don't completely), estimate the difference of the squares of the masses Δm^2. Express your answer in units of eV^2/c^4.

9.14* A particular system has three basis vectors: $|a\rangle, |b\rangle$ & $|c\rangle$, which are eigenstates of the observable O. In this basis, the Hamiltonian is given by

$$\hat{H} \doteq \begin{pmatrix} \Gamma_1 & \alpha & 0 \\ \alpha^* & \Gamma_1 & 0 \\ 0 & 0 & \Gamma_2 \end{pmatrix},$$

where Γ_1 & Γ_2 are real-valued constants, and α is complex.

(a) Find the allowed energies of this system. (b) At $t = 0$ the system is placed in state $|a\rangle$. What is the probability that a measurement of O at time t will yield b?

COMPLEMENT 9.A

Magnetic Resonance

In this complement we'll discuss a quantum system that has a time-dependent Hamiltonian. Because the Hamiltonian is not time independent, many of the results in chapter 9 are not directly applicable. The system we will consider is that of a particle with spin, and hence a magnetic dipole moment, situated in an oscillating magnetic field. As we will see, the spin can be made to oscillate resonantly with the time-dependent field, a condition known as magnetic resonance. In nuclear magnetic resonance (NMR) it is the spin of atomic nuclei that oscillates, while in electron spin resonance (ESR) [or electron paramagnetic resonance (EPR)] it is the spin of unpaired electrons that oscillates.

9.A.1 Schrödinger Equation

A particle with spin-1/2 is placed in a magnetic field that consists of a component along the *z*-direction that is static, and a component along the *x*-direction that oscillates at angular frequency ω. The field is thus

$$\boldsymbol{\mathcal{B}} = \mathcal{B}_z \boldsymbol{u}_z + \mathcal{B}_x \cos\left(\omega t\right) \boldsymbol{u}_z, \tag{9.A.1}$$

and the Hamiltonian is

$$\begin{aligned} \hat{H} &= -\hat{\boldsymbol{\mu}} \bullet \boldsymbol{\mathcal{B}} \\ &= -\gamma \hat{S}_z \mathcal{B}_z - \gamma \hat{S}_x \mathcal{B}_x \cos\left(\omega t\right) \\ &= -\Omega_0 \hat{S}_z - \Omega_1 \cos\left(\omega t\right) \hat{S}_x \ . \end{aligned} \tag{9.A.2}$$

Here we have defined two different Larmor frequencies

$$\Omega_0 = \gamma \mathcal{B}_z, \; \Omega_1 = \gamma \mathcal{B}_x. \tag{9.A.3}$$

The particle starts in state $\left|\psi(0)\right\rangle = \left|+z\right\rangle$, and we want to find $\left|\psi\left(t\right)\right\rangle$. Since the Hamiltonian is time dependent, we need to go back to the original Schrödinger equation

$$\frac{d}{dt}\left|\psi(t)\right\rangle = -\frac{i}{\hbar}\hat{H}\left|\psi(t)\right\rangle. \tag{9.A.4}$$

We can write the state as

$$\begin{aligned}\left|\psi(t)\right\rangle &= a(t)\left|+z\right\rangle + b(t)\left|-z\right\rangle \\ &\doteq \begin{pmatrix} a(t) \\ b(t) \end{pmatrix},\end{aligned} \tag{9.A.5}$$

where $a(t)$ and $b(t)$ are the probability amplitudes that the system is in the states $\left|+z\right\rangle$ and $\left|-z\right\rangle$, respectively. Their square magnitudes are the probabilities that the particle will be measured to have spin-up and spin-down. The initial conditions are

$$\left|\psi(0)\right\rangle \doteq \begin{pmatrix} a(0) \\ b(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{9.A.6}$$

Written in matrix form, the Hamiltonian is

$$\begin{aligned}\hat{H} &\doteq -\frac{\Omega_0\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} - \frac{\Omega_1\hbar}{2}\cos(\omega t)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= -\frac{\hbar}{2}\begin{pmatrix} \Omega_0 & \Omega_1\cos(\omega t) \\ \Omega_1\cos(\omega t) & -\Omega_0 \end{pmatrix},\end{aligned} \tag{9.A.7}$$

and the Schrödinger equation becomes

$$\frac{d}{dt}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \frac{i}{2}\begin{pmatrix} \Omega_0 & \Omega_1\cos\omega t \\ \Omega_1\cos\omega t & -\Omega_0 \end{pmatrix}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix}. \tag{9.A.8}$$

Multiplying, this becomes a set of coupled differential equations:

$$\frac{d}{dt}a(t) = \frac{i\Omega_0}{2}a(t) + \frac{i\Omega_1}{2}\cos(\omega t)\,b(t), \tag{9.A.9}$$

$$\frac{d}{dt}b(t) = \frac{i\Omega_1}{2}\cos(\omega t)\,a(t) - \frac{i\Omega_0}{2}b(t). \tag{9.A.10}$$

When performing magnetic resonance experiments, the strengths of the fields satisfy $\mathcal{B}_z >> \mathcal{B}_x$, which means $\Omega_0 >> \Omega_1$. The dominant terms in our coupled equations are then ones proportional to Ω_0. In the limit that we can ignore the Ω_1 terms completely, the solution to these equations is

$$a(t) = a(0)e^{i\Omega_0 t/2}, \tag{9.A.11}$$

$$b(t) = b(0)e^{-i\Omega_0 t/2}. \tag{9.A.12}$$

This suggests that we should factor out the exponentials, and solve eqs (9.A.9) and (9.A.10) using the substitution

$$a(t) = c(t)e^{i\Omega_0 t/2}. \tag{9.A.13}$$

$$b(t) = d(t)e^{-i\Omega_0 t/2}. \tag{9.A.14}$$

The corresponding initial conditions become

$$|\psi(0)\rangle \doteq \begin{pmatrix} c(0) \\ d(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{9.A.15}$$

Differentiating eq. (9.A.13) yields

$$\begin{aligned} \frac{d}{dt}a(t) &= \frac{i\Omega_0}{2}c(t)e^{i\Omega_0 t/2} + e^{i\Omega_0 t/2}\frac{d}{dt}c(t) \\ &= \frac{i\Omega_0}{2}a(t) + e^{i\Omega_0 t/2}\frac{d}{dt}c(t). \end{aligned} \tag{9.A.16}$$

Substituting eqs. (9.A.14) and (9.A.16) into eq. (9.A.9), we find

$$\begin{aligned} \frac{i\Omega_0}{2}a(t) + e^{i\Omega_0 t/2}\frac{d}{dt}c(t) &= \frac{i\Omega_0}{2}a(t) + \frac{i\Omega_1}{2}\cos(\omega t)d(t)e^{-i\Omega_0 t/2}, \\ \frac{d}{dt}c(t) &= \frac{i\Omega_1}{2}\cos(\omega t)d(t)e^{-i\Omega_0 t} \\ &= \frac{i\Omega_1}{4}\left(e^{i\omega t} + e^{-i\omega t}\right)d(t)e^{-i\Omega_0 t}. \end{aligned} \tag{9.A.17}$$

One last simplification yields

$$\frac{d}{dt}c(t) = \frac{i\Omega_1}{4}\left(e^{i(\omega-\Omega_0)t} + e^{-i(\omega+\Omega_0)t}\right)d(t). \tag{9.A.18}$$

Similarly, eqs. (9.A.13) and (9.A.14) can be substituted into eq. (9.A.10) to yield

$$\frac{d}{dt}d(t) = \frac{i\Omega_1}{4}\left(e^{i(\omega+\Omega_0)t} + e^{-i(\omega-\Omega_0)t}\right)c(t). \tag{9.A.19}$$

While eqs. (9.A.13) and (9.A.14) were motivated by assuming that $\Omega_0 >> \Omega_1$, we have not yet actually made that assumption. Equations (9.A.13) and (9.A.14) are a substitution, and as long as we allow $c(t)$ and $d(t)$ to have any arbitrary time dependence, eqs. (9.A.18) and (9.A.19) are completely general.

Now, however, we'll assume $\omega \cong \Omega_0$. In other words, we'll assume that the frequency of the oscillating magnetic field is nearly resonant with the Larmor frequency associated with the strong static field. The frequency of the applied field is a knob that an experimenter can easily adjust, and this condition is easy to satisfy. If $\omega \cong \Omega_0$, terms in eqs. (9.A.18) and (9.A.19) that oscillate at $\omega - \Omega_0$ will oscillate slowly, while terms that oscillate at $\omega+\Omega_0$ will oscillate rapidly. In a typical nuclear magnetic resonance experiment $\Omega_0 \sim 10^8$ rad/s (the gyromagnetic ratio of a proton is $\gamma_p = 2.68\times10^8$ s^{-1}T^{-1}, and field strengths are 1T or higher), and these rapidly oscillating terms quickly average to 0 and can be ignored. With the approximation $\omega \cong \Omega_0$, which is known as the rotating-wave approximation (RWA), eqs. (9.A.18) and (9.A.19) become

$$\frac{d}{dt}c(t)=\frac{i\Omega_1}{4}e^{i(\omega-\Omega_0)t}d(t). \tag{9.A.20}$$

$$\frac{d}{dt}d(t)=\frac{i\Omega_1}{4}e^{-i(\omega-\Omega_0)t}c(t). \tag{9.A.21}$$

9.A.2 Solution

Equations (9.A.20) and (9.A.21) represent a pair of coupled first-order linear differential equations. They are most easily solved by uncoupling them. Start by solving eq. (9.A.20) for $d(t)$:

$$d(t)=\frac{4}{i\Omega_1}e^{-i(\omega-\Omega_0)t}\frac{d}{dt}c(t). \tag{9.A.22}$$

Now differentiate eq. (9.A.20):

$$\begin{aligned}\frac{d^2}{dt^2}c(t)&=\frac{i\Omega_1}{4}\left[i(\omega-\Omega_0)\right]e^{i(\omega-\Omega_0)t}d(t)+\frac{i\Omega_1}{4}e^{i(\omega-\Omega_0)t}\frac{d}{dt}d(t)\\&=-\frac{\Omega_1}{4}(\omega-\Omega_0)e^{i(\omega-\Omega_0)t}\left[\frac{4}{i\Omega_1}e^{-i(\omega-\Omega_0)t}\frac{d}{dt}c(t)\right]\\&\quad+\frac{i\Omega_1}{4}e^{i(\omega-\Omega_0)t}\left[\frac{i\Omega_1}{4}e^{-i(\omega-\Omega_0)t}c(t)\right]\\&=i(\omega-\Omega_0)\frac{d}{dt}c(t)-\frac{\Omega_1^{\,2}}{4^2}c(t),\end{aligned} \tag{9.A.23}$$

where we’ve used eqs. (9.A.21) and (9.A.22). Rearranging yields

$$\frac{d^2}{dt^2}c(t)-i(\omega-\Omega_0)\frac{d}{dt}c(t)+\frac{\Omega_1^{\,2}}{4^2}c(t)=0\cdot \tag{9.A.24}$$

This is a second-order, linear differential equation, and $c(t)$ has been decoupled from $d(t)$.

We anticipate that the solutions to eq. (9.A.24) will be oscillatory, so let’s assume a solution of the form

$$c(t)=e^{i\omega' t}. \tag{9.A.25}$$

Substituting this into eq. (9.A.24) yields

$$\begin{gathered}(i\omega')^2 c(t) - i(\omega - \Omega_0)(i\omega')c(t) + \frac{\Omega_1{}^2}{4^2}c(t) = 0,\\ -\omega'^2 + (\omega - \Omega_0)\omega' + \frac{\Omega_1{}^2}{4^2} = 0,\end{gathered} \tag{9.A.26}$$

which is a quadratic equation for ω'. The solutions are

$$\begin{aligned}\omega'_\pm &= \frac{(\omega - \Omega_0) \pm \sqrt{(\omega - \Omega_0)^2 + (\Omega_1/2)^2}}{2}\\ &= \frac{(\omega - \Omega_0) \pm \Omega_R}{2},\end{aligned} \tag{9.A.27}$$

where we've defined the generalized Rabi frequency Ω_R (after Isidor Rabi) as[5]

$$\Omega_R \equiv \sqrt{(\omega - \Omega_0)^2 + (\Omega_1/2)^2}. \tag{9.A.28}$$

The solution for $c(t)$ is then

$$c(t) = c_1 e^{i\omega'_+ t} + c_2 e^{i\omega'_- t}, \tag{9.A.29}$$

where the constants c_1 and c_2 are used to satisfy the initial conditions. Before doing this, however, we need to find the solution for $d(t)$. This is most easily done by substituting eq. (9.A.29) into eq. (9.A.22), which yields

$$\begin{aligned}d(t) &= \frac{4}{i\Omega_1} e^{-i(\omega - \Omega_0)t} \frac{d}{dt}\left(c_1 e^{i\omega'_+ t} + c_2 e^{i\omega'_- t}\right)\\ &= \frac{4}{i\Omega_1} e^{-i(\omega - \Omega_0)t}\left(i\omega'_+ c_1 e^{i\omega'_+ t} + i\omega'_- c_2 e^{i\omega'_- t}\right).\end{aligned} \tag{9.A.30}$$

Using eqs. (9.A.29) and (9.A.30), along with the initial conditions [eq. (9.A.15)], we find

$$c_1 + c_2 = 1, \tag{9.A.31}$$

$$\omega'_+ c_1 + \omega'_- c_2 = 0. \tag{9.A.32}$$

The solution to these equations is

$$c_1 = -\frac{\omega'_-}{\omega'_+ - \omega'_-} = -\frac{\omega'_-}{\Omega_R}, \tag{9.A.33}$$

$$c_2 = \frac{\omega'_+}{\omega'_+ - \omega'_-} = \frac{\omega'_+}{\Omega_R}. \tag{9.A.34}$$

---

[5] The Rabi frequency is equal to the generalized Rabi frequency on resonance ($\omega = \Omega_0$).

These can be substituted into eqs. (9.A.29) and (9.A.30) to obtain the final expressions for $c(t)$ and $d(t)$.

9.A.3 Rabi Oscillations

We are interested in the probabilities that the particle will be measured to have either spin-up or spin-down. We assumed that the particle started with spin-up at $t = 0$, so let's examine $P(-z,t)$, which represents the probability that its spin flips to spin-down. Using eqs. (9.A.14) and (9.A.30), we find

$$\begin{aligned} P(-z,t) &= \left|b(t)\right|^2 \\ &= \left|d(t)\right|^2 \\ &= \left|\frac{4}{\Omega_1}\left(\omega'_+ c_1 e^{i\omega'_+ t} + \omega'_- c_2 e^{i\omega'_- t}\right)\right|^2 . \end{aligned} \tag{9.A.35}$$

Substituting eqs. (9.A.27), (9.A.33), and (9.A.34) into this equation yields

$$\begin{aligned} P(-z,t) &= \left|\frac{4}{\Omega_1}\left(-\frac{\omega'_+\omega'_-}{\Omega_R} e^{i\omega'_+ t} + \frac{\omega'_+\omega'_-}{\Omega_R} e^{i\omega'_- t}\right)\right|^2 \\ &= \frac{16\left(\omega'_+\omega'_-\right)^2}{\Omega_1{}^2\Omega_R{}^2}\left|\left(-e^{i\left[(\omega-\Omega_0)+\Omega_R\right]t/2} + e^{i\left[(\omega-\Omega_0)-\Omega_R\right]t/2}\right)\right|^2 \\ &= \frac{\left\{\left[(\omega-\Omega_0)+\Omega_R\right]\left[(\omega-\Omega_0)-\Omega_R\right]\right\}^2}{\Omega_1{}^2\Omega_R{}^2}\left|\left(-e^{i\Omega_R t/2} + e^{-i\Omega_R t/2}\right)\right|^2 \\ &= \frac{\left[(\omega-\Omega_0)^2 - \Omega_R{}^2\right]^2}{\Omega_1{}^2\Omega_R{}^2} 4\sin^2\left(\frac{\Omega_R t}{2}\right) \\ &= P_{\max}(\omega)\sin^2\left(\frac{\Omega_R t}{2}\right) \\ &= P_{\max}(\omega)\frac{1}{2}\left[1-\cos\left(\Omega_R t\right)\right]. \end{aligned} \tag{9.A.36}$$

Here the maximum probability is

$$P_{\max}(\omega) = \frac{4\left[(\omega-\Omega_0)^2 - \Omega_R{}^2\right]^2}{\Omega_1{}^2\Omega_R{}^2}. \tag{9.A.37}$$

Equation (9.A.36) is known as Rabi's formula, and it is plotted as a function of t in fig. 9.A.1. This figure shows that the probability of a spin flip oscillates at the generalized Rabi frequency.

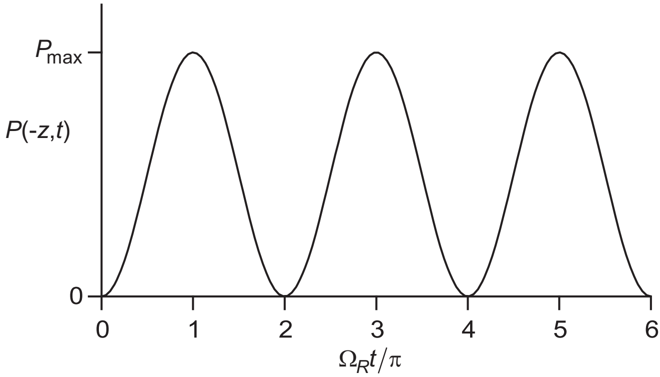


Fig 9.A.1 $P(-z,t)$ is plotted as a function of t.

We can rewrite the expression for $P_{\max}(\omega)$ by using eq. (9.A.28):

$$\begin{aligned} P_{\max}(\omega) &= \frac{4\left((\omega-\Omega_0)^2 - \left[(\omega-\Omega_0)^2 + (\Omega_1/2)^2\right]\right)^2}{\Omega_1{}^2\left[(\omega-\Omega_0)^2 + (\Omega_1/2)^2\right]} \\ &= \frac{4(\Omega_1/2)^4}{\Omega_1{}^2\left[(\omega-\Omega_0)^2 + (\Omega_1/2)^2\right]} \\ &= \frac{(\Omega_1/2)^2}{\left[(\omega-\Omega_0)^2 + (\Omega_1/2)^2\right]} . \end{aligned} \tag{9.A.38}$$

Notice that if the frequency of the driving field is perfectly on resonance ($\omega = \Omega_0$) $P_{\max}(\omega) = 1$. A plot of $P_{\max}(\omega)$ as a function of the drive frequency is shown in fig. 9.A.2. The shape of this curve is known as a Lorentzian.

One way to understand these results is as follows. For the moment assume that there is no oscillating magnetic field, $\mathcal{B}_x = 0$, and there is only the static field $\mathcal{B}_z$. The energy of the particle is given by $E = -\boldsymbol{\mu}\boldsymbol{\cdot}\boldsymbol{\mathcal{B}}$. For a spin-1/2 particle there are two energy eigenstates: in the lower energy state $E = -\mu_z \mathcal{B}_z$ and the dipole moment is parallel to $\boldsymbol{\mathcal{B}}$, while in the higher energy state $E = \mu_z \mathcal{B}_z$ and the dipole moment is antiparallel to $\boldsymbol{\mathcal{B}}$. This represents a two-level quantum system, and the energy difference between the two levels is

$$\begin{aligned} \Delta E &= 2\mu_z \mathcal{B}_z \\ &= 2\gamma S_z \mathcal{B}_z \\ &= \Omega_0 \hbar . \end{aligned} \tag{9.A.39}$$

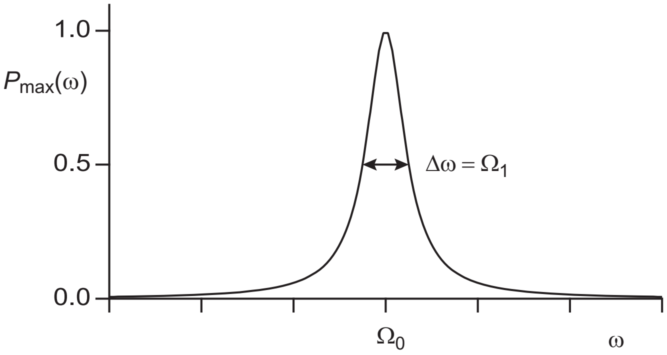


Fig 9.A.2 $P_{\max}\left(\omega\right)$ is plotted as a function of ω.

Without the oscillating field, a particle placed in either one of these energy levels is in an energy eigenstate, and will thus remain in that state. In other words, the spin of the particle will not flip.

In order to get the spin to flip, and the energy to change, the particle must absorb or emit energy. For example, in order to move from the lower energy state to the higher energy state, the particle must absorb energy $E = \Omega_0 \hbar$. This energy comes from the oscillating magnetic field.[6] Since this field oscillates at frequency ω, it contains photons of energy $E = \hbar\omega$. The particle will be most likely to absorb photons if the photon energy is nearly equal to the transition energy, that is, if $\omega \cong \Omega_0$.

If we drive the transition on resonance, then $P_{\max}\left(\omega = \Omega_0\right) = 1$. If the oscillating field is turned on for time $t = \pi / \Omega_R$, and then turned off again, we will have switched from state $\left|+z\right\rangle$ to state $\left|-z\right\rangle$ with 100% certainty, and the particle will remain in state $\left|-z\right\rangle$. A pulse of this duration is referred to as a π-pulse, because $\Omega_R t = \pi$. By choosing pulses of different durations, one can create states in different linear combinations of $\left|+z\right\rangle$ and $\left|-z\right\rangle$.

For a given magnetic field strength $\mathcal{B}_z$, the nucleus of an atom has a corresponding resonance frequency Ω_0. Different nuclei have different resonance frequencies. In nuclear magnetic resonance the oscillation frequency ω is varied, and each of the resonances in a sample is identified. This can be used to identify which atoms are present.

If the static magnetic field strength varies with position (e.g., it has a gradient), then the resonance frequency will vary with position. This can be used to obtain position dependent information about a sample, which forms the basis of magnetic resonance imaging (MRI).

9.A.4 PROBLEMS

9.A.1 What is the resonance frequency of a hydrogen nucleus in a 4 T magnetic field?

---

[6] Oscillating magnetic fields generate electric fields, so it's really an electromagnetic field.

9.A.2 A hydrogen atom is placed in a static magnetic field with a strength of 8.5 T along the z-direction. A field oscillating along the x-direction with strength 10^{-6} T is then pulsed on and off. In order to be 100% certain that the nuclear spin will flip:

(a) What frequency should the field oscillate at?

(b) What should be the duration of the oscillating field pulse?

CHAPTER 10

Position and Momentum

In introductory physics we usually begin by discussing kinematics. We write down equations that describe objects in terms of their positions, velocities, and accelerations. Clearly, we are taking a very different approach to quantum mechanics. You may be surprised that we're in chapter 10, and we're only just now getting to a discussion of position. Why is that?

Recall that we began with a discussion of polarization. Polarization is a two-dimensional (2-D) system, so we can describe polarization states by 2-D vectors, and polarization operators by 2 × 2 matrices. We then moved on to spin-1/2, which is also 2-D, and then to spin-*s*, which is still discrete. Position and momentum are different in that they are continuous, not discrete. In principle a particle can be anywhere, and moving with any momentum. In this chapter we'll extend the formalism we have been developing to continuous variable systems.

10.1 POSITION

For the moment we'll concentrate on position in one dimension. The position operator $\hat{x}$ is Hermitian, and corresponds to an observable; we can measure the position of an object. The eigenstates of $\hat{x}$ correspond to states of definite position:

$$\hat{x}\left|x\right\rangle = x\left|x\right\rangle. \tag{10.1}$$

The states $\left|x\right\rangle$ are eigenstates of a Hermitian operator that corresponds to an observable, so they form a complete set. Because of this, we should be able to represent a general state $\left|\psi\right\rangle$ in the position basis. The complication is that the eigenvalues x are members of the set of all real numbers—they represent a continuous variable. We need to learn how to deal with this.

10.1.1 The Wave Function

What do we mean by the representation of a state $\left|\psi\right\rangle$ in a Hilbert space described by a continuous variable? To answer this, let's first recall what we mean by the representation of a state in a discrete Hilbert space; for example, one whose basis states are $\left|j\right\rangle$. In this space the sum of the projection operators onto the states $\left|j\right\rangle$ is equal to the identity operator:

$$\hat{1} = \sum_j \left|j\right\rangle\left\langle j\right|. \tag{10.2}$$

Applying this operator to an arbitrary state $\left|\psi\right\rangle$ yields

$$\begin{aligned} \hat{1}\left|\psi\right\rangle &= \sum_j \left|j\right\rangle\left\langle j|\psi\right\rangle \\ \left|\psi\right\rangle &= \sum_j \psi_j \left|j\right\rangle . \end{aligned} \tag{10.3}$$

It is the complex coefficients ψ_j that form the representation of $\left|\psi\right\rangle$ in the $\left|j\right\rangle$-basis.

In a continuous basis we must use an integral to express the identity operator. In terms of the position states we have

$$\hat{1} = \int_{-\infty}^{\infty} dx \left|x\right\rangle\left\langle x\right|. \tag{10.4}$$

Applying this to an arbitrary state yields

$$\begin{aligned} \hat{1}\left|\psi\right\rangle &= \int_{-\infty}^{\infty} dx \left|x\right\rangle\left\langle x|\psi\right\rangle , \\ \left|\psi\right\rangle &= \int_{-\infty}^{\infty} dx \left\langle x|\psi\right\rangle \left|x\right\rangle \\ &= \int_{-\infty}^{\infty} dx \psi(x) \left|x\right\rangle , \end{aligned} \tag{10.5}$$

where we've defined

$$\psi(x) \equiv \left\langle x|\psi\right\rangle . \tag{10.6}$$

$\psi(x)$ is a complex function called the wave function. The wave function is the representation of $\left|\psi\right\rangle$ in the $\left|x\right\rangle$-basis. The wave function $\psi(x)$ and the state vector $\left|\psi\right\rangle$ both contain all of the information about the state of the system. Whether $\psi(x)$ or $\left|\psi\right\rangle$ is more useful to you at a given time depends on what you want to calculate.[1]

---

[1] In eq. (10.4) the identity operator is dimensionless, and the differential dx has units of length. This means that the states $\left|x\right\rangle$, and hence the wave functions $\psi(x)$, must have units of $1/\sqrt{\text{length}}$. To eliminate this complication, we could work with dimensionless units. If we define $x = x_{actual}/L$, where L is a characteristic length scale for the problem, then x is dimensionless.

If you've studied quantum mechanics and the Schrödinger equation before, then you should already be familiar with $\psi(x)$. The wave function we're talking about here, and the wave function you learned as a solution to the Schrödinger equation, are the same. This fact will become more obvious in chapter 11.

The square magnitude of the wave function is real, positive, and properly normalized, as long as the original state $|\psi\rangle$ is normalized:

$$\begin{aligned}\langle\psi|\psi\rangle &= \langle\psi|\hat{1}|\psi\rangle \\ &= \langle\psi|\left[\int_{-\infty}^{\infty} dx\,|x\rangle\langle x|\right]|\psi\rangle \\ &= \int_{-\infty}^{\infty} dx\,\langle\psi|x\rangle\langle x|\psi\rangle \\ &= \int_{-\infty}^{\infty} dx\,\psi^*(x)\psi(x) \\ &= \int_{-\infty}^{\infty} dx\,|\psi(x)|^2 \\ &= 1\,.\end{aligned} \tag{10.7}$$

If we apply the identity operator of eq. (10.4) to the position eigenstate $|x_0\rangle$, we find

$$\begin{aligned}|x_0\rangle &= \hat{1}|x_0\rangle \\ &= \left[\int_{-\infty}^{\infty} dx\,|x\rangle\langle x|\right]|x_0\rangle \\ &= \int_{-\infty}^{\infty} dx\,|x\rangle\langle x|x_0\rangle \\ &= \int_{-\infty}^{\infty} dx\,|x\rangle\psi_{x_0}(x)\end{aligned} \tag{10.8}$$

In order for the final integral in this expression to yield the original state, it must be the case that the wave function of a position eigenstate is a delta function:

$$\psi_{x_0}(x) = \langle x|x_0\rangle = \delta(x - x_0). \tag{10.9}$$

The delta function is discussed in complement 10.A.

10.1.2 Expectation Values

The wave function can be used to calculate expectation values. For example, the average position of a particle is given by the expectation value $\langle x \rangle$, and it can be computed with the help of eq. (10.4):

$$\begin{aligned}
\langle x \rangle &= \langle \psi | \hat{x} | \psi \rangle \\
&= \langle \psi | \hat{x}\hat{1} | \psi \rangle \\
&= \langle \psi | \hat{x} \left[\int_{-\infty}^{\infty} dx \, |x\rangle\langle x| \right] | \psi \rangle \\
&= \int_{-\infty}^{\infty} dx \, \langle \psi | \hat{x} | x \rangle \langle x | \psi \rangle \\
&= \int_{-\infty}^{\infty} dx \, \langle \psi | x | x \rangle \langle x | \psi \rangle \\
&= \int_{-\infty}^{\infty} dx \, \langle \psi | x \rangle x \langle x | \psi \rangle \\
&= \int_{-\infty}^{\infty} dx \, \psi^*(x) x \psi(x). \qquad (10.10)
\end{aligned}$$

This can also be written as

$$\langle x \rangle = \int_{-\infty}^{\infty} dx \, x |\psi(x)|^2. \qquad (10.11)$$

Generalizing, the average of a function of *x*, $f(x)$ is

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} dx \, \psi^*(x) f(x) \psi(x). \qquad (10.12)$$

We'll soon see why the functions inside the integral are typically ordered in this way.

Comparing eq. (10.11) to eq. (1.23) tells us that $|\psi(x)|^2$ is a probability density, which is an interpretation originally due to Max Born. This means that $|\psi(x)|^2 \, dx$ is the probability that an object will be found in the interval between x and $x + dx$. To find the probability that an object will be located in a non-infinitesimal interval, for example $a \le x \le b$, we need to integrate over that interval:

$$P(a \le x \le b) = \int_{a}^{b} dx \, |\psi(x)|^2. \qquad (10.13)$$

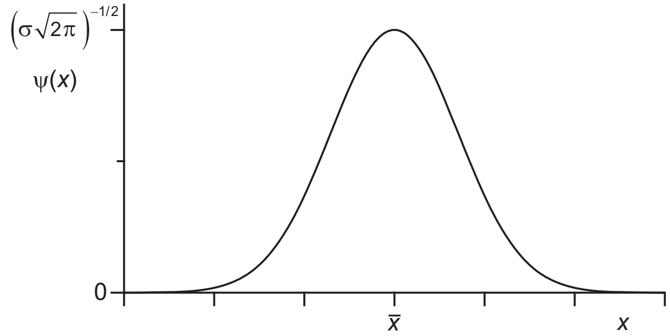


Fig 10.1 A Gaussian wave packet.

10.1.3 Wave Packets

The wave function of a real particle cannot be a delta function, because that would mean that the position of the particle is defined with infinite precision. Real particles are never perfectly localized, and therefore, physically realistic wave functions must have a finite extent. Wave functions that have some spread to them are frequently referred to as wave packets.

The canonical example of a wave packet is the Gaussian wave packet, shown in fig. 10.1. It has the functional form

$$\psi(x) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^{1/2}} e^{-(x-\bar{x})^2/4\sigma^2}, \tag{10.14}$$

where $\bar{x}$ is a parameter that determines the average position (i.e., the location), and σ is a parameter that determines the spread. Using the properties of Gaussian distributions described in complement 10.A, you can show that the expectation value and standard deviation of x, for the wave packet of eq. (10.14), are $\langle x \rangle = \bar{x}$ and $\Delta x = \sigma$. Further properties of wave packets will be explored in complement 11.A.

Let's look at an example which uses a wave function to perform some useful calculations.

EXAMPLE 10.1

For a particle whose wave function is given by

$$\psi(x) \propto \begin{cases} \cos\left(\dfrac{\pi x}{2a}\right) & -a < x < a \\ 0 & \textit{elsewhere} \end{cases}, \tag{10.15}$$

(a) normalize the wave function, (b) calculate the probability that the particle will be found in the interval $0 \le x \le a/2$, and (c) calculate $\langle x \rangle$ and Δx.

(a) The normalization is given by eq. (10.7). Let the normalization constant be c; we then have

$$\begin{aligned}\int_{-\infty}^{\infty} dx\left|\psi(x)\right|^2 &= c^2\int_{-a}^{a} dx\cos^2\left(\frac{\pi x}{2a}\right)\\ &= c^2\frac{1}{2\pi}\left[a\sin\left(\frac{\pi x}{a}\right)+\pi x\right]_{x=-a}^{x=a}\\ &= c^2\frac{1}{2\pi}\left(0+a\pi+0+a\pi\right)\\ &= c^2 a\\ &= 1.\end{aligned} \tag{10.16}$$

So $c = 1/\sqrt{a}$ and

$$\psi(x) = \begin{cases}\dfrac{1}{\sqrt{a}}\cos\left(\dfrac{\pi x}{2a}\right) & -a < x < a\\ 0 & elsewhere\end{cases}. \tag{10.17}$$

(b) This probability can be calculated using eq. (10.13):

$$\begin{aligned}P\left(0\le x\le a/2\right) &= \frac{1}{a}\int_{0}^{a/2} dx\cos^2\left(\frac{\pi x}{2a}\right)\\ &= \frac{1}{2\pi a}\left[a\sin\left(\frac{\pi x}{a}\right)+\pi x\right]_{x=0}^{x=a/2}\\ &= \frac{1}{2\pi}\left(1+\frac{\pi}{2}\right)\\ &= 0.409.\end{aligned} \tag{10.18}$$

(c) To calculate these quantities we can use eq. (10.12)

$$\begin{aligned}\langle x\rangle &= \int_{-\infty}^{\infty} dx\,\psi^*(x)x\psi(x)\\ &= \frac{1}{a}\int_{-a}^{a} dx\,x\cos^2\left(\frac{\pi x}{2a}\right)\\ &= \left.\frac{x^2}{4a}+\frac{a\cos\left(\frac{\pi x}{a}\right)}{2\pi^2}+\frac{x\sin\left(\frac{\pi x}{a}\right)}{2\pi}\right|_{x=-a}^{x=a}\\ &= \left(\frac{a}{4}-\frac{a}{2\pi^2}+0\right)-\left(\frac{a}{4}-\frac{a}{2\pi^2}+0\right)\\ &= 0\end{aligned} \tag{10.19}$$

We could have obtained this same result by noting that the integrand is an odd function, and the integral is symmetric about the origin; therefore, the integral must be zero. To calculate Δx we first find $\left\langle x^2 \right\rangle$:

$$\begin{aligned}
\left\langle x^2 \right\rangle &= \int_{-\infty}^{\infty} dx\, \psi^*(x) x^2 \psi(x) \\
&= \frac{1}{a} \int_{-a}^{a} dx\, x^2 \cos^2\left(\frac{\pi x}{2a}\right) \\
&= \frac{x^3}{6a} + \frac{ax\cos\left(\frac{\pi x}{a}\right)}{\pi^2} + \frac{\left[\pi^2 x^2 - 2a^2\right]\sin\left(\frac{\pi x}{a}\right)}{\pi^3}\Bigg|_{x=-a}^{x=a} \qquad (10.20) \\
&= \left(\frac{a^2}{6} - \frac{a^2}{\pi^2} + 0\right) - \left(-\frac{a^2}{6} + \frac{a^2}{\pi^2} + 0\right) \\
&= 2a^2\left(\frac{1}{6} - \frac{1}{\pi^2}\right) = (0.131)a^2.
\end{aligned}$$

Using this:

$$\Delta x = \left(\left\langle x^2 \right\rangle - \left\langle x \right\rangle^2\right)^{1/2} = (0.131)^{1/2}\, a = (0.362)a. \tag{10.21}$$

10.2 MOMENTUM

10.2.1 The Translation Operator

To translate a state from one location to another we can use the translation operator $\hat{T}(D)$, which translates the state by the distance D. Applied to a position eigenstate this operator yields

$$\hat{T}(D)\left|x\right\rangle = \left|x + D\right\rangle. \tag{10.22}$$

Applied to a general state the translation operator yields

$$\begin{aligned}
\left|\varphi\right\rangle &= \hat{T}(D)\left|\psi\right\rangle \\
&= \int_{-\infty}^{\infty} dx'\, \hat{T}(D)\left|x'\right\rangle\left\langle x'\middle|\psi\right\rangle \qquad (10.23) \\
&= \int_{-\infty}^{\infty} dx' \left|x' + D\right\rangle \psi(x'),
\end{aligned}$$

where we've used eq. (10.4). The wave function of this translated state is

$$\begin{aligned}\varphi(x) &= \langle x|\varphi\rangle \\ &= \int_{-\infty}^{\infty} dx' \langle x|x'+D\rangle \psi(x') \\ &= \int_{-\infty}^{\infty} dx' \delta\left[x-(x'+D)\right]\psi(x') \\ &= \int_{-\infty}^{\infty} dx' \delta\left[-(x'-x+D)\right]\psi(x').\end{aligned} \tag{10.24}$$

Using the fact that $\delta(-x)=\delta(x)$ [eq. (10.A.13), complement 10.A], we can simplify this as

$$\begin{aligned}\varphi(x) &= \int_{-\infty}^{\infty} dx' \delta\left[x'-(x-D)\right]\psi(x') \\ &= \psi(x-D).\end{aligned} \tag{10.25}$$

The new wave function has the same functional form as the old, but it has been shifted by *D*. For example, if *D* is positive then $\psi(x-D)$ is simply $\psi(x)$ shifted to the right by *D*, as shown in fig. 10.2.

Following the examples of our treatment of the rotation operator $\hat{R}(\theta,\boldsymbol{u}_n)$ (sec. 7.5), and the time evolution operator $\hat{U}(t)$ (secs. 9.1 and 9.2), we'll express the operator that performs an infinitesimal translation *dx* as

$$\hat{T}(dx)=\hat{1}-\frac{i}{\hbar}\hat{p}_x dx. \tag{10.26}$$

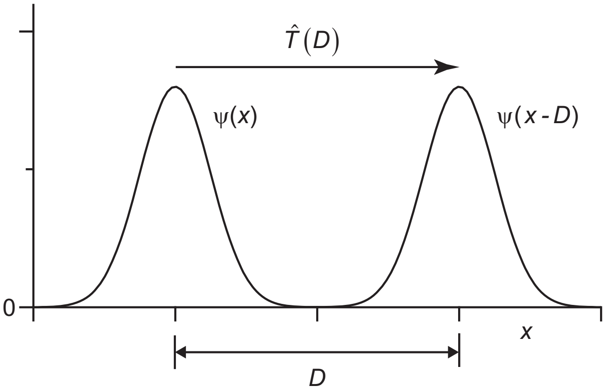


Fig 10.2 The translation operator shifts the position of a wave function, but leaves its shape unaltered.

Here $\hat{p}_x$ is the operator corresponding to the x-component of momentum, and we use it as the generator of translation because momentum is the generator of translation in classical physics. $\hat{p}_x$ is Hermitian (as our previous generators were), and corresponds to an observable. In analogy with secs. 7.5 and 9.2, we can use eq. (10.26) to find the translation operator for a macroscopic translation D, which you'll prove in the problems to be

$$\hat{T}(D) = e^{-i\hat{p}_x D/\hbar}. \tag{10.27}$$

In the problems you'll also prove that the translation operator is unitary, and that

$$\hat{T}^{\dagger}(D) = \hat{T}(-D). \tag{10.28}$$

10.2.2 The Position-Momentum Commutator

Let's examine the commutation relationship between $\hat{x}$ and $\hat{T}(dx)$. We'll do this by applying the commutator to the state $|x\rangle$:

$$\begin{aligned}\left[\hat{x},\hat{T}(dx)\right]|x\rangle &= \hat{x}\hat{T}(dx)|x\rangle - \hat{T}(dx)\hat{x}|x\rangle \\ &= \hat{x}|x+dx\rangle - \hat{T}(dx)x|x\rangle \\ &= (x+dx)|x+dx\rangle - x|x+dx\rangle \\ &= dx|x+dx\rangle \\ &\cong dx|x\rangle,\end{aligned} \tag{10.29}$$

where the last line is accurate to first order in dx (i.e., any corrections are of order dx^2). Since eq. (10.29) holds for any arbitrary position state, and the position states form a complete set, this equation must hold for the operators. In other words, it must be the case that

$$\left[\hat{x},\hat{T}(dx)\right] = dx\,. \tag{10.30}$$

If we substitute eq. (10.26) into eq. (10.30) we find

$$\begin{aligned}\left[\hat{x},\hat{1}-\frac{i}{\hbar}\hat{p}_x dx\right] &= \underbrace{\cancel{\left[\hat{x},\hat{1}\right]}}_{=0} - \left[\hat{x},\frac{i}{\hbar}\hat{p}_x dx\right] \\ &= -\frac{i}{\hbar}dx\left[\hat{x},\hat{p}_x\right] \\ &= dx\,.\end{aligned} \tag{10.31}$$

Rearranging this yields

$$\left[\hat{x},\hat{p}_x\right]=i\hbar. \tag{10.32}$$

In general, the operators corresponding to position along one axis, and momentum along a parallel axis, do not commute; for example

$$\left[\hat{y},\hat{p}_y\right]=\left[\hat{z},\hat{p}_z\right]=i\hbar. \tag{10.33}$$

In contrast, operators corresponding to position along one axis, and momentum along a perpendicular axis, do commute, for example,

$$\left[\hat{x},\hat{p}_y\right]=\left[\hat{x},\hat{p}_z\right]=0. \tag{10.34}$$

We know that there must be an indeterminacy relation for noncommuting observables. Using eqs. (5.21) and (10.32) we find that

$$\Delta x\Delta p_x \geq \frac{1}{2}\left|\left\langle\left[\hat{x},\hat{p}_x\right]\right\rangle\right| = \frac{\hbar}{2}. \tag{10.35}$$

This is the famous Heisenberg indeterminacy principle, which says that the position and the momentum of a particle are not simultaneously well defined.[2]

10.2.3 Position Representation of the Momentum Operator

Let's apply the operator for an infinitesimal translation to an arbitrary state:

$$\begin{aligned}\hat{T}(dx)\left|\psi\right\rangle &= \int_{-\infty}^{\infty} dx'\left|x'\right\rangle\left\langle x'\right|\hat{T}(dx)\left|\psi\right\rangle \\ &= \int_{-\infty}^{\infty} dx'\left|x'\right\rangle\left\langle x'\right|\hat{T}^{\dagger}(-dx)\left|\psi\right\rangle \\ &= \int_{-\infty}^{\infty} dx'\left|x'\right\rangle\left\langle x'-dx\right|\psi\right\rangle \\ &= \int_{-\infty}^{\infty} dx'\left|x'\right\rangle\psi\left(x'-dx\right),\end{aligned} \tag{10.36}$$

where we've used eq. (10.28). To first order in *dx,* we can expand the wave function $\psi\left(x'-dx\right)$ in a power series as

---

[2] More commonly this is referred to as the Heisenberg uncertainty principle. See sec 5.5 for a discussion of my reason for using "indeterminacy" instead of "uncertainty."

$$\psi\left(x'-dx\right)=\psi\left(x'\right)-dx\frac{\partial}{\partial x'}\psi\left(x'\right). \tag{10.37}$$

Substituting eqs. (10.26) and (10.37) into eq. (10.36) yields

$$\begin{aligned}
\left(\hat{1}-\frac{i}{\hbar}\hat{p}_x dx\right)\left|\psi\right\rangle &= \int_{-\infty}^{\infty} dx'\left|x'\right\rangle\left[\psi\left(x'\right)-dx\frac{\partial}{\partial x'}\psi\left(x'\right)\right],\\
\left\langle x\right|\left(\hat{1}-\frac{i}{\hbar}\hat{p}_x dx\right)\left|\psi\right\rangle &= \int_{-\infty}^{\infty} dx'\left\langle x|x'\right\rangle\left[\psi\left(x'\right)-dx\frac{\partial}{\partial x'}\psi\left(x'\right)\right],\\
\psi\left(x\right)-\frac{i}{\hbar}dx\left\langle x\right|\hat{p}_x\left|\psi\right\rangle &= \int_{-\infty}^{\infty} dx'\,\delta\left(x-x'\right)\left[\psi\left(x'\right)-dx\frac{\partial}{\partial x'}\psi\left(x'\right)\right]\\
&=\psi\left(x\right)-dx\frac{\partial}{\partial x}\psi\left(x\right),
\end{aligned} \tag{10.38}$$

where we've also used eqs. (10.9) and (10.A.8) (from complement 10.A). We can simplify this further as

$$\begin{aligned}
\left\langle x\right|\hat{p}_x\left|\psi\right\rangle &= \frac{\hbar}{i}\frac{\partial}{\partial x}\psi\left(x\right)\\
&= -i\hbar\frac{\partial}{\partial x}\left\langle x|\psi\right\rangle .
\end{aligned} \tag{10.39}$$

Since this is true for any state, it means

$$\left\langle x\right|\hat{p}_x = -i\hbar\frac{\partial}{\partial x}\left\langle x\right| . \tag{10.40}$$

We can use this equation to calculate the expectation value of p_x as

$$\begin{aligned}
\left\langle p_x\right\rangle &= \left\langle\psi\right|\hat{p}_x\left|\psi\right\rangle\\
&= \int_{-\infty}^{\infty} dx\left\langle\psi|x\right\rangle\left\langle x\right|\hat{p}_x\left|\psi\right\rangle\\
&= \int_{-\infty}^{\infty} dx\,\psi^*\left(x\right)\left(-i\hbar\frac{\partial}{\partial x}\right)\left\langle x|\psi\right\rangle\\
&= \int_{-\infty}^{\infty} dx\,\psi^*\left(x\right)\left(-i\hbar\frac{\partial}{\partial x}\right)\psi\left(x\right).
\end{aligned} \tag{10.41}$$

Furthermore, repeated applications of eq (10.40) tell us that

$$\left\langle p_x^n \right\rangle = \int_{-\infty}^{\infty} dx\, \psi^*(x) \left(-i\hbar \frac{\partial}{\partial x}\right)^n \psi(x). \tag{10.42}$$

The ordering inside the integral is important here. Equations (10.40)–(10.42) tell us that $\hat{p}_x$ is represented in the position basis as

$$\hat{p}_x \doteq -i\hbar \frac{\partial}{\partial x}. \tag{10.43}$$

EXAMPLE 10.2

Does the wave function of eq. (10.17) satisfy the indeterminacy principle?

From example 10.1 we know that for this wave function $\Delta x = (0.362)a$, so we need to calculate Δp_x. Using eq. (10.41) we find,

$$\begin{aligned} \left\langle p_x \right\rangle &= \frac{1}{a} \int_{-a}^{a} dx \cos\left(\frac{\pi x}{2a}\right) \left(-i\hbar \frac{\partial}{\partial x}\right) \cos\left(\frac{\pi x}{2a}\right) \\ &= \frac{i\hbar\pi}{2a^2} \int_{-a}^{a} dx \cos\left(\frac{\pi x}{2a}\right) \sin\left(\frac{\pi x}{2a}\right) \\ &= 0\,. \end{aligned} \tag{10.44}$$

This integral is 0 because the integrand is an odd function, and the integral is symmetric about the origin.

We can find $\left\langle p_x^2 \right\rangle$ using eq. (10.42):

$$\begin{aligned} \left\langle p_x^2 \right\rangle &= \frac{1}{a} \int_{-a}^{a} dx \cos\left(\frac{\pi x}{2a}\right) \left(-i\hbar \frac{\partial}{\partial x}\right)^2 \cos\left(\frac{\pi x}{2a}\right) \\ &= \frac{\hbar^2\pi^2}{4a^2} \frac{1}{a} \int_{-a}^{a} dx \cos^2\left(\frac{\pi x}{2a}\right) \\ &= \frac{\hbar^2\pi^2}{4a^2} \int_{-\infty}^{\infty} dx \left|\psi(x)\right|^2 \\ &= \frac{\hbar^2\pi^2}{4a^2}\,. \end{aligned} \tag{10.45}$$

The standard deviation of p_x becomes

$$\Delta p_x = \left(\left\langle p_x^2 \right\rangle - \left\langle p_x \right\rangle^2 \right)^{1/2} = \frac{\hbar\pi}{2a}, \tag{10.46}$$

and the uncertainty product is thus

$$\Delta x \Delta p_x = (0.362)a \frac{\hbar\pi}{2a} = (1.137)\frac{\hbar}{2} > \frac{\hbar}{2}. \tag{10.47}$$

The indeterminacy principle is satisfied, as it must be.

In this example we found that $\langle p_x \rangle = 0$. We could have obtained this result simply by looking at the wave function, without doing any calculation. That's because if we examine the expression for $\langle p_x \rangle$ in eq. (10.41), we notice that it contains a factor of *i*. If the wave function is real, eq. (10.41) implies that $\langle p_x \rangle$ will be purely imaginary. However, p_x is an observable, so all measurements of it yield real values, and $\langle p_x \rangle$ must be real. The only way to reconcile these seemingly contradictory facts is if $\langle p_x \rangle = 0$ when the wave function $\psi(x)$ is real.

10.3 THE MOMENTUM BASIS

10.3.1 Momentum States

The momentum operator $\hat{p}$ is Hermitian and corresponds to an observable, therefore it must have eigenstates with real eigenvalues.[3] These states correspond to states of definite linear momentum:

$$\hat{p}|p\rangle = p|p\rangle. \tag{10.48}$$

The momentum states also form a complete set

$$\hat{1} = \int_{-\infty}^{\infty} dp\,|p\rangle\langle p|, \tag{10.49}$$

and, like the position states, they are orthogonal

$$\langle p|p'\rangle = \delta(p - p'). \tag{10.50}$$

The wave function in momentum space is defined as

$$\tilde{\psi}(p) = \langle p|\psi\rangle. \tag{10.51}$$

Here we've added a tilde to distinguish the momentum basis wave function from the wave function in the position basis. This wave function represents a probability amplitude, and $|\tilde{\psi}(p)|^2\, dp$ is the probability that the particle will have momentum between p and $p + dp$.[4]

The wave function of a momentum eigenstate $|p\rangle$ is given in the position basis as

$$\psi_p(x) = \langle x|p\rangle. \tag{10.52}$$

---

[3] From now on, unless there is the potential for confusion, we'll assume that we're working in one dimension and write $\hat{p}$ instead of $\hat{p}_x$.

[4] Note that the momentum states and wave function have units of $1/\sqrt{\text{momentum}}$.

To calculate this we can apply $\langle x|$ to the left of eq. (10.48):

$$\begin{aligned}
\langle x|\hat{p}|p\rangle &= \langle x|p|p\rangle, \\
-i\hbar\frac{\partial}{\partial x}\langle x|p\rangle &= p\langle x|p\rangle, \\
-i\hbar\frac{\partial}{\partial x}\psi_p(x) &= p\psi_p(x),
\end{aligned} \tag{10.53}$$

where we've also used eq. (10.40). Equation (10.53) is a differential equation for $\psi_p(x)$, and the solution is

$$\psi_p(x) = ce^{ipx/\hbar}, \tag{10.54}$$

where c is a normalization constant. We can determine the normalization constant from the orthogonality condition, eq. (10.50):

$$\begin{aligned}
\delta(p-p') &= \langle p|p'\rangle \\
&= \int_{-\infty}^{\infty} dx\, \langle p|x\rangle\langle x|p'\rangle \\
&= \int_{-\infty}^{\infty} dx\, \psi_p^*(x)\psi_{p'}(x) \\
&= |c|^2 \int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar} e^{ip'x/\hbar} \\
&= |c|^2 \int_{-\infty}^{\infty} dx\, e^{-i(p-p')x/\hbar}\ .
\end{aligned} \tag{10.55}$$

Comparing this to eq. (10.A.35) of complement 10.A tells us that $c = 1/\sqrt{2\pi\hbar}$, so the properly normalized wave function of the momentum eigenstate is

$$\psi_p(x) = \langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{ipx/\hbar}. \tag{10.56}$$

We'll discuss the physical meaning of this wave function below.

10.3.2 Changing Representations

Given the wave function in the position basis $\psi(x)$, how can we determine the wave function in the momentum basis $\tilde{\psi}(p)$? One way is to use eq. (10.4), and we find

$$\begin{aligned}\tilde{\psi}(p) &= \langle p|\psi\rangle \\ &= \int_{-\infty}^{\infty} dx\, \langle p|x\rangle\langle x|\psi\rangle \\ &= \int_{-\infty}^{\infty} dx\, \psi_p^*(x)\psi(x) \\ &= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar}\psi(x),\end{aligned} \tag{10.57}$$

which means that $\tilde{\psi}(p)$ is a Fourier transform of $\psi(x)$ (see complement 10.A). To change in the other direction, you'll show in the problems that

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dp\, e^{ipx/\hbar}\tilde{\psi}(p); \tag{10.58}$$

$\psi(x)$ is the inverse Fourier transform of $\tilde{\psi}(p)$.

10.3.3 Physical Interpretation

Consider a particle that's in a state of definite momentum $|p_0\rangle$. Its momentum basis wave function is given by

$$\langle p|p_0\rangle = \tilde{\psi}_{p_0}(p) = \delta(p - p_0), \tag{10.59}$$

and its position basis wave function is

$$\langle x|p_0\rangle = \psi_{p_0}(x) = \frac{1}{\sqrt{2\pi\hbar}}e^{ip_0x/\hbar}. \tag{10.60}$$

Since the momentum basis wave function is a delta function, it is infinitely narrow, and $\Delta p = 0$. The indeterminacy relation of eq. (10.35) then tells us that $\Delta x \to \infty$. The probability density for finding the particle at a particular position is given by the square magnitude of eq. (10.60), which is

$$\left|\psi_{p_0}(x)\right|^2 = \frac{1}{2\pi\hbar}. \tag{10.61}$$

Notice that this probability density is constant over all space, which is consistent with an infinite standard deviation.

The wave function of eq. (10.60) corresponds to what we call a traveling wave. It has a definite momentum, so it's "traveling," but it exists everywhere in space—it's not localized anywhere. Of course, the idea of a particle with a perfectly defined momentum and an infinite uncertainty in its position is an idealization. Real particles must have some spread to their momenta, and be at least partly localized.

The indeterminacy principle tells us that the more localized the particle is (smaller Δx), the larger the spread of the momentum must be (larger Δp). The fact that $\psi(x)$ and $\tilde{\psi}(p)$ are Fourier transform pairs enforces the indeterminacy principle. We know that if one member of a Fourier transform pair becomes "narrower" the other becomes "fatter"; this is a property of the Fourier transform. Other properties of the Fourier transform can be found in complement 10.A.

EXAMPLE 10.3

Calculate the momentum representation of the wave function of eq. (10.17). Show that $\langle p \rangle$ and $\langle p^2 \rangle$ calculated using this representation agree with the values obtained in example 10.2.

Using eq. (10.57)

$$\begin{aligned}
\tilde{\psi}(p) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar} \psi(x) \\
&= \frac{1}{\sqrt{2\pi a\hbar}} \int_{-a}^{a} dx\, e^{-ipx/\hbar} \cos\left(\frac{\pi x}{2a}\right) \\
&= \sqrt{\frac{2a\hbar}{\pi}} e^{-ipx/\hbar} \left. \frac{\pi\hbar \sin\left(\frac{\pi x}{2a}\right) - i2ap\cos\left(\frac{\pi x}{2a}\right)}{\pi^2\hbar^2 - 4a^2p^2} \right|_{x=-a}^{x=a} \qquad (10.62) \\
&= \sqrt{\frac{2a\hbar}{\pi}} \frac{1}{\pi^2\hbar^2 - 4a^2p^2} \left(e^{-ipa/\hbar}\pi\hbar + e^{ipa/\hbar}\pi\hbar \right) \\
&= \sqrt{\frac{2a\pi}{\hbar}} \frac{2}{\pi^2 - \left(\frac{2pa}{\hbar}\right)^2} \cos\left(\frac{pa}{\hbar}\right)
\end{aligned}$$

The average of the momentum can be calculated from the momentum basis wave function using

$$\begin{aligned}
\langle p \rangle &= \langle \psi | \hat{p} | \psi \rangle \\
&= \int_{-\infty}^{\infty} dp \langle \psi | \hat{p} | p \rangle \langle p | \psi \rangle \qquad (10.63) \\
&= \int_{-\infty}^{\infty} dp\, \tilde{\psi}^*(p)\, p \tilde{\psi}(p).
\end{aligned}$$

Substituting for $\tilde{\psi}(p)$ using eq. (10.62) yields

$$\begin{aligned}\langle p\rangle &= \frac{8a\pi}{\hbar}\int_{-\infty}^{\infty} dp\, p\left(\frac{\cos\left(\frac{pa}{\hbar}\right)}{\pi^2-\left(\frac{2pa}{\hbar}\right)^2}\right)^2 \\ &= 0\,.\end{aligned} \tag{10.64}$$

This integral is 0 because the integrand is an odd function. We can find $\langle p^2\rangle$ in a similar manner:

$$\langle p^2\rangle = \frac{8a\pi}{\hbar}\int_{-\infty}^{\infty} dp\, p^2\left(\frac{\cos\left(\frac{pa}{\hbar}\right)}{\pi^2-\left(\frac{2pa}{\hbar}\right)^2}\right)^2 . \tag{10.65}$$

This integral is most easily calculated using a program such as Maple or Mathematica. The result is

$$\langle p^2\rangle = \frac{\hbar^2\pi^2}{4a^2}, \tag{10.66}$$

so our values for $\langle p\rangle$ and $\langle p^2\rangle$ agree with those from Example 10.2.

10.4 PROBLEMS

10.1 For a particle whose wave function is given by

$$\psi(x) \propto \text{rect}\left(\frac{x}{a}\right), \tag{10.67}$$

(a) normalize the wave function, (b) calculate $\langle x\rangle$ and Δx. The function $\text{rect}(x)$ is defined in complement 10.A [eq. (10.A.1)].

10.2 For a particle whose wave function is given by

$$\psi(x) \propto \begin{cases} 0 & x<0 \\ e^{-\gamma x} & x>0 \end{cases}, \tag{10.68}$$

(a) normalize the wave function, (b) calculate the probability $P\left[0 \le x \le (1/\gamma)\right]$, (c) calculate $\langle x\rangle$ and Δx.

10.3 For a particle whose wave function is given by

$$\psi(x) \propto \frac{1}{a^2+x^2}, \tag{10.69}$$

(a) normalize the wave function, (b) calculate $P(-a \le x \le a)$, (c) calculate $\langle x \rangle$ and Δx.

10.4 For a particle whose wave function is given by

$$\psi(x) \propto \operatorname{sech}(x/a), \tag{10.70}$$

(a) normalize the wave function, (b) calculate $P(0 \le x \le a)$, (c) calculate $\langle x \rangle$ and Δx.

10.5* Prove that $\hat{T}(D) = e^{-i\hat{p}_x D/\hbar}$, given the form of the infinitesimal translation operator $\hat{T}(dx)$ in eq. (10.26).

10.6* Prove that the translation operator is unitary.

10.7* Prove that $\hat{T}^{\dagger}(D) = \hat{T}(-D)$.

10.8 Prove that if $\left[\hat{A}, \hat{B}\right] = c$, where c is a constant, then

$$\left[\hat{A}, \hat{B}^n\right] = cn\hat{B}^{n-1}. \tag{10.71}$$

[Hint: Eq. (7.8) will be useful. Start by looking at $n = 2$, $n = 3$, etc.]

10.9 Prove that if $\left[\hat{A}, \hat{B}\right] = c$, where c is a constant, then

$$\left[\hat{A}, f\left(\hat{B}\right)\right] = cf'\left(\hat{B}\right), \tag{10.72}$$

where f is a function, and f' is its derivative. [Hint: Expand f as a power series, and use eq. (10.71).]

10.10 For a particle in state $|\psi\rangle$, $\langle x \rangle = \overline{x}$ and $\langle p \rangle = \overline{p}$. This state is then translated: $|\varphi\rangle = \hat{T}(D)|\psi\rangle$. Compute $\langle x \rangle$ and $\langle p \rangle$ for state $|\varphi\rangle$, and compare to the results for the untranslated state. You may find eq. (10.72) useful.

10.11* Prove that $\langle p \rangle = 0$ for the particle in problem 10.2 by explicitly demonstrating that the integral is zero. You need to be very careful about the discontinuity at $x = 0$, and you will need some results from complement 10.A.

10.12 Verify the indeterminacy relation for the particle in problem 10.3.

10.13 Verify the indeterminacy relation for the particle in problem 10.4.

10.14* Prove eq. (10.58).

10.15* Find the momentum representation of a particle in a position eigenstate. Discuss how the indeterminacy principle applies to this state.

10.16* Find the momentum representation of a particle in the Gaussian wave packet state of eq. (10.14). Use this representation to calculate $\langle p \rangle$ and Δp, and to verify the indeterminacy relation.

10.17 Find the momentum representation of the state of the particle in problem 10.3. Use this representation to calculate $\langle p \rangle$ and Δp, and to verify the indeterminacy relation.

COMPLEMENT 10.A

Useful Mathematics

10.A.1 Delta Functions

Consider the rectangle function, which is defined as

$$\mathrm{rect}(x)=\begin{cases}1 & |x|<1/2\\ 0 & \mathit{elsewhere}\end{cases}, \qquad (10.A.1)$$

and is depicted in fig. 10.A.1(a). If we let

$$f(x)=\frac{1}{a}\mathrm{rect}\left(\frac{x-x_0}{a}\right), \qquad (10.A.2)$$

then we get the function shown in fig. 10.A.1(b). This function has a width of a, and a height of $1/a$, so its area is 1. In the limit $a\to 0$, $f(x)$ becomes infinitely narrow, but infinitely tall, and maintains an area of 1; thus, in this limit $f(x)$ becomes the delta function (also known as the Dirac delta function):

$$\delta(x-x_0)=\lim_{a\to 0}\frac{1}{a}\mathrm{rect}\left(\frac{x-x_0}{a}\right). \qquad (10.A.3)$$

A delta function diverges where its argument is equal to 0, and is 0 everywhere else. Since the "area" of a delta function is 1, we know that

$$\int_{-\infty}^{\infty}dx\,\delta(x-x_0)=1. \qquad (10.A.4)$$

Actually, the limits don't need to extend all the way from $-\infty$ to ∞. As long as the limits of integration include the delta function, the integral will be 1, while if the limits do not include the delta function, the integral is 0. Other functions that approach the delta function as their width approaches zero, and their area is held fixed, are:

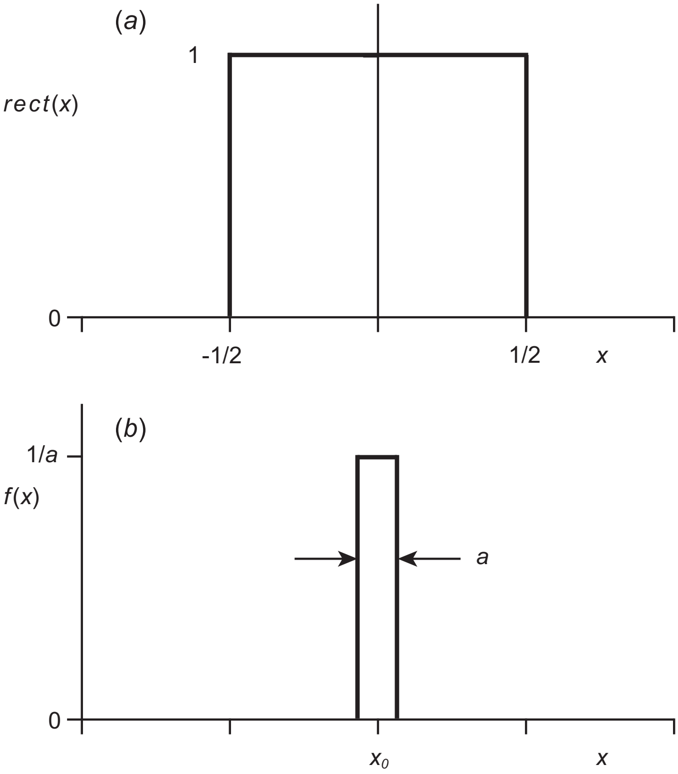


Fig 10.A.1 (a) The rectangle function, and (b) the function $f\left(x\right)$ of eq. (10.A.2).

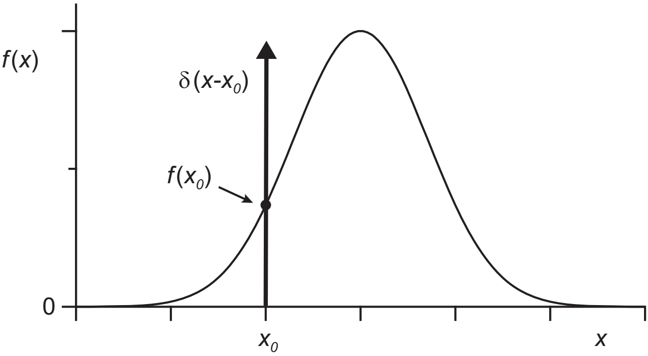


Fig 10.A.2 If a function $f\left(x\right)$ is multiplied by a delta function and integrated, the integral is simply the value of the function evaluated at the location of the delta function.

$$\delta(x) = \lim_{a\to 0} \frac{1}{a\sqrt{\pi}} e^{-x^2/a^2} \tag{10.A.5}$$

$$\delta(x) = \lim_{a\to 0} \frac{1}{\pi} \frac{a}{x^2 + a^2} \tag{10.A.6}$$

$$\delta(x) = \lim_{a\to 0} \frac{1}{a\pi} \frac{\sin(x/a)}{x/a}. \tag{10.A.7}$$

The delta function by itself it doesn't make a lot of sense, but it is perfectly well behaved inside an integral. For example, suppose we multiply the function $f(x)$ by $\delta(x - x_0)$, and then integrate over some interval that includes the delta function. In regions where the delta function is zero, the product is zero, so these regions do not contribute to the integral. The only point that contributes to the integral is at $x = x_0$, so we must evaluate $f(x)$ at that point. The end result is that

$$\int_{-\infty}^{\infty} dx f(x) \delta(x - x_0) = f(x_0). \tag{10.A.8}$$

The delta function picks out the value of $f(x)$ at the point where the delta function is located (see fig. 10.A.2).

Since the variable of integration in eq. (10.A.8) is x, the delta function must take as its argument x minus something. If the delta function is not of that form, then it must be manipulated to place it in that form for eq. (10.A.8) to be useful. For example, if instead the integral is of the form

$$\int_{-\infty}^{\infty} dx f(x) \delta(ax - x_0), \tag{10.A.9}$$

we need to make the substitution

$$u = ax,\ du = adx. \tag{10.A.10}$$

Then the integral becomes

$$\frac{1}{|a|} \int_{-\infty}^{\infty} du f(u/a) \delta(u - x_0) = \frac{1}{|a|} f(x_0/a). \tag{10.A.11}$$

The absolute value symbol ensures that the integration goes from $-\infty$ to ∞, independent of whether a is positive or negative. This procedure is equivalent to

$$\delta(ax - x_0) = \frac{1}{|a|} \delta\left(x - \frac{x_0}{a}\right). \tag{10.A.12}$$

An important special case is

$$\delta(-x) = \delta(x). \tag{10.A.13}$$

The integral of the delta function is

$$\int_{-\infty}^{x} dx'\delta(x') = \begin{cases} 0 & x<0 \\ 1 & x>0 \end{cases}. \tag{10.A.14}$$

To make this integral symmetric about zero, is standard to define it to be equal to 1/2 for $x = 0$. With this definition

$$\int_{-\infty}^{x} dx'\delta(x') = \Theta(x), \tag{10.A.15}$$

where the step function $\Theta(x)$ is defined as

$$\Theta(x) = \begin{cases} 1 & x>0 \\ 1/2 & x=0. \\ 0 & x<0 \end{cases} \tag{10.A.16}$$

Equation (10.A.15) tells us that the integral of the delta function is the step function; this means that the derivative of the step function must be the delta function

$$\frac{d}{dx}\Theta(x) = \delta(x). \tag{10.A.17}$$

The derivative of the delta function $\delta'(x)$ can be defined by taking the limit as $a \to 0$ of the derivative of one of the functions that we have used to define the delta function [e.g., eqs. (10.A.5)–(10.A.7)]. If we multiply $f(x)$ by $\delta'(x)$ and integrate, we can compute the integral using integration by parts, and find

$$\int_{-\infty}^{\infty} dx f(x)\delta'(x) = f(x)\delta(x)\Big|_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} dx f'(x)\delta(x) = -f'(0). \tag{10.A.18}$$

A number of other properties of the delta function can be found in ref. [10.A.1].

10.A.2 Gaussian Integrals

Define the Gaussian wave function $\psi(x)$ by

$$\psi(x) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^{1/2}} e^{-(x-\bar{x})^2/4\sigma^2}. \tag{10.A.19}$$

This wave function is normalized, and the moments of x are given by

$$\begin{aligned} \langle x^n \rangle &= \int_{-\infty}^{\infty} dx\, x^n \left|\psi(x)\right|^2 \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, x^n\, e^{-(x-\bar{x})^2/2\sigma^2}. \end{aligned} \tag{10.A.20}$$

Some of the lower order moments are

$$\langle x\rangle = \bar{x}, \tag{10.A.21}$$

$$\langle x^2\rangle = \bar{x}^2 + \sigma^2, \tag{10.A.22}$$

$$\langle x^3\rangle = \bar{x}\left(\bar{x}^2 + 3\sigma^2\right), \tag{10.A.23}$$

$$\langle x^4\rangle = \bar{x}^4 + 6\bar{x}^2\sigma^2 + 3\sigma^4. \tag{10.A.24}$$

In the case that $\bar{x} = 0$, the integral for the moments simplifies to

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\, x^n e^{-x^2/2\sigma^2} = \begin{cases}(n-1)!!\sigma^n & n \textit{ even integer}\\ 0 & n \textit{ odd integer}\end{cases}, \tag{10.A.25}$$

where the double factorial, for even n, is given by

$$(n-1)!! = (n-1)(n-3)\ldots(3)(1). \tag{10.A.26}$$

10.A.3 Fourier Transforms

We will define the Fourier transform of $\psi(x)$ as

$$\tilde{\psi}(p) = \mathfrak{F}\{\psi(x)\} = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar}\psi(x). \tag{10.A.27}$$

The inverse transform is

$$\psi(x) = \mathfrak{F}^{-1}\{\tilde{\psi}(p)\} = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dp\, e^{ipx/\hbar}\tilde{\psi}(p). \tag{10.A.28}$$

Notice that the only difference between these two definitions is the sign of the argument of the exponential. We'll assume that the functions we are dealing with are well enough behaved so that the integrals exist.[5]

Fourier transforms always come in pairs. For example, the Fourier transform of a rectangle function is a sinc function (see problem 10.A.1). This means that the inverse transform of a sinc is a rectangle. However, since the transform and inverse transform integrals are the same (apart from a sign) it also means that the transform of a sinc is a rectangle. The widths of the functions that make up a Fourier transform pair are inversely related; for example, if $\psi(x)$ becomes broader, $\tilde{\psi}(p)$ becomes narrower.

---

[5] We have included the factor of $\hbar$ in the definition of the Fourier transform to be consistent with the relationship between the position and momentum wave functions. To obtain the definition of the Fourier transform used in many math texts, you may take $\hbar = 1$.

The Fourier transform is linear:

$$\mathfrak{F}\left\{A_1\psi_1(x)+A_2\psi_2(x)\right\}=A_1\mathfrak{F}\left\{\psi_1(x)\right\}+A_2\mathfrak{F}\left\{\psi_2(x)\right\}, \tag{10.A.29}$$

and the *Shift Theorem* for Fourier transforms says that

$$\mathfrak{F}\left\{e^{ip_0x/\hbar}\psi(x)\right\}=\tilde{\psi}(p-p_0). \tag{10.A.30}$$

We can prove this using the definition of the Fourier transform in eq. (10.A.27):

$$\begin{aligned}\mathfrak{F}\left\{e^{ip_0x/\hbar}\psi(x)\right\}&=\frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}dx\,e^{-ipx/\hbar}e^{ip_0x/\hbar}\psi(x)\\&=\frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}dx\,e^{-i(p-p_0)x/\hbar}\psi(x)\\&=\tilde{\psi}(p-p_0).\end{aligned} \tag{10.A.31}$$

The magnitude of the Fourier transform of a real function is symmetric about the origin. Specifically, if $\mathfrak{F}\left\{\psi(x)\right\}=\tilde{\psi}(p)$, and $\psi(x)$ is a real function, then $\tilde{\psi}(p)=\tilde{\psi}^*(-p)$ and $\left|\tilde{\psi}(p)\right|=\left|\tilde{\psi}(-p)\right|$. You'll prove this in problem 10.A.2.

We can substitute eq. (10.A.28) into eq. (10.A.27) to obtain the following identity:

$$\tilde{\psi}(p)=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx\,e^{-ipx/\hbar}\int_{-\infty}^{\infty}dp'\,e^{ip'x/\hbar}\tilde{\psi}(p'). \tag{10.A.32}$$

We can rewrite this expression as

$$\tilde{\psi}(p)=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dp'\int_{-\infty}^{\infty}dx\,e^{-i(p-p')x/\hbar}\tilde{\psi}(p'). \tag{10.A.33}$$

We also know that

$$\tilde{\psi}(p)=\int_{-\infty}^{\infty}dp'\,\delta(p'-p)\tilde{\psi}(p'). \tag{10.A.34}$$

These last two equations, and the fact that $\delta(-x)=\delta(x)$, tells us that

$$\delta(p-p')=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx\,e^{-i(p-p')x/\hbar}. \tag{10.A.35}$$

We can rewrite this as

$$\delta(p-p')=\frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}dx\,e^{-ipx/\hbar}\left(\frac{1}{\sqrt{2\pi\hbar}}e^{ip'x/\hbar}\right), \tag{10.A.36}$$

which says that the Fourier transform of a complex exponential is a delta function.

10.A.4 References

[10.A.1] C. Cohen-Tannoudji, B. Diu, and Franck Laloë, *Quantum Mechanics* (Wiley, New York, 1977), Appendix II.

10.A.5 PROBLEMS

10.A.1 Show that

$$\mathfrak{F}\left\{\frac{1}{\sqrt{a}}\operatorname{rect}\left(\frac{x}{a}\right)\right\}=\sqrt{\frac{a}{2\pi\hbar}}\operatorname{sinc}\left(\frac{pa}{2\hbar}\right), \tag{10.A.37}$$

where the rectangle function is defined in eq. (10.A.1), and

$$\operatorname{sinc}(p)\equiv\frac{\sin(p)}{p}. \tag{10.A.38}$$

10.A.2 Prove that if $\mathfrak{F}\{\psi(x)\}=\tilde{\psi}(p)$, and $\psi(x)$ is a real function, then $\tilde{\psi}(p)=\tilde{\psi}^{*}(-p)$ and $\left|\tilde{\psi}(p)\right|=\left|\tilde{\psi}(-p)\right|$.

10.A.3 Prove that if $\mathfrak{F}\{\psi(x)\}=\tilde{\psi}(p)$, then $\mathfrak{F}\{\psi(x-x_0)\}=e^{-ipx_0/\hbar}\tilde{\psi}(p)$.

CHAPTER 11

Wave Mechanics and the Schrödinger Equation

In chapter 9 we described how the Schrödinger equation determines the dynamical evolution of quantum systems, and in chapter 10 we saw that the wave function is the position basis representation of a quantum state. Now it's time to join these ideas together to solve some new types of problems. Here we'll discuss the temporal evolution of systems whose potential energy varies as a function of position. The wave-like nature of particles will become much more readily apparent as we do this, so this description of quantum mechanics is often referred to as wave mechanics.

11.1 THE SCHRÖDINGER EQUATION REVISITED

11.1.1 The Schrödinger Equation in the Position Basis

Previously we discussed some of the general properties of the Schrödinger equation,

$$\hat{H}\left|\psi(t)\right\rangle = i\hbar\frac{d}{dt}\left|\psi(t)\right\rangle. \tag{11.1}$$

The Hamiltonian $\hat{H}$ tells us the energy of a particle, and in general consists of the sum of the kinetic and potential energies. For the moment we'll work in one spatial dimension, so the kinetic energy is

$$K = \frac{1}{2}mv^2 = \frac{p^2}{2m}. \tag{11.2}$$

We'll assume that the potential energy $V(x)$ is independent of time, and depends on the position of the particle, but not its momentum. With these assumptions, the Hamiltonian operator becomes

$$\hat{H}=\frac{\hat{p}^2}{2m}+V(\hat{x}), \tag{11.3}$$

and the Schrödinger equation is

$$\left(\frac{\hat{p}^2}{2m}+V(\hat{x})\right)|\psi(t)\rangle=i\hbar\frac{d}{dt}|\psi(t)\rangle. \tag{11.4}$$

By applying $\langle x|$ to this equation we can obtain its representation in the position basis

$$\langle x|\left(\frac{\hat{p}^2}{2m}+V(\hat{x})\right)|\psi(t)\rangle=\langle x|i\hbar\frac{d}{dt}|\psi(t)\rangle. \tag{11.5}$$

The wave function at time *t* is defined as[1]

$$\Psi(x,t)\equiv\langle x|\psi(t)\rangle. \tag{11.6}$$

Using the position-basis representation of the momentum operator [eq. (10.43)], we can write eq. (11.5) as

$$\left(\frac{1}{2m}\left[-i\hbar\frac{\partial}{\partial x}\right]^2+V(x)\right)\Psi(x,t)=i\hbar\frac{\partial}{\partial t}\Psi(x,t), \tag{11.7}$$

or

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\Psi(x,t)+V(x)\Psi(x,t)=i\hbar\frac{\partial}{\partial t}\Psi(x,t). \tag{11.8}$$

This is the one-dimensional Schrödinger wave equation, which is simply the Schrödinger equation written in the position basis. If you've studied quantum mechanics before, this equation should be familiar to you.

Although the form of eq. (11.8) is different from the equation for waves on a string or electromagnetic waves, the solutions do take the form of waves. This is why the function $\Psi(x,t)$ is called the wave function.

11.1.2 The Time-Independent Schrödinger Equation

Equation (11.8) is a partial differential equation, involving both space and time derivatives. One way to solve it is to use the mathematical technique of separation of variables. In separation of variables we write the function $\Psi(x,t)$ as a product of two functions, one of which depends on space, and the other of which depends on time:

---

[1] I'll follow the convention that the capitalized wave function $\Psi(x,t)$ has both time and space dependence, whereas the lowercase wave function $\psi(x)$ is independent of time.

$$\Psi(x,t) = \psi(x)\phi(t). \tag{11.9}$$

Clearly, not all functions of space and time can be so factorized, and it may not be obvious that a solution of this form will be guaranteed to solve eq. (11.8). However, as long as the potential energy $V(x)$ is independent of time, eq. (11.8) has solutions of the form in eq. (11.9). A key point here is that there are multiple solutions. Since the Schrödinger wave equation is linear, any linear combination of these solutions is also a solution. It can be proved that it is always possible to obtain a unique solution to eq. (11.8), with its associated boundary and initial conditions, by using a proper linear combination of solutions of the form in eq. (11.9):

$$\Psi(x,t) = \sum_n c_n \psi_n(x)\phi_n(t). \tag{11.10}$$

We'll discuss the technique of separation of variables in detail later, in conjunction with our discussion of systems that have more than one spatial dimension. For now, the separation of eq. (11.8) using eq. (11.9) is left as problem 11.1.

Another way to separate the spatial and temporal parts of eq. (11.8) is to use the fact that we've already solved the time-dependent problem in chapter 9! We are considering a time-independent Hamiltonian, so the time dependence of the state is given by eq. (9.13):

$$\left|\psi(t)\right\rangle = e^{-i\hat{H}t/\hbar}\left|\psi(0)\right\rangle. \tag{11.11}$$

Recall that time-dependent problems are most easily dealt with using the eigenstates $\left|\psi_n\right\rangle$ of the Hamiltonian,

$$\hat{H}\left|\psi_n\right\rangle = E_n\left|\psi_n\right\rangle, \tag{11.12}$$

where the eigenvalues E_n are the allowed energies of the system. We can write the initial state as a linear combination of energy eigenstates

$$\left|\psi(0)\right\rangle = \sum_n c_n\left|\psi_n\right\rangle, \tag{11.13}$$

and the wave function at $t = 0$, $\Psi(x,0)$, is then given by

$$\begin{aligned}\Psi(x,0) &= \left\langle x\middle|\psi(0)\right\rangle \\ &= \sum_n c_n\left\langle x\middle|\psi_n\right\rangle \\ &= \sum_n c_n\psi_n(x).\end{aligned} \tag{11.14}$$

The coefficients c_n are determined by the initial conditions:

$$\begin{aligned} c_n &= \langle \psi_n | \psi(0) \rangle \\ &= \int_{-\infty}^{\infty} dx \langle \psi_n | x \rangle \langle x | \psi(0) \rangle \\ &= \int_{-\infty}^{\infty} dx\, \psi_n^*(x) \Psi(x,0). \end{aligned} \tag{11.15}$$

From eqs. (11.11) and (11.13), the state at future times is

$$\begin{aligned} |\psi(t)\rangle &= \sum_n c_n e^{-i\hat{H}t/\hbar} |\psi_n\rangle \\ &= \sum_n c_n e^{-iE_n t/\hbar} |\psi_n\rangle \\ &= \sum_n c_n e^{-i\omega_n t} |\psi_n\rangle, \end{aligned} \tag{11.16}$$

where we've used Planck's relation $E_n = \hbar\omega_n$. In the position basis this equation becomes

$$\begin{aligned} \langle x | \psi(t) \rangle &= \sum_n c_n e^{-i\omega_n t} \langle x | \psi_n \rangle, \\ \Psi(x,t) &= \sum_n c_n e^{-i\omega_n t} \psi_n(x). \end{aligned} \tag{11.17}$$

Thus, the time dependence is determined, once we find the time-independent wave functions $\psi_n(x)$. We can find them by projecting eq. (11.12) into the position basis:

$$\begin{aligned} \langle x | \hat{H} | \psi_n \rangle &= \langle x | E_n | \psi_n \rangle, \\ \langle x | \left(\frac{\hat{p}^2}{2m} + V(\hat{x}) \right) | \psi_n \rangle &= E_n \psi_n(x), \\ \left(\frac{1}{2m} \left[-i\hbar \frac{d}{dx} \right]^2 + V(x) \right) \psi_n(x) &= E_n \psi_n(x). \end{aligned} \tag{11.18}$$

This can be simplified as

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_n(x) + V(x) \psi_n(x) = E_n \psi_n(x), \tag{11.19}$$

which is known as the time-independent Schrödinger equation. The solutions to this equation determine the full time-dependent wave function via eq. (11.17).

Note that if we compare eq. (11.17) to eq. (11.10), we learn that when using separation of variables to solve the Schrödinger equation, the solution for the time dependence must be

$$\phi_n(t) = e^{-i\omega_n t}. \tag{11.20}$$

11.1.3 Boundary Conditions

What boundary conditions must we impose on wave functions? Assume that $V(x)$ has a discontinuity at $x = x_0$. Rearrange eq. (11.19), and integrate it about a small region centered on $x = x_0$:[2]

$$\int_{x_0-\eta}^{x_0+\eta} dx\, \psi_n''(x) = \psi_n'(x_0+\eta) - \psi_n'(x_0-\eta) = \frac{2m}{\hbar^2} \int_{x_0-\eta}^{x_0+\eta} dx\left[V(x) - E_n\right]\psi_n(x). \tag{11.21}$$

If the discontinuity is finite [$V(x)$ is bounded] the integrand in the integral on the right is finite, so the integral will approach 0 in the limit that $\eta \to 0$. This means that $\psi_n'(x)$ is continuous in the case of a finite discontinuity in $V(x)$ [clearly, $\psi_n'(x)$ is also continuous if $V(x)$ is continuous]. However, if the potential energy discontinuity is infinite, $\psi_n'(x)$ may be discontinuous. In all cases we require $\psi_n(x)$ to be continuous. We've also stated before that real particles cannot be completely unlocalized, which means that $\psi_n(x)$ must be normalizable.

11.1.4 Putting It All Back Together

To summarize, the procedure for solving the Schrödinger wave equation [eq. (11.8)], given a form of the potential energy $V(x)$, and a set of boundary and initial conditions, is then:

1. The potential energy function and the boundary conditions determine the allowed energies E_n and the time-independent wave functions $\psi_n(x)$ by solving eq. (11.19).
2. The unique solution to a given problem at $t = 0$ is given by a linear combination of the $\psi_n(x)$'s [eq. (11.14)]. The coefficients in this linear combination are determined by the initial wave function, $\Psi(x,0)$, using eq. (11.15).
3. The full time-dependent wave function $\Psi(x,t)$ is given by eq. (11.17).

Let's see how to apply these ideas, by looking at a few examples involving different potential energy functions.

---

[2] Here the primes refer to derivatives with respect to x.

11.2 CONSTANT POTENTIAL–THE FREE PARTICLE

The time-independent Schrödinger equation takes on its simplest form when the potential energy is constant throughout all of space, $V(x) = V_0$. In this case the equation is[3]

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x) + V_0\psi(x) = E\psi(x). \tag{11.22}$$

We can rewrite this as

$$\frac{d^2}{dx^2}\psi(x) = -\frac{2m(E - V_0)}{\hbar^2}\psi(x). \tag{11.23}$$

You can verify for yourself that the solutions to this equation are

$$\psi(x) = Ae^{ikx}, \tag{11.24}$$

with

$$k = \left(\frac{2m(E - V_0)}{\hbar^2}\right)^{1/2}. \tag{11.25}$$

We can rewrite this equation to express the energy as

$$E = \frac{\hbar^2 k^2}{2m} + V_0. \tag{11.26}$$

The Schrödinger equation is a second-order differential equation, so to satisfy the boundary conditions we need two, linearly independent solutions. Both positive and negative roots for k in eq. (11.25) are allowed, and these are the linearly independent solutions. To make this more explicit, we write the solution as

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \tag{11.27}$$

and confine ourselves to the positive root in eq. (11.25).

Ordinarily at this point we would apply boundary conditions, but since the potential energy is assumed to be constant everywhere, there are no boundary conditions; we are examining the behavior of a free particle. The wave functions of eq. (11.27) extend throughout all space, and k can take on any positive real value ($k > 0$). From eq. (11.26), we see that this implies $E > V_0$, which makes sense physically because the total energy must be larger than the potential energy.

From eq. (11.26) we find that the kinetic energy K is given by

$$K = E - V_0 = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m}, \tag{11.28}$$

---

[3] Here we've suppressed the subscripts that index the energies and wave functions.

where we've used eq. (11.2). This is consistent with the de Broglie relation $p = \hbar k$. For every value of k there is a corresponding energy, and for every energy there is a corresponding angular frequency $\omega = E/\hbar$. We can put the time dependence into eq. (11.27) using the method discussed in the previous section, and obtain

$$\Psi(x,t) = \left(Ae^{ikx} + Be^{-ikx}\right)e^{-i\omega t} = Ae^{i(kx-\omega t)} + Be^{-i(kx+\omega t)}. \tag{11.29}$$

The wave with amplitude A is traveling in the positive direction, while that with amplitude B is traveling in the negative direction. The unique solution to the full time-dependent problem, which fits the initial conditions, will be a linear combination of solutions of the form in eq. (11.29), with different values of k.

Comparing eq. (11.24) to the wave function of a momentum eigenstate, eq. (10.56), we find that these wave functions are the same if $k = p/\hbar$, which is again equivalent to the de Broglie relation, and if $A = 1/\sqrt{2\pi\hbar}$. Thus, the free particle solutions we have obtained in eq. (11.24) correspond to solutions which have well-defined momenta.

From our discussion in section 10.1, we know that the unlocalized solutions of eq. (11.24) are unphysical. Any real particle must be at least partially localized, so a physical solution will be represented by a wave packet. For a discussion of wave packet solutions to eq. (11.22), and their time evolution, see complement 11.A.

11.3 POTENTIAL STEP

Another example is that of the potential step, which is described by

$$V(x) = \begin{cases} 0 & x < 0 \\ V_0 & x > 0 \end{cases}, \tag{11.30}$$

and is pictured in fig. 11.1(a). Since the potential is constant in each of the two regions (to the left and right of the origin) we can use the results of the previous section to determine the solution to the time-independent Schrödinger equation in each region. To the left of the origin the solution is

$$\psi(x) = A_1 e^{ik_1 x} + B_1 e^{-ik_1 x} \qquad x < 0, \tag{11.31}$$

where

$$k_1 = \left(\frac{2mE}{\hbar^2}\right)^{1/2}, \tag{11.32}$$

and E is the energy of the particle. Similarly, to the right of the origin we have

$$\psi(x) = A_2 e^{ik_2 x} + B_2 e^{-ik_2 x} \qquad x > 0, \tag{11.33}$$

where

$$k_2 = \left(\frac{2m(E - V_0)}{\hbar^2}\right)^{1/2}. \tag{11.34}$$

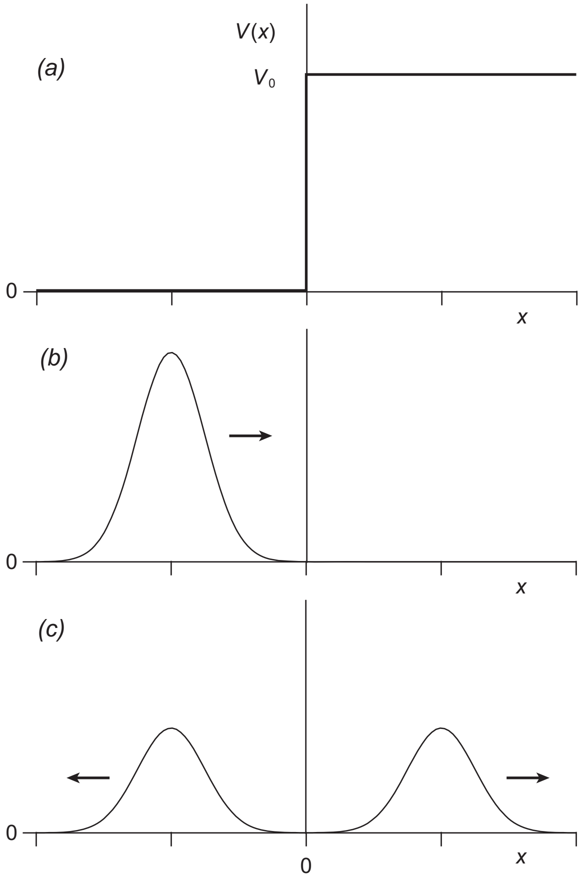


Fig 11.1 (a) A potential step. (b) A particle (represented as a wave packet) is incident on the step from the left. (c) The particle is either transmitted or reflected at the step.

As in the previous section, we take the positive square root in eqs. (11.32) and (11.34).

Before going any further with the mathematics, let's think more carefully about the physics. The step function potential of fig. 11.1(a) represents a barrier, and fig. 11.1(b) shows a particle incident on it. As shown in fig. 11.1(c), the interesting physical question is, "What is the probability that the particle will be transmitted, or reflected?"

In fig. 11.1 the incident wave is on the left of the boundary and traveling to the right, so it is the wave of amplitude A_1 in eq. (11.31). The reflected wave is on the left of the boundary and traveling to the left, so it is the wave of amplitude B_1 in this same equation. The transmitted wave is on the right of the boundary and traveling to the right, so it is the wave of amplitude A_2 in eq. (11.33). There is no wave on the right of the boundary traveling to the left, so we have $B_2 = 0$ in eq. (11.33).

There are two different solutions, depending on whether the particle energy is larger or smaller than the height of the potential step. We'll examine each of these separately below. Before we do that, however, we need to explore the concept of the probability flux.

11.3.1 Probability Flux

The square magnitude of the wave function represents a probability density. This remains true when we add time dependence: $\left|\Psi(x,t)\right|^2$ is a time-dependent probability density. Taking the time derivative of $\left|\Psi(x,t)\right|^2$, and using the Schrödinger equation [eq. (11.8)], yields

$$\begin{aligned}\frac{\partial}{\partial t}\left|\Psi(x,t)\right|^2 &= \frac{\partial}{\partial t}\left[\Psi^*\Psi\right]\\ &= \Psi^*\frac{\partial}{\partial t}\Psi + \Psi\frac{\partial}{\partial t}\Psi^*\\ &= \Psi^*\left(\frac{-\hbar}{i2m}\frac{\partial^2}{\partial x^2}\Psi + \frac{1}{i\hbar}V(x)\Psi\right) + \Psi\left(\frac{\hbar}{i2m}\frac{\partial^2}{\partial x^2}\Psi^* - \frac{1}{i\hbar}V(x)\Psi^*\right)\\ &= -\frac{\hbar}{i2m}\left(\Psi^*\frac{\partial^2}{\partial x^2}\Psi - \Psi\frac{\partial^2}{\partial x^2}\Psi^*\right). \end{aligned} \tag{11.35}$$

Let's define

$$j_x(x,t) \equiv \frac{\hbar}{i2m}\left(\Psi^*\frac{\partial}{\partial x}\Psi - \Psi\frac{\partial}{\partial x}\Psi^*\right). \tag{11.36}$$

The spatial derivative of $j_x(x,t)$ is

$$\begin{aligned}\frac{\partial}{\partial x}j_x(x,t) &= \frac{\hbar}{i2m}\left[\left(\frac{\partial}{\partial x}\Psi^*\right)\left(\frac{\partial}{\partial x}\Psi\right) + \Psi^*\frac{\partial^2}{\partial x^2}\Psi - \left(\frac{\partial}{\partial x}\Psi\right)\left(\frac{\partial}{\partial x}\Psi^*\right) - \Psi\frac{\partial^2}{\partial x^2}\Psi^*\right]\\ &= \frac{\hbar}{i2m}\left(\Psi^*\frac{\partial^2}{\partial x^2}\Psi - \Psi\frac{\partial^2}{\partial x^2}\Psi^*\right), \end{aligned} \tag{11.37}$$

which means that

$$\frac{\partial}{\partial t}\left|\Psi(x,t)\right|^2 = -\frac{\partial}{\partial x}j_x(x,t). \tag{11.38}$$

In three dimensions, $j_x(x,t)$ is the x-component of the vector $\boldsymbol{j}(\boldsymbol{r},t)$,

$$\boldsymbol{j}(\boldsymbol{r},t) \equiv \frac{\hbar}{i2m}\left(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*\right) = \frac{\hbar}{m}\mathrm{Im}\left(\Psi^*\nabla\Psi\right), \tag{11.39}$$

and eq (11.38) becomes

$$\frac{\partial}{\partial t}\left|\Psi(\boldsymbol{r},t)\right|^2 = -\nabla\bullet\boldsymbol{j}(\boldsymbol{r},t). \tag{11.40}$$

If we integrate eq. (11.40) over some region in space Ω, we find

$$\begin{aligned}\iiint_{\Omega} d^3r\,\frac{\partial}{\partial t}\left|\Psi(\boldsymbol{r},t)\right|^2 &= -\iiint_{\Omega} d^3r\,\nabla\cdot\boldsymbol{j}(\boldsymbol{r},t),\\ \frac{\partial}{\partial t}\left(\iiint_{\Omega} d^3r\,\left|\Psi(\boldsymbol{r},t)\right|^2\right) &= -\oiint_{S} d\boldsymbol{a}\cdot\boldsymbol{j}(\boldsymbol{r},t).\end{aligned} \tag{11.41}$$

Here we've used the divergence theorem to replace the volume integral over Ω by an integral over the closed surface S that bounds Ω. The integral in parentheses on the left is the probability that the particle is found inside Ω, while the expression on the right represents the inward flux of $\boldsymbol{j}(\boldsymbol{r},t)$ through S (the integral itself is the outward flux, while the minus sign changes it to the inward flux). Taken as a whole, eq. (11.41) says that the rate of change of probability in some region of space is equal to the inward flux of $\boldsymbol{j}(\boldsymbol{r},t)$ through the surface bounding that region. For this reason we refer to $\boldsymbol{j}(\boldsymbol{r},t)$ as the probability flux.[4]

11.3.2 $E > V_0$

Back to the potential step. Assume that $E > V_0$, so k_2 [eq. (11.34)] is real and positive. We now need to apply the boundary conditions. The wave function must be continuous at $x = 0$, so eqs. (11.31) and (11.33) yield (for $B_2 = 0$, as described above):

$$A_1 + B_1 = A_2. \tag{11.42}$$

The derivative of the wave function must also be continuous at this point, because the discontinuity is finite, so the same equations tell us that

$$ik_1A_1 - ik_1B_1 = ik_2A_2. \tag{11.43}$$

We can solve these equations for B_1 and A_2 in terms of the incident amplitude A_1. You'll show in problem 11.2 that the solutions are

$$B_1 = \frac{(k_1 - k_2)}{(k_1 + k_2)}A_1,\quad A_2 = \frac{2k_1}{(k_1 + k_2)}A_1. \tag{11.44}$$

---

[4] Compare eq. (11.40) to the continuity equation of electromagnetism, which says

$$\frac{\partial}{\partial t}\rho(\boldsymbol{r},t) = -\nabla\cdot\boldsymbol{J}(\boldsymbol{r},t),$$

Here $\rho(\boldsymbol{r},t)$ is the charge density and $\boldsymbol{J}(\boldsymbol{r},t)$ is the current density [11.1]. The continuity equation is a statement about the conservation of charge. Because of the similarity, some people refer to $\boldsymbol{j}(\boldsymbol{r},t)$ as the probability current.

We'll denote the probability flux for the wave with amplitude A_1 by $j_x^{A_1}(x,t)$, and it is given by

$$\begin{aligned} j_x^{A_1}(x,t) &= \frac{\hbar}{m}\operatorname{Im}\left(A_1^* e^{-i(k_1 x - \omega_1 t)} \frac{\partial}{\partial x} A_1 e^{i(k_1 x - \omega_1 t)} \right) \\ &= \frac{\hbar k_1}{m}|A_1|^2 \\ &= v_1 |A_1|^2 , \end{aligned} \tag{11.45}$$

where

$$\frac{\hbar k_1}{m} = \frac{p_1}{m} = v_1, \tag{11.46}$$

and v_1 is the speed of the particle to the left of the step. Similarly, we find that $j_x^{B_1}(x,t) = -v_1|B_1|^2$, where the negative sign indicates that the wave with amplitude B_1 is moving to the left.

The probability that the particle will be found to reflect, R, is given by

$$R = \left|\frac{j_x^{B_1}(x,t)}{j_x^{A_1}(x,t)}\right| = \frac{v_1|B_1|^2}{v_1|A_1|^2} = \frac{|B_1|^2}{|A_1|^2} = \frac{(k_1 - k_2)^2}{(k_1 + k_2)^2} = \frac{\left(\sqrt{E} - \sqrt{E - V_0}\right)^2}{\left(\sqrt{E} + \sqrt{E - V_0}\right)^2}. \tag{11.47}$$

It is necessary to take the magnitude because the probability must be positive. If a particle is not reflected, it must be transmitted, so the probability it will be found to transmit T is

$$\begin{aligned} T &= 1 - R \\ &= 1 - \frac{(k_1 - k_2)^2}{(k_1 + k_2)^2} \\ &= \frac{\left(k_1^2 + 2k_1k_2 + k_1^2\right) - \left(k_1^2 - 2k_1k_2 + k_1^2\right)}{(k_1 + k_2)^2} \\ &= \frac{4k_1k_2}{(k_1 + k_2)^2} \\ &= \frac{4\sqrt{E(E - V_0)}}{\left(\sqrt{E} + \sqrt{E - V_0}\right)^2} . \end{aligned} \tag{11.48}$$

Let's examine the limiting behavior. If $E >> V_0$ the particle shouldn't even notice the barrier, and for all intents and purposes we have a constant potential. We thus expect $R \cong 0$ and $T \cong 1$, and this is exactly what we find from eqs. (11.47) and (11.48) in this limit. If $E \approx V_0$ (but E is still greater than V_0) Eqs. (11.47) and (11.48) predict the opposite behavior: $R \cong 1$ and $T \cong 0$; the particle is found to reflect with almost 100% certainty. This is in stark contrast to the classical predictions. Classically $R = 0$ and $T = 1$ if $E > V_0$ (even by just a little bit); the particle will always be transmitted as long as it has enough energy to get over the barrier.

You might wonder why we used probability fluxes in the above, and not simply $R = |B_1|^2 / |A_1|^2$ and $T = |A_2|^2 / |A_1|^2$. Equation (11.47) shows that these definitions will work for R, but in the problems you'll show that they don't work for T. The reason they don't work is that the particle speed is different on the two sides of the boundary. To get the correct answer in all cases, it's necessary to use the probability flux.

Keep in mind that the wave function of the particle exists on both sides of the barrier. Initially the particle doesn't reflect or transmit, it does *both*. The particle is not localized on one side of the barrier or the other until after a measurement is performed that determines where it is.

11.3.3 $E < V_0$

Now consider the case where the energy of the incident particle is less than the step height, that is, $E < V_0$. To the left of the step ($x < 0$) the solutions are still given by eqs. (11.31) and (11.32). However, eq. (11.34) becomes

$$k_2 = \left(-\frac{2m(V_0 - E)}{\hbar^2} \right)^{1/2} = i\left(\frac{2m(V_0 - E)}{\hbar^2} \right)^{1/2} = i\alpha, \tag{11.49}$$

where α is real and positive. The solution to the right of the barrier thus becomes

$$\psi(x) = A_2 e^{-\alpha x} + B_2 e^{\alpha x} \qquad x > 0. \tag{11.50}$$

We must have $B_2 = 0$ in eq. (11.50), otherwise the wave function would diverge as x increases, and not be normalizable. Applying the boundary conditions that the wave function and its derivative must be continuous at $x = 0$ yields

$$A_1 + B_1 = A_2 \tag{11.51}$$

$$ik_1 A_1 - ik_1 B_1 = -\alpha A_2. \tag{11.52}$$

A little algebra reveals that that the wave amplitudes determined by these equations are

$$B_1 = \frac{(ik_1 + \alpha)}{(ik_1 - \alpha)} A_1, \; A_2 = \frac{2ik_1}{(ik_1 - \alpha)} A_1 . \tag{11.53}$$

The reflection probability is

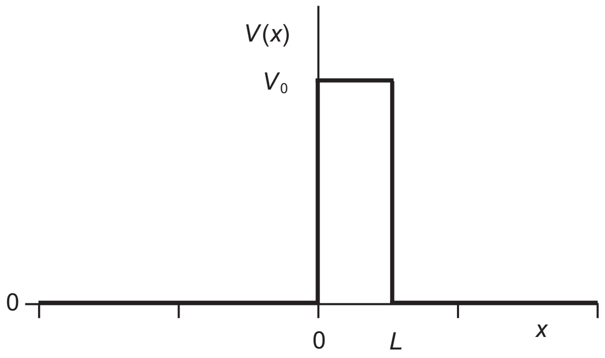


Fig 11.2 A potential barrier with a finite width.

$$R = \left| \frac{j_x^{B_1}(x,t)}{j_x^{A_1}(x,t)} \right| = \frac{k_1 |B_1|^2}{k_1 |A_1|^2} = \left| \frac{(ik_1 + \alpha)}{(ik_1 - \alpha)} \right|^2 = 1, \tag{11.54}$$

so clearly the transmission probability is $T = 0$.

If the energy of the particle is less than the potential energy of the step, it reflects with 100% certainty, a result that agrees with the classical prediction. The particle reflects, even though its wave function extends into the classically forbidden region. The simple traveling wave model we're using here doesn't show the motion of a localized particle. If we were to do a full calculation, with a moving localized particle, we would see that the wave function of the particle extends beyond the barrier when the particle is close to the barrier. However, as the particle reflects and moves away, the wave function inside the barrier decays away. To observe this time-dependent behavior for yourself, see the Java applet in ref. [11.2].

11.4 TUNNELING

What happens if the potential barrier has a finite width? Consider a potential barrier of height V_0 and width L, as shown in fig. 11.2. The solution to the Schrödinger equation for a barrier such as this is most interesting when the particle energy is less than the barrier height ($E < V_0$). In this case the solution is

$$\psi(x) = \begin{cases} A_1 e^{ikx} + B_1 e^{-ikx} & x < 0 \\ A_2 e^{-\alpha x} + B_2 e^{\alpha x} & 0 < x < L, \\ A_3 e^{ikx} & x > L \end{cases} \tag{11.55}$$

with

$$k = \left(\frac{2mE}{\hbar^2}\right)^{1/2}, \tag{11.56}$$

and

$$\alpha = \left(\frac{2m\left(V_0 - E\right)}{\hbar^2}\right)^{1/2}. \tag{11.57}$$

In eq. (11.55) we've assumed that the particle is incident on the barrier from the left, and there is no leftward traveling wave on the right of the barrier.

By applying boundary conditions at $x = 0$ and $x = L$, it is possible to determine the four wave amplitudes B_1, A_2, B_2, and A_3, in terms of the amplitude of the incident wave A_1. From this one can calculate the probability that the particle will be found to the right of the barrier (see problem 11.8). Such a particle is said to "tunnel," because it does not have enough energy to get over the barrier, yet somehow it manages to appear on the other side.

A qualitative solution is easy to obtain in the limit that the barrier is very wide ($L >> \alpha^{-1}$). In this case the particle will be reflected with near 100% probability, so $|B_1| \sim |A_1|$. The growing exponential term inside of the barrier must be small, and we'll make the approximation that it is 0 ($B_2 \cong 0$). At the left side of the barrier we then have $|A_2| \sim |A_1|$ and at the right side of the barrier we have $|A_3| \sim |A_2|e^{-\alpha L} \sim |A_1|e^{-\alpha L}$. The transmission probability is then

$$T = \left|\frac{j_x^{A_3}\left(x,t\right)}{j_x^{A_1}\left(x,t\right)}\right| = \frac{|A_3|^2}{|A_1|^2} \sim e^{-2\alpha L}. \tag{11.58}$$

The tunneling probability decreases exponentially with the width of the barrier, in the limit of a wide barrier.

There are some other interesting aspects to tunneling. Inside the barrier $E < V_0$, and eq. (11.49) tells us that the wave vector is purely imaginary, which means that the momentum of the particle is purely imaginary. This implies that the particle does not "propagate" through the barrier in any traditional sense. The amount of time it takes for the particle to travel through the barrier is an interesting, and difficult, question. There is some experimental evidence that particles can propagate through the barrier in times that would suggest that they are propagating at superluminal speeds (faster than c) in the barrier [11.3]. However, this does not violate special relativity because no useful information is transmitted—the probability of any particle getting through the barrier is very small, so no information propagates faster than c.

Tunneling and superluminal velocities are not limited to quantum mechanics, as classical waves can display similar behavior. If a classical electromagnetic wave propagating in a medium with index of refraction n_1 (e.g., glass) is incident on a boundary with a medium of index of refraction n_2 (e.g., air) at a large angle, and $n_1 > n_2$, the light can experience total internal reflection: 100% of the light reflects into the glass, and the air represents a barrier that the light cannot propagate into. However, there is an electric

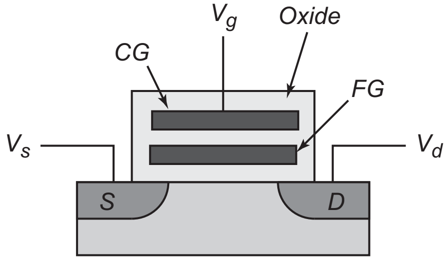


Fig 11.3 One bit of flash memory. Here S denotes the source, D denotes the drain, CG denotes the control gate, and FG denotes the floating gate. The source, control gate, and drain voltages are V_s, V_g and V_d, respectively.

field, known as the evanescent field, which extends into the air, but decays exponentially away from the boundary. If a second block of glass is brought very close to the boundary, so that some of the evanescent field extends into it, the field can "jump" the air gap, and continue propagating in the second block of glass. This effect is known as frustrated total internal reflection.

Superluminal propagation also occurs with classical waves. Indeed, it is possible for the transit time to be negative—the peak of the signal comes out before it goes in. For a simple experiment demonstrating this effect using electrical circuits, see ref. [11.4].

11.4.1 Microscopes and Memory

There are some important practical uses of tunneling. In a scanning tunneling microscope (STM) an extremely fine conducting tip, with ideally only a single atom at its point, is brought close (<1 nm) to a conducting surface. A bias voltage is applied between the surface and the tip, and electrons can tunnel across the gap between them. The tip is scanned across the surface and the tunneling current is measured. Since the probability of tunneling (and hence the tunneling current) is extremely sensitive to the distance between the tip and the surface, a profile of the surface can be obtained. It is possible to achieve atomic-scale resolution of the surface with an STM.

The flash memory in your portable music player or digital camera uses tunneling in its erasure process. One bit of flash memory is depicted schematically in fig. 11.3; it's similar to a metal-oxide-semiconductor field-effect-transistor (MOSFET), only the floating gate (FG) is absent in a standard MOSFET. In a MOSFET the oxide behaves as an insulator, and applying a positive voltage to the control gate (CG) creates an electric field in the channel between the source and the drain, which allows electrons to flow between them.[5]

---

[5] Semiconductor physics is beyond the scope of this book, but a brief explanation is as follows. Electric fields (due to charged impurity atoms) within the channel region between the source and drain ordinarily impede the flow of both electrons and holes. (A hole is essentially the lack of an electron, and behaves as a mobile positive charge carrier.) Applying a positive voltage to the control gate drives holes out of the channel, while pulling electrons in. This increases the conductivity of the channel, and allows an electron current to flow between the source and drain.

Assume that there is no charge on the FG, which corresponds to a memory state of binary 1. In this case the memory cell acts as an ordinary MOSFET, and electrons will flow from the source to the drain with the application of a positive voltage (~5V) to the CG. The state of the memory is thus read by applying 5V to the CG; if electrons flow from the source to the drain the memory state is 1.

Binary 0 is written by applying an elevated positive voltage (~12V) to the CG, which creates a higher electron current between the source and the drain. Some of these electrons have high enough kinetic energy to cross the oxide barrier, and collect on the FG. This process is known as hot-electron injection. If power is removed, electrons stay on the FG because it is surrounded by the insulating oxide, so the memory state is maintained even in the absence of power (the memory is nonvolatile). This state corresponds to binary 0 because the electrons on the floating gate shield the voltage applied to the CG; if 5V is applied to the CG, electrons will not flow between the source and drain.

The memory state is changed from 0 back to 1 (erased) by applying a large negative voltage to the CG. This creates an electric field that allows the electrons on the FG to tunnel through the oxide layer to the source. With the electrons now removed from the FG, the memory state is back at 1.

11.5 INFINITE SQUARE WELL

One of the features that makes quantum mechanics different from classical mechanics is the fact that quantum systems have discrete energy levels in situations where classical systems would allow a continuum of energies. So far in this chapter, however, the systems we have examined do not have quantized energies. Why is that? It's because particles typically have quantized energies when they are confined, and the particles we've examined so far have been free. In this section we'll study the behavior of a particle in a confining potential—an infinite square well.

An infinite square well potential is given by

$$V(x)=\begin{cases}0 & 0<x<L\\ \infty & elsewhere\end{cases}, \tag{11.59}$$

and is depicted in fig. 11.4. Clearly this is an idealized situation, but our results will closely approximate the physical solutions for a particle whose energy is much less than the potential energy outside the well. The particle must be confined to the inside of the well, because a particle with finite energy cannot be located where the potential energy is infinite. Thus, the wave function must be 0 outside of the well. Inside the well the potential energy is 0, so the solution to the time-independent Schrödinger equation of eq. (11.19) will once again be a linear combination of right-traveling and left-traveling waves. The solution is thus

$$\psi(x)=\begin{cases}Ae^{ikx}+Be^{-ikx} & 0<x<L\\ 0 & elsewhere\end{cases}, \tag{11.60}$$

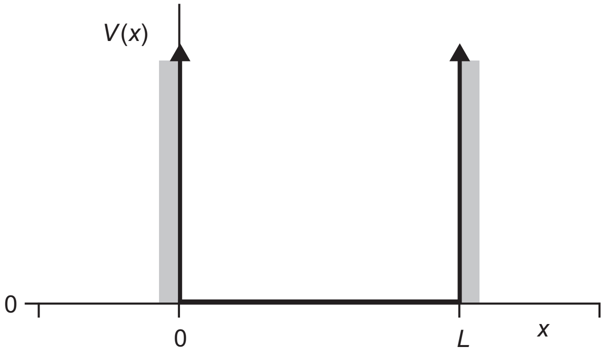


Fig 11.4 An infinite square well potential.

where

$$k = \left(\frac{2mE}{\hbar^2} \right)^{1/2}, \tag{11.61}$$

and we again take the positive square root.

The boundary conditions require the wave function to be continuous at the edges of the well.[6] At $x = 0$ this means

$$\begin{aligned} A + B &= 0\,, \\ B &= -A\,, \end{aligned} \tag{11.62}$$

and inside the well the wave function is

$$\psi(x) = A\left(e^{ikx} - e^{-ikx}\right) = C\sin(kx), \tag{11.63}$$

with $C = 2iA$. The boundary condition at $x = L$ tells us that

$$C\sin(kL) = 0. \tag{11.64}$$

To satisfy this equation we can't choose $C = 0$, because that would mean $\psi(x) = 0$ throughout all space, and the particle would not be anywhere. The only other possibility is that the sine is 0, which occurs when

$$kL = n\pi \quad n = 1, 2, 3, \ldots. \tag{11.65}$$

Here the value $n = 0$ is excluded, because it would also mean $\psi(x) = 0$ throughout all space. Since L is determined by the width of the well, eq. (11.65) determines the allowed values of the wave vector, and these values depend on n. With this in mind, we'll rewrite eq. (11.65) as

---

[6] Since there is an infinite discontinuity at the edges of the well, the derivative of the wave function need *not* be continuous at these points.

$$k_n = \frac{n\pi}{L} \quad n = 1, 2, 3, \ldots, \tag{11.66}$$

where the subscript distinguishes the different allowed wave vectors.

The wave vector and the energy are related by eq. (11.61), and we can use this equation to express the allowed energies as

$$E_n = \frac{k_n^2 \hbar^2}{2m} = \frac{n^2 \pi^2 \hbar^2}{2mL^2} \quad n = 1, 2, 3, \ldots. \tag{11.67}$$

Note that the lowest energy E_1, called the ground-state energy, is *not* 0. Why is that? An energy of 0 would imply a momentum of 0, with no uncertainty. The Heisenberg indeterminacy relation would then require the particle to be completely unlocalized (infinite uncertainty in position), which we know cannot be true because the particle must be in the well. In problem 11.11 you'll examine what the indeterminacy relation implies about the ground-state energy.

We can use eq. (11.66) to rewrite the wave functions inside the well as

$$\psi_n(x) = C_n \sin(k_n x) = C_n \sin\left(\frac{n\pi}{L} x\right). \tag{11.68}$$

We want our wave functions to be normalized, so the coefficients C_n are determined by

$$\begin{aligned} \int_{-\infty}^{\infty} dx \left|\psi(x)\right|^2 &= C_n^2 \int_0^L dx \sin^2\left(\frac{n\pi}{L} x\right) \\ &= C_n^2 \left[\frac{x}{2} - \frac{\sin\left(\frac{2n\pi}{L} x\right)}{\frac{4n\pi}{L}} \right]_{x=0}^{x=L} \\ &= C_n^2 \frac{L}{2} \\ &= 1, \end{aligned} \tag{11.69}$$

and $C_n = \sqrt{2/L}$. Thus, we obtain as the wave functions for a particle in an infinite square well

$$\psi_n(x) = \begin{cases} \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi}{L} x\right) & 0 < x < L \\ 0 & \textit{elsewhere} \end{cases} \quad n = 1, 2, 3, \ldots. \tag{11.70}$$

The first three of these wave functions and their square magnitudes (corresponding probability densities) are shown in fig. 11.5.

The states $\left|\psi_n\right\rangle$ are the eigenstates of the Hamiltonian, which is a Hermitian operator, so they must form a complete, orthonormal, basis set. Likewise, the functions

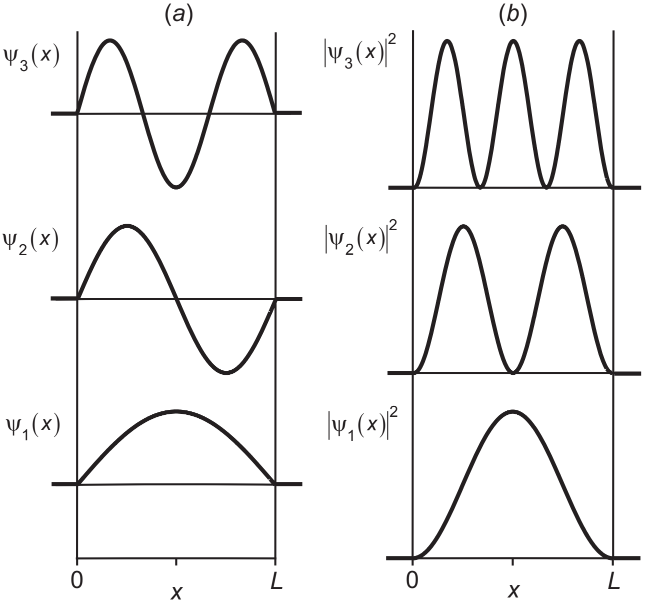


Fig 11.5 (a) The wave functions corresponding to the lowest three energies of an infinite square well potential. (b) The squares of these wave functions, which correspond to the probability density of finding a particle at a particular position in the well.

$\psi_n(x)$ form a complete orthonormal set, so any allowable wave function for the infinite square well can be written as a linear combination of $\psi_n(x)$'s.

Let's do an example to see how these results are useful.

EXAMPLE 11.1

At $t = 0$ a particle in a potential well described by eq. (11.59) is known to be localized in the left half of the well. Determine (a) $\Psi(x,t)$, and (b) the probability that a measurement of the energy at time t will yield the value E_n.

(a) First we need to determine $\Psi(x,0)$. We'll assume that the particle is equally likely to be anywhere in the left half of the well, so its properly normalized wave function is

$$\Psi(x,0) = \begin{cases} \sqrt{\dfrac{2}{L}} & 0 < x < L/2 \\ 0 & elsewhere \end{cases}, \tag{11.71}$$

which is depicted in fig. 11.6. Note that this is technically an illegal wave function, as it has discontinuities at $x = 0$ and $x = L/2$; we could smooth it out, but the results obtained here would not be substantially different.

To find $\Psi(x,t)$ we'll use the prescription described in sec. 11.1.4. We've already accomplished step (1), which is to find the allowed energies and their corresponding wave functions $\psi_n(x)$. Step (2) tells us to use eq. (11.15) to find the coefficients c_n in the linear expansion of $\Psi(x,0)$:

$$\begin{aligned} c_n &= \int_{-\infty}^{\infty} dx\, \psi_n^*(x)\Psi(x,0) \\ &= \frac{2}{L}\int_0^{L/2} dx \sin\left(\frac{n\pi x}{L}\right) \\ &= \frac{-2}{n\pi}\cos\left(\frac{n\pi x}{L}\right)\Bigg|_{x=0}^{x=L/2} \\ &= -\frac{2}{n\pi}\left[\cos\left(\frac{n\pi}{2}\right)-1\right]. \end{aligned} \tag{11.72}$$

This can be rewritten as

$$c_n = \begin{cases} \dfrac{2}{n\pi} & n \text{ odd} \\ \dfrac{4}{n\pi} & n = 2,6,10,\ldots. \\ 0 & n = 4,8,12,\ldots \end{cases} \tag{11.73}$$

Finally, step (3) is to determine $\Psi(x,t)$ using eq. (11.17). Remembering that $\omega_n = E_n/\hbar$, and using the energies in eq. (11.67), yields

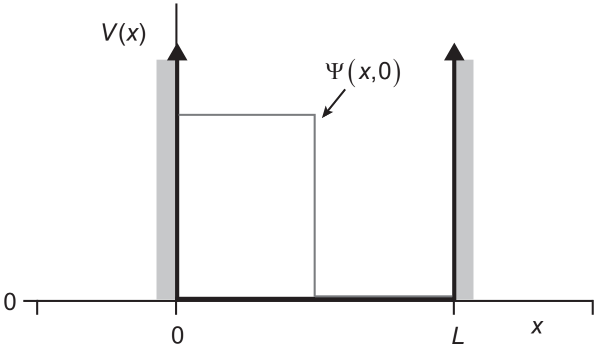


Fig 11.6 The wave function of a particle localized to the left half of an infinite square well potential.

$$
\begin{aligned}
\Psi(x,t) &= \sum_n c_n \psi_n(x) e^{-i\omega_n t} \\
&= \frac{2}{\pi}\sqrt{\frac{2}{L}} \sum_{n=1,3,5,\ldots} \frac{1}{n}\sin\left(\frac{n\pi x}{L}\right) e^{-i\left(n^2\pi^2\hbar/2mL^2\right)t} \\
&+ \frac{4}{\pi}\sqrt{\frac{2}{L}} \sum_{n=2,6,10,\ldots} \frac{1}{n}\sin\left(\frac{n\pi x}{L}\right) e^{-i\left(n^2\pi^2\hbar/2mL^2\right)t} .
\end{aligned}
\tag{11.74}
$$

(b) The probability of measuring E_n at time t is $P(E_n,t) = \left|\langle \psi_n | \psi(t) \rangle\right|^2$. Using eq. (11.16) we can write the inner product as

$$
\begin{aligned}
\langle \psi_n | \psi(t) \rangle &= \sum_m c_m e^{-i\omega_m t} \langle \psi_n | \psi_m \rangle \\
&= \sum_m c_m e^{-i\omega_m t} \delta_{nm} \\
&= c_n e^{-i\omega_n t} .
\end{aligned}
$$

The probability we seek is then

$$
P(E_n,t) = |c_n|^2 = \begin{cases} \dfrac{4}{(n\pi)^2} & n \text{ odd} \\ \dfrac{16}{(n\pi)^2} & n = 2,6,10,\ldots. \\ 0 & n = 4,8,12,\ldots \end{cases}
\tag{11.75}
$$

These probabilities are independent of time.

11.5.1 Time Evolution

Clearly the time evolution of the wave function described in eq. (11.74) is quite complicated, and is best calculated using a computer. You can visualize the time evolution of quantum states in an infinite square well potential (and other potentials) using the Java applets of refs. [11.5] and [11.6].

In example 11.1 we found that the probability of measuring any particular energy did not change with time, which means that the expectation value of the energy is time-independent as well. This is consistent with our discussion in sec. 9.3, and with conservation of energy.

Let's assume that we perform a measurement of the energy of the particle in example 11.1 and obtain E_1. What happens then? Postulate III(b) in sec. 5.2 tells us that the system is left in the eigenstate $|\psi_1\rangle$ [wave function $\psi_1(x)$] corresponding to E_1. Since $|\psi_1\rangle$ is an eigenstate of the Hamiltonian, the particle will remain in this state (as discussed in sec. 9.2). Future measurements of the energy with a particle in this state will yield E_1 with 100% probability.

A measurement of the position of the particle will change the state of the system. However, after a position measurement, the state of the particle is *not* a position eigenstate. This is because position eigenstates are defined with infinite precision, whereas real measurements have only a finite precision. The state of a particle after a position measurement is described by a wave packet. For example, an idealized measurement which finds the particle to be in the left-hand side of a well leaves us with the situation in example 11.1, with the state of the particle approximately described by eq. (11.71). A more realistic position measurement would most likely leave the particle in a state approximated by the Gaussian wave packet of eq. (10.14); the width of the wave packet being determined by resolution of the measurement. To observe the results of various measurements performed on particles in different potentials, see the Java applets of refs. [11.2] and [11.6].

11.5.2 Classical Comparison

If you think the wave functions in fig. 11.5 look just like the modes of a vibrating string tied at both ends, you're correct; the modes of a vibrating string are also described by eq. (11.68). Indeed, at $t = 0$ the problems of finding the series expansions of the initial wave function $\Psi(x,0)$, and the initial displacement of the string $Y(x,0)$, are essentially the same.

However, the time evolutions of the wave function and the vibrating string are different, because they have different dispersion relations. A dispersion relation is the relationship between ω and k. For a quantum particle in an infinite square well the dispersion relation is

$$\omega_n = \frac{E_n}{\hbar} = \frac{k_n^2\hbar}{2m} \quad n = 1,2,3,\ldots \text{ (infinite well)}, \tag{11.76}$$

while for the vibrating string it is

$$\omega_n = vk_n \quad n = 1,2,3,\ldots \text{ (string)}, \tag{11.77}$$

with v being the velocity of the wave on the string. The important difference that leads to different time evolutions is that for the classical wave $\omega_n \propto k_n$, while for the quantum wave $\omega_n \propto k_n^2$.

The time behavior of particles in other potentials is also dependent upon the corresponding dispersion relation. For example, the dispersion relation for a quantum free particle causes it to propagate in a manner that is different from a classical free particle (see complement 11.A).

While the detailed time evolution of a quantum particle can be quite different from its classical counterpart, the time evolution of expectation values (averages) for quantum and classical particles have some similarities. The temporal behavior of the expectation value of an observable in a time-independent Hamiltonian is given by eq. (9.19). Applying this equation to the position of a particle tells us that

$$\begin{aligned}\frac{d}{dt}\langle x\rangle &= \frac{i}{\hbar}\langle\psi(t)|\left[\hat{H},\hat{x}\right]|\psi(t)\rangle \\ &= \frac{i}{\hbar}\langle\psi(t)|\left[\left(\frac{\hat{p}^2}{2m}+V(\hat{x})\right),\hat{x}\right]|\psi(t)\rangle \\ &= \frac{i}{2m\hbar}\langle\psi(t)|\left[\hat{p}^2,\hat{x}\right]|\psi(t)\rangle \\ &= \frac{i}{2m\hbar}\langle\psi(t)|\left(\hat{p}\left[\hat{p},\hat{x}\right]+\left[\hat{p},\hat{x}\right]\hat{p}\right)|\psi(t)\rangle \\ &= \frac{1}{m}\langle\psi(t)|\hat{p}|\psi(t)\rangle,\end{aligned} \tag{11.78}$$

where we have used eqs. (7.8) and (10.32). Rearranging, this becomes

$$\langle p\rangle = m\frac{d}{dt}\langle x\rangle, \tag{11.79}$$

which says that the expectation values of position and momentum obey the corresponding classical equation.

In the problems you'll show that

$$\frac{d}{dt}\langle p\rangle = \left\langle -\frac{d}{dx}V(x)\right\rangle = \langle F(x)\rangle, \tag{11.80}$$

where the force on the particle $F(x)$ is defined in the usual way as the negative of the gradient of the potential energy. These equations may lead you to believe that expectation values in quantum mechanics always mimic classical results, but this is *not* true, because in general

$$\frac{d}{dt}\langle p\rangle = \langle F(x)\rangle \neq F(\langle x\rangle). \tag{11.81}$$

Indeed, even if $\langle F(x)\rangle = F(\langle x\rangle)$ quantum particles can display nonclassical temporal evolution, as we'll see in chapter 12.

Equations (11.79) and (11.80) are known as Ehrenfest's theorem. This theorem is quite general, as we have not assumed anything about the potential, other than that it is time independent.

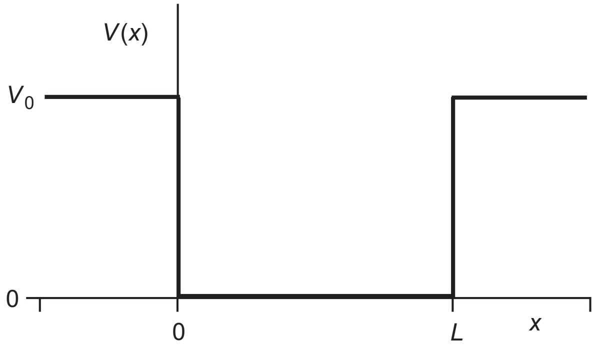


Fig 11.7 Finite square well potential.

11.5.3 Finite Well

As stated above, the infinite potential well is an idealization, as a real well will have a finite potential energy V_0 on the outside (fig. 11.7). Particles with $E > V_0$ have free-particle solutions of the form in eq. (11.27), and can have any energy. Particles with $E < V_0$ will be bound. On the inside of the well they have wave functions that oscillate, similar to those in fig. 11.5, and the allowed energies are quantized. However, these wave functions will not be 0 outside the well, but decay exponentially away from the boundary, as was the case for the potential step with $E < V_0$ in sec. 11.3.3. The difference between the well and the step is that the well has two sides, and we must apply boundary conditions at both. You can work out the details of this in problem 11.21.

11.5.4 Semiconductor Laser Diodes

While it may seem that the potential well is a theoretical curiosity with no real-world applications, that would be incorrect. Using modern fabrication techniques, such as molecular beam epitaxy (MBE) or metal-organic chemical vapor deposition (MOCVD), it is possible to create devices with dimensions that can be controlled on an atomic scale. When the size of such a device approaches the de Broglie wavelength of the particles, quantum size effects become important. In particular, one is not completely limited by the energy level structure of materials handed to us by nature; it is possible to engineer the energy level structure of a device by changing its size L [eq. (11.67)].

For example, the majority of semiconductor laser diodes use a layered structure called a quantum well, to tune and improve their performance. A very thin (~10 nm) layer of a material (such as gallium-arsenide, GaAs) is sandwiched between two layers of another material (such as aluminum-gallium-arsenide, AlGaAs) that has different electronic properties. The detailed physics of what happens in the semiconductor materials is beyond our interests here; suffice it to say that electrons and holes are confined along one dimension in the GaAs layer, which behaves as a potential well. By altering the thickness of this layer one can control the energy level structure, and hence the laser wavelength. In addition to wavelength tunability, quantum wells have other properties, such as higher laser gain, which improve device performance.

The 3-D analog of a 1-D quantum well is called a quantum dot, and is discussed in complement 13.A.

11.6 References

[11.1] D.J. Griffiths, *Introduction to Electrodynamics, 3rd ed.* (Prentice Hall, Upper Saddle River, New Jersey, 1999), p. 214.

[11.2] The PhET Project, *Quantum Tunneling and Wave Packets*, http://phet.colorado.edu/en/simulation/quantum-tunneling.

[11.3] A.M. Steinberg, P.G. Kwiat, and R.Y. Chiao, "Measurement of the single-photon tunneling time," Phys. Rev. Lett. **71**. 708 (1993).

[11.4] T. Nakanishi, K. Sugiyama, and M. Kitano, "Demonstration of negative group delays in a simple electronic circuit," Am. J. Phys. **70**, 1117 (2002).

[11.5] P. Falstad, *1-D Quantum States Applet*, http://www.falstad.com/qm1d/; The PhET Project, *Quantum Bound States*, http://phet.colorado.edu/en/simulation/bound-states.

[11.6] M. Belloni and W. Christian, *QM Measurement Package*, http://www.compadre.org/Repository/document/ServeFile.cfm?ID=9773&DocID=1461.

11.7 PROBLEMS

11.1 Use separation of variables to split the partial differential equation of eq. (11.8) into two ordinary differential equations. The steps for doing this are:

(i) Substitute $\Psi_n(x,t) = \psi_n(x)\phi_n(t)$ into eq. (11.8).

(ii) Divide the resulting equation by $\psi_n(x)\phi_n(t)$, and separate the temporal and spatial dependencies on opposite sides of the = sign.

(iii) Note that time changes cannot affect the spatial part of the equation, and vice versa. The only way for the temporal and spatial parts to be equal to each other is if they are both constants. Set both the temporal and spatial parts equal to the same constant, E_n, yielding two equations.

(iv) Rearrange the spatial equation to obtain eq. (11.19).

(v) Solve the temporal equation. Compare your solution to eq. (11.20).

11.2* Verify eq. (11.44).

11.3* (a) Verify eq. (11.48) by using the appropriate ratio of probability fluxes.

(b) Show that $T = |A_2|^2 / |A_1|^2$ does *not* give the correct solution.

11.4* Verify eq. (11.53).

11.5 For a particle incident on a potential step with $E < V_0$, show that $T = 0$ using the appropriate ratio of probability fluxes.

11.6 For a particle incident on a potential step with $E < V_0$, show that the magnitudes of the amplitudes of the incident and reflected waves functions are the same. Find the phase shift that the wave function acquires on reflection.

11.7* A particle experiences the potential shown in fig. 11.1(a). It has $E > V_0$, is to the right of the origin ($x > 0$), and is moving to the left. In other words, it doesn't strike a potential barrier, but falls off a potential cliff. Calculate the probabilities that the particle will be found to be reflected or transmitted by the cliff.

11.8 A particle is incident from the left on the potential barrier shown in fig. 11.2 with $E < V_0$. (a) Solve the Schrödinger equation, with appropriate boundary conditions, to find the spatial wave function throughout all of space. (b) Show that the probability T of measuring the particle to tunnel through the barrier is

$$T = \left(1 + \frac{V_0^2}{4E\left(V_0 - E\right)} \sinh^2\left(\alpha L\right)\right)^{-1}. \qquad (11.82)$$

11.9 Take the limit that $\alpha L \gg 1$ in eq. (11.82), verifying that $T \sim e^{-2\alpha L}$ in this limit.

11.10 Show that the wave functions $\psi_n\left(x\right)$ of the infinite square well are orthogonal.

11.11* Use the Heisenberg indeterminacy relation to estimate the ground state energy of a particle in an infinite square well potential.

11.12* Determine whether or not the wave functions $\psi_n\left(x\right)$ of the infinite square well satisfy the Heisenberg indeterminacy relation.

11.13 At $t = 0$ a particle in an infinite square well potential [eq. (11.59)] is in a state described by the wave function

$$\Psi\left(x,0\right) = \begin{cases} \sqrt{\dfrac{2}{3L}} \sin\left(\dfrac{2\pi}{L}x\right) + i\sqrt{\dfrac{4}{3L}} \sin\left(\dfrac{3\pi}{L}x\right) & 0 < x < L \\ 0 & \textit{elsewhere} \end{cases}. \qquad (11.83)$$

Determine (a) the probability $P\left(E_n,t\right)$ that a measurement of the energy will yield the value E_n, (b) $\langle E\rangle\left(t\right)$, (c) $\langle x\rangle\left(t\right)$, (d) $\langle p\rangle\left(t\right)$.

11.14 At $t = 0$ a particle in an infinite square well potential [eq. (11.59)] is in a state that is well approximated by the wave function

$$\Psi\left(x,0\right) = \begin{cases} \dfrac{2}{\sqrt{L}} \sin\left(\dfrac{2\pi}{L}x\right) & 0 < x < L/2 \\ 0 & \textit{elsewhere} \end{cases}, \qquad (11.84)$$

which is depicted in fig. 11.8 (technically this is an illegal wave function, as the derivative is not continuous at $x = L/2$). Determine (a) $\Psi\left(x,t\right)$, (b) the probability $P\left(E_n,t\right)$ that a measurement of the energy at time t will yield the value E_n.

11.15 At $t = 0$ a particle in an infinite square well potential [eq. (11.59)] is in a state described by the wave function

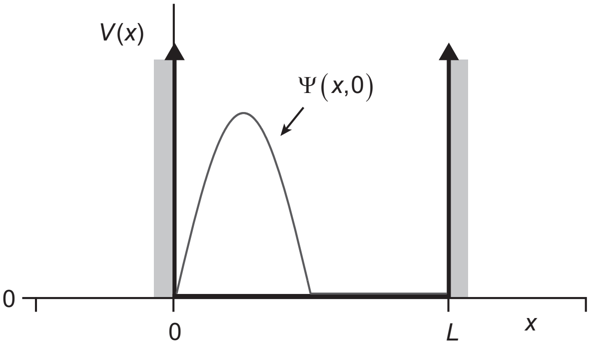


Fig 11.8 The wave function of eq. (11.84).

$$\Psi(x,0)=\begin{cases}\sqrt{\dfrac{30}{L^5}}\,x(L-x) & 0<x<L\\ 0 & elsewhere\end{cases}, \tag{11.85}$$

which is depicted in fig. 11.9. Determine
(a) $\Psi(x,t)$, (b) the probability $P(E_n,t)$ that a measurement of the energy at time t will yield the value E_n, (c) $\langle E\rangle(t)$.

11.16 For the particle in problem 11.15, calculate $\langle x\rangle(t)$.

11.17 For the particle in problem 11.15, calculate $\langle p\rangle(t)$. (Hint: This is straightforward if you use the results of problem 11.16.)

11.18 At $t=0$ a particle in an infinite square well potential [eq. (11.59)] is in a state that is well approximated by the wave function

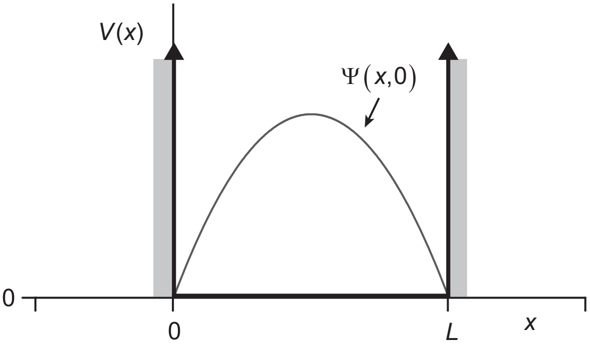


Fig 11.9 The wave function of eq. (11.85).

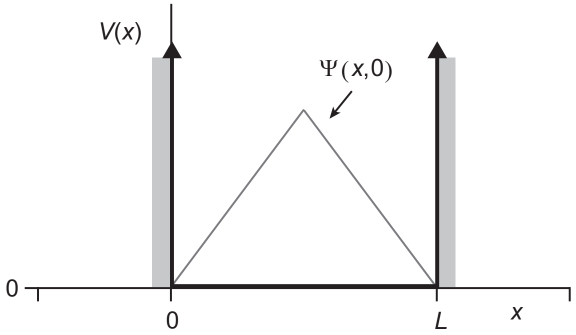


Fig 11.10 The wave function of eq. (11.86).

$$\Psi(x,0)=\begin{cases}\sqrt{\dfrac{12}{L^3}}x & 0<x<L/2\\ \sqrt{\dfrac{12}{L^3}}(L-x) & L/2<x<L,\\ 0 & elsewhere\end{cases} \tag{11.86}$$

which is depicted in fig. 11.10 (technically this is an illegal wave function, as the derivative is not continuous at $x=L/2$). Determine (a) $\Psi(x,t)$, (b) $\langle E\rangle(t)$.

11.19 Verify eq. (11.80). [Hint: Eq. (10.72) may be useful.]

11.20* When can we say that $\langle F(x)\rangle \cong F(\langle x\rangle)$, so that quantum mechanical expectation values behave as their classical counterparts? (Hint: use a power series.)

11.21 For a particle in the finite square well potential of fig. 11.7, show that if $E<V_0$ the allowed energies are given by the solutions to the equation

$$\tan(kL)=\frac{2\alpha k}{(k^2-\alpha^2)}, \tag{11.87}$$

where

$$k=\left(\frac{2mE}{\hbar^2}\right)^{1/2}, \alpha=\left(\frac{2m(V_0-E)}{\hbar^2}\right)^{1/2}. \tag{11.88}$$

COMPLEMENT 11.A

Free Particle Propagation

In sec. 11.2 we discussed the solutions to the Schrödinger equation for a free particle in a constant potential. The solutions we obtained there [eq. (11.24)], however, are not localized, and hence are unphysical. Here we'll consider localized, wave-packet solutions to the constant potential Schrödinger equation [eq. (11.22)]. We'll assume that $V_0 = 0$, because a constant shift in the potential energy has no physical significance anyway.

11.A.1 The Initial Wave Packet

We know from our discussions in chapter 10 that localized particles must have a spread of momenta. Since $k = p/\hbar$, a spread of momenta means a spread of wave vectors. We know that the constant potential Schrödinger equation allows solutions consisting of a spread of wave vectors, because the solutions that we found in eq. (11.24) were valid for any value of k. Furthermore, the Schrödinger equation is linear, so any superposition of solutions with different k's is also a solution.

The shape of the wave packet at $t = 0$, $\Psi(x,0)$, is determined by the initial conditions. Let's assume that initially our free particle is described by a wave packet that is a Gaussian centered about the origin, so

$$\Psi(x,0) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^{1/2}} e^{-x^2/4\sigma^2} e^{ip_0 x/\hbar}, \tag{11.A.1}$$

where σ and p_0 are constants. The parameter σ determines the initial width of the wave packet, and p_0 is the momentum of the particle (see the problems). In sec. 10.3 we described how to represent position-basis wave functions in terms of momentum-basis wave functions. We can use eq. (10.58) to write our initial wave function as

$$\Psi(x,0) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp\, e^{ipx/\hbar} \tilde{\psi}(p). \tag{11.A.2}$$

We can think of this as a superposition of momentum-state wave functions $e^{ipx/\hbar}$, where amplitude of each of the wave functions is given by $\tilde{\psi}(p)$. Equation (11.A.2) says that $\Psi(x,0)$ is the inverse Fourier transform of $\tilde{\psi}(p)$. However, we know $\Psi(x,0)$, and we want to find $\tilde{\psi}(p)$, so we use the Fourier transform relationship of eq. (10.57), which tells us that

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar} \Psi(x,0). \tag{11.A.3}$$

Substituting in our expression for $\Psi(x,0)$, eq. (11.A.1), we see that

$$\begin{aligned} \tilde{\psi}(p) &= \frac{1}{\sqrt{2\pi\hbar}\left(\sigma\sqrt{2\pi}\right)^{1/2}} \int_{-\infty}^{\infty} dx\, e^{-ipx/\hbar} e^{-x^2/4\sigma^2} e^{ip_0x/\hbar} \\ &= \frac{1}{\sqrt{2\pi\hbar}\left(\sigma\sqrt{2\pi}\right)^{1/2}} \int_{-\infty}^{\infty} dx\, e^{-i(p-p_0)x/\hbar} e^{-x^2/4\sigma^2} \\ &= \sqrt{\frac{\sigma}{\hbar}} \left(\frac{2}{\pi}\right)^{1/4} e^{-(p-p_0)^2\sigma^2/\hbar^2} \quad . \end{aligned} \tag{11.A.4}$$

11.A.2 Propagation

In the case of a discrete spectrum of energies, E_n, the time evolution of the wave function is described by eq. (11.17). However, for a free particle we have a continuous energy spectrum, so we need to use an integral instead of a sum. Rather than trying to turn the sum of eq. (11.17) into an integral, the solution will be more transparent if we go back and look at the time dependence of the state vector.

The momentum states $|p\rangle$ are eigenstates of the free-particle Hamiltonian, which is

$$\hat{H} = \frac{\hat{p}^2}{2m}. \tag{11.A.5}$$

Thus, we can use them to write the time-dependent state as

$$\begin{aligned} |\psi(t)\rangle &= e^{-i\hat{H}t/\hbar} |\psi(0)\rangle \\ &= \int_{-\infty}^{\infty} dp\, e^{-i\hat{H}t/\hbar} |p\rangle\langle p|\psi(0)\rangle \\ &= \int_{-\infty}^{\infty} dp\, e^{-i\hat{p}^2t/2m\hbar} |p\rangle \tilde{\psi}(p) \\ &= \int_{-\infty}^{\infty} dp\, e^{-ip^2t/2m\hbar} |p\rangle \tilde{\psi}(p), \end{aligned} \tag{11.A.6}$$

where we have used eq. (10.49). $\tilde{\psi}(p)$ is the momentum representation of the state at $t = 0$ [eq. (11.A.4)]. The time-dependent wave function can then be written as

$$
\begin{aligned}
\Psi(x,t) &= \langle x|\psi(t)\rangle \\
&= \int_{-\infty}^{\infty} dp \langle x|p\rangle e^{-ip^2 t/2m\hbar}\tilde{\psi}(p) \qquad (11.A.7)\\
&= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp\, e^{ipx/\hbar} e^{-ip^2 t/2m\hbar}\tilde{\psi}(p),
\end{aligned}
$$

where we've used eq. (10.56).

We can make eq. (11.A.7) look a little nicer if we recognize that the angular frequency of a free particle can be written as

$$
\omega = \frac{E}{\hbar} = \frac{k^2\hbar^2}{2m\hbar} = \frac{p^2}{2m\hbar}, \qquad (11.A.8)
$$

where we've used eqs. (11.26) (with $V_0 = 0$). Substituting this into eq. (11.A.7) gives

$$
\Psi(x,t) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp\, e^{ipx/\hbar} e^{-i\omega t}\tilde{\psi}(p). \qquad (11.A.9)
$$

Note, however, that we can't pull the factor of $e^{-i\omega t}$ outside of the integral, because ω is a function of p. Equation (11.A.9) is the continuous variable analog of eq. (11.17).

We can substitute eq. (11.A.4) into eq. (11.A.7), to obtain

$$
\begin{aligned}
\Psi(x,t) &= \frac{1}{\sqrt{2\pi\hbar}} \sqrt{\frac{\sigma}{\hbar}} \left(\frac{2}{\pi}\right)^{1/4} \int_{-\infty}^{\infty} dp\, e^{ipx/\hbar} e^{-ip^2 t/2m\hbar} e^{-(p-p_0)^2\sigma^2/\hbar^2} \\
&= \left(\frac{2}{\pi}\right)^{1/4} \left(2\sigma + i\frac{t\hbar}{\sigma m}\right)^{-1/2} \exp\left[\frac{-mx^2\hbar + i\left(-2\sigma^2 p_0^2 t + 4mx\sigma^2 p_0\right)}{2\hbar\left(2\sigma^2 m + it\hbar\right)}\right],
\end{aligned}
\qquad (11.A.10)
$$

where exp[ ] is the exponential function. Equation (11.A.10) can't be written much more simply, but we can write the magnitude of $\Psi(x,t)$ as

$$
|\Psi(x,t)| = \left(\frac{2}{\pi}\right)^{1/4} \left(4\sigma^2 + \frac{t^2\hbar^2}{\sigma^2 m^2}\right)^{-1/4} \exp\left[\frac{-\left(x - v_0 t\right)^2}{\left(4\sigma^2 + \dfrac{t^2\hbar^2}{m^2\sigma^2}\right)}\right], \qquad (11.A.11)
$$

where $v_0 = p_0 / m$ is the particle velocity.

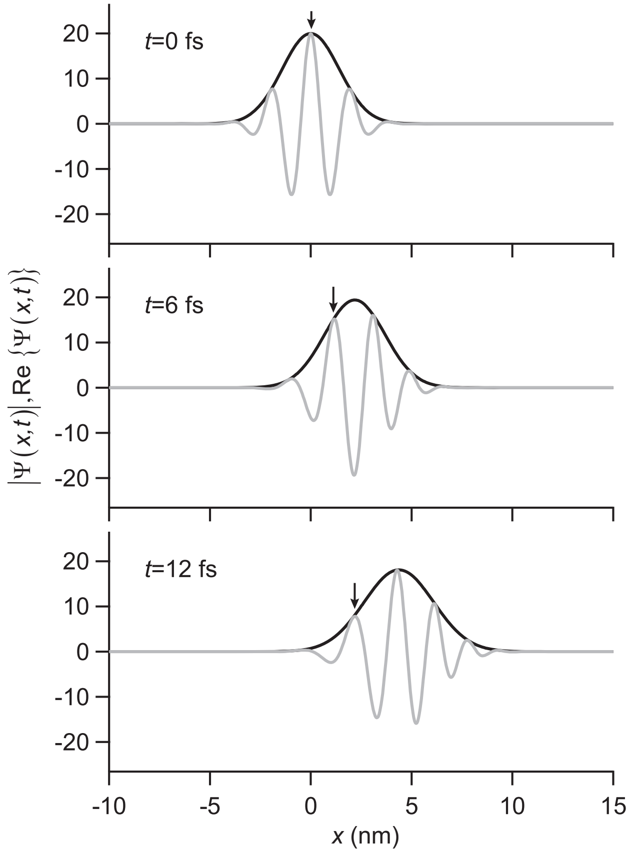


Fig 11.A.1 An electron wave packet described by eq. (11.A.10), with $\sigma = 2\,\text{nm}$, $\lambda = 1\,\text{nm}$, and $p_0 = h/\lambda$, is plotted at different times. The magnitude of the wave function $|\Psi(x,t)|$ is plotted in black, while the real part $\text{Re}\{\Psi(x,t)\}$ is plotted in gray. The arrow follows the peak of one of the oscillations of $\text{Re}\{\Psi(x,t)\}$.

Plots of the time evolution of an electron wave packet, described by eq. (11.A.10), are shown in fig. 11.A.1. The solid black line represents the magnitude of the wave function, and it can be seen moving to the right because of its initial momentum. That the wave packet moves with velocity v_0 can be seen by calculating the expectation value of the position, which for the wave function of eq. (11.A.10) yields $\langle x \rangle = v_0 t$.

The particle velocity is the velocity with which $|\Psi(x,t)|$ moves. This velocity is known as the group velocity of the wave v_g, which is given by $v_g = d\omega/dk$. In the problems you'll show that $v_g = v_0$, as we would expect.

However, note in fig. 11.A.1 that the envelope of the wave function, $|\Psi(x,t)|$, and the oscillations underneath it do not propagate at the same speed. The arrow indicating one of the oscillation peaks is clearly moving slower than the peak of the envelope. The oscillations propagate at the phase velocity $v_p = \omega/k$, which you'll find in the problems is $v_p = v_0/2$. The phase velocity is half of the particle velocity for a quantum mechanical free particle.

Longer-term evolution of the wave packet is shown in fig. 11.A.2. It can be seen that the wave packet spreads as it propagates. Additionally, the wave packet becomes "chirped", which means that the wavelengths (frequencies) contained in the pulse spread out. Notice that at $t = 60$ fs the leading edge of the packet contains shorter wavelengths, and the back edge of the packet contains longer wavelengths. This is because the shorter wavelengths correspond to larger momenta, which move faster. The term "chirp" comes from the fact that if you listen to a sound whose frequency varies in time, it sounds like a chirp.

The wave packet in fig. 11.A.2 spreads rapidly because it is very narrow to start with. A wave packet that is many wavelengths wide at $t = 0$ would still spread, but not nearly as rapidly. To create your own wave packets, and see how they propagate in time, see the Java applet at ref. [11.A.1].

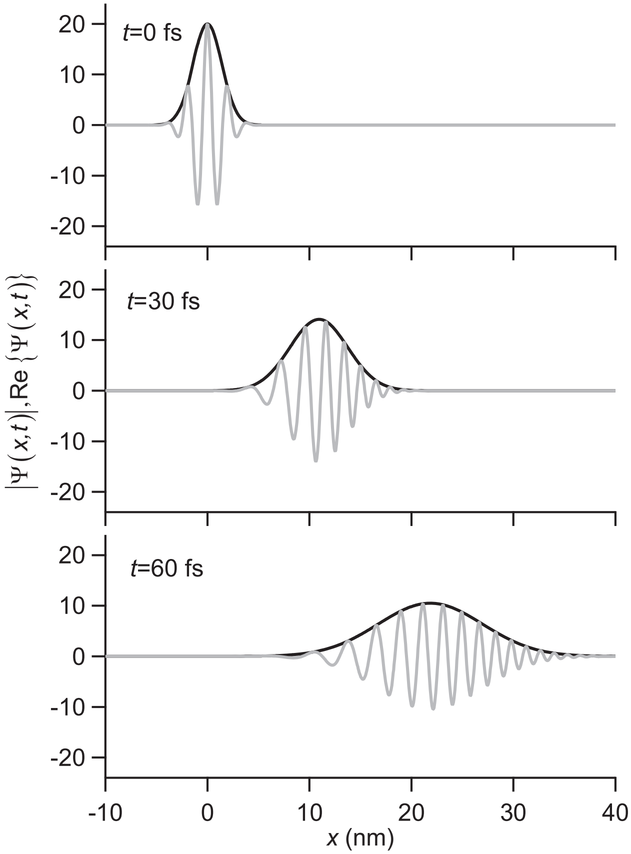


Fig 11.A.2 An electron wave packet with the same parameters as fig. 11.A.1 is plotted at different times. The magnitude of the wave function $|\Psi(x,t)|$ is plotted in black, while the real part $\mathrm{Re}\{\Psi(x,t)\}$ is plotted in gray.

11.A.3 References

[11.A.1] The PhET Project, *Quantum Tunneling and Wave Packets*, http://phet.colorado.edu/en/simulation/quantum-tunneling.

11.A.4 PROBLEMS

11.A.1* Show that the wave function in eq. (11.A.1) has $\langle p \rangle = p_0$. Do this by expressing $\hat{p}$ in the position basis.

11.A.2 Show that the wave function in eq. (11.A.4) has $\langle p \rangle = p_0$. Do this by expressing $\hat{p}$ in the momentum basis.

11.A.3* Show that for a free particle in a constant potential, $v_g = v_0$ and $v_p = v_0 / 2$.

CHAPTER 12

The Harmonic Oscillator

In the previous chapter we discussed several systems where potential energy depends on position. Another system with a position-dependent potential is the harmonic oscillator, and we'll discuss it here. We'll describe the properties of the quantum harmonic oscillator, and look at the similarities and differences between it and its classical counterpart.

12.1 WHY STUDY THE HARMONIC OSCILLATOR?

Think about a particle placed in the arbitrary one-dimensional potential shown in fig 12.1. Can we say anything general about the behavior of this particle? Let's assume that we "cool" the particle, to lower its energy. The particle will then seek the location of the potential minimum, which in fig 12.1 is at $x = x_0$. We can expand the potential energy in a Taylor series about this point as

$$V(x) = V(x_0) + V'(x_0)(x - x_0) + \frac{1}{2}V''(x_0)(x - x_0)^2 + \ldots . \tag{12.1}$$

The first term in eq. (12.1) is constant, and subtracting it merely shifts the minimum of the potential to 0, with no effect on the behavior of the particle. The second term in eq. (12.1) is 0, because we are expanding about the minimum of the potential. Given these facts, the expansion of the potential energy function becomes

$$V(x) = \frac{1}{2}\kappa(x - x_0)^2 + \ldots, \tag{12.2}$$

where $\kappa = V''(x_0)$. The lowest order term in the expansion is quadratic in displacement, and is shown in fig. 12.1. Since a simple harmonic oscillator also has a potential energy that is quadratic, a particle in this potential will behave similarly to a harmonic oscillator; it will oscillate about the location of the minimum potential energy. As long

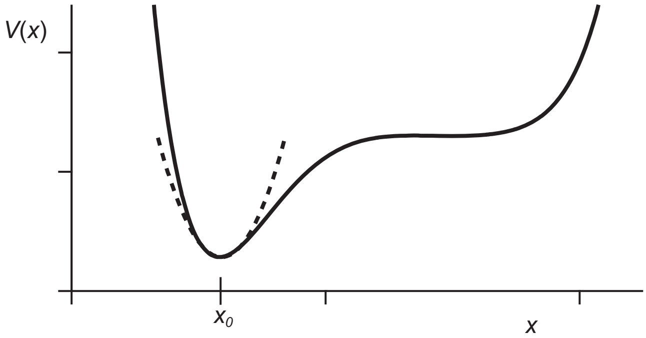


Fig 12.1 An arbitrary one-dimensional potential energy is plotted as a solid line. A second-order Taylor-series approximation to the potential at $x = x_0$ is plotted as a dashed line.

as the amplitude of the oscillations is not large, the behavior of the particle will be well approximated by that of a simple harmonic oscillator, with an effective "spring constant" of κ.[1]

12.2 CREATION, ANNIHILATION, AND NUMBER OPERATORS

The potential energy of a harmonic oscillator whose equilibrium position is at $x = 0$ can be written as

$$V(x) = \frac{1}{2} m\omega^2 x^2, \tag{12.3}$$

where m is the mass, and $\omega = \sqrt{\kappa/m}$ is the resonance angular frequency. With this potential energy, the Hamiltonian of the system is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2} m\omega^2 \hat{x}^2 = \frac{1}{2} m\omega^2 \left[\hat{x}^2 + \frac{1}{(m\omega)^2} \hat{p}^2 \right]. \tag{12.4}$$

Given this Hamiltonian, one can follow the procedure described in chap. 11 to find the allowed energies and corresponding wave functions. That is, write down the time-independent Schrödinger wave equation, and solve it. We'll use that approach in complement 12.A, but here we'll use an algebraic approach instead.

We used an algebraic approach to find the eigenvalues and eigenstates when discussing angular momentum in chap. 7. Where we defined the raising $\hat{J}_+$and lowering $\hat{J}_-$operators as [eq. (7.19)]

$$\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y. \tag{12.5}$$

---

[1] We're using κ instead of k for the spring constant, so as not to confuse it with the wave vector.

When applied to angular momentum states, $\hat{J}_+$ ($\hat{J}_-$) increases (decreases) by 1 the quantum number corresponding to the z-component of angular momentum. In complement 7.B these operators played a crucial role in determining the eigenvalues and eigenstates of angular momentum. We'll now see that the harmonic oscillator has analogous operators, which allow us to determine its eigenvalues and eigenstates, without solving any differential equations.

Consider the operator $\hat{a}$, which is defined as

$$\hat{a} \equiv \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} + \frac{i}{m\omega}\hat{p}\right). \tag{12.6}$$

The constant factor in front of the parentheses is chosen to make $\hat{a}$ dimensionless. The operator $\hat{a}$ is referred to as the annihilation operator, for reasons that will become obvious soon. This operator is not Hermitian; its adjoint is given by

$$\hat{a}^\dagger \equiv \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} - \frac{i}{m\omega}\hat{p}\right), \tag{12.7}$$

and $\hat{a}^\dagger$ is called the creation operator. We can also define the dimensionless number operator $\hat{n}$ as

$$\hat{n} \equiv \hat{a}^\dagger \hat{a}. \tag{12.8}$$

Using eqs. (12.6) and (12.7), this becomes

$$\begin{aligned} \hat{n} &= \frac{m\omega}{2\hbar}\left(\hat{x} - \frac{i}{m\omega}\hat{p}\right)\left(\hat{x} + \frac{i}{m\omega}\hat{p}\right) \\ &= \frac{m\omega}{2\hbar}\left[\hat{x}^2 + \frac{1}{(m\omega)^2}\hat{p}^2 + \frac{i}{m\omega}\left(\hat{x}\hat{p} - \hat{p}\hat{x}\right)\right] \\ &= \frac{1}{\hbar\omega}\left[\frac{1}{2}m\omega^2\hat{x}^2 + \frac{\hat{p}^2}{2m}\right] + \frac{i}{2\hbar}\left[\hat{x}, \hat{p}\right] \\ &= \frac{1}{\hbar\omega}\hat{H} - \frac{1}{2}, \end{aligned} \tag{12.9}$$

which can be rearranged to express the Hamiltonian as

$$\hat{H} = \hbar\omega\left(\hat{n} + \frac{1}{2}\right) = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right). \tag{12.10}$$

Because $\hat{H}$ is Hermitian, it is straightforward to see that $\hat{n}$ is as well.

As was the case for angular momentum, commutation relations play an important role in determining the eigenvalues and eigenstates of the harmonic oscillator. The first commutator we are interested in is that for $\hat{a}$ and $\hat{a}^\dagger$, which you'll show in the problems to be

$$\left[\hat{a}, \hat{a}^\dagger\right] = 1. \tag{12.11}$$

The commutator of $\hat{n}$ and $\hat{a}$ is

$$\begin{aligned}\left[\hat{n},\hat{a}\right]&=\left[\hat{a}^{\dagger}\hat{a},\hat{a}\right]\\&=\hat{a}^{\dagger}\left[\hat{a},\hat{a}\right]+\left[\hat{a}^{\dagger},\hat{a}\right]\hat{a}\\&=-\hat{a},\end{aligned}\tag{12.12}$$

where we've used eq. (7.8). This means

$$\begin{aligned}\left[\hat{n},\hat{a}\right]=\hat{n}\hat{a}-\hat{a}\hat{n}&=-\hat{a},\\\hat{n}\hat{a}&=\hat{a}\hat{n}-\hat{a},\end{aligned}\tag{12.13}$$

which we'll find useful below. Similarly, you can show that

$$\left[\hat{n},\hat{a}^{\dagger}\right]=\hat{a}^{\dagger}.\tag{12.14}$$

12.2.1 Eigenvalues and Eigenstates

The eigenvalues and eigenstates of the number operator $\hat{n}$ are n and $|n\rangle$, respectively:

$$\hat{n}|n\rangle=n|n\rangle.\tag{12.15}$$

Since $\hat{n}$ is Hermitian, its eigenvalues must be real. By examining eq. (12.10), we see that the states $|n\rangle$ are also eigenstates of the Hamiltonian,

$$\begin{aligned}\hat{H}|n\rangle&=\hbar\omega\left(\hat{n}+\frac{1}{2}\right)|n\rangle,\\E_n|n\rangle&=\hbar\omega\left(n+\frac{1}{2}\right)|n\rangle.\end{aligned}\tag{12.16}$$

The states $|n\rangle$ form an orthonormal basis that we can use to express any other state of the harmonic oscillator. I'll foreshadow the allowed values of n by telling you that the states $|n\rangle$ are called the number states, or the Fock states.

We begin our search for the eigenvalues and eigenstates of $\hat{n}$ by realizing that the form of the Hamiltonian in eq. (12.4) indicates that the energies of the harmonic oscillator must be positive. From eq. (12.16) we then know that

$$\begin{aligned}\hbar\omega\left(n+\frac{1}{2}\right)&\geq 0,\\n&\geq-\frac{1}{2},\end{aligned}\tag{12.17}$$

and there is a lower bound on n, n_{min}.

If we apply the annihilation operator $\hat{a}$ to a state $|n\rangle$, we will in general obtain a new state $|\psi\rangle$:

$$\hat{a}|n\rangle=|\psi\rangle.\tag{12.18}$$

Applying $\hat{n}$ to this, and using eq. (12.13), yields,

$$\begin{aligned}\hat{n}|\psi\rangle &= \hat{n}\hat{a}|n\rangle \\ &= \left(\hat{a}\hat{n}-\hat{a}\right)|n\rangle \\ &= \left(\hat{a}n-\hat{a}\right)|n\rangle \\ &= \left(n-1\right)\hat{a}|n\rangle \\ &= \left(n-1\right)|\psi\rangle,\end{aligned} \tag{12.19}$$

which means that the state $|\psi\rangle$ is an eigenstate of $\hat{n}$, with eigenvalue $n-1$. Using this in eq. (12.18) yields

$$\hat{a}|n\rangle = c_-|n-1\rangle, \tag{12.20}$$

where the constant c_- needs to be determined. The operator $\hat{a}$ lowers the value of n by 1, so we can see now why it's called the annihilation operator. Similarly, you'll show in the problems that the creation operator increases the value of n by 1:

$$\hat{a}^\dagger|n\rangle = c_+|n+1\rangle. \tag{12.21}$$

Thus, the operation of the creation and annihilation operators on harmonic oscillator states is similar to the operation of the raising and lowering operators on angular momentum states.

If we apply the annihilation operator to $|n_{\min}\rangle$, we can't lower n anymore (it's at its minimum). For consistency, we must then have

$$\hat{a}|n_{\min}\rangle = 0. \tag{12.22}$$

We can show that $n_{\min}=0$ by interrogating $|n_{\min}\rangle$ with $\hat{n}$:

$$\begin{aligned}\hat{n}|n_{\min}\rangle &= n_{\min}|n_{\min}\rangle, \\ \hat{a}^\dagger\hat{a}|n_{\min}\rangle &= n_{\min}|n_{\min}\rangle, \\ \hat{a}^\dagger(0) &= n_{\min}|n_{\min}\rangle, \\ 0 &= n_{\min}.\end{aligned} \tag{12.23}$$

Knowing the minimum value of n, we can use the creation operator to generate the states with other allowed values. Starting from $n=0$, eq. (12.21) tells us that the allowed values of n are the positive integers, which is why we refer to $\hat{n}$ as the number operator. Using this information in eq (12.16), we find that the allowed energies (eigenvalues of the Hamiltonian) for the harmonic oscillator are

$$E_n = \hbar\omega\left(n+\frac{1}{2}\right) \quad n = 0,1,2,\ldots. \tag{12.24}$$

Note that the ground-state energy E_0 is not 0. The reason for this is same as it was for the infinite well in sec. 11.5; a ground state energy of 0 is not allowed by the Heisenberg indeterminacy relation. Also note that the energy levels of the harmonic oscillator are evenly spaced, with a separation of

$$\Delta E = E_{n+1} - E_n = \hbar\omega. \tag{12.25}$$

12.2.2 Expectation Values

In chapter 11 we calculated expectation values such as $\langle x\rangle$ and $\langle p\rangle$ using integrals of wave functions. We can do this for the harmonic oscillator as well, however, for the harmonic oscillator we can also calculate these expectation values algebraically, without resorting to integrals or wave functions.

To do this, we must first determine the constants c_- and c_+ in eqs. (12.20) and (12.21). Taking the adjoint of eq. (12.20), we find

$$\langle n|\hat{a}^\dagger = \langle n-1|c_-^*. \tag{12.26}$$

Now using eqs. (12.20) and (12.26), we find

$$\begin{aligned} n &= \langle n|\hat{n}|n\rangle \\ &= \langle n|\hat{a}^\dagger\hat{a}|n\rangle \\ &= \langle n-1|c_-^*c_-|n-1\rangle \\ &= |c_-|^2 . \end{aligned} \tag{12.27}$$

Taking c_- to be real, $c_- = \sqrt{n}$, and eq. (12.20) becomes

$$\hat{a}|n\rangle = \sqrt{n}\,|n-1\rangle. \tag{12.28}$$

We can apply the same technique to find the constant c_+ in eq. (12.21), yielding

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}\,|n+1\rangle. \tag{12.29}$$

Next, we invert eqs. (12.6) and (12.7), to obtain the representations of $\hat{x}$ and $\hat{p}$ in terms of the creation and annihilation operators:

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega}}\left(\hat{a}+\hat{a}^\dagger\right) = \frac{1}{\beta\sqrt{2}}\left(\hat{a}+\hat{a}^\dagger\right). \tag{12.30}$$

$$\hat{p} = -i\sqrt{\frac{m\omega\hbar}{2}}\left(\hat{a}-\hat{a}^\dagger\right) = \frac{\hbar\beta}{i\sqrt{2}}\left(\hat{a}-\hat{a}^\dagger\right). \tag{12.31}$$

In these equations we have defined the quantity β,

$$\beta \equiv \sqrt{\frac{m\omega}{\hbar}}, \tag{12.32}$$

which has units of inverse length. We can use eqs. (12.30) and (12.31) to determine the operation of $\hat{x}$ and $\hat{p}$ on the number states. Since the number states form a basis, we can determine how $\hat{x}$ and $\hat{p}$ will operate on any harmonic oscillator state. Let's look at an example of this.

EXAMPLE 12.1

Calculate the standard deviation of the position Δx for a harmonic oscillator in state $|n\rangle$.

We begin by calculating $\langle x\rangle$ and $\langle x^2\rangle$:

$$\begin{aligned}\langle x\rangle &= \langle n|\hat{x}|n\rangle \\ &= \frac{1}{\beta\sqrt{2}}\langle n|\left(\hat{a}+\hat{a}^\dagger\right)|n\rangle \\ &= \frac{1}{\beta\sqrt{2}}\left[\langle n|\hat{a}|n\rangle+\langle n|\hat{a}^\dagger|n\rangle\right] \\ &= \frac{1}{\beta\sqrt{2}}\left[\sqrt{n}\,\langle n|n-1\rangle+\sqrt{n+1}\,\langle n|n+1\rangle\right] \\ &= 0,\end{aligned} \tag{12.33}$$

$$\begin{aligned}\langle x^2\rangle &= \langle n|\hat{x}^2|n\rangle \\ &= \frac{1}{2\beta^2}\langle n|\left(\hat{a}+\hat{a}^\dagger\right)^2|n\rangle \\ &= \frac{1}{2\beta^2}\left[\langle n|\hat{a}\hat{a}|n\rangle+\langle n|\hat{a}\hat{a}^\dagger|n\rangle+\langle n|\hat{a}^\dagger\hat{a}|n\rangle+\langle n|\hat{a}^\dagger\hat{a}^\dagger|n\rangle\right] \\ &= \frac{1}{2\beta^2}\left[\sqrt{n}\,\langle n|\hat{a}|n-1\rangle+\sqrt{n+1}\,\langle n|\hat{a}|n+1\rangle\right. \\ &\quad \left.+\sqrt{n}\,\langle n|\hat{a}^\dagger|n-1\rangle+\sqrt{n+1}\,\langle n|\hat{a}^\dagger|n+1\rangle\right] \\ &= \frac{1}{2\beta^2}\left[\sqrt{n(n-1)}\,\langle n|n-2\rangle+(n+1)\langle n|n\rangle\right. \\ &\quad \left.+n\langle n|n\rangle+\sqrt{(n+1)(n+2)}\,\langle n|n+2\rangle\right] \\ &= \frac{\hbar}{2m\omega}(2n+1).\end{aligned} \tag{12.34}$$

The fact that the expectation value of the position is 0 makes sense, because the potential energy is symmetric about $x=0$. We can now calculate the standard deviation of the position as

$$\Delta x=\left(\langle x^2\rangle-\langle x\rangle^2\right)^{1/2}=\left[\frac{\hbar}{m\omega}\left(n+\frac{1}{2}\right)\right]^{1/2}=\sqrt{\frac{E_n}{m\omega^2}}\,. \tag{12.35}$$

12.3 WAVE FUNCTIONS

Example 12.1 tells us that for the number states, as n increases the uncertainty in the position of the oscillator increases as well. To get a better feel for this, we can look at the wave functions. Complement 12.A illustrates how to solve the Schrödinger equation directly to find both the allowed energies and corresponding wave functions.

Once again, we'll take a different approach here.

Let's start by finding the ground state wave function. If we apply the annihilation operator to the ground state, the result is

$$\hat{a}|0\rangle = 0. \tag{12.36}$$

We can project this equation into the position basis, to obtain a differential equation for the ground state wave function $\psi_0(x) = \langle x|0\rangle$:

$$\begin{aligned}\langle x|\hat{a}|0\rangle &= \langle x|\frac{\beta}{\sqrt{2}}\left(\hat{x}+\frac{i}{m\omega}\hat{p}\right)|0\rangle \\ &= \frac{\beta}{\sqrt{2}}\left[x\psi_0(x)+\frac{i}{m\omega}\left(-i\hbar\frac{\partial}{\partial x}\right)\psi_0(x)\right] \\ &= 0.\end{aligned} \tag{12.37}$$

Here we have used eqs. (10.43), (12.6), and (12.32). Rearranging, we find that

$$\frac{\partial}{\partial x}\psi_0(x) = -\beta^2 x\psi_0(x). \tag{12.38}$$

In the problems you'll verify that the properly normalized solution to this differential equation is

$$\psi_0(x) = \left(\frac{\beta^2}{\pi}\right)^{1/4} e^{-\beta^2 x^2/2}, \tag{12.39}$$

which is a Gaussian wave packet.

Previously, we used the creation operator to determine the higher-energy states from the ground state. We can use the same basic technique to find the higher-order wave functions as well:

$$\begin{aligned}\hat{a}^\dagger|n\rangle &= \sqrt{n+1}|n+1\rangle, \\ \frac{\beta}{\sqrt{2}}\left(\hat{x}-\frac{i}{m\omega}\hat{p}\right)|n\rangle &= \sqrt{n+1}|n+1\rangle, \\ \frac{\beta}{\sqrt{2(n+1)}}\langle x|\left(\hat{x}-\frac{i}{m\omega}\hat{p}\right)|n\rangle &= \langle x|n+1\rangle, \\ \frac{\beta}{\sqrt{2(n+1)}}\left[x\psi_n(x)-\frac{1}{\beta^2}\frac{\partial}{\partial x}\psi_n(x)\right] &= \psi_{n+1}(x).\end{aligned} \tag{12.40}$$

This is not an equation we need to solve, it is a formula which allows us to determine $\psi_{n+1}(x)$ given $\psi_n(x)$. For example, if $n = 0$ we see that

$$\begin{aligned}\psi_1(x) &= \frac{\beta}{\sqrt{2}}\left[x\psi_0(x) - \frac{1}{\beta^2}\frac{\partial}{\partial x}\psi_0(x)\right]\\ &= \left(\frac{\beta^2}{\pi}\right)^{1/4}\frac{\beta}{\sqrt{2}}\left[xe^{-\beta^2x^2/2} - \frac{1}{\beta^2}\frac{\partial}{\partial x}e^{-\beta^2x^2/2}\right]\\ &= \left(\frac{\beta^2}{\pi}\right)^{1/4}\frac{\beta}{\sqrt{2}}\left[xe^{-\beta^2x^2/2} - \frac{1}{\beta^2}\left(\frac{-2x\beta^2}{2}\right)e^{-\beta^2x^2/2}\right]\\ &= \left(\frac{\beta^2}{\pi}\right)^{1/4}\frac{\beta}{\sqrt{2}}2x\,e^{-\beta^2x^2/2}.\end{aligned} \tag{12.41}$$

From $\psi_1(x)$ we can get $\psi_2(x)$, etc.

The general expression for the harmonic oscillator wave functions can be written as

$$\psi_n(x) = \left(\frac{\beta^2}{\pi}\right)^{1/4}\frac{1}{\sqrt{2^n n!}}H_n(\beta x)e^{-\beta^2x^2/2}, \tag{12.42}$$

where we recall that β is defined in Eq. (12.32). In eq. (12.42) the $H_n(x)$'s are the Hermite polynomials, the first few of which are given in table 12.1. The Hermite polynomials can be found using the recursion formula:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x). \tag{12.43}$$

Before moving on, let's recall something that we know about the classical harmonic oscillator. The total energy of a particle is given by the sum of its kinetic and potential energies. Since the kinetic energy is positive, we must have $E \geq V(x)$. The plot in fig. 12.2 shows the quadratic potential energy for a harmonic oscillator, as well as a horizontal line representing its total energy. There is a region in the center of the well where $E \geq V(x)$, and the particle is allowed to be. The points where $E = V(x)$ are where the kinetic energy is zero, and the particle turns around as it reaches its furthest excursions from the origin; they are called the classical turning points. We can calculate the turning

Table 12.1 The first six Hermite polynomials, $H_n(x)$.

$H_0(x) = 1$

$H_1(x) = 2x$

$H_2(x) = 4x^2 - 2$

$H_3(x) = 8x^3 - 12x$

$H_4(x) = 16x^4 - 48x^2 + 12$

$H_5(x) = 32x^5 - 160x^3 + 120x$

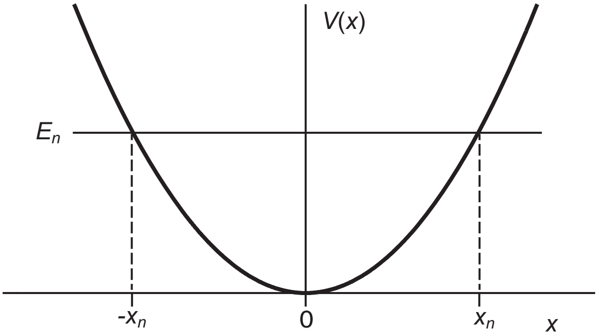


Fig 12.2 A quadratic potential energy corresponding to a harmonic oscillator. The total energy E_n is a horizontal line, and the places where the total energy intersects the potential energy correspond to the classical turning points $\pm x_n$.

points x_n for a classical particle, whose energy E_n is given by the quantum value in eq. (12.24):

$$\begin{aligned} E_n &= V(x_n) \\ \hbar\omega\left(n+\frac{1}{2}\right) &= \frac{1}{2}m\omega^2 x_n^2 \\ x_n^2 &= 2\frac{\hbar}{m\omega}\left(n+\frac{1}{2}\right) \\ x_n &= \pm\frac{1}{\beta}\sqrt{(2n+1)} \qquad n = 0,1,2,\ldots \end{aligned} \tag{12.44}$$

Figure 12.3 plots the three lowest order wave functions and their square magnitudes (corresponding probability densities), and figure 12.4 plots the probability density of a higher-order wave function. The probability density corresponding to the wave function $\psi_n(x)$ has $n+1$ maxima. As n increases, the wave function gets broader, consistent with eq. (12.35). Between the classical turning points the wave function is oscillatory. While the wave function extends beyond the turning points, it decays rapidly in this classically forbidden region.

12.4 FOCK STATES AND PHOTONS

Let's discuss the Fock states, $|n\rangle$, of the harmonic oscillator in more detail. We'll begin by noting that the time-dependent wave function for the state $|n\rangle$ is

$$\Psi(x,t) = \psi_n(x)e^{-iE_nt/\hbar}. \tag{12.45}$$

and its corresponding probability density is

$$|\Psi(x,t)|^2 = \left(\psi_n^*(x)e^{iE_nt/\hbar}\right)\left(\psi_n(x)e^{-iE_nt/\hbar}\right) = |\psi_n(x)|^2, \tag{12.46}$$

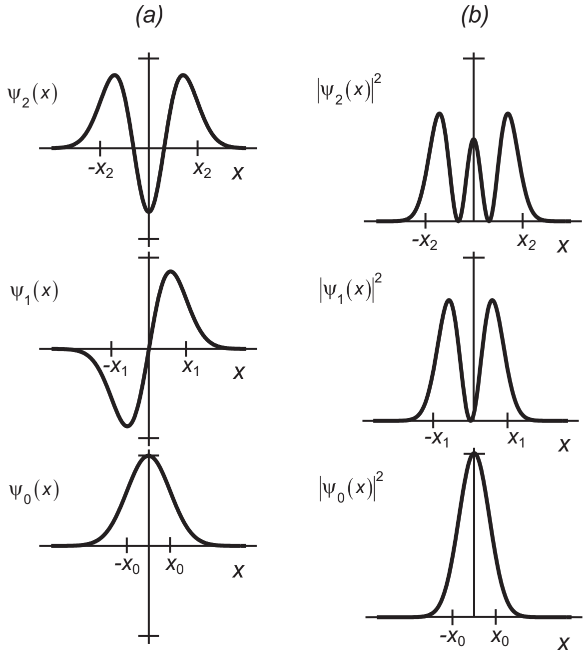


Fig 12.3 (a) The wave functions corresponding to the three lowest energies of a harmonic oscillator. (b) The squares of these wave functions, which correspond to the probability density of finding a particle at a particular position. The classical turning points $\pm x_n$ are indicated.

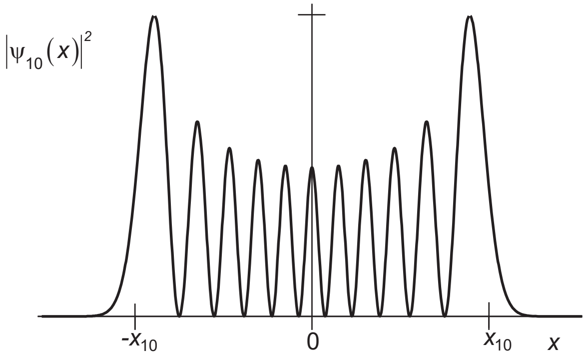


Fig 12.4 The probability density corresponding to the $n = 10$ state of the harmonic oscillator. The classical turning points $\pm x_{10}$ are indicated.

This probability is independent of time. The oscillator doesn't oscillate! Why not?

As discussed in sec. 9.2, eigenstates of the Hamiltonian are time-independent; they acquire a time-dependent overall phase shift, but the state itself is unchanged. In order for a state to evolve in time, it must be in a superposition of two, or more, energy eigenstates. If a harmonic oscillator is in a superposition of states corresponding to two adjacent energy levels, it will oscillate at angular frequency ω, as you'll see in the problems.

Ordinarily, we think of the classical limit of quantum mechanics as applying when the energy of an object becomes large. For a harmonic oscillator this means that n is large. However, no matter how large n gets, eq. (12.46) tells us that the oscillator won't oscillate. The Fock states $|n\rangle$ are nonclassical states, no matter how large n is. From a practical standpoint, however, as n gets large the energy of the particle grows ($E_n \simeq n\hbar\omega$), while the separation between the energy levels remains constant at $\Delta E = \hbar\omega$; the ratio of the energy separation to the total energy becomes small. As n increases, it becomes increasingly difficult to place a particle in state $|n\rangle$, without any contributions from other states. Indeed, even for small n it's usually experimentally challenging to place a quantum system into a Fock state.

An example of a harmonic oscillator that can be placed in a Fock state is a single atom in a trap. An rf Paul trap uses radio-frequency fields applied to three electrodes (a center ring shaped electrode, and two end caps) to confine ions in three dimensions. Along each direction the potential energy seen by the ion can be approximated by a harmonic oscillator potential, and using lasers it is possible to excite different states of atomic motion along the trap axis. In ref. [12.1] the vibrations of a $^9\mathrm{Be}^+$ ion were excited into Fock states of up to $n = 16$.

Another quantum oscillator is described in ref. [12.2]. Here, a microscopic mechanical oscillator is cooled to its ground state, and then placed into an oscillation corresponding to the Fock state $|1\rangle$. A picture of this oscillator is shown in fig. 12.5(a), and while it is considered microscopic, it's clearly VERY large compared to a single atom or molecule. Indeed, this is one of the largest objects I'm aware of that has been demonstrated to exhibit nonclassical behavior (along with the diamond crystals described in complement 8.A). Note that the oscillation is not an up-and-down oscillation of the cantilever, but rather a thickness oscillation of the suspended mass, as shown in fig. 12.5(b). This is referred to as a dilatational mode of oscillation.

A single mode of the electromagnetic field is another example of a harmonic oscillator that can be placed in a Fock state.[2] That the field is described by a harmonic oscillator will be discussed in detail in chapter 16. For now, suffice it to say that in this case it is not the position of an individual particle that is oscillating, but the field itself.

Recall that the allowed energies of a harmonic oscillator are

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) \quad n = 0,1,2,\dots . \tag{12.47}$$

A single photon of angular frequency ω has an energy of $\hbar\omega$, so this equation indicates that for an electromagnetic field, n tells us the number of photons in that field.[3] A field with no photons ($n = 0$) is said to be a vacuum field. The energy of a vacuum field is nonzero. The contribution to the field energy by the vacuum is called the vacuum energy, or the zero-point energy (more on this in chapter 16).

---

[2] Each mode of an electromagnetic field can be labeled by its wave vector and polarization. A plane wave field [eg, eq. (2.10)] is a single-mode field.

[3] Remember that before any measurement, the field will in general be in a superposition of states with different values of n, and the number of photons is not well defined. After we perform a measurement, however, we can say that the field was measured to contain n photons.

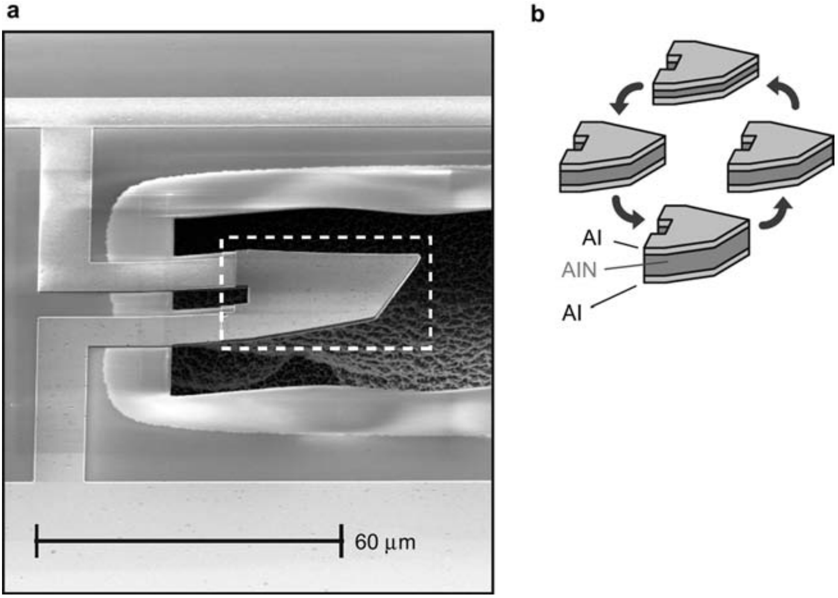


Fig 12.5 (a) A suspended micromechanical oscillator consisting of, from bottom to top, 150 nm SiO_2, 130 nm Al, 330 nm AlN, and 130 nm Al. The AlN is a piezoelectric material, and a voltage applied between the Al electrodes causes the AlN to expand or contract (alternatively, if the AlN expands or contracts, a voltage is generated). Because of this, the oscillator can be coupled to an electronic circuit, which can be used to drive and/or measure the oscillations. (b) An illustration of the dilatational mode of oscillation; the oscillation frequency is 6.175 GHz. Reprinted by permission from Macmillan Publishers Ltd: A.D. O'Connell et al., Nature **464**, 697, copyright (2010).

In the laboratories described at the end of this book we use spontaneous parametric down conversion to create a field containing a single photon, which means that it is in the Fock state $|1\rangle$. This is done by splitting a single photon from a pump beam into two photons (the signal and idler photons) using a crystal, as described in sec. 8.1 and lab 1. Detection of an idler photon projects the signal beam into a single-photon state.

12.5 COHERENT STATES

The Fock states of the harmonic oscillator are nonclassical, even in the limit of large n. However, we know that classical harmonic oscillators exist. What states can we use to describe their motion?

The quantum mechanical states that best describe the behavior of a classical harmonic oscillator are called the coherent states, $|\alpha\rangle$. These states are the eigenstates of the annihilation operator $\hat{a}$:

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle. \tag{12.48}$$

Recall that $\hat{a}$ is not a Hermitian operator, so its eigenvalues α will in general be complex numbers. As we'll see, the magnitude of α determines the amplitude of the oscillation, while the phase of α determines the phase of the oscillation. Note that eq. (12.48) also implies that

$$\langle\alpha|\hat{a}^{\dagger} = \langle\alpha|\alpha^{*}. \tag{12.49}$$

Let's write the state $|\alpha\rangle$ as a linear combination of Fock states:

$$|\alpha\rangle = \sum_{n=0}^{\infty} c_n |n\rangle. \tag{12.50}$$

Applying the annihilation operator to this state, and using eq. (12.28), we find

$$\begin{aligned} \hat{a}|\alpha\rangle &= \sum_{n=0}^{\infty} c_n \hat{a}|n\rangle \\ &= \sum_{n=1}^{\infty} c_n \sqrt{n}\,|n-1\rangle. \end{aligned} \tag{12.51}$$

We also know that

$$\begin{aligned} \hat{a}|\alpha\rangle &= \alpha|\alpha\rangle \\ &= \alpha \sum_{n'=0}^{\infty} c_{n'} |n'\rangle \\ &= \sum_{n'=0}^{\infty} c_{n'}\alpha |n'\rangle. \end{aligned} \tag{12.52}$$

Make the substitution $n' = n-1$, and this equation becomes

$$\hat{a}|\alpha\rangle = \sum_{n=1}^{\infty} c_{n-1}\alpha |n-1\rangle. \tag{12.53}$$

Since the series representation of $|\alpha\rangle$ must be unique, each of the terms in eqs. (12.51) and (12.53) must be equal. Equating the coefficients, we find that

$$\begin{aligned} c_n\sqrt{n} &= c_{n-1}\alpha, \\ c_n &= \frac{\alpha}{\sqrt{n}} c_{n-1}. \end{aligned} \tag{12.54}$$

We pick up a factor of $\alpha/\sqrt{n}$ each time we increase n. In terms of c_0, we can write this as

$$c_n = \frac{\alpha^n}{\sqrt{n!}} c_0. \tag{12.55}$$

Substituting this into eq. (12.50), we find that

$$|\alpha\rangle = c_0 \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \tag{12.56}$$

The factorial in the denominator ensures that the series converges for all values of α.

The factor of c_0 in eq. (12.56) is used for normalization:

$$\begin{aligned}
\langle\alpha|\alpha\rangle &= |c_0|^2 \left[\sum_{n'=0}^{\infty} \frac{\left(\alpha^*\right)^{n'}}{\sqrt{n'!}} \langle n'|\right] \left[\sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle\right] \\
&= |c_0|^2 \sum_{n=0}^{\infty} \sum_{n'=0}^{\infty} \frac{\left(\alpha^*\right)^{n'}}{\sqrt{n'!}} \frac{\alpha^n}{\sqrt{n!}} \langle n'|n\rangle \\
&= |c_0|^2 \sum_{n=0}^{\infty} \frac{\left(\alpha^*\right)^{n}}{\sqrt{n!}} \frac{\alpha^n}{\sqrt{n!}} \\
&= |c_0|^2 \sum_{n=0}^{\infty} \frac{\left(|\alpha|^2\right)^{n}}{n!} \\
&= |c_0|^2 e^{|\alpha|^2} \\
&= 1.
\end{aligned} \tag{12.57}$$

Taking c_0 to be real, this means

$$c_0 = e^{-|\alpha|^2/2}, \tag{12.58}$$

and our final expression for the series representation of the coherent states is

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \tag{12.59}$$

Here α can be any complex number. In the case that $\alpha = 0$, we see that

$$|\alpha = 0\rangle = |n = 0\rangle.$$

The coherent state with $\alpha = 0$ is equal to the Fock state with $n = 0$, which is the ground state of the harmonic oscillator.

12.5.1 Expectation Values

Let's define a set of dimensionless operators as

$$\hat{X} \equiv \frac{\beta}{\sqrt{2}} \hat{x} = \frac{1}{2}\left(\hat{a} + \hat{a}^\dagger\right), \tag{12.60}$$

$$\hat{P} \equiv \frac{1}{\hbar\beta\sqrt{2}}\hat{p} = \frac{1}{2i}\left(\hat{a} - \hat{a}^{\dagger}\right), \tag{12.61}$$

where β is defined in eq. (12.32). These new variables are simply scaled position and momentum operators, and in terms of them we can write the creation and annihilation operators as

$$\hat{a} = \left(\hat{X} + i\hat{P}\right), \tag{12.62}$$

$$\hat{a}^{\dagger} = \left(\hat{X} - i\hat{P}\right). \tag{12.63}$$

We can use eqs. (12.48) and (12.49) to write the expectation value of X for a system in the state $|\alpha\rangle$ as

$$\begin{aligned}\langle X\rangle &= \langle\alpha|\hat{X}|\alpha\rangle \\ &= \frac{1}{2}\langle\alpha|\left(\hat{a} + \hat{a}^{\dagger}\right)|\alpha\rangle \\ &= \frac{1}{2}\left(\langle\alpha|\hat{a}|\alpha\rangle + \langle\alpha|\hat{a}^{\dagger}|\alpha\rangle\right) \\ &= \frac{1}{2}\left(\alpha + \alpha^{*}\right)\langle\alpha|\alpha\rangle \\ &= \mathrm{Re}(\alpha).\end{aligned} \tag{12.64}$$

Furthermore,

$$\begin{aligned}\left\langle X^{2}\right\rangle &= \langle\alpha|\hat{X}^{2}|\alpha\rangle \\ &= \frac{1}{4}\langle\alpha|\left(\hat{a} + \hat{a}^{\dagger}\right)\left(\hat{a} + \hat{a}^{\dagger}\right)|\alpha\rangle \\ &= \frac{1}{4}\langle\alpha|\left(\hat{a}^{2} + \hat{a}^{\dagger 2} + \hat{a}\hat{a}^{\dagger} + \hat{a}^{\dagger}\hat{a}\right)|\alpha\rangle.\end{aligned} \tag{12.65}$$

The commutator in eq. (12.11) tells us that

$$\begin{aligned}\left[\hat{a}, \hat{a}^{\dagger}\right] &= \hat{a}\hat{a}^{\dagger} - \hat{a}^{\dagger}\hat{a} = 1, \\ \hat{a}\hat{a}^{\dagger} &= \hat{a}^{\dagger}\hat{a} + 1,\end{aligned} \tag{12.66}$$

which means we can rewrite eq. (12.65) as

$$\begin{aligned}\left\langle X^{2}\right\rangle &= \frac{1}{4}\langle\alpha|\left(\hat{a}^{2} + \hat{a}^{\dagger 2} + 2\hat{a}^{\dagger}\hat{a} + 1\right)|\alpha\rangle \\ &= \frac{1}{4}\left(\alpha^{2} + \alpha^{*^{2}} + 2\alpha^{*}\alpha + 1\right) \\ &= \frac{1}{4}\left(\alpha + \alpha^{*}\right)^{2} + \frac{1}{4} \\ &= \left[\mathrm{Re}(\alpha)\right]^{2} + \frac{1}{4}.\end{aligned} \tag{12.67}$$

The standard deviation of X is then

$$\Delta X = \sqrt{\left\langle X^2 \right\rangle - \left\langle X \right\rangle^2} = \frac{1}{2}. \tag{12.68}$$

Similarly, it's straightforward to show that

$$\left\langle P \right\rangle = \mathrm{Im}(\alpha), \tag{12.69}$$

and

$$\Delta P = \frac{1}{2}. \tag{12.70}$$

Because α is a complex number, graphically we can depict it as a point in the complex plane. Equations (12.64) and (12.69) suggest that we should think of X and P as representing the real and imaginary axes of this plane, as shown in fig. 12.6. The mean values of X and P are determined by α, but there is also some uncertainty in these quantities. This uncertainty is represented in fig. 12.6 as the gray circle. For different values of α, the center of the circle moves to a different location in the complex plane. However, the uncertainty is constant, so the size of the circle is constant.

The position and momentum of a harmonic oscillator correspond to scaled versions of X and P, so their uncertainty is also constant for coherent states. In the problems you'll show that the coherent states represent minimum uncertainty states of position and momentum for the harmonic oscillator.

12.5.2 Time Dependence

Assume that the system starts out in the state $\left|\psi(0)\right\rangle = \left|\alpha_0\right\rangle$. Since the harmonic oscillator Hamiltonian is time independent, the state at a future time is given by

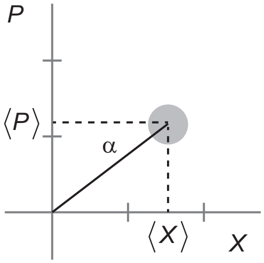


Fig 12.6 A schematic representation of the coherent state amplitude α in the complex plane. Here $\left\langle X \right\rangle = \mathrm{Re}(\alpha)$ and $\left\langle P \right\rangle = \mathrm{Im}(\alpha)$. The gray circle represents uncertainties in X and P.

$$
\begin{aligned}
\left|\psi(t)\right\rangle &= e^{-i\hat{H}t/\hbar}\left|\alpha_0\right\rangle \\
&= e^{-\left|\alpha_0\right|^2/2}\sum_{n=0}^{\infty}\frac{\alpha_0{}^n}{\sqrt{n!}}e^{-i\hat{H}t/\hbar}\left|n\right\rangle \\
&= e^{-\left|\alpha_0\right|^2/2}\sum_{n=0}^{\infty}\frac{\alpha_0{}^n}{\sqrt{n!}}e^{-i\omega(n+1/2)t}\left|n\right\rangle \\
&= e^{-i\omega t/2}e^{-\left|\alpha_0\right|^2/2}\sum_{n=0}^{\infty}\frac{\left(\alpha_0 e^{-i\omega t}\right)^n}{\sqrt{n!}}\left|n\right\rangle \\
&= e^{-i\omega t/2}\left|\alpha_0 e^{-i\omega t}\right\rangle \\
&= e^{-i\omega t/2}\left|\alpha(t)\right\rangle .
\end{aligned}
\tag{12.71}
$$

The exponential in front is an overall phase factor that is not physically significant, and which can be safely ignored. What *is* physically significant is the fact that the state in eq. (12.71) remains a coherent state at all times. The eigenvalue of $\hat{a}$ associated with this state is $\alpha(t) = \alpha_0 e^{-i\omega t}$, which has magnitude $\left|\alpha(t)\right| = \left|\alpha_0\right|$. So, if a harmonic oscillator starts out in the state $\left|\alpha_0\right\rangle$, at future times it will be in the state $\left|\alpha(t)\right\rangle = \left|\alpha_0 e^{-i\omega t}\right\rangle$. The magnitude of the coherent state does not change, only its phase changes. In terms of the representation of the coherent state pictured in fig. 12.6, $\alpha(t)$ always remains the same distance from the origin, but rotates counterclockwise at angular frequency ω.

If the state at $t = 0$ is $\left|\alpha_0\right\rangle$, the time dependence of the expectation value of the position of the oscillator can be obtained from eqs.(12.60) and (12.64) as

$$
\begin{aligned}
\left\langle x\right\rangle(t) &= \frac{\sqrt{2}}{\beta}\left\langle X\right\rangle(t) \\
&= \sqrt{\frac{2\hbar}{m\omega}}\,\mathrm{Re}\left[\alpha(t)\right] \\
&= \sqrt{\frac{2\hbar}{m\omega}}\,\mathrm{Re}\left(\left|\alpha_0\right|e^{i\phi_0}e^{-i\omega t}\right) \\
&= \sqrt{\frac{2\hbar}{m\omega}}\left|\alpha_0\right|\cos\left(-\omega t + \phi_0\right) \\
&= A\cos\left(\omega t - \phi_0\right).
\end{aligned}
\tag{12.72}
$$

Here $A = \left|\alpha_0\right|\sqrt{2\hbar/m\omega}$ represents the amplitude of the oscillation. For the momentum of the oscillator, we find from eqs. (12.61) and (12.69) that

$$
\begin{aligned}
\left\langle p\right\rangle(t) &= \hbar\beta\sqrt{2}\left\langle P\right\rangle(t) \\
&= \sqrt{2m\omega\hbar}\,\mathrm{Im}\left[\alpha(t)\right] \\
&= -\sqrt{2m\omega\hbar}\left|\alpha_0\right|\sin\left(\omega t - \phi_0\right) \\
&= -m\omega A\sin\left(\omega t - \phi_0\right).
\end{aligned}
\tag{12.73}
$$

These expectation values are consistent with those that we would find for a classical harmonic oscillator. As stated above, the magnitude of α_0 determines the amplitude of the oscillation, and the phase of α_0 determines the phase of the oscillation.

12.5.3 Wave Functions

It is possible to write down the wave functions of the coherent states. In the problems you'll show that if the initial state is $|\alpha_0\rangle$, these wave functions are given by

$$\begin{aligned}\Psi_\alpha(x,t) &= \langle x|\alpha(t)\rangle \\ &= \left(\frac{\beta^2}{\pi}\right)^{1/4} e^{-\beta^2(x-\langle x\rangle(t))^2/2} e^{i\langle p\rangle(t)x/\hbar}.\end{aligned} \tag{12.74}$$

The time-dependent position and momentum expectation values are those of eqs. (12.72) and (12.73). The wave function $\Psi_\alpha(x,t)$ corresponds to a Gaussian wave packet with a shape that does not change in time, but a mean position that oscillates sinusoidally according to eq. (12.72). The probability density $|\Psi_\alpha(x,t)|^2$ is also Gaussian, and its mean clearly oscillates in the same manner.

As stated in sec.12.4, the electric field can be described by a harmonic oscillator, and as such can exist in a coherent state. Indeed, it can be shown that the output of an ideal laser is a coherent state field, typically with a very large amplitude. If care is taken, and the field is attenuated so that the amplitude is small, the output of a real laser well approximates a coherent state.

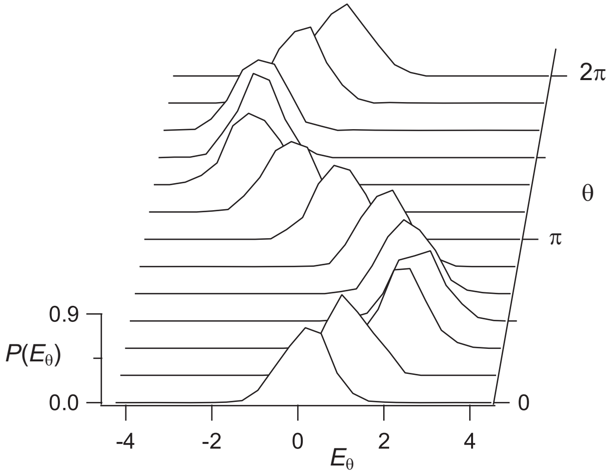


Fig 12.7 Probability distributions (normalized histograms) corresponding to measurements of the amplitude of the electric field E_θ, for a field in a state that is approximately a coherent state. The means of the distributions oscillate as the relative phase θ between the measured field and a reference field is varied, but the width of the distributions remains approximately constant. (These data were acquired by A.M. Dawes.)

Figure 12.7 shows measured probability distributions (normalized histograms) of the electric field amplitude E_θ for an ensemble of measurements performed on a field prepared in nearly a coherent state. Here E_θ is equivalent to the "position" of the oscillator. The measured field is seen to have a roughly Gaussian distribution of amplitudes. The width of the field amplitude distribution is approximately constant, but the mean oscillates sinusoidally.[4] The oscillations in this figure are not truly time resolved, but are displayed as a function of the phase difference θ between the measured field and a reference field.[5] The measured field amplitude depends on this phase difference, hence the subscript on the field amplitude E_θ.

As stated above, the coherent states are the closest quantum mechanical analog to classical states of motion for a harmonic oscillator. Other quantum systems, besides the electromagnetic field, have been placed in coherent states. For example, the trapped atom of ref. [12.1] and the micromechanical oscillator of ref. [12.2] were also placed in states that were well described by coherent states.

12.6 References

[12.1] D. M. Meekhof et al., "Generation of nonclassical motional states of a trapped atom," Phys Rev. Lett. **76**, 1796 (1996).

[12.2] A.D. O'Connell et al., "Quantum ground state and single-phonon control of a mechanical resonator," Nature **464**, 697 (2010).

12.7 PROBLEMS

12.1* Show that $\left[\hat{a},\hat{a}^\dagger\right]=1$.

12.2 Verify that $\left[\hat{n},\hat{a}^\dagger\right]=\hat{a}^\dagger$.

12.3* Show that $\hat{a}^\dagger\left|n\right\rangle=\sqrt{n+1}\left|n+1\right\rangle$.

12.4* Verify that the number states satisfy the Heisenberg indeterminacy relation.

12.5 Determine the matrix representation of the operator $\hat{a}$ in the number-state basis.

12.6 Determine the matrix representation of the operator $\hat{a}^\dagger$ in the number-state basis.

12.7 For a harmonic oscillator in the state $\left|\psi\right\rangle=\frac{1}{\sqrt{2}}\left(\left|1\right\rangle+\left|2\right\rangle\right)$, verify that the Heisenberg indeterminacy relation is satisfied.

12.8 For a harmonic oscillator in the state $\left|\psi\right\rangle=\frac{1}{\sqrt{2}}\left(\left|2\right\rangle+\left|4\right\rangle\right)$, verify that the Heisenberg indeterminacy relation is satisfied.

12.9 Verify that a harmonic oscillator in a Fock state satisfies Ehrenfest's theorem.

---

[4] The normalization of the field amplitude in fig. 12.7 is such that for a coherent state field, the displayed probability distributions would be expected to have a standard deviation of $\Delta E_\theta=1/2$.

[5] This phase difference is proportional to the time delay between the measured field and the reference field. In this sense, the oscillations in fig. 12.7 can be considered to represent oscillations in time. For more details about how the data in fig. 12.7 were measured, see sec. 16.4 and problem 16.26.

12.10* Show that $\psi_0(x)$ in eq. (12.39) is a solution to eq. (12.38). Verify that $\psi_0(x)$ is normalized.

12.11 Given the Hermite polynomials $H_0(x)=1$ and $H_1(x)=2x$, generate the Hermite polynomials for $n = 2, 3$, and 4, using the recursion formula.

12.12 The Rodrigues formula for the Hermite polynomials is

$$H_n(x)=(-1)^n e^{x^2}\left(\frac{d}{dx}\right)^n e^{-x^2}. \tag{12.75}$$

Use this formula to generate the first 3 polynomials.

12.13 Prove that the wave functions corresponding to the number states of the harmonic oscillator satisfy the following recursion formula:

$$\sqrt{n+1}\,\psi_{n+1}(x)=\beta x\sqrt{2}\,\psi_n(x)-\sqrt{n}\,\psi_{n-1}(x). \tag{12.76}$$

12.14* For a harmonic oscillator in the state $|\psi\rangle=\frac{1}{\sqrt{2}}(|1\rangle+|2\rangle)$, calculate $\langle x\rangle(t)$, $\langle p\rangle(t)$, and $\langle H\rangle(t)$.

12.15* For a harmonic oscillator in the state $|\psi\rangle=\frac{1}{\sqrt{2}}(|n\rangle+|n'\rangle)$, calculate $\langle x\rangle(t)$. Under what circumstances does this expectation value oscillate?

12.16 Verify that a harmonic oscillator in a coherent state satisfies Ehrenfest's theorem.

12.17* Show that the coherent states are minimum uncertainty states of position and momentum for the harmonic oscillator.

12.18 The displacement operator is defined as

$$\hat{D}(\alpha)=e^{\alpha\hat{a}^\dagger-\alpha^*\hat{a}}. \tag{12.77}$$

Calculate the result of applying the displacement operator to the ground state of the harmonic oscillator. To do this, you will probably find the following relationship useful: if the operators $\hat{A}$ and $\hat{B}$ both commute with their commutator, then

$$e^{\hat{A}}e^{\hat{B}}=e^{\hat{A}+\hat{B}}e^{\frac{1}{2}[\hat{A},\hat{B}]}. \tag{12.78}$$

This relationship is referred to as Glauber's formula, or the Baker-Hausdorf-Campbell theorem.

12.19* Project the eigenvalue equation for the coherent states $\hat{a}|\alpha(t)\rangle=\alpha(t)|\alpha(t)\rangle$ into the position basis. Verify that the wave functions for the coherent states $\Psi_\alpha(x,t)$, given in eq. (12.74), are solutions to the resulting equation.

12.20 Verify that the expectation values of position and momentum are indeed $\langle x\rangle(t)$ and $\langle p\rangle(t)$, for a harmonic oscillator whose wave function is $\Psi_\alpha(x,t)$ [eq. (12.74)].

12.21* Calculate the probability that a measurement of the photon number will yield n, $P(n)$, for a field in a coherent state.

Complement 12.A

Solving the Schrödinger Equation Directly

In secs. 12.2 and 12.3 we were able to find the eigenvalues and eigenstates of the harmonic oscillator, and their corresponding wave functions, without directly solving the Schrödinger equation. Here we will show that it is possible to obtain the same solutions with a direct solution.

12.A.1 Solving the Differential Equation

The time-independent Schrödinger equation for the harmonic oscillator is

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x)+\frac{1}{2}m\omega^2x^2\psi(x)=E\psi(x). \tag{12.A.1}$$

Define the dimensionless length χ as

$$\chi\equiv\sqrt{\frac{m\omega}{\hbar}}\,x=\beta x. \tag{12.A.2}$$

Substituting this into eq. (12.A.1), the Schrödinger equation becomes

$$-\frac{\hbar\omega}{2}\frac{d^2}{d\chi^2}\psi(\chi)+\frac{\hbar\omega}{2}\chi^2\psi(\chi)=E\psi(\chi). \tag{12.A.3}$$

We can further simplify things by defining the dimensionless energy ε as

$$\varepsilon\equiv\frac{2E}{\hbar\omega}, \tag{12.A.4}$$

which allows us to rewrite eq. (12.A.3) as

$$\frac{d^2}{d\chi^2}\psi(\chi)=\left(\chi^2-\varepsilon\right)\psi(\chi). \tag{12.A.5}$$

In the limit that $\chi \to \infty$, eq. (12.A.5) becomes

$$\frac{d^2}{d\chi^2}\psi(\chi) \cong \chi^2\psi(\chi). \qquad (12.A.6)$$

The solutions to this equation are

$$\psi(\chi) \cong Ae^{-\chi^2/2} + Be^{\chi^2/2}, \qquad (12.A.7)$$

but we know that we must have $B = 0$ for $\psi(\chi)$ to be normalizable. Thus, in the limit that $\chi \to \infty$ we know that the wave function satisfies

$$\psi(\chi) \xrightarrow[\chi\to\infty]{} Ae^{-\chi^2/2}. \qquad (12.A.8)$$

This suggests that we make the substitution

$$\psi(\chi) = h(\chi)e^{-\chi^2/2} \qquad (12.A.9)$$

in eq. (12.A.5). In the problems, you'll show that this substitution yields

$$\frac{d^2}{d\chi^2}h(\chi) - 2\chi\frac{d}{d\chi}h(\chi) + (\varepsilon - 1)h(\chi) = 0. \qquad (12.A.10)$$

We'll write the solutions to eq. (12.A.10) as a power series:

$$h(\chi) = \sum_{j=0}^{\infty} a_j\chi^j. \qquad (12.A.11)$$

Substituting this series into eq. (12.A.10) yields

$$\sum_{j=0}^{\infty} j(j-1)a_j\chi^{j-2} - 2\chi\sum_{j=0}^{\infty} ja_j\chi^{j-1} + (\varepsilon-1)\sum_{j=0}^{\infty} a_j\chi^j = 0. \qquad (12.A.12)$$

The first two terms in the first sum are 0, and we can pull the χ into the second sum, so we can rewrite this as

$$\sum_{j=2}^{\infty} j(j-1)a_j\chi^{j-2} - 2\sum_{j=0}^{\infty} ja_j\chi^{j} + (\varepsilon-1)\sum_{j=0}^{\infty} a_j\chi^j = 0. \qquad (12.A.13)$$

Let $j' = j - 2$ in the first sum, and this equation becomes

$$\sum_{j'=0}^{\infty} (j'+2)(j'+1)a_{j'+2}\chi^{j'} + \sum_{j=0}^{\infty} \left[-2j + (\varepsilon-1)\right]a_j\chi^j = 0. \qquad (12.A.14)$$

Now let $j' = j$, and we obtain

$$\sum_{j=0}^{\infty} \left\{(j+2)(j+1)a_{j+2} - \left[2j - (\varepsilon-1)\right]a_j\right\}\chi^j = 0. \qquad (12.A.15)$$

Because the power series must be unique, each term in this expansion must vanish for the series to sum to 0. We are then left with the recursion relation

$$a_{j+2} = \frac{2j+1-\varepsilon}{(j+2)(j+1)} a_j. \tag{12.A.16}$$

As long as this recursion relationship is satisfied, the series for $h(\chi)$ in eq. (12.A.11) is a solution to the differential equation in eq. (12.A.10). Note that a_j determines a_{j+2}, which means that the even and the odd terms are separate. All of the values of a_j for even j are derived from a_0, while all of the a_j's for odd j are derived from a_1.

Remember, however, that we are ultimately interested in finding $\psi(\chi)$, not just $h(\chi)$. $h(\chi)$ must be sufficiently well-behaved for $\psi(\chi)$ to be normalizable, so we need to examine the behavior of the series solution for $h(\chi)$ given by eqs. (12.A.11) and (12.A.16). In the limit that $j \to \infty$, the recursion relation of eq. (12.A.16) tells us that

$$\frac{a_{j+2}}{a_j} \xrightarrow[j\to\infty]{} \frac{2}{j}. \tag{12.A.17}$$

To understand what this implies, look at the power series representation of e^{χ^2}, which is

$$e^{\chi^2} = \sum_{n=0}^{\infty} \frac{1}{n!}\chi^{2n} = \sum_{\substack{j=0 \\ j\ even}}^{\infty} b_j \chi^j, \tag{12.A.18}$$

where we've set $2n = j$. For this series,

$$\frac{b_{j+2}}{b_j} = \frac{\left(\frac{j}{2}\right)!}{\left(\frac{j}{2}+1\right)!}, \tag{12.A.19}$$

which in the limit $j \to \infty$ becomes

$$\frac{b_{j+2}}{b_j} \xrightarrow[j\to\infty]{} \frac{2}{j}. \tag{12.A.20}$$

In the limit of large χ, the series solution for $h(\chi)$ given in eq. (12.A.11) is determined by the behavior of the coefficients a_j in the limit of large j. From the above discussion, we see that in the limit of large j, the coefficients a_j behave the same as the coefficients in the power series expansion of e^{χ^2}. This means that for large χ, $h(\chi) \simeq e^{\chi^2}$, and $\psi(\chi) \cong e^{\chi^2/2}$ [from eq. (12.A.9)]. Thus, the series solution for $h(\chi)$, given by eqs. (12.A.11) and (12.A.16), grows too rapidly for $\psi(\chi)$ to be normalizable.

We can fix this problem by never allowing j to get large. In other words, suppose the series in eqs. (12.A.11) is truncated at $j = n$, so that

$$h_n(\chi) = \sum_{j=0}^{n} a_j \chi^j. \tag{12.A.21}$$

The function $h_n(\chi)$ is an n^{th}-order polynomial, so the wave function

$$\psi_n(\chi) = h_n(\chi) e^{-\chi^2/2} \tag{12.A.22}$$

is guaranteed to be normalizable. The coefficients a_j must still satisfy the recursion relation of eq. (12.A.16) in order for $h_n(\chi)$ to be a solution to eq. (12.A.10).

We can force the series to truncate at $j = n$ by having $a_j = 0$ for $j > n$. There are two requirements to ensure that this is the case. First, examining eq. (12.A.16), we see that having

$$\varepsilon_n = 2n + 1 \tag{12.A.23}$$

ensures that $a_{j+2} = 0$ when $j = n$; this will also ensure that all higher-order terms are 0 as well. Remember, however, that the even and the odd terms are separate. As we use the recursion relation, and increase j by 2 in going to higher-order terms, we need to ensure that we arrive at $j = n$ in order for the series to truncate. Thus, the second requirement is that if n is odd the allowed values for j are odd, whereas if n is even the allowed values for j are even. For example, assume that n is even; in this case we must have $a_0 \neq 0$ and $a_1 = 0$, which ensures that the polynomial contains only even-order powers of χ.

Substituting eq. (12.A.23) into eq. (12.A.4), we find that the allowed energies of the harmonic oscillator are

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) \quad n = 0, 1, 2, \ldots, \tag{12.A.24}$$

which agrees with eq. (12.24). Thus, the boundary condition that the wave functions must be normalizable leads to energy quantization of the harmonic oscillator.

With proper choice of a_0 and a_1, the polynomials $h_n(\chi)$ are the Hermite polynomials described in sec. 12.3, so after normalization the wave functions given by eq. (12.A.22) are the same as those we obtained previously in eq. (12.42).

12.A.2 PROBLEMS

12.A.1 Verify that substituting $\psi(\chi) = h(\chi) e^{-\chi^2/2}$ into eq. (12.A.5) yields eq. (12.A.10).

12.A.2 Show that for $n = 0$ through $n = 3$, the functions $h_n(x)$ are the Hermite polynomials.

CHAPTER 13

Wave Mechanics in Three Dimensions

So far we have been discussing the behavior of particles in one spatial dimension. However, the world around us is three dimensional. In order to treat some important problems, we need to take what we've learned about wave mechanics in one dimension and generalize it to three dimensions.

13.1 THE SCHRÖDINGER EQUATION IN THREE DIMENSIONS

The Hamiltonian is the sum of the kinetic and potential energies, and in three dimensions it can be written as

$$\hat{H} = \frac{\hat{p}^2}{2m} + V\left(\hat{\boldsymbol{r}}\right) = \frac{\hat{p}_x^2}{2m} + \frac{\hat{p}_y^2}{2m} + \frac{\hat{p}_z^2}{2m} + V\left(\hat{\boldsymbol{r}}\right), \tag{13.1}$$

where the three-dimensional (3-D) position and momentum operators are

$$\hat{\boldsymbol{r}} = \hat{x}\boldsymbol{u}_x + \hat{y}\boldsymbol{u}_y + \hat{z}\boldsymbol{u}_z, \tag{13.2}$$

$$\hat{\boldsymbol{p}} = \hat{p}_x\boldsymbol{u}_x + \hat{p}_y\boldsymbol{u}_y + \hat{p}_z\boldsymbol{u}_z. \tag{13.3}$$

Following the procedure in sec. 11.1, we can project the Schrödinger equation [eq. (11.1)] into the position basis using the 3-D position eigenstates $\left|\boldsymbol{r}\right\rangle = \left|x, y, z\right\rangle$:

$$
\begin{aligned}
&\langle \boldsymbol{r}|\hat{H}|\psi(t)\rangle = \langle \boldsymbol{r}|i\hbar\frac{d}{dt}|\psi(t)\rangle, \\
&\langle \boldsymbol{r}|\left[\frac{\hat{p}_x^2}{2m}+\frac{\hat{p}_y^2}{2m}+\frac{\hat{p}_z^2}{2m}+V(\hat{\boldsymbol{r}})\right]|\psi(t)\rangle = \langle \boldsymbol{r}|i\hbar\frac{d}{dt}|\psi(t)\rangle, \\
&\left[\frac{(-i\hbar)^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V(\boldsymbol{r})\right]\Psi(\boldsymbol{r},t) = i\hbar\frac{d}{dt}\Psi(\boldsymbol{r},t),
\end{aligned}
\tag{13.4}
$$

where we have generalized the position-basis representation of the momentum operator [eq. (10.43)] into three dimensions. In eq. (13.4) the terms in () are the Laplacian ∇^2 in Cartesian coordinates, so we can rewrite eq. (13.4) in a coordinate independent form:

$$
\frac{-\hbar^2}{2m}\nabla^2\Psi(\boldsymbol{r},t)+V(\boldsymbol{r})\Psi(\boldsymbol{r},t) = i\hbar\frac{d}{dt}\Psi(\boldsymbol{r},t). \tag{13.5}
$$

Note that the time dependence in eq. (13.5) is the same as in the one-dimensional Schrödinger wave equation of eq. (11.8), so we can treat it in the same manner as we did in sec. 11.1. We can separate out the time dependence, and write the solution to eq. (13.5) as

$$
\begin{aligned}
\Psi(\boldsymbol{r},t) &= \sum_n c_n\psi_n(\boldsymbol{r})\phi_n(t) \\
&= \sum_n c_n e^{-iE_n t/\hbar}\psi_n(\boldsymbol{r}) \\
&= \sum_n c_n e^{-i\omega_n t}\psi_n(\boldsymbol{r}).
\end{aligned}
\tag{13.6}
$$

The wave functions $\psi_n(\boldsymbol{r})$ are solutions to the time-independent, 3-D Schrödinger wave equation:

$$
\frac{-\hbar^2}{2m}\nabla^2\psi(\boldsymbol{r})+V(\boldsymbol{r})\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}), \tag{13.7}
$$

and the normalization of the three-dimensional wave function is

$$
\iiint\limits_{all\ space} d^3r\,|\psi(\boldsymbol{r})|^2 = 1. \tag{13.8}
$$

EXAMPLE 13.1

Solve the Schrödinger equation [eq. (13.7)] for the "particle in a box." Assume a rectangular box, with sides of length L_x, L_y, and L_z; the potential energy is 0 inside the box, and infinite outside.

Inside the box, the Schrödinger equation in Cartesian coordinates is

$$\frac{-\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi(\boldsymbol{r})=E\psi(\boldsymbol{r}). \tag{13.9}$$

Since the potential is infinite outside the box, the wave function must be 0 there. Place one corner of the box at the origin. The wave function must go to zero at the boundaries of the box: $x=0$ and $x=L_x$ (and similarly for the y and z coordinates).

A standard technique for solving partial differential equations is separation of variables. The idea is to look for solutions that are products of functions of the individual variables. If we can find such a solution, and if it satisfies the boundary conditions, then we know that the solution must be unique. Here we'll look for solutions of the form

$$\psi(\boldsymbol{r})=X(x)Y(y)Z(z). \tag{13.10}$$

Substituting this into eq. (13.9), we obtain

$$\begin{aligned}&\frac{-\hbar^2}{2m}\left(Y(y)Z(z)\frac{\partial^2}{\partial x^2}X(x)+X(x)Z(z)\frac{\partial^2}{\partial y^2}Y(y)+X(x)Y(y)\frac{\partial^2}{\partial z^2}Z(z)\right)\\&\quad=EX(x)Y(y)Z(z).\end{aligned} \tag{13.11}$$

Divide this equation by $X(x)Y(y)Z(z)$, and we find

$$\frac{-\hbar^2}{2m}\left(\frac{1}{X(x)}\frac{\partial^2}{\partial x^2}X(x)+\frac{1}{Y(y)}\frac{\partial^2}{\partial y^2}Y(y)+\frac{1}{Z(z)}\frac{\partial^2}{\partial z^2}Z(z)\right)=E. \tag{13.12}$$

Notice that the first term in eq. (13.12) depends only on x, the second term depends only on y, the third depends only on z, and the three terms add to a constant. For this equation to be true for all values of x, y, and z, then each term must be constant. (If the first term changed with x, for example, the other two terms could not compensate to keep the sum constant, because they have no x dependence.) This means that we must have

$$\frac{-\hbar^2}{2m}\frac{1}{X(x)}\frac{d^2}{dx^2}X(x)=E_x, \tag{13.13}$$

and two similar equations for $Y(y)$ and $Z(z)$. The separation constants satisfy

$$E_x+E_y+E_z=E. \tag{13.14}$$

We have taken a 3-D, partial differential equation, and converted it into three, separate, ordinary differential equations—one for each variable.

Multiply eq. (13.13) by $X(x)$, and we have the one-dimensional Schrödinger equation for a constant, zero potential [eq. (11.22) with $V_0=0$]. The boundary conditions on eq. (13.13) are the same as those for the one-dimensional infinite square well in sec. 11.5, so the solutions for the energies E_x and the wave functions $X(x)$ are the same as we found there:

$$E_{n_x} = \frac{n_x^2 \pi^2 \hbar^2}{2mL^2},\ n_x = 1, 2, 3, \ldots, \tag{13.15}$$

$$X_{n_x}(x) = \begin{cases} \sqrt{\dfrac{2}{L_x}} \sin\left(\dfrac{n_x \pi}{L_x} x\right) & 0 < x < L_x \\ 0 & \textit{elsewhere} \end{cases}. \tag{13.16}$$

Since the differential equations and boundary conditions for $Y(y)$ and $Z(z)$ are the same as those for $X(x)$, the energies and wave functions will also be the same. The final solution is thus

$$E_{n_x n_y n_z} = \frac{\pi^2 \hbar^2}{2m} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2} \right) \quad n_x, n_y, n_z = 1, 2, 3, \ldots, \tag{13.17}$$

$$\psi_{n_x n_y n_z}(\boldsymbol{r}) = \sqrt{\frac{8}{L_x L_y L_z}} \sin\left(\frac{n_x \pi}{L_x} x\right) \sin\left(\frac{n_y \pi}{L_y} y\right) \sin\left(\frac{n_z \pi}{L_z} z\right) \quad \begin{matrix} 0 < x < L_x \\ 0 < y < L_y. \\ 0 < z < L_z \end{matrix} \tag{13.18}$$

The wave function is 0 elsewhere.

13.2 CENTRAL POTENTIALS

A number of interesting problems involve a potential energy $V(\boldsymbol{r})$ that depends only on the distance r from a particular point. If that point is placed at the origin of the coordinate system, the potential energy depends only on the magnitude of $\boldsymbol{r}$, so the potential becomes $V(r)$. A potential energy of this form is called a central potential, and the classic example is the Coulomb potential energy of two point charges. Before looking at that specific problem, we'll look at some general properties of solutions to the Schrödinger equation with a central potential.

Central potentials are spherically symmetric, so we'll use spherical coordinates, as shown in fig. 13.1. To transform from Cartesian coordinates to spherical coordinates, the following equations are used:

$$r = \sqrt{x^2 + y^2 + z^2},\ \theta = \cos^{-1}\left(\frac{z}{\sqrt{x^2 + y^2 + z^2}} \right),\ \phi = \tan^{-1}\left(\frac{y}{x}\right), \tag{13.19}$$

The inverse transformation is

$$x = r \sin\theta \cos\phi,\ y = r \sin\theta \sin\phi,\ z = r \cos\theta. \tag{13.20}$$

In these coordinates, the 3-D volume element is

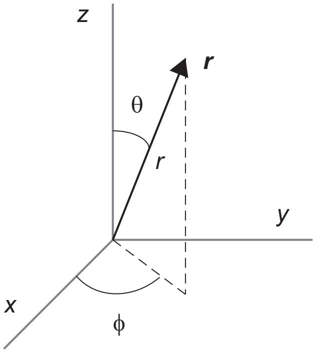


Fig 13.1 Spherical coordinates.

$$d^3r = r^2 \sin\theta dr d\theta d\phi, \tag{13.21}$$

and the normalization integral for the wave function is thus

$$\int_0^\infty dr \int_0^\pi d\theta \int_0^{2\pi} d\phi r^2 \sin\theta \left|\psi(\boldsymbol{r})\right|^2 = 1. \tag{13.22}$$

The Laplacian in spherical coordinates is

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial}{\partial\theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial\phi^2}, \tag{13.23}$$

and the time-independent Schrödinger equation is thus

$$\begin{aligned} &\frac{-\hbar^2}{2m}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\psi(\boldsymbol{r})\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial}{\partial\theta}\psi(\boldsymbol{r})\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial\phi^2}\psi(\boldsymbol{r})\right] \\ &\quad + V(r)\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}). \end{aligned} \tag{13.24}$$

For more information about differential operators in spherical coordinates, see sec. 2.5 of ref. [13.1].

13.2.1 Separation of Variables

We’ll use separation of variables, following the general procedure of example 13.1, to solve eq. (13.24). Begin by separating the radial and angular dependencies, and writing the wave function as

$$\psi(\boldsymbol{r}) = R(r)Y(\theta,\phi). \tag{13.25}$$

Substituting this into eq. (13.24) yields

$$\frac{-\hbar^2}{2m}\left[\frac{Y(\theta,\phi)}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}R(r)\right)+\frac{R(r)}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}Y(\theta,\phi)\right)+\frac{R(r)}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}Y(\theta,\phi)\right]$$
$$+V(r)R(r)Y(\theta,\phi)=ER(r)Y(\theta,\phi). \tag{13.26}$$

Multiplying through by $\left(-2mr^2\right)/\left[\hbar^2R(r)Y(\theta,\phi)\right]$, and rearranging, we find

$$\begin{aligned}&\frac{1}{R(r)}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}R(r)\right)-\frac{2m}{\hbar^2}\left[V(r)-E\right]r^2\\&+\frac{1}{Y(\theta,\phi)}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}Y(\theta,\phi)\right)+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}Y(\theta,\phi)\right]=0.\end{aligned} \tag{13.27}$$

The first two terms of eq. (13.27) depend only on r, while the last term depends only on the angular variables. The only way that the radial and angular pieces can add to 0, for all values of r, θ and ϕ, is if each piece is constant. Here we'll choose this is separation constant to be $l(l+1)$, for reasons that will become apparent shortly. (There is no loss in generality in choosing the constant in this manner, because at this point l can be any complex number.) With this choice, eq. (13.27) yields the two equations

$$\frac{1}{R(r)}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}R(r)\right)-\frac{2m}{\hbar^2}\left[V(r)-E\right]r^2=l(l+1), \tag{13.28}$$

$$\frac{1}{Y(\theta,\phi)}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}Y(\theta,\phi)\right)+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}Y(\theta,\phi)\right]=-l(l+1). \tag{13.29}$$

To place eq. (13.28) in a more useful form, we multiply through by $R(r)/r^2$ and rearrange, to obtain

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}R(r)\right)-\frac{2m}{\hbar^2}\left[V(r)+\frac{\hbar^2l(l+1)}{2mr^2}-E\right]R(r)=0. \tag{13.30}$$

This equation is called the radial equation, and to solve it we must specify $V(r)$. We'll tackle that problem later, but now we'll solve the angular equation.

13.2.2 The Angular Equation

If we multiply eq. (13.29) by $\sin^2\theta Y(\theta,\phi)$, we can rewrite it as

$$\sin\theta\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}Y(\theta,\phi)\right)+l(l+1)\sin^2\theta Y(\theta,\phi)+\frac{\partial^2}{\partial\phi^2}Y(\theta,\phi)=0. \tag{13.31}$$

This equation is independent of the potential. As such, the angular dependence of the solutions will be the same for *any* problem involving a central potential. One frequently comes across eq. (13.31) in physics, not just in quantum mechanics. It arises when solving second-order differential equations in spherical coordinates. We'll solve it here by again using separation of variables.

Write the functions $Y(\theta,\phi)$ as

$$Y(\theta,\phi)=\Theta(\theta)\Phi(\phi). \tag{13.32}$$

Substitute this into eq. (13.31), and then divide through by $\Theta(\theta)\Phi(\phi)$. The result is

$$\frac{\sin\theta}{\Theta(\theta)}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\Theta(\theta)\right)+l(l+1)\sin^2\theta+\frac{1}{\Phi(\phi)}\frac{\partial^2}{\partial\phi^2}\Phi(\phi)=0. \tag{13.33}$$

The first two terms depend only on θ, while the last term depends only on ϕ. Once again, in order for the two pieces to add to 0, for all possible values of θ and ϕ, each piece must be constant. Call the separation constant m_l^2, and we end up with

$$\frac{1}{\Phi(\phi)}\frac{\partial^2}{\partial\phi^2}\Phi(\phi)=-m_l^2, \tag{13.34}$$

$$\frac{\sin\theta}{\Theta(\theta)}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\Theta(\theta)\right)+l(l+1)\sin^2\theta=m_l^2. \tag{13.35}$$

13.2.3 The Azimuthal Equation

The azimuthal equation, eq. (13.34), can be rewritten as

$$\frac{\partial^2}{\partial\phi^2}\Phi(\phi)=-m_l^2\Phi(\phi). \tag{13.36}$$

This equation has solutions

$$\Phi_{m_l}(\phi)=e^{im_l\phi}, \tag{13.37}$$

where the subscript on $\Phi_{m_l}(\phi)$ indicates that the solutions depend on m_l. At this point m_l is any complex number. However, we now need to apply the boundary conditions. In spherical coordinates, if we increase the value of ϕ by 2π, we return to the same point in space (see fig. 13.1). The boundary condition on $\Phi_{m_l}(\phi)$ is that it must be single valued when we do this. Thus, we must have

$$\Phi_{m_l}(\phi)=\Phi_{m_l}(\phi+2\pi). \tag{13.38}$$

This means

$$e^{im_l\phi} = e^{im_l(\phi+2\pi)} = e^{im_l\phi}e^{im_l 2\pi}, \tag{13.39}$$

$$e^{im_l 2\pi} = 1. \tag{13.40}$$

This constrains m_l to be an integer, and the final solution to the azimuthal equation is

$$\Phi_{m_l}(\phi) = e^{im_l\phi} \quad m_l = 0, \pm 1, \pm 2, \ldots . \tag{13.41}$$

The functions $\Phi_{m_l}(\phi)$ are orthogonal on the interval $0 \le \phi \le 2\pi$:

$$\int_0^{2\pi} d\phi\, \Phi^*_{m_l}(\phi)\Phi_{m'_l}(\phi) = 2\pi\delta_{m_l\, m'_l}, \tag{13.42}$$

13.2.4 The Polar Equation

The polar equation, eq. (13.35), can be rewritten as

$$\sin\theta\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\Theta(\theta)\right) + \left[l(l+1)\sin^2\theta - m_l^2\right]\Theta(\theta) = 0. \tag{13.43}$$

We'll just state the solutions to this equation.[1] In order for the solutions to be finite over the interval $0 \le \theta \le \pi$, l must be a nonnegative integer. With this constraint, the solutions are

$$\Theta_l^{m_l}(\theta) = P_l^{m_l}(\cos\theta), \tag{13.44}$$

where the functions $P_l^{m_l}(x)$ are the associated Legendre functions. They are associated with the Legendre polynomials, $P_l(x)$, and for nonnegative m_l the $P_l^{m_l}(x)$'s are given by

$$P_l^{m_l}(x) = \left(1-x^2\right)^{m_l/2}\left(\frac{d}{dx}\right)^{m_l} P_l(x) \quad m_l \ge 0. \tag{13.45}$$

The Rodrigues formula for the Legendre polynomials is

$$P_l(x) = \frac{1}{2^l\, l!}\left(\frac{d}{dx}\right)^l \left(x^2-1\right)^l. \tag{13.46}$$

---

[1] For a derivation of the solutions, see ref. [13.2] (especially problem 2 of that reference).

Combining the last two equations, we find that

$$P_l^{m_l}\left(x\right)=\frac{1}{2^l l!}\left(1-x^2\right)^{m_l/2}\left(\frac{d}{dx}\right)^{l+m_l}\left(x^2-1\right)^l. \tag{13.47}$$

Equation (13.47) for $P_l^{m_l}\left(x\right)$ allows for negative values of m_l, as long as $\left(l+m_l\right)\geq 0$. It can also be shown that the $P_l^{m_l}\left(x\right)$'s for negative values of m_l can be obtained from those with positive values by using

$$P_l^{-m_l}\left(x\right)=\left(-1\right)^{m_l}\frac{\left(l-m_l\right)!}{\left(l+m_l\right)!}P_l^{m_l}\left(x\right). \tag{13.48}$$

Calculating $P_l^{m_l}\left(x\right)$ using eq. (13.47) involves taking the $\left(l+m_l\right)$-*th* derivative of a $2l$-*th*-order polynomial, which yields 0 if $m_l > l$; $P_l^{m_l}\left(x\right)=0$ means that the wave function is 0 everywhere, which is not allowed. Equation (13.47) thus constrains the allowed values of m_l to be $\left|m_l\right|\leq l$. For each value of m_l, the associated Legendre functions form an orthogonal set on the interval $-1\leq x\leq 1$. The orthogonality relationship is

$$\int_{-1}^{1} dx\, P_l^{m_l}\left(x\right)P_{l'}^{m_l}\left(x\right)=\int_{0}^{\pi} d\theta\sin\theta P_l^{m_l}\left(\cos\theta\right)P_{l'}^{m_l}\left(\cos\theta\right)=\frac{2}{2l+1}\frac{\left(l+m_l\right)!}{\left(l-m_l\right)!}\delta_{ll'}. \tag{13.49}$$

To summarize, the allowed values of l are

$$l=0,1,2,\ldots, \tag{13.50}$$

and for a given value of l, the allowed values for m_l are

$$m_l=-l,-l+1,\ldots,l-1,l. \tag{13.51}$$

Does this look familiar? We'll come back to that question, and discuss the physical interpretation of these results, in sec. 13.3.

13.2.5 Spherical Harmonics

We can combine the functions $\Phi_{m_l}\left(\phi\right)$ and $\Theta_l^{m_l}(\theta)$, as in eq. (13.32), and obtain the solutions $Y_l^{m_l}\left(\theta,\phi\right)$ of the full angular differential equation [eq. (13.31)]:

$$Y_l^{m_l}\left(\theta,\phi\right)=\left(-1\right)^{m_l}\left[\frac{\left(2l+1\right)\left(l-m_l\right)!}{4\pi\left(l+m_l\right)!}\right]^{1/2}P_l^{m_l}\left(\cos\theta\right)e^{im_l\phi}. \tag{13.52}$$

Here we've also included a normalization factor. The functions $Y_l^{m_l}(\theta,\phi)$ are called the spherical harmonics.[2] Combining eqs. (13.48) and (13.52), we can see that the $Y_l^{m_l}(\theta,\phi)$'s satisfy

$$Y_l^{-m_l}(\theta,\phi) = (-1)^{m_l}\, Y_l^{m_l *}(\theta,\phi). \tag{13.53}$$

The spherical harmonics are a complete set of orthonormal functions over $0 \le \phi \le 2\pi$ and $0 \le \theta \le \pi$; thus, they satisfy the orthogonality relation

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta\, Y_l^{m_l *}(\theta,\phi)\, Y_{l'}^{m_l'}(\theta,\phi) = \delta_{ll'}\delta_{m_l m_l'}. \tag{13.54}$$

The factor of $\sin\theta$ in this equation comes from the volume element in spherical coordinates [eq. (13.21)].

Table 13.1 lists the formulas for some of the low-order spherical harmonics. Figure 13.2 plots the real and imaginary parts of these same functions. To observe plots of higher-order spherical harmonics, see ref. [13.3].

13.3 ORBITAL ANGULAR MOMENTUM

I hope the allowed values for l and m_l [eqs. (13.50) and (13.51)] remind you of the angular momentum quantum numbers we discussed in chapter 7 (without the half integer values for l, a difference that will be discussed below). It shouldn't be too surprising that angular momentum would come up. We have been finding the eigenstates of the

Table 13.1 The spherical harmonics for $l=0$ through $l=3$.

$$Y_0^0(\theta,\phi) = \frac{1}{\sqrt{4\pi}}$$

$$Y_1^0(\theta,\phi) = \sqrt{\frac{3}{4\pi}}\cos\theta$$

$$Y_1^{\pm 1}(\theta,\phi) = \mp\sqrt{\frac{3}{8\pi}}\sin\theta e^{\pm i\phi}$$

$$Y_2^0(\theta,\phi) = \sqrt{\frac{5}{16\pi}}\left(3\cos^2\theta - 1\right)$$

$$Y_2^{\pm 1}(\theta,\phi) = \mp\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{\pm i\phi}$$

$$Y_2^{\pm 2}(\theta,\phi) = \sqrt{\frac{15}{32\pi}}\sin^2\theta e^{\pm i2\phi}$$

$$Y_3^0(\theta,\phi) = \sqrt{\frac{7}{16\pi}}\left(5\cos^3\theta - 3\cos\theta\right)$$

$$Y_3^{\pm 1}(\theta,\phi) = \mp\sqrt{\frac{21}{64\pi}}\sin\theta\left(5\cos^2\theta - 1\right)e^{\pm i\phi}$$

$$Y_3^{\pm 2}(\theta,\phi) = \sqrt{\frac{105}{32\pi}}\sin^2\theta\cos\theta e^{\pm i2\phi}$$

$$Y_3^{\pm 3}(\theta,\phi) = \mp\sqrt{\frac{35}{64\pi}}\sin^3\theta e^{\pm i3\phi}$$

---

[2] There are several, slightly different, definitions for the spherical harmonics, and the definition here is the same as that in sec. 11.5 of ref. [13.1]. The factor of $(-1)^{m_l}$ ensures the proper phase relationship between the $Y_l^{m_l}(\theta,\phi)$'s (see problem 13.12).

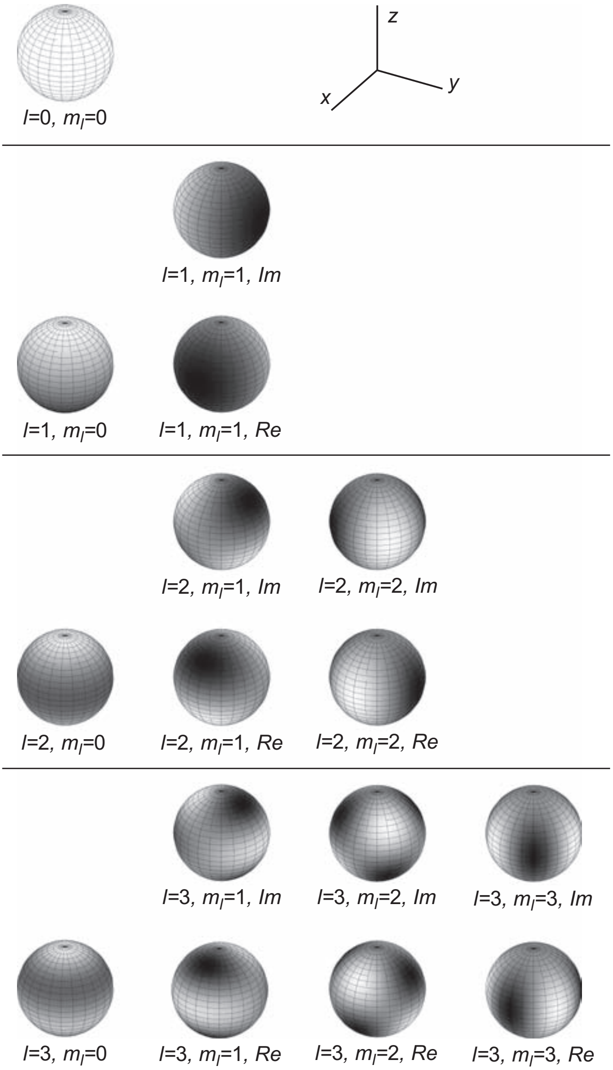


Fig 13.2 The real and imaginary parts of the spherical harmonics for $l = 0$ through $l = 3$. Here lighter shading (more white) corresponds to more positive values, while darker shading (more black) corresponds to more negative values.

Hamiltonian, which describes the energy of a quantum mechanical particle. An object which is rotating (i.e., has orbital angular momentum) has kinetic energy. Indeed, we can express the Hamiltonian of eq. (13.1) in terms of the orbital angular momentum operator $\hat{\boldsymbol{L}}$ as

$$\hat{H} = \frac{\hat{p}_r^2}{2m} + \frac{\hat{L}^2}{2I} + V(\hat{r}), \tag{13.55}$$

where I is the moment of inertia of the particle about the origin. The first term in this equation represents the linear (radial) kinetic energy, while the second represents the rotational kinetic energy. Let's show that eq. (13.55) yields the Schrödinger equation in spherical coordinates [eq. (13.24)].

In analogy with classical physics, we know that the orbital angular momentum operator is given by

$$\hat{\boldsymbol{L}} = \hat{\boldsymbol{r}} \times \hat{\boldsymbol{p}}. \tag{13.56}$$

In the problems you'll show that the components of $\hat{\boldsymbol{L}}$ can be represented in spherical coordinates as

$$\hat{L}_x \doteq i\hbar\left(\sin\phi\frac{\partial}{\partial\theta} + \cot\theta\cos\phi\frac{\partial}{\partial\phi}\right), \tag{13.57}$$

$$\hat{L}_y \doteq i\hbar\left(-\cos\phi\frac{\partial}{\partial\theta} + \cot\theta\sin\phi\frac{\partial}{\partial\phi}\right), \tag{13.58}$$

and

$$\hat{L}_z \doteq -i\hbar\frac{\partial}{\partial\phi}. \tag{13.59}$$

From these equations for the components, it can be shown that

$$\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 \doteq -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]. \tag{13.60}$$

Next, we need to find the radial momentum operator $\hat{p}_r$. Since $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{p}}$ do not commute, it's not obvious whether we should use $\hat{p}_r = (\hat{\boldsymbol{r}}/r)\cdot\hat{\boldsymbol{p}}$, or $\hat{p}_r = \hat{\boldsymbol{p}}\cdot(\hat{\boldsymbol{r}}/r)$ (here $\hat{\boldsymbol{r}}/r$ is the radial "unit" vector operator). The solution to this conundrum is to use a symmetric combination of both possibilities:

$$\hat{p}_r \equiv \frac{1}{2}\left[\left(\frac{\hat{\boldsymbol{r}}}{r}\right)\cdot\hat{\boldsymbol{p}} + \hat{\boldsymbol{p}}\cdot\left(\frac{\hat{\boldsymbol{r}}}{r}\right)\right]. \tag{13.61}$$

In the problems you'll prove that the position representation of this operator can be written as

$$\hat{p}_r \doteq -i\hbar\left(\frac{\partial}{\partial r} + \frac{1}{r}\right). \tag{13.62}$$

Squaring this, we obtain

$$\begin{aligned}\hat{p}_r^2 &\doteq \left[-i\hbar\left(\frac{\partial}{\partial r}+\frac{1}{r}\right)\right]^2 \\ &= -\hbar^2\left(\frac{\partial^2}{\partial r^2}+\frac{2}{r}\frac{\partial}{\partial r}\right) \\ &= -\hbar^2\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right).\end{aligned} \tag{13.63}$$

To verify that this equation is correct, it is best to explicitly place a trial function to the right of the operator.[3]

The moment of inertial of a point mass about the origin is $I = mr^2$, so the eigenvalue equation for the Hamiltonian of eq. (13.55) is

$$\begin{gathered}\langle \boldsymbol{r}|\hat{H}|\psi\rangle = \langle \boldsymbol{r}|E|\psi\rangle, \\ \langle \boldsymbol{r}|\left[\frac{\hat{p}_r^2}{2m}+\frac{\hat{L}^2}{2mr^2}+V(\hat{r})\right]|\psi\rangle = E\psi(\boldsymbol{r}).\end{gathered} \tag{13.64}$$

Substituting eqs. (13.60) and (13.63) into this, we obtain eq. (13.24) (see the problems).

13.3.1 Eigenstates of Orbital Angular Momentum

The operator $\hat{\boldsymbol{L}}$ is an angular momentum operator, so it must satisfy all of the properties of angular momentum operators that we described in chapter 7. For example, its components must satisfy the commutation relations

$$\left[\hat{L}_x,\hat{L}_y\right] = i\hbar\hat{L}_z,\ \left[\hat{L}_y,\hat{L}_z\right] = i\hbar\hat{L}_x,\ \left[\hat{L}_z,\hat{L}_x\right] = i\hbar\hat{L}_y, \tag{13.65}$$

and each of the components must commute with $\hat{L}^2$:

$$\left[\hat{L}^2,\hat{L}_i\right] = 0, \tag{13.66}$$

where $i = x, y, z$. We also know that the eigenvalues and eigenstates of orbital angular momentum are given by

$$\hat{L}^2|l,m_l\rangle = l(l+1)\hbar^2|l,m_l\rangle, \tag{13.67}$$

---

[3] Note that $\hat{p}_r$ is not well defined at $r = 0$. As defined in eq. (13.61), the operator $\hat{p}_r$ is Hermitian, but does not correspond to an observable [13.4].

$$\hat{L}_z \left| l, m_l \right\rangle = m_l \hbar \left| l, m_l \right\rangle. \tag{13.68}$$

As discussed in chapter 7, because the component operators do not commute with each other, it is not possible to know with certainty the value of more than one individual component of orbital angular momentum at one time.

To show that eqs. (13.67) and (13.68) are consistent with what we've learned so far in this chapter, we can use eq. (13.60) to express eq. (13.67) in the position basis:

$$\begin{gathered}
\left\langle \boldsymbol{r} \right| \hat{L}^2 \left| l, m_l \right\rangle = \left\langle \boldsymbol{r} \right| l(l+1)\hbar^2 \left| l, m_l \right\rangle, \\
-\hbar^2 \left[\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right] Y_l^{m_l}(\theta,\phi) = \hbar^2 l(l+1) Y_l^{m_l}(\theta,\phi), \\
\sin\theta \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} Y_l^{m_l}(\theta,\phi) \right) + \frac{\partial^2}{\partial\phi^2} Y_l^{m_l}(\theta,\phi) = -l(l+1)\sin^2\theta\, Y_l^{m_l}(\theta,\phi).
\end{gathered} \tag{13.69}$$

Here we have made the assignment $\left\langle \boldsymbol{r} \middle| l, m_l \right\rangle = Y_l^{m_l}(\theta,\phi)$. We can do this because eq. (13.69) is equivalent to eq. (13.31).

As a final reassurance that everything is consistent, eq. (13.68) can be expressed in the position basis using eq. (13.59), and the result is

$$\begin{gathered}
\left\langle \boldsymbol{r} \right| \hat{L}_z \left| l, m_l \right\rangle = \left\langle \boldsymbol{r} \right| m_l \hbar \left| l, m_l \right\rangle \\
-i\hbar \frac{\partial}{\partial\phi} Y_l^{m_l}(\theta,\phi) = m_l \hbar Y_l^{m_l}(\theta,\phi).
\end{gathered} \tag{13.70}$$

Using the definitions of the spherical harmonics in eq. (13.52), it is straightforward to verify that this equation is satisfied.

The above discussion confirms that the states $\left| l, m_l \right\rangle$ are states of definite angular momentum, and satisfy all of the properties of angular momentum states that we described in chapter 7. The one difference is that the orbital angular momentum quantum number l may only take on integer values [eq. (13.50)], while the spin and total angular momentum quantum numbers s and j may also take on half integer values [eqs. (7.13) and (7.25)]. As described in sec. 13.2, this is because the wave functions corresponding to orbital angular momentum $Y_l^{m_l}(\theta,\phi)$ are only well behaved for integer values of l.

13.4 THE HYDROGEN ATOM

Nearly everything we have described thus far in this chapter applies generally to any central potential $V(r)$. One of the most important problems in physics is that of the hydrogen atom; since it involves a central potential, let's examine it now.

13.4.1 The Radial Equation

A hydrogen atom consists of a proton and an electron held together by the electromagnetic force. If we place the proton at the origin of our coordinate system, the potential energy of the system is given by the Coulomb potential:

$$V(r) = -\frac{e^2}{4\pi\varepsilon_0 r}, \tag{13.71}$$

where $e = 1.60 \times 10^{-19}$ C is the magnitude of the electron charge, and $\varepsilon_0 = 8.85 \times 10^{-12}$ C^2/Nm^2 is the permittivity of free space. Substituting this potential into the radial part of the Schrödinger equation [eq. (13.30)] yields

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r} R(r)\right) - \frac{2m}{\hbar^2}\left[-\frac{e^2}{4\pi\varepsilon_0 r} + \frac{\hbar^2 l(l+1)}{2mr^2} - E\right] R(r) = 0. \tag{13.72}$$

The mass here is the mass of the electron, $m = m_e = 9.11 \times 10^{-31}$ kg, since it is the electron that is moving, while the proton is assumed to be fixed at the origin.[4]

To simplify eq. (13.72), we'll begin by defining a dimensionless radial coordinate ρ, which is given by

$$\rho = r\frac{\sqrt{-8mE}}{\hbar}. \tag{13.73}$$

At first glance it may appear that ρ is imaginary, but on the contrary, the negative sign in the square root is chosen to ensure that ρ is real. This is because the potential energy in eq. (13.71) is everywhere negative, which means that the bound-state solutions must have negative total energy. If we substitute eq. (13.73) into eq. (13.72), and expand the derivative, we obtain

$$\frac{\partial^2}{\partial \rho^2} R(\rho) + \frac{2}{\rho}\frac{\partial}{\partial \rho} R(\rho) + \left[\frac{\lambda}{\rho} - \frac{l(l+1)}{\rho^2} - \frac{1}{4}\right] R(\rho) = 0. \tag{13.74}$$

Here we've also defined the dimensionless parameter λ, which is given by

$$\lambda = \frac{e^2}{\hbar 4\pi\varepsilon_0}\sqrt{\frac{m}{-2E}}. \tag{13.75}$$

To simplify eq. (13.74), look at the asymptotic behavior of the differential equation in the limits of large and small ρ. In the limit $\rho \to \infty$, eq. (13.74) is given approximately by

---

[4] A slightly more accurate treatment places the origin of the coordinate system at the center of mass, in which case the mass of the electron is replaced by the reduced mass $\mu = m_e m_p / (m_e + m_p)$. For more details about center of mass coordinates, see ref. [13.5].

$$\frac{\partial^2}{\partial \rho^2} R(\rho) = \frac{1}{4} R(\rho), \tag{13.76}$$

which has a solution of the form

$$R(\rho) = Ae^{-\rho/2} + Be^{\rho/2}. \tag{13.77}$$

The growing exponential is not normalizable, so $B = 0$; this means

$$R(\rho) \xrightarrow[\rho \to \infty]{} e^{-\rho/2}. \tag{13.78}$$

In the limit $\rho \to 0$, the approximate expression for eq. (13.74) is

$$\frac{\partial^2}{\partial \rho^2} R(\rho) + \frac{2}{\rho} \frac{\partial}{\partial \rho} R(\rho) - \frac{l(l+1)}{\rho^2} R(\rho) = 0. \tag{13.79}$$

Look for solutions to this equation in terms of powers of ρ: $R(\rho) = \rho^j$. Substituting in a solution of this form, eq. (13.79) becomes

$$\begin{gathered} j(j-1)\rho^{j-2} + \frac{2}{\rho} j\rho^{j-1} - \frac{l(l+1)}{\rho^2} \rho^j = 0, \\ j(j-1) + 2j - l(l+1) = 0, \\ j(j+1) - l(l+1) = 0. \end{gathered} \tag{13.80}$$

Clearly $j = l$ is a solution to this equation. However, there are two solutions to this second-order equation, and the other is $j = -(l+1)$. We know from the angular solutions that $l \geq 0$. In the limit $\rho \to 0$, $\rho^{-(l+1)}$ diverges, which is not allowed. We thus have

$$R(\rho) \xrightarrow[\rho \to 0]{} \rho^l. \tag{13.81}$$

The above discussion suggests that we should make the substitution

$$R(\rho) = \rho^l e^{-\rho/2} w(\rho). \tag{13.82}$$

This is a substitution, not an approximation. In the problems you'll show that substituting eq. (13.82) into eq. (13.74) yields the following equation for $w(\rho)$:

$$\rho \frac{\partial^2}{\partial \rho^2} w(\rho) + \left[2(l+1) - \rho\right] \frac{\partial}{\partial \rho} w(\rho) + (\lambda - l - 1) w(\rho) = 0. \tag{13.83}$$

At this point there are two ways to proceed. One way is to look for solutions for $w(\rho)$ in the form of a power series—complement 13.B follows this approach. The other way is to realize that eq. (13.83) is a well-studied differential equation, whose solutions are known. If we make the substitutions

$$j = \lambda - l - 1 \text{ and } k = 2l + 1, \tag{13.84}$$

eq. (13.83) becomes

$$\rho \frac{\partial^2}{\partial \rho^2} w(\rho) + [k + 1 - \rho] \frac{\partial}{\partial \rho} w(\rho) + j w(\rho) = 0. \tag{13.85}$$

The solutions to this equation are given by the associated Laguerre polynomials (ref. [13.1], sec. 13.2):

$$w(\rho) = L_j^k(\rho). \tag{13.86}$$

The associated Laguerre polynomials can be determined from their Rodrigues representation:

$$L_j^k(\rho) = \frac{e^{\rho} \rho^{-k}}{j!} \frac{d^j}{d\rho^j} \left(e^{-\rho} \rho^{j+k} \right). \tag{13.87}$$

They also satisfy the recursion relation

$$(j+1) L_{j+1}^k(\rho) = (2j + k + 1 - \rho) L_j^k(\rho) - (j + k) L_{j-1}^k(\rho). \tag{13.88}$$

13.4.2 Energies

If j is not 0 or a positive integer, $w(\rho)$ diverges as e^{ρ} as $\rho \to \infty$, which means that $R(\rho)$ diverges as $e^{\rho/2}$, and this is not allowed because $R(\rho)$ would not be normalizable. Since j must be an integer, eq. (13.84) tells us that λ must be an integer, so it is customary to set $\lambda = n$; n is called the principal quantum number. Since $j \geq 0$, eq. (13.84) requires that $n \geq l + 1$, and since $l \geq 0$ [eq. (13.50)], we must have $n = 1, 2, 3, \ldots$.

Setting $\lambda = n$ in eq. (13.75), and solving for the energy, yields

$$E_n = -\frac{m}{2} \left(\frac{e^2}{\hbar 4\pi\varepsilon_0} \right)^2 \frac{1}{n^2} = \frac{E_1}{n^2} = \frac{-13.6\,\text{eV}}{n^2}, \quad n = 1, 2, 3, \ldots, \tag{13.89}$$

The energies of the hydrogen atom are depicted in fig. 13.3. They depend on the principal quantum number n (but not on l, even though the wave functions depend on l), and agree very well with experimental measurements. In the experiments the frequencies (wavelengths) of light emitted or absorbed by hydrogen atoms are measured. These

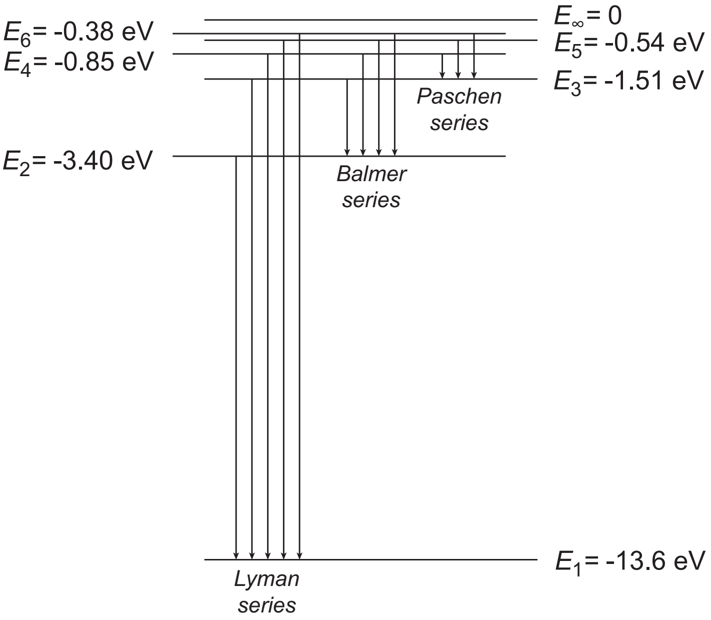


Fig 13.3 A partial energy level structure for hydrogen, indicating some transitions between levels.

measurements do not directly determine the energy levels, but rather the differences in energies between levels; this is because absorption or emission of a photon corresponds to a change in energy. Figure 13.3 shows downward arrows indicating possible energy level transitions from higher energy states to lower energy states. In an experiment measuring an emission spectrum, each of these transitions would show up as a distinct measured frequency. In terms of the energies of the initial and final states, the frequency of the emitted photon is given by

$$f = \frac{\Delta E}{h} = \frac{E_i - E_f}{h}. \tag{13.90}$$

Spectroscopists often group transitions together into “series.” Each series corresponds to transitions that end at a particular energy level. The Lyman series corresponds to transitions that end in the $n = 1$ state, while the Balmer and Paschen series transitions end in the $n = 2$ and $n = 3$ states, respectively. The Balmer series has several visible lines, and these are the lines you see if you look at a hydrogen discharge tube through a diffraction grating. The most prominent of these is the $n = 3$ to $n = 2$ transition, which corresponds to a wavelength of 656 nm, and appears quite red.

The energies of the hydrogen atom we have just derived [eq. (13.89)] agree perfectly with those predicted by the Bohr model of the atom.[5] The difference is that quantum

---

[5] For a discussion of the Bohr model, see, for example, ref. [13.6].

mechanics represents a full theory, whereas the Bohr model is simply that, a model. These energies are correct, given the Hamiltonian that we have been considering (i.e., that the potential energy is due to the Coulomb interaction of the electron with the nucleus). However, there are other interactions that we have ignored. For example, we have ignored the spins of both the proton and the electron, and these spins turn out to play a role in determining the energies. In chapter 14 we'll examine some of the corrections to the energy level structure of hydrogen. For now we'll note that the energies we have derived here are correct to about 1 part in 10^4, if we replace the mass of the electron by the reduced mass (see note 4, on page 315).

13.4.3 Radial Wave Functions

When describing the wave functions of the hydrogen atom, it is convenient to use the parameter a_0, which is the Bohr radius:

$$a_0 = \frac{4\pi\varepsilon_0\hbar^2}{me^2} = 0.529\times 10^{-10}\,\mathrm{m} = 0.529\,\mathrm{Å}. \tag{13.91}$$

We can write the energies of the hydrogen atom [eq. (13.89)] in terms of the Bohr radius as

$$E_n = -\frac{\hbar^2}{2ma_0^2}\frac{1}{n^2},\ n = 1,2,3,\ldots. \tag{13.92}$$

The dimensionless radial coordinate ρ of eq. (13.73) is then given by

$$\rho = \frac{2r}{na_0}. \tag{13.93}$$

We can now use eqs. (13.82), (13.86), and (13.93) to reassemble the radial wave functions; the result is

$$R_{nl}(r) = \frac{2}{n^2}\sqrt{\frac{(n-l-1)!}{a_0^3(n+l)!}}\left(\frac{2r}{na_0}\right)^l e^{-r/na_0} L_{n-l-1}^{2l+1}\left(\frac{2r}{na_0}\right), \tag{13.94}$$

where we have also included a normalization factor. A list of some of the low-order $R_{nl}(r)$'s is given in table 13.2, and plots of these functions are shown in fig. 13.4. For a given value of l, the radial wave functions corresponding to different n values are orthogonal. The orthogonality relation is

$$\int_0^\infty dr\, r^2 R_{nl}(r) R_{n'l}(r) = \delta_{nn'}. \tag{13.95}$$

This equation implies that the radial probability density $p_{nl}(r)$ is given by

Table 13.2 The radial wave functions for $n=1$ through $n=3$.

$R_{10}(r)=2\sqrt{\frac{1}{a_0^3}}e^{-r/a_0}$	$R_{30}(r)=\frac{2}{9}\sqrt{\frac{1}{3a_0^3}}\left[\frac{2}{9}\left(\frac{r}{a_0}\right)^2-2\frac{r}{a_0}+3\right]e^{-r/3a_0}$
$R_{20}(r)=\frac{1}{2}\sqrt{\frac{1}{2a_0^3}}\left(-\frac{r}{a_0}+2\right)e^{-r/2a_0}$	$R_{31}(r)=\frac{4}{27}\sqrt{\frac{1}{24a_0^3}}\left[-\frac{2}{3}\frac{r}{a_0}+4\right]\left(\frac{r}{a_0}\right)e^{-r/3a_0}$
$R_{21}(r)=\frac{1}{2}\sqrt{\frac{1}{6a_0^3}}\left(\frac{r}{a_0}\right)e^{-r/2a_0}$	$R_{32}(r)=\frac{8}{81}\sqrt{\frac{1}{120a_0^3}}\left(\frac{r}{a_0}\right)^2e^{-r/3a_0}$

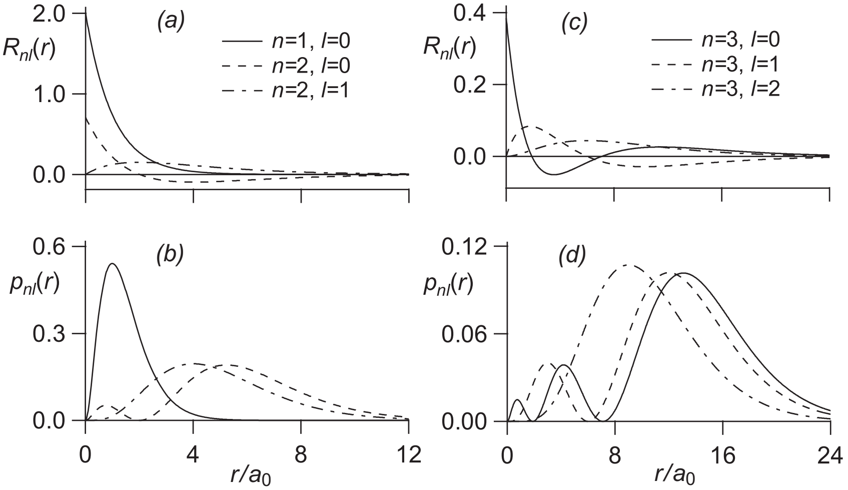


Fig 13.4 (a) Plots of the radial wave functions $R_{nl}(r)$ for $n=1$ and $n=2$, and (b) the corresponding radial probability densities $p_{nl}(r)$; the legend in (a) is also applicable to (b). (c) Plots of the radial wave functions $R_{nl}(r)$ for $n=3$, and (d) the corresponding radial probability densities $p_{nl}(r)$; the legend in (c) is also applicable to (d).

$$p_{nl}(r)=r^2\left|R_{nl}(r)\right|^2. \tag{13.96}$$

The geometrical factor of r^2 comes from the volume element in spherical coordinates [eq. (13.21)].

From fig. 13.4 we note a few features of the radial probability density $p_{nl}(r)$. The first is that $p_{nl}(r)$ contains $n-l$ maxima. The second is that as n increases, the electron is most likely to be found further from the origin. This is most easily quantified for states with $l=n-1$, for which $p_{nl}(r)$ has only a single maximum. For these states, you'll show in the problems that the maximum in the radial probability density occurs at

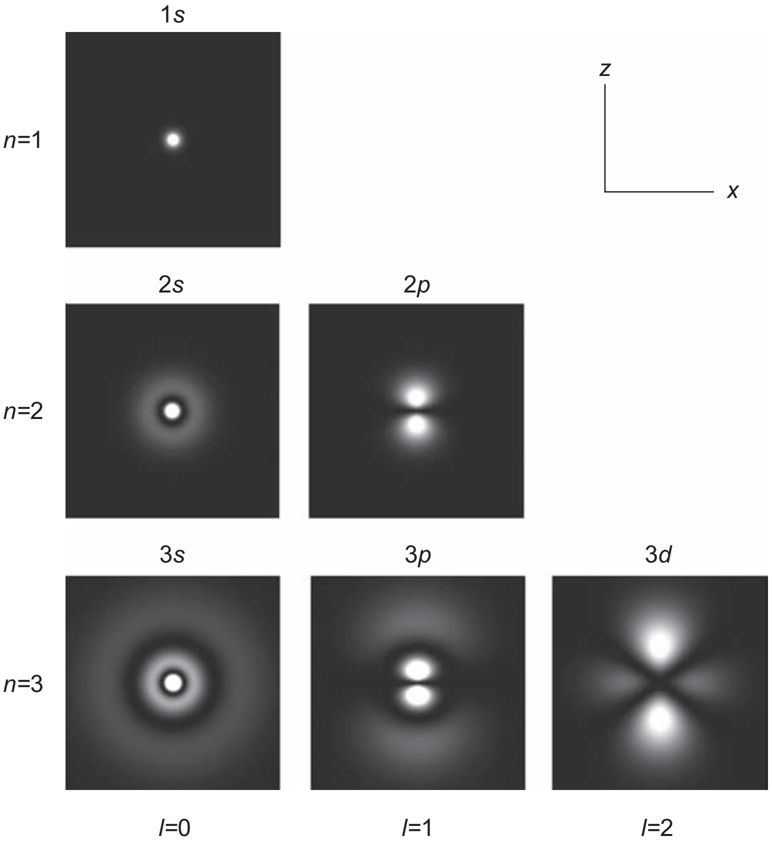


Fig 13.5 Depictions of the probability density $\left|\psi_{nl0}(\boldsymbol{r})\right|^2$ for states through $n=3$; lighter corresponds to higher probability. These distributions are rotationally symmetric about the *z*-axis. The spectroscopic designation for each of the states is shown above each figure.

$$r = n^2 a_0. \tag{13.97}$$

Thus, the size of an atom grows rapidly with increasing *n*; this can also be seen in fig. 13.5.

13.4.5 Eigenstates

The states of the hydrogen atom are are $\left|n,l,m_l\right\rangle$. The principal quantum number, *n*, determines the energy, *l* determines the orbital angular momentum, and m_l determines the *z*-component of the orbital angular momentum. These states are simultaneous eigenstates of $\hat{H}$, $\hat{L}^2$, and $\hat{L}_z$:

$$\hat{H}\left|n,l,m_l\right\rangle = E_n\left|n,l,m_l\right\rangle, \tag{13.98}$$

$$\hat{L}^2\left|n,l,m_l\right\rangle = l\left(l+1\right)\hbar^2\left|n,l,m_l\right\rangle, \tag{13.99}$$

$$\hat{L}_z\left|n,l,m_l\right\rangle = m_l\hbar\left|n,l,m_l\right\rangle. \tag{13.100}$$

The energies E_n are given in eq. (13.89). In order for $\hat{H}$, $\hat{L}^2$, and $\hat{L}_z$ to have simultaneous eigenstates, they must all commute with each other. We know that $\hat{L}^2$ and $\hat{L}_z$ commute, and in the problems you'll prove that $\hat{H}$ commutes with these operators as well. The observables H, L^2, and L_z are said to form a complete set of commuting observables for the hydrogen atom, because the eigenstates of their corresponding operators uniquely specify the state of the system (ignoring spin).

The allowed values for n are $n = 1,2,3,\ldots$. For a given value of n, the allowed values of l are $l = 0,1,\ldots,n-1$, and for a given value of l, the allowed values of m_l are $m_l = 0, \pm 1, \pm 2, \ldots, \pm l$. This means that for each value of n, there are n^2 allowed states. Since the energy of the hydrogen atom depends only on n, each energy level has a degeneracy of n^2.

Frequently the states are labeled by spectroscopic notation. In this notation there is a number that specifies the value of n, and a letter that specifies the value of l. This lettering scheme is: s for $l = 0$, p for $l = 1$, d for $l = 2$, f for $l = 3$, g for $l = 4$, and the letters continue alphabetically for higher l values. The spectroscopic notation for some states is shown in fig. 13.5.

The full wave functions for the hydrogen atom are given by

$$\left\langle \boldsymbol{r}\middle|n,l,m_l\right\rangle = \psi_{nlm_l}\left(\boldsymbol{r}\right) = R_{nl}\left(r\right)Y_l^{m_l}\left(\theta,\phi\right). \tag{13.101}$$

Plots of some of the wave functions are shown in fig. 13.5.[6] The states are orthogonal,

$$\left\langle n,l,m_l\middle|n',l',m_l'\right\rangle = \delta_{nn'}\delta_{ll'}\delta_{m_lm_l'}, \tag{13.102}$$

and so are the wave functions:

$$\int_0^\infty dr\int_0^\pi d\theta\int_0^{2\pi} d\phi\, r^2\sin\theta\,\psi^*_{nlm_l}\left(\boldsymbol{r}\right)\psi_{n'l'm_l'}\left(\boldsymbol{r}\right) = \delta_{nn'}\delta_{ll'}\delta_{m_lm_l'}. \tag{13.103}$$

The orthonormality of the wave functions is guaranteed by Eqs. (13.54) and (13.95).

---

[6] To view three-dimensional representations of these and other wave functions, see ref. [13.7].

13.5 MULTIELECTRON ATOMS

Hydrogen is the simplest atom, because it has only a single electron. Before describing how multiple electrons arrange themselves within more complicated atoms, we first need to discuss some general properties of identical quantum particles.

13.5.1 Identical Particles

In classical physics particles may look the same, but they are never indistinguishable; in principle we can tag them in some way so as to tell them apart. However, electrons and other fundamental particles are identical (indistinguishable)—there is no way, even in principle, to tell them apart. This turns out to have a profound effect on the behavior of multiparticle systems.

Of course, if the wave functions of two particles are localized, and do not overlap, we can distinguish between them—one is here, and the other is there. If particle 1 is known to be in state $\left|m\right\rangle$, and particle 2 is known to be in state $\left|n\right\rangle$, we can describe the state of the two particle system as

$$\left|\psi\right\rangle_{12} = \left|m\right\rangle_1 \left|n\right\rangle_2 . \tag{13.104}$$

However, if the wave functions of the particles overlap, as is the case for the electrons in an atom, we need to account for the fact that the particles are indistinguishable when describing the state of the two-particle system. Since we can't tell the particles apart, how do we know that it's particle 1 in state $\left|m\right\rangle$, and particle 2 in state $\left|n\right\rangle$, and not the other way around? We don't, so we have to allow for both possibilities. There are two possible combinations of the two-particle states. One is symmetric,

$$\left|\psi_s\right\rangle_{12} = \frac{1}{\sqrt{2}}\left(\left|m\right\rangle_1 \left|n\right\rangle_2 + \left|n\right\rangle_1 \left|m\right\rangle_2\right), \tag{13.105}$$

and the other is antisymmetric,

$$\left|\psi_a\right\rangle_{12} = \frac{1}{\sqrt{2}}\left(\left|m\right\rangle_1 \left|n\right\rangle_2 - \left|n\right\rangle_1 \left|m\right\rangle_2\right). \tag{13.106}$$

If the particles are exchanged, the symmetric and antisymmetric combinations satisfy:

$$\left|\psi_s\right\rangle_{12} = \left|\psi_s\right\rangle_{21}, \tag{13.107}$$

$$\left|\psi_a\right\rangle_{12} = -\left|\psi_a\right\rangle_{21}. \tag{13.108}$$

Generalizing to larger numbers of particles, you'll see in the problems that the possible combinations are more complicated, but that it is still possible to create symmetric and antisymmetric combinations. For the symmetric combination, if any two particles are exchanged the state of the system is unchanged [as in eq. (13.107)]; for the antisymmetric combination, the state acquires a minus sign on the exchange of any two particles [as in eq. (13.108)].

In relativistic quantum mechanics, it can be proven that the state of a collection of bosons (particles with integer spin) must be symmetric, while the state of a collection of fermions (particles with half-integer spin) must be antisymmetric. This is the spin-statistics theorem, and it results in an important consequence for the allowed states of fermions. Consider the antisymmetric combination of fermions described by eq. (13.106). If both particles are in the same state (i.e., $\left|m\right\rangle = \left|n\right\rangle$), then the total state would be 0, which is not allowed. This leads to the Pauli exclusion principle, which says that no two fermions can be in the same state.

13.5.2 Atomic Structure

With the Pauli exclusion principle in hand, we can now discuss how electrons arrange themselves in multielectron atoms. We'll limit ourselves to the ground states of atoms. The first thing to note is that for the ground state, electrons fill states starting with the lowest energies, working toward the highest. Thus, the ground state of hydrogen (atomic number $Z = 1$) consists of a single electron in the 1*s* state.

Now consider helium ($Z = 2$). Because electrons are spin-1/2 particles they are fermions, and the exclusion principle would seem to imply that we could not place a second electron in the 1*s* state. However, up to now we've been ignoring spin. When we take spin into account, we note that an electron can have either spin-up or spin-down, which correspond to different states. Thus, it is possible to place two electrons in the 1*s* state, but only as long as they have opposite spins. This is what happens in the ground state of helium.

The first two electrons in lithium ($Z = 3$) go into the 1*s* energy level. The third electron cannot go into that level because it is full, so this electron must move up to $n = 2$. At the level of approximation we have been using here, the $l = 0$ (2*s*) and $l = 1$ (2*p*) states of a one-electron atom have the same energy. However, for multi-electron atoms this is not the case, because the electrons in the 1*s* state remain close to the nucleus, as seen in fig. 13.4(b). The negative charge of these electrons tends to "screen" the positive nuclear charge from electrons in the $n = 2$ states. Note, however, in fig. 13.4(b) that an electron in the 2*s* state has a higher probability of being very close to the nucleus than an electron in the 2*p* state. The 2*s* electron "feels" more of the positive nuclear charge, and will thus have a lower energy. This means that the third electron in lithium goes into the 2*s* state. Beryllium ($Z = 4$) has four electrons, and the fourth electron also goes into the 2*s* state, with a spin opposite the third electron.

The fifth electron in boron ($Z = 5$) must go into one of the 2*p* states. The 2*p* energy level can accommodate six electrons ($l = 1$, so there are three allowed values of m_l, each with two possible spin states), so the atoms boron through neon ($Z = 10$) have outermost electrons that fill up the 2*p* states. Elements with higher atomic numbers have electrons that continue to fill up levels of increasing energy. It is not always obvious at first which level will be the next highest (e.g., the 4*s* level fills before the 3*d* level), but it is possible to calculate the energies, and hence determine into which state electrons will go. The energy level structure of atoms determines their properties, and hence the arrangement of the periodic table. For more details, see ref. [13.8].

For atomic physicists the alkali metals, which reside in the first column of the periodic table, play an important role. This is because these elements have electrons that fill lower-energy, inner-core states, leaving a single outer-most electron in an $l = 0$ state. Because the inner electrons are tightly bound, it is this single outer electron that largely determines the chemical and spectroscopic properties of these elements. Because there is only a single electron that determines these properties, they are comparatively easy to understand.

13.5.3 Rydberg Atoms and Electron Wave Packets

Atoms in which an electron is excited to a very high state (typically $n = 40$ or higher) are referred to as Rydberg atoms. The alkali metals are most often used in experiments involving Rydberg atoms.

Rydberg atoms have some very interesting properties. First, from eq. (13.97) we see that Rydberg atoms are much larger than typical ground state atoms; a hydrogen atom excited to $n = 300$ has a diameter of nearly 10 μm. Second, for an electron in a state with large n, there are a very large number of available l and m_l states. It is possible to place an electron in a superposition of these atomic states to create an electron wave packet within the atom.

An electron in an atomic eigenstate is not well localized. Indeed, as shown in fig. 13.5, the electron may be very spread out. However, an electron in a wave packet state of a Rydberg atom may be well localized. It may display some of the behavior of a "classical" Bohr atom; it may orbit the nucleus in much the same way that a planet orbits the sun. For an accessible introduction to electron wave packets, see ref. [13.9].

13.6 References

[13.1] H.J. Weber and G.B. Arfken, *Essential Mathematical Methods for Physicists* (Academic Press, San Diego, 2004).
[13.2] M.L. Boas, *Mathematical Methods in the Physical Sciences, 3rd ed*. (John Wiley and Sons, Hoboken, NJ, 2006), Ch. 12, Sec. 10.
[13.3] B. Reid, *Particle on a Sphere—Spherical Harmonics*, http://www.bpreid.com/poas.php.
[13.4] A. Messiah, *Quantum Mechanics* (North Holland, Amsterdam, 1961), Ch. IX.
[13.5] C. Cohen-Tannoudji, B. Diu, and Franck Laloë, *Quantum Mechanics*, (John Wiley and Sons, New York, 1977), Sec. VII.B.
[13.6] P.A. Tipler and R.A Llewellyn, *Modern Physics 5th ed*. (W.H. Freeman, New York, 2008), Sec. 4.3.
[13.7] P. Falstad, *Hydrogen Atom Applet*, http://www.falstad.com/qmatom/.
[13.8] A.P. French and E.F. Taylor, *An Introduction to Quantum Physics* (W.W. Norton, New York, 1978), Ch 13.

[13.9] M. Nauenberg, C. Stroud and J. Yeazell, “The classical limit of an atom,” Scientific American **270**, 44 (1994).

13.7 PROBLEMS

13.1 Assume that the box in example 13.1 is a cube: $L_x = L_y = L_z = L$. Find all of the allowed energies between 0 and $21\pi^2\hbar^2/2mL^2$. What is the degeneracy of each of the energy levels?

13.2 Use separation of variables in *Cartesian* coordinates to solve the Schrödinger equation [eq. (13.7)] for a particle in a symmetric harmonic oscillator potential:

$$V(\boldsymbol{r}) = \frac{1}{2}m\omega^2 r^2 = \frac{1}{2}m\omega^2\left(x^2+y^2+z^2\right). \tag{13.109}$$

Show that the allowed energies are

$$E_{n_x n_y n_z} = \hbar\omega\left(n_x + n_y + n_z + \frac{3}{2}\right) \quad n_x, n_y, n_z = 0,1,2,\ldots. \tag{13.110}$$

13.3 Show that

$$\frac{\partial}{\partial x} = \sin\theta\cos\phi\frac{\partial}{\partial r} + \frac{\cos\theta\cos\phi}{r}\frac{\partial}{\partial\theta} - \frac{\sin\phi}{r\sin\theta}\frac{\partial}{\partial\phi}, \tag{13.111}$$

$$\frac{\partial}{\partial y} = \sin\theta\sin\phi\frac{\partial}{\partial r} + \frac{\cos\theta\sin\phi}{r}\frac{\partial}{\partial\theta} + \frac{\cos\phi}{r\sin\theta}\frac{\partial}{\partial\phi}, \tag{13.112}$$

and

$$\frac{\partial}{\partial z} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta}, \tag{13.113}$$

Hint: Use the chain rule in 3 dimensions.

13.4 Show that $Y_l^{-m_l}(\theta,\phi) = (-1)^{m_l} Y_l^{m_l *}(\theta,\phi)$.

13.5 Verify eq. (13.54).

13.6 Show that the orbital angular momentum operator can be represented as

$$\hat{\boldsymbol{L}} \doteq -i\hbar\left(\boldsymbol{r}\times\nabla\right), \tag{13.114}$$

13.7* Prove eqs. (13.57)–(13.59). [Hint: Use eqs. (13.111)–(13.113).]

13.8* Verify that $\hat{L}^2$ can be represented using eq. (13.60).

13.9 Show that the angular momentum operator can be represented as

$$\hat{\boldsymbol{L}} \doteq -i\hbar\left(-\boldsymbol{u}_\theta\frac{1}{\sin\theta}\frac{\partial}{\partial\phi} + \boldsymbol{u}_\phi\frac{\partial}{\partial\theta}\right) \tag{13.115}$$

Hint: The gradient operator is represented in spherical coordinates as

$$\nabla = \boldsymbol{u}_r \frac{\partial}{\partial r} + \boldsymbol{u}_\theta \frac{1}{r}\frac{\partial}{\partial \theta} + \boldsymbol{u}_\phi \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}. \qquad (13.116)$$

13.10 Show that

$$\langle \boldsymbol{r} | \hat{p}_r | \psi \rangle = -i\hbar\left(\frac{\partial}{\partial r} + \frac{1}{r}\right)\langle \boldsymbol{r} | \psi \rangle, \qquad (13.117)$$

and hence that eq. (13.62) is correct. Hint: The divergence in spherical coordinates is given by

$$\nabla \cdot \boldsymbol{f} = \frac{1}{r^2 \sin\theta}\left[\sin\theta \frac{\partial}{\partial r}\left(r^2 f_r\right) + r\frac{\partial}{\partial \theta}\left(\sin\theta f_\theta\right) + r\frac{\partial}{\partial \phi} f_\phi\right]. \qquad (13.118)$$

13.11* Prove that eq. (13.24) can be derived from eq. (13.64), assuming a central potential.

13.12* Express the raising operator $\hat{L}_+$ in the position basis. Show that it can be used to generate $Y_l^{m_l+1}(\theta,\phi)$ from $Y_l^{m_l}(\theta,\phi)$.

13.13* Show that by substituting eq. (13.82) into eq. (13.74), one obtains eq. (13.83).

13.14* Show that $\hat{H}$ for a central potential commutes with $\hat{L}^2$ and $\hat{L}_z$. (Hint: Work in Cartesian coordinates, and begin by examining the commutator of $\hat{L}_z$ with the components of $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{p}}$.)

13.15 What is the shortest wavelength of light emitted by a hydrogen atom making a transition in the Lyman series?

13.16 What is the longest wavelength of light emitted by a hydrogen atom making a transition in the Paschen series?

13.17 Show that the maximum of the radial probability density for a hydrogen atom in a state with $l = n-1$ is located at $r = n^2 a_0$.

13.18* What is the probability that an electron in a $2p$ state of a hydrogen atom will be found in the region $r < 3a_0$?

13.19 What is the probability that an electron in the $3s$ state of a hydrogen atom will be found inside the nucleus? Take the nuclear radius to be $\sim 10^{-15}$ m. (Hint: An excellent approximation is obtained by assuming that the nucleus is much smaller than the atom.)

13.20* Find $\langle r \rangle$, $\langle r^2 \rangle$, $\langle z \rangle$, and $\langle z^2 \rangle$ for an electron in the ground state of hydrogen.

13.21 At $t = 0$ the state of a hydrogen atom is

$$|\psi(0)\rangle = \frac{1}{\sqrt{2}}\left(|1,0,0\rangle + |2,1,0\rangle\right). \qquad (13.119)$$

Find $\langle z \rangle(t)$.

13.22 For an electron in the ground state of hydrogen, find the expectation value of the potential energy. What must the expectation value of the kinetic energy be?

13.23 Three particles (1,2,3) can be in each of the three states $|\psi_1\rangle$, $|\psi_2\rangle$, or $|\psi_3\rangle$. Determine the symmetric and antisymmetric wave function combinations for the particles. To find the antisymmetric combination, you can use the use Slater determinant, which for N particles takes the form:

$$\left|\psi_a\right\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix} \left|\psi_1\right\rangle_1 & \cdots & \left|\psi_N\right\rangle_1 \\ \vdots & \ddots & \vdots \\ \left|\psi_1\right\rangle_N & \cdots & \left|\psi_N\right\rangle_N \end{vmatrix}. \tag{13.120}$$

13.24 Let $\langle K\rangle$ be the expectation value of the kinetic energy. The virial theorem tells us that for stationary (time independent) states,

$$2\langle K\rangle = \left\langle \boldsymbol{r}\bullet\nabla V(\boldsymbol{r})\right\rangle. \tag{13.121}$$

Apply the virial theorem to the hydrogen atom. [The gradient in spherical coordinates is given in eq. (13.116).] How do the kinetic and potential energies relate to the total energy for stationary states?

COMPLEMENT 13.A

Quantum Dots

In chap. 13 we described some of the properties of solutions to the Schrödinger equation for a central potential. Here we'll discuss a specific example, that of the infinite spherical well, for which the potential is zero inside a sphere of radius *R*, and infinite outside:

$$V(\mathbf{r}) = \begin{cases} 0 & r < R \\ \infty & r \geq R \end{cases}. \tag{13.A.1}$$

Understanding the infinite spherical well will help us to understand the behavior of quantum dots. Quantum dots are small structures, typically a few nm in diameter, and that are usually made from semiconducting materials. In sec. 11.5.4 we discussed quantum wells, in which electrons and holes are confined in one dimension. Quantum dots are similar, only the confinement is in all three dimensions, so we need a 3-D model to more fully understand them.

13.A.1 The Infinite Spherical Well

Inside the well the radial wave equation [eq. (13.30)] is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}R(r)\right) - \frac{2m}{\hbar^2}\left[\frac{\hbar^2 l(l+1)}{2mr^2} - E\right]R(r) = 0. \tag{13.A.2}$$

The boundary conditions on this equation are:

$$R(r) = 0 \quad r \geq R, \tag{13.A.3}$$

and $R(r)$ must be normalizable for $r < R$.

We can simplify eq. (13.A.2) by making the substitution

$$\rho = kr = \frac{\sqrt{2mE}}{\hbar} r, \tag{13.A.4}$$

which yields

$$\frac{\partial^2}{\partial\rho^2} R(\rho) + \frac{2}{\rho}\frac{\partial}{\partial\rho} R(\rho) + \left[1 - \frac{l(l+1)}{\rho^2}\right] R(\rho) = 0. \tag{13.A.5}$$

This is a well-known differential equation, and the normalizable solutions are given by the spherical Bessel functions:

$$R(\rho) = j_l(\rho), \tag{13.A.6}$$

where l is a nonnegative integer. These functions can be expressed using the Rayleigh formula:

$$j_l(\rho) = (-\rho)^l \left(\frac{1}{\rho}\frac{d}{d\rho}\right)^l \left(\frac{\sin\rho}{\rho}\right), \tag{13.A.7}$$

The first few spherical Bessel functions are listed in table 13.A.1. For more information about the spherical Bessel functions, see ref. [13.A.1].

To find the allowed energies for a particle in the well, we need to satisfy the boundary condition

$$R_{nl}(k_{nl}R) = j_l(k_{nl}R) = j_l(\beta_{nl}) = 0. \tag{13.A.8}$$

Here we have defined the n^{th} zero of $j_l(\rho)$ as β_{nl}. As can be seen from table 13.A.1, the $j_l(\rho)$'s are oscillatory functions, so each has an infinite number of zeros. From the form of $j_0(\rho)$, we can see that its zeros are simply multiples of π. However, the zeros of the other spherical Bessel functions must be computed numerically. Some of the low-order zeros are given in table 13.A.2; for a more extensive table, see ref. [13.A.2].

The equation for the allowed energies, from eqs. (13.A.4) and (13.A.8), is

$$k_{nl}R = \frac{\sqrt{2mE_{nl}}}{\hbar} R = \beta_{nl}. \tag{13.A.9}$$

Solving for the energies, we find

$$E_{nl} = \frac{\beta_{nl}^2 \hbar^2}{2mR^2}. \tag{13.A.10}$$

Note that this looks similar to the equation for the energies of a one-dimensional infinite square well [eq. (11.67)].

The angular dependence of the wave functions is given by the spherical harmonics, as described in sec. 13.2, so the full wave functions are

$$\langle \boldsymbol{r} | n, l, m_l \rangle = \psi_{nlm_l}(\boldsymbol{r}) = c_{nl}\, j_l\!\left(\frac{\beta_{nl} r}{R}\right) Y_l^{m_l}(\theta, \phi). \tag{13.A.11}$$

Table 13.A.1 The spherical Bessel functions for $l = 0$ through $l = 2$.

$$j_0(\rho) = \frac{\sin\rho}{\rho} = \operatorname{sinc}(\rho)$$

$$j_1(\rho) = \frac{\sin\rho}{\rho^2} - \frac{\cos\rho}{\rho}$$

$$j_2(\rho) = \left(\frac{3}{\rho^3} - \frac{1}{\rho}\right)\sin\rho - \frac{3}{\rho^2}\cos\rho$$

Table 13.A.2 The n^th^ zero of $j_l(\rho)$, β_{nl}, for $n = 1$ through $n = 3$, and $l = 0$ through $l = 2$.

β_{nl}	n=1	n=2	n=3
l=0	π	2π	3π
l=1	4.49	7.73	10.90
l=2	5.76	9.10	12.32

The constants c_{nl} are determined by the normalization integral

$$\int_0^R dr \int_0^\pi d\theta \int_0^{2\pi} d\phi\, r^2 \sin\theta \left|\psi_{nlm_l}(\boldsymbol{r})\right|^2 = c_{nl}^2 \int_0^R dr\, r^2 \left[j_l\left(\frac{\beta_{nl} r}{R}\right)\right]^2 = 1. \quad (13.A.12)$$

The integral simplifies in this manner because we know that the spherical harmonics are normalized [eq. (13.54)].

13.A.2 Semiconductors

As stated above, quantum dots are typically made from semiconducting materials, and confine electrons and holes in three dimensions. Before discussing how the confinement affects the behavior of the dot, let's talk briefly about bulk solids.

What happens if we bring two atoms of the same type close enough together, so that the wave functions of their outer electrons start to overlap? When the atoms are far apart their energy level structures are identical. This means that the energy levels of the two-atom system are degenerate. As the wave functions of the atoms begin to overlap, the atoms become coupled to each other, and each of the degenerate energy levels split into two levels, with a small energy separation, as shown in fig. 13.A.1.

If we bring N atoms together, each energy level splits into N levels. If N is large, as in a solid, there are so many closely spaced levels that each energy level becomes a "band" of allowed energies, as in fig. 13.A.1. Each band extends over some range of energies, and any energy within the band is allowed. For this reason, when discussing solids we often refer to the band structure of the material.

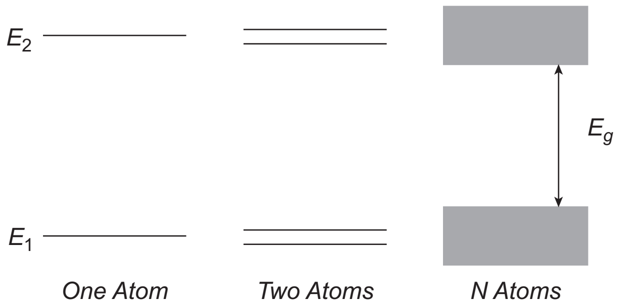


Fig 13.A.1 The energy level structure that results when atoms are brought together. For two atoms, each energy level splits into two levels. For N atoms, the energy levels split into N closely spaced levels, which effectively form a band of allowed energies. The band gap energy E_g is shown.

The band structure tells us the allowed energy levels for the electrons, but in which levels do the electrons actually reside? As we saw in sec. 13.5, when discussing multi-electron atoms, electrons fill the lowest energy states first. In a bulk semiconductor at absolute zero, the low-lying states are filled with electrons, all the way to the top of what is called the valence band—the valence band is full. When electrons are in the filled valence band they are tightly bound to the individual atoms in the solid, and are not free to move. Because of this, a material with all of its electrons in the valence band won't conduct electricity, and is an insulator.

The allowed band with the next highest energy is called the conduction band, and it is separated from the valence band by the band gap energy E_g, as depicted in fig. 13.A.1. If an electron acquires enough energy to move to the conduction band, it is no longer tightly bound. If sufficient numbers of electrons are in the conduction band, the material readily conducts electricity. Semiconductors are good insulators at low temperature, but their band gap is relatively small. Through an increase in temperature, or illumination by light, electrons absorb energy and move to the conduction band, which makes the material conducting (as happens in solar cells).[7]

For the purposes of our discussion here, we are interested in the band structure of a semiconductor in order to understand what energy photons it will absorb and emit. Consider the optical process in which a valence-band electron is promoted to an available state in the conduction band through the absorption of a photon. The incident photon must have at least E_g worth of energy in order to excite the electron from the valence band to the conduction band. During this process, the electron leaves behind an empty state in the valence band, which is called a hole. As described previously, a hole is a "lack of an electron," and behaves in many ways as a mobile positive charge carrier.

The properties of electrons and holes in solids are strongly influenced by the properties of the bulk material. For example, electrons and holes have "effective masses" that we use to describe their behavior in a solid. We denote the effective masses of electrons and holes as m_e^* and m_h^*, respectively. For much more information about solids, see, for example, ref. [13.A.3].

---

[7] It is also possible to dope impurities into semiconductors, which can make them more readily conducting.

13.A.3 Quantum Dots

Suppose that a semiconducting material is formed into a quantum dot, which we will treat as a sphere of radius R. What would be the minimum energy needed to create an electron-hole pair in this dot?

First, we know that it needs to absorb at least E_g of energy, because this is the energy needed to excite the electron into the conduction band, and to create an electron-hole pair. If this occurred in bulk material, not a quantum dot, this would be the whole story. However, for a quantum dot we also have to worry about the energy associated with the confinement. For our model of the dot, the confinement energy is given by eq. (13.A.10); since we're interested in the lowest energy we'll take the lowest allowed value of β_{nl}, which is $\beta_{nl} = \pi$. We have two confined particles, the electron and the hole, and we need to account for the confinement energy of each. The lowest energy of an electron-hole pair in a quantum dot is thus

$$E_{eh} = E_g + \frac{\pi^2 \hbar^2}{2m_e^* R^2} + \frac{\pi^2 \hbar^2}{2m_h^* R^2} . \tag{13.A.13}$$

Equation (13.A.13) is a reasonable first approximation, but it is of course an approximation. One thing it ignores is the energy of the Coulomb interaction between the electron and the hole. If the radius of the dot is small enough, the confinement energy dominates the Coulomb energy of the electron and the hole, and the approximation improves. However, if the dot is too small there are not many atoms, and the value of E_g for the dot is no longer well described by the value of the bulk material.

Quantum dots are technologically useful precisely because their energy levels can be controlled by varying their size, as described by eq. (13.A.13). Dots fabricated from the same material can be made to absorb or emit light at dramatically different wavelengths. For more information on the energy level structure of semiconductor quantum dots, see ref. [13.A.4].

13.A.4 References

[13.A.1] H.J. Weber and G.B. Arfken, *Essential Mathematical Methods for Physicists* (Academic Press, San Diego, 2004), Sec. 12.4.

[13.A.2] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1965), Table 10.6.

[13.A.3] C. Kittel, *Introduction to Solid State Physics, 8th ed.* (John Wiley and Sons, Hoboken, NJ, 2005).

[13.A.4] Z.M. Schultz and J.M. Essick, "Investigation of exciton ground state in quantum dots via Hamiltonian diagonalization method," Am. J. Phys. **76**, 241 (2008).

13.A.5 PROBLEMS

13.A.1 A collection of quantum dots is made from CdSe. Bulk CdSe has $E_g = 1.75$ eV, $m_e^* = 0.13\, m_e$ and $m_h^* = 0.45\, m_e$. What dot radius would be necessary to produce light with wavelengths of 610 nm and shorter?

COMPLEMENT 13.B

Series Solution to the Radial Equation

In sec. 13.4 we found solutions to the radial part of the Schrödinger equation for the hydrogen atom in terms of the associated Laguerre polynomials. Here we will show that it is possible to obtain these same solutions in terms of a power series.

We were able to write the radial wave functions as [eq. (13.82)]

$$R(\rho) = \rho^l e^{-\rho/2} w(\rho), \tag{13.B.1}$$

where the function $w(\rho)$ is a solution to eq. (13.83):

$$\rho \frac{\partial^2}{\partial\rho^2} w(\rho) + \left[2(l+1) - \rho\right] \frac{\partial}{\partial\rho} w(\rho) + (\lambda - l - 1) w(\rho) = 0. \tag{13.B.2}$$

We'll look for a power series solution to this equation, so we'll write

$$w(\rho) = \sum_{j=0} c_j \rho^j. \tag{13.B.3}$$

Substitute this into eq. (13.B.2), and we obtain

$$\rho \sum_{j=0} c_j j(j-1)\rho^{j-2} + \left[2(l+1) - \rho\right] \sum_{j=0} c_j j\rho^{j-1} + (\lambda - l - 1) \sum_{j=0} c_j \rho^j = 0. \tag{13.B.4}$$

Grouping together like powers of ρ, and noticing that some of the $j = 0$ terms are 0, we can rewrite this as

$$\sum_{j=1} c_j j\left[(j-1) + 2(l+1)\right]\rho^{j-1} + \sum_{j=0} c_j (\lambda - l - 1 - j)\rho^j = 0. \tag{13.B.5}$$

Making the substitution $j' = j - 1$ yields:

$$\sum_{j'=0} c_{j'+1} (j'+1)\left[j' + 2(l+1)\right]\rho^{j'} + \sum_{j=0} c_j (\lambda - l - 1 - j)\rho^j = 0. \tag{13.B.6}$$

Now making the substitution $j' = j$, and combining the two series, gives

$$\sum_{j=0} \left\{c_{j+1}(j+1)\left[j+2(l+1)\right]+c_j(\lambda-l-1-j)\right\}\rho^j = 0. \tag{13.B.7}$$

In order for this equation to be true, every term in the series must be 0, which yields the recursion relation

$$c_{j+1} = -\frac{c_j(\lambda-l-1-j)}{(j+1)\left[j+2(l+1)\right]} = \frac{j-(\lambda-l-1)}{(j+1)\left[j+2(l+1)\right]}c_j. \tag{13.B.8}$$

Given the coefficient c_0, which is used for normalization, the higher-order coefficients are determined.

We need our solutions to be normalizable, so the series cannot diverge too rapidly as $\rho \to \infty$. The large ρ behavior of the series is determined by the behavior of the coefficients c_j in the limit of large j. In the large j limit, we have

$$\frac{c_{j+1}}{c_j} \xrightarrow[j\to\infty]{} \frac{j}{j^2} = \frac{1}{j}. \tag{13.B.9}$$

The power series representation of e^ρ is

$$e^\rho = \sum_{j=0}^{\infty} \frac{1}{j!}\rho^j = \sum_{j=0}^{\infty} b_j\rho^j. \tag{13.B.10}$$

and for this series we have

$$\frac{b_{j+1}}{b_j} = \frac{j!}{(j+1)!} = \frac{1}{j+1} \xrightarrow[j\to\infty]{} \frac{1}{j}, \tag{13.B.11}$$

which behaves the same as eq. (13.B.9). Thus, the series of eq. (13.B.3), with coefficients given by eq. (13.B.8), will diverge as e^ρ in the limit of large ρ. This means that the radial wave functions $R(\rho)$ will diverge as $e^{\rho/2}$, which is not normalizable, and hence is not allowed.

The only way to stop the series in eq. (13.B.3) from diverging is to truncate it at $j = j_{\text{max}}$. To guarantee that the series truncates, it is necessary to have $c_{j_{\text{max}}+1} = 0$, which will then ensure that all higher terms will be 0 as well. From the recursion relation in eq. (13.B.8), we see that $c_{j_{\text{max}}+1} = 0$ as long as $j_{\text{max}} = \lambda - l - 1$. In order for this to occur, it is necessary for λ to be an integer, so we set $\lambda = n$.

Note that we have now reproduced results of sec. 13.4. Requiring that $\lambda = n$ be an integer yields the quantized energies of eq. (13.89). It also ensures that the series of eq. (13.B.3) truncates, so that the functions $w(\rho)$ are a set of polynomials of order $j_{\text{max}} = n - l - 1$. These polynomials, whose coefficients are given by eq. (13.B.8), are exactly the associated Laguerre polynomials described in sec. 13.4.

13.B.1 PROBLEMS

13.B.1* Solve the Schrödinger equation for a particle in a symmetric harmonic oscillator potential,

$$V(\mathbf{r}) = \frac{1}{2} m\omega^2 r^2, \tag{13.B.12}$$

using spherical coordinates. For the radial equation, factor out the limiting behaviors as $r \to \infty$ and $r \to 0$, and then look for a series solution. Show that the energies are given by

$$E_{nl} = \hbar\omega\left(2n + l + \frac{3}{2}\right), \; n, l = 0, 1, 2, \ldots. \tag{13.B.13}$$

(Hint: Use the dimensionless parameters $\rho = \sqrt{m\omega/\hbar}\, r$, and $\lambda = 2E/\hbar\omega$.)

13.B.2* Show that the energies for the symmetric harmonic oscillator are the same for solutions in both spherical coordinates [eq. (13.B.13)] and Cartesian coordinates [eq. (13.110)], as they must be. Show that the degeneracy of the first three energy levels is the same for both spherical and Cartesian coordinates. (The degeneracy of all the levels must be the same.)

CHAPTER 14

Time-Independent Perturbation Theory

So far we have concentrated on finding exact solutions to problems. In order to do this we've had to simplify many of the problems to make them tractable. An example is the hydrogen atom. The solutions we obtained in chap. 13 are exact, assuming that the interaction between electron and the nucleus is fully characterized by the Coulomb force. However, we've ignored several effects, such as those due to the spins of the electron and the nucleus. In this chapter we'll use perturbation theory to look at how some of the things we've ignored modify the energy level structure of quantum systems, such as hydrogen. Perturbation theory is one technique for obtaining approximate solutions to more complicated problems. The idea is that we start with a problem for which we know the exact solutions, and then add a level of complication that modifies them. Perturbation theory is valid if the perturbation is small enough, so that the solutions are largely unchanged.

14.1 NONDEGENERATE THEORY

We want to find the energies and eigenstates of a Hamiltonian $\hat{H}$, which consists of the sum of two pieces:

$$\hat{H} = \hat{H}_0 + \hat{H}_p. \tag{14.1}$$

The first piece, $\hat{H}_0$, is a Hamiltonian for which we know the exact energies and eigenstates:

$$\hat{H}_0 \left|\psi_n^{(0)}\right\rangle = E_n^{(0)} \left|\psi_n^{(0)}\right\rangle. \tag{14.2}$$

The superscript (0) indicates that these are the zeroth-order (unperturbed) energies and states. In this section we are considering nondegenerate perturbation theory, so each energy $E_n^{(0)}$ corresponds to a unique eigenstate $\left|\psi_n^{(0)}\right\rangle$. The second piece of the Hamiltonian, $\hat{H}_p$, is called the perturbing Hamiltonian. It changes the energies and states from their unperturbed values, and the goal of perturbation theory is to determine how.

We'll start by slightly modifying the Hamiltonian of eq. (14.1), so that it becomes

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}_p. \tag{14.3}$$

Here λ is a dimensionless parameter, which can take on values between 0 and 1. There are two uses for λ. The first is that it allows us to change the strength of the perturbation. By setting λ equal to 0 we return to our original problem, while setting it equal to 1 yields and the full problem we're interested in. For values of λ between 0 and 1, we transition smoothly from one extreme to the other. The second use of λ is as a bookkeeping parameter; it allows us to keep track of the orders of perturbation, as we'll see.

The ultimate goal is to find the full energies and eigenstates of $\hat{H}$:

$$\hat{H}\left|\psi_n\right\rangle = E_n\left|\psi_n\right\rangle. \tag{14.4}$$

Write these energies and eigenstates as a power series in λ:

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \ldots, \tag{14.5}$$

$$\left|\psi_n\right\rangle = \left|\psi_n^{(0)}\right\rangle + \lambda\left|\psi_n^{(1)}\right\rangle + \lambda^2\left|\psi_n^{(2)}\right\rangle + \ldots. \tag{14.6}$$

The terms beyond the zeroth-order solutions are the corrections to the energies and wave functions we're looking for. If we set $\lambda = 0$ in eqs. (14.3), (14.5), and (14.6), then the eigenvalue problem of eq. (14.4) reduces to the unperturbed eigenvalue problem of eq. (14.2), as it should. If λ is small enough, the power series in eqs. (14.5) and (14.6) will converge. However, to ensure the validity of perturbation theory we would like these series to converge for $\lambda = 1$, and this will only happen if higher-order corrections get successively smaller.

If we substitute eqs. (14.3), (14.5), and (14.6) into eq. (14.4) we obtain

$$\begin{aligned}
&\left(\hat{H}_0 + \lambda\hat{H}_p\right)\left(\left|\psi_n^{(0)}\right\rangle + \lambda\left|\psi_n^{(1)}\right\rangle + \lambda^2\left|\psi_n^{(2)}\right\rangle + \ldots\right)\\
&\quad = \left(E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \ldots\right)\left(\left|\psi_n^{(0)}\right\rangle + \lambda\left|\psi_n^{(1)}\right\rangle + \lambda^2\left|\psi_n^{(2)}\right\rangle + \ldots\right).
\end{aligned} \tag{14.7}$$

Collecting together terms with like powers of λ yields

$$\begin{aligned}
&\left(\hat{H}_0\left|\psi_n^{(0)}\right\rangle\right) + \lambda\left(\hat{H}_0\left|\psi_n^{(1)}\right\rangle + \hat{H}_p\left|\psi_n^{(0)}\right\rangle\right) + \lambda^2\left(\hat{H}_0\left|\psi_n^{(2)}\right\rangle + \hat{H}_p\left|\psi_n^{(1)}\right\rangle\right) + \ldots\\
&\quad = \left(E_n^{(0)}\left|\psi_n^{(0)}\right\rangle\right) + \lambda\left(E_n^{(0)}\left|\psi_n^{(1)}\right\rangle + E_n^{(1)}\left|\psi_n^{(0)}\right\rangle\right)\\
&\qquad + \lambda^2\left(E_n^{(0)}\left|\psi_n^{(2)}\right\rangle + E_n^{(1)}\left|\psi_n^{(1)}\right\rangle + E_n^{(2)}\left|\psi_n^{(0)}\right\rangle\right) + \ldots.
\end{aligned} \tag{14.8}$$

In order for this equation to hold for an arbitrary value of λ, terms of like powers of λ on opposite sides of the equal sign must be equal. The zeroth-order terms, proportional to λ^0, yield eq. (14.2), which involves only the unperturbed energies and states.

The first-order terms in eq. (14.8) yield

$$\hat{H}_0\left|\psi_n^{(1)}\right\rangle+\hat{H}_p\left|\psi_n^{(0)}\right\rangle=E_n^{(0)}\left|\psi_n^{(1)}\right\rangle+E_n^{(1)}\left|\psi_n^{(0)}\right\rangle, \tag{14.9}$$

and the second-order terms yield

$$\hat{H}_0\left|\psi_n^{(2)}\right\rangle+\hat{H}_p\left|\psi_n^{(1)}\right\rangle=E_n^{(0)}\left|\psi_n^{(2)}\right\rangle+E_n^{(1)}\left|\psi_n^{(1)}\right\rangle+E_n^{(2)}\left|\psi_n^{(0)}\right\rangle. \tag{14.10}$$

It is possible to collect higher-order terms, but in practice perturbation theory is usually not the best approach to solving a problem if terms higher than second-order are needed. Now that we've identified the orders of perturbation, we no longer have need of the parameter λ, and can set it equal to 1.

14.1.1 First-Order Corrections

If we project $\left\langle\psi_n^{(0)}\right|$ onto eq. (14.9), we obtain

$$\begin{gathered}\left\langle\psi_n^{(0)}\right|\hat{H}_0\left|\psi_n^{(1)}\right\rangle+\left\langle\psi_n^{(0)}\right|\hat{H}_p\left|\psi_n^{(0)}\right\rangle=\left\langle\psi_n^{(0)}\right|E_n^{(0)}\left|\psi_n^{(1)}\right\rangle+\left\langle\psi_n^{(0)}\right|E_n^{(1)}\left|\psi_n^{(0)}\right\rangle,\\ E_n^{(0)}\left\langle\psi_n^{(0)}\middle|\psi_n^{(1)}\right\rangle+\left\langle\psi_n^{(0)}\right|\hat{H}_p\left|\psi_n^{(0)}\right\rangle=E_n^{(0)}\left\langle\psi_n^{(0)}\middle|\psi_n^{(1)}\right\rangle+E_n^{(1)},\end{gathered} \tag{14.11}$$

where we've used the fact that $\hat{H}_0$ is Hermitian. The first terms on either side of the equal sign cancel, and we find

$$E_n^{(1)}=\left\langle\psi_n^{(0)}\right|\hat{H}_p\left|\psi_n^{(0)}\right\rangle. \tag{14.12}$$

This is a straightforward result—the first-order correction to the energy is given by the expectation value of the perturbing Hamiltonian, using the unperturbed state. Since $E_n^{(1)}$ is the first-order correction, the total energy is $E_n=E_n^{(0)}+E_n^{(1)}$.

Now we'd like to find the first-order correction to the state, $\left|\psi_n^{(1)}\right\rangle$. The unperturbed states form a complete set, so we can write the first-order correction as a linear combination of the unperturbed states:

$$\left|\psi_n^{(1)}\right\rangle=\sum_m c_{mn}^{(1)}\left|\psi_m^{(0)}\right\rangle. \tag{14.13}$$

Note, however, that our ultimate goal is to find the eigenstates $\left|\psi_n\right\rangle$ of the full Hamiltonian $\hat{H}$. These states are given by the expansion of eq. (14.6) (with $\lambda=1$), which already contains $\left|\psi_n^{(0)}\right\rangle$. Thus, there is no need to include this $\left|\psi_n^{(0)}\right\rangle$ contribution in the expansion of eq. (14.13). By assuming that $\left\langle\psi_n^{(0)}\middle|\psi_n^{(1)}\right\rangle=0$, the expansion for $\left|\psi_n^{(1)}\right\rangle$ becomes

$$\left|\psi_n^{(1)}\right\rangle = \sum_{m\neq n} c_{mn}^{(1)}\left|\psi_m^{(0)}\right\rangle. \tag{14.14}$$

The coefficients in this expansion are given by

$$c_{mn}^{(1)} = \left\langle\psi_m^{(0)}\middle|\psi_n^{(1)}\right\rangle. \tag{14.15}$$

Applying $\left\langle\psi_m^{(0)}\right|$, with $m \neq n$, to eq. (14.9) yields

$$\left\langle\psi_m^{(0)}\middle|\hat{H}_0\middle|\psi_n^{(1)}\right\rangle + \left\langle\psi_m^{(0)}\middle|\hat{H}_p\middle|\psi_n^{(0)}\right\rangle = \left\langle\psi_m^{(0)}\middle|E_n^{(0)}\middle|\psi_n^{(1)}\right\rangle + \left\langle\psi_m^{(0)}\middle|E_n^{(1)}\middle|\psi_n^{(0)}\right\rangle. \tag{14.16}$$

The unperturbed states are orthogonal, so the last term is 0, and this reduces to

$$E_m^{(0)}\left\langle\psi_m^{(0)}\middle|\psi_n^{(1)}\right\rangle + \left\langle\psi_m^{(0)}\middle|\hat{H}_p\middle|\psi_n^{(0)}\right\rangle = E_n^{(0)}\left\langle\psi_m^{(0)}\middle|\psi_n^{(1)}\right\rangle. \tag{14.17}$$

Rearranging, we find that

$$c_{mn}^{(1)} = \left\langle\psi_m^{(0)}\middle|\psi_n^{(1)}\right\rangle = \frac{\left\langle\psi_m^{(0)}\middle|\hat{H}_p\middle|\psi_n^{(0)}\right\rangle}{E_n^{(0)} - E_m^{(0)}}. \tag{14.18}$$

We are assuming nondegenerate states, so no two of the energies are the same, and we don't need to worry about the denominator in this equation being zero. Substituting the coefficients of eq. (14.18) into the expansion of eq. (14.14) yields the first-order correction to the state:

$$\left|\psi_n^{(1)}\right\rangle = \sum_{m\neq n} \frac{\left\langle\psi_m^{(0)}\middle|\hat{H}_p\middle|\psi_n^{(0)}\right\rangle}{E_n^{(0)} - E_m^{(0)}}\left|\psi_m^{(0)}\right\rangle. \tag{14.19}$$

Let's do an example, to better see how all this works.

EXAMPLE 14.1

A uniform electric field $\boldsymbol{\mathcal{E}} = \mathcal{E}\boldsymbol{u}_x$ is applied to a particle with a charge of q that is situated in an infinite potential well. The potential energy is given by

$$V(x) = \begin{cases} -q\mathcal{E}x & 0 < x < L \\ \infty & elsewhere \end{cases}, \tag{14.20}$$

and is depicted in fig. 14.1. To first-order, determine the allowed energies of this particle.

The unperturbed energies and wave functions are those of the infinite potential well [eqs. (11.67) and (11.70)]:

$$E_n^{(0)} = \frac{n^2\pi^2\hbar^2}{2mL^2} \quad n = 1, 2, 3, \ldots, \tag{14.21}$$

$$\psi_n^{(0)}(x) = \begin{cases} \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi}{L}x\right) & 0 < x < L \\ 0 & elsewhere \end{cases}. \tag{14.22}$$

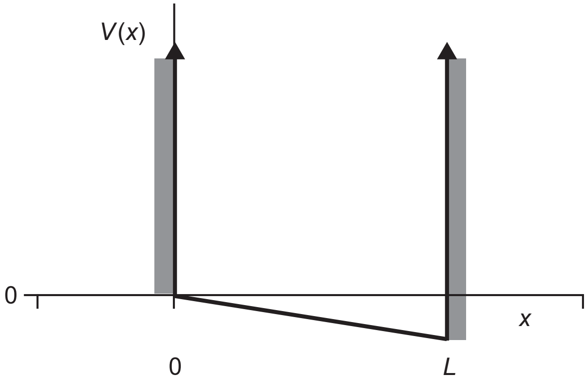


Fig 14.1 The potential energy of a charged particle in an infinite well with an applied electric field.

The perturbing Hamiltonian is

$$\hat{H}_p\left(\hat{x}\right)=-q\mathcal{E}\,\hat{x}. \tag{14.23}$$

The first-order correction to the energy is given by eq. (14.12):

$$\begin{aligned}E_n^{(1)}&=\left\langle\psi_n^{(0)}\middle|\hat{H}_p\middle|\psi_n^{(0)}\right\rangle\\&=\frac{2}{L}\int_0^L dx\sin^2\left(\frac{n\pi}{L}x\right)\left(-q\mathcal{E}x\right)\\&=-\frac{q\mathcal{E}L}{2}.\end{aligned} \tag{14.24}$$

Since this is a constant, all of the energy levels are shifted by the same amount. To first order, the total energies are

$$E_n=\frac{n^2\pi^2\hbar^2}{2mL^2}-\frac{q\mathcal{E}L}{2}\qquad n=1,2,3,\ldots. \tag{14.25}$$

Given the form of the potential shown in fig. 14.1, it's not surprising that the energies decrease for a positive charge.

14.1.2 Second-Order Corrections

To obtain the second-order corrections to the energy, we project $\left\langle\psi_n^{(0)}\right|$ onto eq. (14.10), yielding

$$E_n^{(0)}\left\langle\psi_n^{(0)}\middle|\psi_n^{(2)}\right\rangle+\left\langle\psi_n^{(0)}\middle|\hat{H}_p\middle|\psi_n^{(1)}\right\rangle=E_n^{(0)}\left\langle\psi_n^{(0)}\middle|\psi_n^{(2)}\right\rangle+E_n^{(1)}\left\langle\psi_n^{(0)}\middle|\psi_n^{(1)}\right\rangle+E_n^{(2)}. \tag{14.26}$$

Canceling like terms, using $\left\langle\psi_n^{(0)}\middle|\psi_n^{(1)}\right\rangle=0$, and solving for $E_n^{(2)}$ leaves us with

$$E_n^{(2)} = \left\langle \psi_n^{(0)} \middle| \hat{H}_p \middle| \psi_n^{(1)} \right\rangle$$

$$= \sum_{m \neq n} \left\langle \psi_n^{(0)} \middle| \hat{H}_p \middle| \psi_m^{(0)} \right\rangle \frac{\left\langle \psi_m^{(0)} \middle| \hat{H}_p \middle| \psi_n^{(0)} \right\rangle}{E_n^{(0)} - E_m^{(0)}} \tag{14.27}$$

$$= \sum_{m \neq n} \frac{\left| \left\langle \psi_m^{(0)} \middle| \hat{H}_p \middle| \psi_n^{(0)} \right\rangle \right|^2}{E_n^{(0)} - E_m^{(0)}},$$

where we have used the first-order corrections to the states, eq. (14.19). In principle we could continue on and obtain the second-order corrections to the states, but in practice this is not often useful.[1]

EXAMPLE 14.2

A uniform electric field $\boldsymbol{\mathcal{E}} = \mathcal{E}\boldsymbol{u}_x$ is applied to a particle with a charge of q that is situated in a harmonic oscillator potential. Determine the energies of the particle to second order, and the energy eigenstates to first order.

Before doing any calculations, let's think about the physics of the problem. The potential energy with the applied field is given by

$$V(x) = \frac{1}{2} m\omega^2 x^2 - q\mathcal{E}x. \tag{14.28}$$

The unperturbed and the perturbed potential energies are plotted in fig. 14.2. The shape of the potential is not changed, but the perturbation shifts the minimum of the potential in space by $q\mathcal{E}/m\omega^2$, and shifts it in energy by $-q^2\mathcal{E}^2/2m\omega^2$. Thus, we anticipate that the energies will be shifted by $-q^2\mathcal{E}^2/2m\omega^2$, and the wave functions will be shifted in space by $q\mathcal{E}/m\omega^2$. (An exact solution bears this out.)

For the perturbation solution, the perturbing Hamiltonian is $\hat{H}_p(x) = -q\mathcal{E}\,\hat{x}$. The eigenstates of the unperturbed Hamiltonian are the Fock states $|n\rangle$, and their energies are [eq. (12.24)]

$$E_n^{(0)} = \hbar\omega\left(n + \frac{1}{2}\right) \quad n = 0,1,2,.... \tag{14.29}$$

The matrix elements of the perturbing Hamiltonian are

$$\begin{aligned} \left\langle \psi_j^{(0)} \middle| \hat{H}_p \middle| \psi_n^{(0)} \right\rangle &= -q\mathcal{E} \langle j|\hat{x}|n\rangle \\ &= -q\mathcal{E}\sqrt{\frac{\hbar}{2m\omega}} \langle j|\left(\hat{a} + \hat{a}^\dagger\right)|n\rangle \\ &= -q\mathcal{E}\sqrt{\frac{\hbar}{2m\omega}} \left(\sqrt{n}\langle j|n-1\rangle + \sqrt{n+1}\langle j|n+1\rangle\right) \\ &= -q\mathcal{E}\sqrt{\frac{\hbar}{2m\omega}} \left(\sqrt{n}\,\delta_{j\,n-1} + \sqrt{n+1}\,\delta_{j\,n+1}\right), \end{aligned} \tag{14.30}$$

---

[1] We said above that for perturbation theory to work, the perturbation must be small, but how small is small? From eqs. (14.19) and (14.27) it's apparent that the matrix elements of the perturbing Hamiltonian must be smaller than the energy differences of the corresponding states.

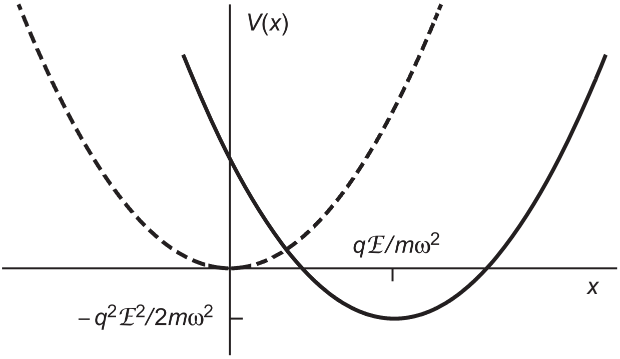


Fig 14.2 When a uniform electric field is applied to a charged particle in a harmonic oscillator potential, the potential minimum shifts.

where we've used eqs. (12.28)–(12.30). The first-order corrections to the energies are given by the diagonal matrix elements,

$$E_n^{(0)} = \left\langle \psi_n^{(0)} \middle| \hat{H}_p \middle| \psi_n^{(0)} \right\rangle = 0. \tag{14.31}$$

Thus, to first-order, there is no change in the energies. The second-order correction to the energies is given by eq. (14.27):

$$\begin{aligned} E_n^{(2)} &= \sum_{j \neq n} \frac{\left| \left\langle \psi_j^{(0)} \middle| \hat{H}_p \middle| \psi_n^{(0)} \right\rangle \right|^2}{E_n^{(0)} - E_j^{(0)}} \\ &= \frac{q^2 \mathcal{E}^2 \hbar}{2m\omega} \sum_{j \neq n} \frac{\left(\sqrt{n}\, \delta_{j\,n-1} + \sqrt{n+1}\, \delta_{j\,n+1} \right)^2}{E_n^{(0)} - E_j^{(0)}} \\ &= \frac{q^2 \mathcal{E}^2 \hbar}{2m\omega} \left(\frac{n}{E_n^{(0)} - E_{n-1}^{(0)}} + \frac{n+1}{E_n^{(0)} - E_{n+1}^{(0)}} \right) \\ &= \frac{q^2 \mathcal{E}^2 \hbar}{2m\omega} \left(\frac{n}{\hbar\omega} - \frac{n+1}{\hbar\omega} \right) \\ &= -\frac{q^2 \mathcal{E}^2}{2m\omega^2} . \end{aligned} \tag{14.32}$$

We see that the second-order correction to the energy yields the anticipated energy shift.

The first-order correction to the states is given by eq. (14.19):

$$\begin{aligned}\left|\psi_n^{(1)}\right\rangle &= \sum_{j\neq n} \frac{\left\langle\psi_j^{(0)}\right|\hat{H}_p\left|\psi_n^{(0)}\right\rangle}{E_n^{(0)}-E_j^{(0)}}\left|\psi_j^{(0)}\right\rangle \\ &= -q\mathcal{E}\sqrt{\frac{\hbar}{2m\omega}}\sum_{j\neq n}\frac{\left(\sqrt{n}\,\delta_{j\,n-1}+\sqrt{n+1}\,\delta_{j\,n+1}\right)}{E_n^{(0)}-E_j^{(0)}}\left|\psi_j^{(0)}\right\rangle \\ &= -q\mathcal{E}\sqrt{\frac{\hbar}{2m\omega}}\left(\frac{\sqrt{n}}{E_n^{(0)}-E_{n-1}^{(0)}}\left|\psi_{n-1}^{(0)}\right\rangle+\frac{\sqrt{n+1}}{E_n^{(0)}-E_{n+1}^{(0)}}\left|\psi_{n+1}^{(0)}\right\rangle\right) \\ &= -q\mathcal{E}\sqrt{\frac{1}{2m\hbar\omega^3}}\left(\sqrt{n}\left|n-1\right\rangle-\sqrt{n+1}\left|n+1\right\rangle\right).\end{aligned} \tag{14.33}$$

To first order, the perturbation couples states of adjacent energies. It's not obvious that this correction is consistent with a shift in the position of the wave functions, but you'll show that this is indeed the case in problem 14.6.

14.2 DEGENERATE THEORY

In the previous section we assumed nondegenerate energies. However, by examining eqs. (14.19) and (14.27) we see that degenerate energies would lead to 0's in the denominator, which could be a problem. We need to go back and reexamine what we've done, in order to allow for degeneracy. As we'll see below, perturbations tend to split energy levels; states that were degenerate no longer are. In other words, the energy corrections are not unique. States of the same unperturbed energies will not in general have the same energy corrections.

Assume that the energy $E_n^{(0)}$ is N-fold degenerate, and the corresponding eigenstates are $\left|\psi_{n,j}^{(0)}\right\rangle$, where $j = 1,\dots,N$. Degenerate eigenstates all share the same energy, which means that any linear combination of degenerate eigenstates will also have the same energy. Thus, for a particular linear combination it is the case that

$$\hat{H}_0\sum_{j=1}^{N} b_{n,j}\left|\psi_{n,j}^{(0)}\right\rangle = E_n^{(0)}\sum_{j=1}^{N} b_{n,j}\left|\psi_{n,j}^{(0)}\right\rangle. \tag{14.34}$$

The linear combination is specified by the coefficients $b_{n,j}$, which form an N-dimensional vector $\boldsymbol{b}_n$. Comparing eq. (14.34) to eq. (14.2), we see that we have made the substitution

$$\left|\psi_n^{(0)}\right\rangle \rightarrow \sum_{j=1}^{N} b_{n,j}\left|\psi_{n,j}^{(0)}\right\rangle. \tag{14.35}$$

Applying this substitution to eq. (14.9), and then operating on the resulting equation with $\left\langle\psi_{n,i}^{(0)}\right|$, yields

$$\begin{aligned}&\left\langle\psi_{n,i}^{(0)}\right|\hat{H}_0\left|\psi_n^{(1)}\right\rangle+\left\langle\psi_{n,i}^{(0)}\right|\hat{H}_p\sum_{j=1}^{N} b_{n,j}\left|\psi_{n,j}^{(0)}\right\rangle \\ &\quad=\left\langle\psi_{n,i}^{(0)}\right|E_n^{(0)}\left|\psi_n^{(1)}\right\rangle+\left\langle\psi_{n,i}^{(0)}\right|E_n^{(1)}\sum_{j=1}^{N} b_{n,j}\left|\psi_{n,j}^{(0)}\right\rangle.\end{aligned} \tag{14.36}$$

The first terms cancel, and we're left with

$$\sum_{j=1}^{N} b_{n,j} \left\langle \psi_{n,i}^{(0)} \middle| \hat{H}_p \middle| \psi_{n,j}^{(0)} \right\rangle = E_n^{(1)} b_{n,i} \; , \tag{14.37}$$

where we've used the orthogonality of the eigenstates. Let's define the matrix elements of $\hat{H}_p$, in the N-dimensional subspace of the degenerate states, to be

$$H_{p\,ij} \equiv \left\langle \psi_{n,i}^{(0)} \middle| \hat{H}_p \middle| \psi_{n,j}^{(0)} \right\rangle. \tag{14.38}$$

Using this, we can rewrite eq. (14.37) as

$$\sum_{j=1}^{N} H_{p\,ij} b_{n,j} = E_n^{(1)} b_{n,i}. \tag{14.39}$$

Comparing this to eq. (1.41), we find that the sum is equivalent to a matrix multiplying a vector, so we can rewrite it as

$$\bar{\boldsymbol{H}}_p \boldsymbol{b}_n = E_n^{(1)} \boldsymbol{b}_n. \tag{14.40}$$

Equation (14.40) is an eigenvalue equation within the subspace of degenerate states, and the eigenvalues are the first-order corrections to the energy. Thus, the procedure for finding the energy corrections and proper eigenstates for perturbations of an energy level with an N-fold degeneracy are as follows:

1. Create the $N \times N$-matrix $\bar{\boldsymbol{H}}_p$, whose matrix elements are found using the degenerate eigenstates $\left| \psi_{n,j}^{(0)} \right\rangle$ [Eq. (14.38)].
2. The eigenvalues of $\bar{\boldsymbol{H}}_p$ yield the first-order corrections to the energy $E_n^{(1)}$.
3. The corresponding eigenvectors $\boldsymbol{b}_n$ tell us which linear combinations of the degenerate, unperturbed eigenstates correspond to which first-order energy corrections.

This may sound confusing, but it's actually straightforward, as the following example will show.

EXAMPLE 14.3

A particle in a cubic box (example 13.1) with sides of length L is subjected to a perturbing potential of the form $H_p(\boldsymbol{r}) = \gamma xy$, where γ is a positive constant. This corresponds to a force which tends to push the particle toward the z-axis. To first order, calculate the effects of this perturbation on the ground state and first-excited state energies.

The unperturbed states for the 3-D box are $\left| n_x, n_y, n_z \right\rangle$. The energies are [eq. (13.17)]

$$E_{n_x n_y n_z}^{(0)} = \frac{\pi^2 \hbar^2}{2mL^2}\left(n_x^2 + n_y^2 + n_z^2\right) \quad n_x, n_y, n_z = 1, 2, 3, \ldots, \tag{14.41}$$

and inside the box the wave functions are [eq. (13.18)]

$$\begin{aligned}\psi^{(0)}_{n_x n_y n_z}(\boldsymbol{r}) &= X_{n_x}(x)Y_{n_y}(y)Z_{n_z}(z)\\ &= \sqrt{\frac{8}{L^3}}\sin\left(\frac{n_x\pi}{L}x\right)\sin\left(\frac{n_y\pi}{L}y\right)\sin\left(\frac{n_z\pi}{L}z\right).\end{aligned} \tag{14.42}$$

The wave functions are 0 outside the box. The ground state $|1,1,1\rangle$ is nondegenerate, so we can use eq. (14.12) to calculate its first-order energy correction:

$$\begin{aligned}E^{(1)}_{111} &= \langle 1,1,1|\hat{H}_p|1,1,1\rangle\\ &= \gamma\int_0^L\int_0^L\int_0^L dxdydzX_1(x)Y_1(y)Z_1(z)xyX_1(x)Y_1(y)Z_1(z)\\ &= \gamma\frac{4}{L^2}\int_0^L dx\sin^2\left(\frac{\pi}{L}x\right)x\int_0^L dy\sin^2\left(\frac{\pi}{L}y\right)y\\ &= \gamma\frac{L^2}{4}.\end{aligned} \tag{14.43}$$

The integral over z is 1, since it is just the normalization integral for $Z_1(z)$. We see that the energy of the ground state is increased.

The first excited state is 3-fold degenerate, so to find the first-order energy corrections, and corresponding eigenstates, we need to diagonalize the matrix

$$\bar{\boldsymbol{H}}_p = \begin{pmatrix}\langle 2,1,1|\hat{H}_p|2,1,1\rangle & \langle 2,1,1|\hat{H}_p|1,2,1\rangle & \langle 2,1,1|\hat{H}_p|1,1,2\rangle\\ \langle 1,2,1|\hat{H}_p|2,1,1\rangle & \langle 1,2,1|\hat{H}_p|1,2,1\rangle & \langle 1,2,1|\hat{H}_p|1,1,2\rangle\\ \langle 1,1,2|\hat{H}_p|2,1,1\rangle & \langle 1,1,2|\hat{H}_p|1,2,1\rangle & \langle 1,1,2|\hat{H}_p|1,1,2\rangle\end{pmatrix}. \tag{14.44}$$

The matrix elements are given by

$$\begin{aligned}&\left\langle n'_x,n'_y,n'_z\middle|\hat{H}_p\middle|n_x,n_y,n_z\right\rangle\\ &= \gamma\int_0^L\int_0^L\int_0^L dxdydzX_{n'_x}(x)Y_{n'_y}(y)Z_{n'_z}(z)xyX_{n_x}(x)Y_{n_y}(y)Z_{n_z}(z)\\ &= \gamma\int_0^L dxX_{n'_x}(x)xX_{n_x}(x)\int_0^L dyY_{n'_y}(y)yY_{n_y}(y)\delta_{n'_z n_z}.\end{aligned} \tag{14.45}$$

These matrix elements are 0 if $n_z \neq n'_z$, so the matrix simplifies to

$$\bar{\boldsymbol{H}}_p = \begin{pmatrix}\langle 2,1,1|\hat{H}_p|2,1,1\rangle & \langle 2,1,1|\hat{H}_p|1,2,1\rangle & 0\\ \langle 1,2,1|\hat{H}_p|2,1,1\rangle & \langle 1,2,1|\hat{H}_p|1,2,1\rangle & 0\\ 0 & 0 & \langle 1,1,2|\hat{H}_p|1,1,2\rangle\end{pmatrix}. \tag{14.46}$$

We can immediately see that the state $|1,1,2\rangle$ is an eigenstate of this matrix, with eigenvalue $\langle 1,1,2|\hat{H}_p|1,1,2\rangle$. The integral for $\langle 1,1,2|\hat{H}_p|1,1,2\rangle$ takes on exactly the same form as the integral for $\langle 1,1,1|\hat{H}_p|1,1,1\rangle$ [Eq. (14.43)], so the first-order energy correction for the state $|1,1,2\rangle$ is

$$E_{112}^{(1)} = \langle 1,1,2|\hat{H}_p|1,1,2\rangle = \gamma\frac{L^2}{4}. \tag{14.47}$$

Our problem is now reduced to diagonalizing the matrix

$$\bar{\boldsymbol{H}}_p = \begin{pmatrix} \langle 2,1,1|\hat{H}_p|2,1,1\rangle & \langle 2,1,1|\hat{H}_p|1,2,1\rangle \\ \langle 1,2,1|\hat{H}_p|2,1,1\rangle & \langle 1,2,1|\hat{H}_p|1,2,1\rangle \end{pmatrix}. \tag{14.48}$$

In eq. (14.45) for the matrix elements, the x- and y-integrals have the same form, which is

$$\int_0^L dx X_{n'_x}(x)\, x X_{n_x}(x) = \frac{2}{L}\int_0^L dx \sin\left(\frac{n'_x\pi}{L}x\right) x \sin\left(\frac{n_x\pi}{L}x\right) = \begin{cases} \dfrac{L}{2} & n_x = n'_x \\ \left(-\dfrac{4Ln_x n'_x\left[1+(-1)^{1+n_x+n'_x}\right]}{\pi^2\left(n_x^2-n_x'^2\right)^2}\right) & n_x \neq n'_x\,. \end{cases} \tag{14.49}$$

Using this, we find that the diagonal matrix elements are

$$\langle 2,1,1|\hat{H}_p|2,1,1\rangle = \langle 1,2,1|\hat{H}_p|1,2,1\rangle = \gamma\frac{L^2}{4}, \tag{14.50}$$

and the off-diagonal elements are

$$\langle 2,1,1|\hat{H}_p|1,2,1\rangle = \langle 1,2,1|\hat{H}_p|2,1,1\rangle = \gamma\left(-\frac{16L}{9\pi^2}\right)^2 = \frac{256\gamma L^2}{81\pi^4}. \tag{14.51}$$

The perturbation matrix is thus

$$\bar{\boldsymbol{H}}_p = \gamma\frac{L^2}{4}\begin{pmatrix} 1 & \dfrac{1024}{81\pi^4} \\ \dfrac{1024}{81\pi^4} & 1 \end{pmatrix}. \tag{14.52}$$

The final two first-order energy corrections are the eigenvalues of this matrix, which you can show to be

$$E_{\pm}^{(1)} = \gamma\frac{L^2}{4}\left(1 \pm \frac{1024}{81\pi^4}\right). \tag{14.53}$$

The corresponding eigenstates are

$$\left|\psi_{\pm}^{(1)}\right\rangle = \frac{1}{\sqrt{2}}\left(|2,1,1\rangle \pm |1,2,1\rangle\right). \tag{14.54}$$

Figure 14.3 shows the resulting energies, and we see that the perturbation splits the first-excited state into three distinct energy levels.

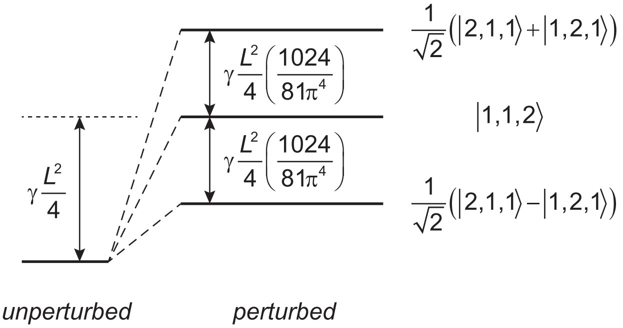


Fig 14.3 Splitting of the three-fold degenerate, first-excited state of a particle in a box (not to scale). The states corresponding to the perturbed energy levels are indicated.

Equation (14.40) asks us to find the eigenvalues and eigenstates of the perturbing Hamiltonian $\hat{H}_p$, in the subspace of fixed n (the states $\left|\psi_{n,j}^{(0)}\right\rangle$ form a basis for this subspace). Our job is simplified greatly if we already know what the eigenstates in this subspace are. Under what circumstances might we know them? Suppose that $\hat{A}$ corresponds to an observable, and it commutes with $\hat{H}_0$ and $\hat{H}_p$. Furthermore, suppose that that the states $\left|\psi_{n,j}^{(0)}\right\rangle$ are nondegenerate eigenstates of $\hat{A}$. In problem 14.22 you'll prove that these states are the correct eigenstates of $\hat{H}_p$ to use in perturbation theory [they are the solutions to Eq. (14.40)]. We can calculate the first-order energy corrections using nondegenerate perturbation theory [eq. (14.12)], as long as we use these eigenstates. We'll soon see that this simplification (looking for an observable whose corresponding operator commutes with $\hat{H}_0$ and $\hat{H}_p$, and whose eigenstates are nondegenerate in the subspace of fixed n) is very useful.[2]

14.3 FINE STRUCTURE OF HYDROGEN

In chapter 13 we assumed that the potential energy of the interaction between the proton and the electron in a hydrogen atom was described completely by the Coulomb potential. In reality the potential has other terms, but they are small in comparison, so we can treat them as perturbations. In this section we will examine some of these corrections to the energy of a hydrogen atom. We'll look at three perturbations, all of which are relativistic in nature, and all of which have energy corrections on the same scale. Taken together, these corrections constitute what is known as the fine structure of hydrogen.

---

[2] If a single observable is insufficient to find nondegenerate eigenstates, one can look for multiple observables whose simultaneous eigenstates are nondegenerate.

14.3.1 Relativistic Kinetic Energy

The relativistic kinetic energy of a particle is given by

$$K = \sqrt{m^2c^4 + p^2c^2} - mc^2 = mc^2\left(1 + \frac{p^2c^2}{m^2c^4}\right)^{1/2} - mc^2. \tag{14.55}$$

If the kinetic energy is much less than the rest energy, the particle is nonrelativistic, and we can use the binomial expansion to approximate the kinetic energy as

$$K \simeq \frac{p^2}{2m} - \frac{p^4}{8m^3c^2} + \dots. \tag{14.56}$$

The first term is the Newtonian kinetic energy that we use in our unperturbed Hamiltonian, $\hat{H}_0$. The second term is the relativistic correction, which is the perturbing Hamiltonian:

$$\hat{H}_R = -\frac{\hat{p}^4}{8m^3c^2} = -\frac{1}{2mc^2}\left(\frac{\hat{p}^2}{2m}\right)^2. \tag{14.57}$$

Here we've written the perturbing Hamiltonian in terms of the nonrelativistic kinetic energy, for reasons that will soon become apparent. For hydrogen it is the electron that is moving, so the appropriate mass is the electron mass: $m = m_e$.

Calculating the first-order energy corrections involves computing expectation values of $\hat{H}_R$, using the unperturbed states of the hydrogen atom $\left|n,l,m_l\right\rangle$. Apart from the ground state the energies of these states are degenerate, so we need to use degenerate perturbation theory. However, here's where the simplification described at the end of sec. 14.2 applies, and we can avoid diagonalizing any matrices.

We know that $\hat{L}^2$ and $\hat{L}_z$ commute with $\hat{H}_0$. As part of problem 13.14 you showed that $\left[\hat{p}^2, \hat{\boldsymbol{L}}\right] = 0$; from this it follows that $\left[\hat{p}^4, \hat{\boldsymbol{L}}\right] = 0$, and hence $\hat{H}_R$ commutes with $\hat{L}^2$ and $\hat{L}_z$. Thus, as described at the end of sec. 14.2, we've found operators, $\hat{L}^2$ and $\hat{L}_z$, that commute with $\hat{H}_0$ and $\hat{H}_R$. Within the subspace of fixed n their eigenstates $\left|n,l,m_l\right\rangle$ are nondegenerate, so there's no need to diagonalize the perturbation Hamiltonian. We can use nondegenerate perturbation theory to calculate the first-order energy corrections [eq. (14.12)], as long as we calculate the expectation values of $\hat{H}_R$ using the states $\left|n,l,m_l\right\rangle$.

Note from eq. (14.57) that we need to calculate expectation values of $\hat{p}^4$, which turns out to be difficult. However, there's a trick that makes this task easier. Recall the original, unperturbed, Hamiltonian for a three-dimensional system [eq. (13.1)]; rearranging that equation, we find that

$$\frac{\hat{p}^2}{2m} = \left(\hat{H}_0 - V\left(\hat{\boldsymbol{r}}\right)\right). \tag{14.58}$$

Substituting this into eq. (14.57), we can rewrite the perturbing Hamiltonian as

$$\hat{H}_R = -\frac{1}{2mc^2}\left(\hat{H}_0^2 - \hat{H}_0 V\left(\hat{\boldsymbol{r}}\right) - V\left(\hat{\boldsymbol{r}}\right)\hat{H}_0 + V^2\left(\hat{\boldsymbol{r}}\right)\right). \tag{14.59}$$

Here $V(\hat{\mathbf{r}})$ is given by the Coulomb potential [eq. (13.71)], so the first-order, relativistic kinetic energy corrections are then

$$\begin{aligned} E_R^{(1)} &= \langle n,l,m_l | \hat{H}_R | n,l,m_l \rangle \\ &= -\frac{1}{2mc^2}\left[\left(E_n^{(0)}\right)^2 + 2E_n^{(0)}\frac{e^2}{4\pi\varepsilon_0}\langle n,l,m_l | \left(\frac{1}{\hat{r}}\right) | n,l,m_l \rangle \right. \\ &\quad \left. + \left(\frac{e^2}{4\pi\varepsilon_0}\right)^2 \langle n,l,m_l | \left(\frac{1}{\hat{r}^2}\right) | n,l,m_l \rangle \right]. \end{aligned} \tag{14.60}$$

There are a few different techniques for calculating the expectation values that we need. The virial theorem can be used to calculate $\langle \hat{r}^{-1} \rangle$ [problem 14.8], and $\langle \hat{r}^{-1} \rangle$ and $\langle \hat{r}^{-2} \rangle$ can be calculated directly by integration [problem 14.9]. Finally, there are some tricks that use the fact that the radial part of the unperturbed Hamiltonian contains both r^{-1} and r^{-2} terms [see eq. (13.72)].[3] By whichever means they are calculated, these expectation values are

$$\left\langle \frac{1}{\hat{r}} \right\rangle = \frac{1}{n^2 a_0}, \tag{14.61}$$

and

$$\left\langle \frac{1}{\hat{r}^2} \right\rangle = \frac{1}{n^3 a_0^2 (l+1/2)}, \tag{14.62}$$

where a_0 is the Bohr radius. Substituting these expressions into eq. (14.60), and doing some algebra, we find the first-order energy corrections to be

$$E_R^{(1)} = -\frac{\left(E_n^{(0)}\right)^2}{2mc^2}\left(\frac{4n}{(l+1/2)} - 3\right). \tag{14.63}$$

This energy correction depends on the orbital angular momentum quantum number l, so the relativistic correction at least partially breaks the degeneracy of the energy level structure of hydrogen.

The fine-structure constant α is defined as

$$\alpha \equiv \frac{e^2}{4\pi\varepsilon_0 \hbar c} \simeq \frac{1}{137}. \tag{14.64}$$

In terms of α, the unperturbed hydrogen energies can be expressed as

$$E_n^{(0)} = -\frac{mc^2\alpha^2}{2n^2}, \tag{14.65}$$

and the relativistic corrections are

---

[3] See sec. 17.3 of ref. [14.1] for more details.

$$E_R^{(1)} = -\frac{mc^2\alpha^4}{8n^4}\left(\frac{4n}{(l+1/2)} - 3\right). \tag{14.66}$$

Comparing these last two equations, we see that the relativistic corrections are smaller than the unperturbed energies by a factor of approximately α^2, validating our use of perturbation theory.

14.3.2 The Spin-Orbit Interaction

Previously we have ignored the spin of the electron in the calculation of the hydrogen energy spectrum, but now it's time to include it. Because of its spin, the electron has an intrinsic magnetic dipole moment $\boldsymbol{\mu}_e = \gamma_e \boldsymbol{S}$ [eq. (6.4)], where γ_e is the gyromagnetic ratio of the electron. The gyromagnetic ratio can be written as (problem 6.2)

$$\gamma_e = \frac{g_e q_e}{2m} = -\frac{g_e e}{2m}, \tag{14.67}$$

where g_e is the g factor of the electron.

In its own rest frame, the electron sees the proton moving about it. This moving charge creates a magnetic field in the rest frame of the electron, which interacts with the electron's magnetic moment. To calculate the strength of this field, treat the proton as moving in a circle of radius r. We know that a current loop produces a magnetic field at its center with a magnitude of

$$\mathcal{B} = \frac{\mu_0 I}{2r} = \frac{I}{\varepsilon_0 c^2 2r}, \tag{14.68}$$

where μ_0 is the permeability of free space ($\mu_0\varepsilon_0 = 1/c^2$). The effective current of a proton with an orbital period of T is $I = e/T$, so the magnitude of the field strength is

$$\mathcal{B} = \frac{e}{\varepsilon_0 c^2 2rT} = \frac{e}{\varepsilon_0 c^2 2r(2\pi r/v)} = \frac{ev}{4\pi\varepsilon_0 c^2 r^2}, \tag{14.69}$$

with v being the speed of the proton.

Switching to the rest frame of the proton, the proton and the electron must agree on their relative speeds, so we know that v is also the speed of the electron about the proton. The speed is related to the magnitude of the angular momentum of the electron by $v = L/mr$. Finally, a little thought shows that the magnetic field experienced by the electron is parallel to the angular momentum of the electron (if the proton moves clockwise about the electron, the electron also moves clockwise about the proton), so we have

$$\boldsymbol{\mathcal{B}} = \frac{e}{4\pi\varepsilon_0 mc^2 r^3}\boldsymbol{L}. \tag{14.70}$$

The potential energy of interaction between the field and the dipole moment of the electron is then given by

$$V = -\boldsymbol{\mu}\cdot\boldsymbol{\mathcal{B}} = \frac{g_e e^2}{8\pi\varepsilon_0 m^2 c^2 r^3}\boldsymbol{S}\cdot\boldsymbol{L}. \tag{14.71}$$

This simple calculation is nearly in agreement with the full relativistic calculation. The problem is that the electron is in an accelerated reference frame, which causes a precession of its magnetic moment, known as the Thomas precession. The correct result subtracts 1 from g_e (see ref. [14.2]), so the Hamiltonian turns out to be

$$\hat{H}_{SO} = \frac{(g_e - 1)e^2}{8\pi\varepsilon_0 m^2 c^2 r^3}\hat{\boldsymbol{S}}\boldsymbol{\cdot}\hat{\boldsymbol{L}} = \frac{e^2}{8\pi\varepsilon_0 m^2 c^2 r^3}\hat{\boldsymbol{S}}\boldsymbol{\cdot}\hat{\boldsymbol{L}}, \tag{14.72}$$

where we've used the fact that $g_e \cong 2$.[4] The subscript on $\hat{H}_{SO}$ refers to the fact that it is known as the spin-orbit Hamiltonian, because it couples the spin and orbital angular momenta of the electron.

The total angular momentum is given by $\hat{\boldsymbol{J}} = \hat{\boldsymbol{S}} + \hat{\boldsymbol{L}}$ (sec. 7.2), and since $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{L}}$ commute, we can write

$$\hat{J}^2 = \hat{\boldsymbol{J}}\boldsymbol{\cdot}\hat{\boldsymbol{J}} = \left(\hat{\boldsymbol{S}} + \hat{\boldsymbol{L}}\right)\boldsymbol{\cdot}\left(\hat{\boldsymbol{S}} + \hat{\boldsymbol{L}}\right) = \hat{S}^2 + \hat{L}^2 + 2\left(\hat{\boldsymbol{S}}\boldsymbol{\cdot}\hat{\boldsymbol{L}}\right). \tag{14.73}$$

Solving for the dot product, and substituting into eq. (14.72), we find that the spin-orbit Hamiltonian can be written as

$$\hat{H}_{SO} = \frac{e^2}{16\pi\varepsilon_0 m^2 c^2 r^3}\left(\hat{J}^2 - \hat{L}^2 - \hat{S}^2\right). \tag{14.74}$$

The operators $\hat{J}^2$, $\hat{L}^2$ and $\hat{S}^2$ all commute, and eq. (14.74) tells us that their simultaneous eigenstates are also eigenstates of $\hat{H}_{SO}$. Another operator that commutes with $\hat{H}_0$ and $\hat{H}_{SO}$ is $\hat{J}_z$ (problem 14.11), and $\hat{J}_z$ also commutes with $\hat{J}^2$, $\hat{L}^2$ and $\hat{S}^2$. Within a subspace of fixed *n*, the simultaneous eigenstates of the operators $\hat{J}^2$, $\hat{L}^2$, $\hat{S}^2$ and $\hat{J}_z$ are nondegenerate. Since these operators all commute with $\hat{H}_{SO}$, their eigenstates $\left|j,l,s,m_j\right\rangle$ (for fixed *n*) are also eigenstates of $\hat{H}_{SO}$; we can use these states to calculate the first-order energy corrections for the spin-orbit interaction, without resorting to matrix diagonalization. For fixed values of *l* and *s*, the allowed values for *j* are

$$j = l+s, l+s-1, \ldots, |l-s|. \tag{14.75}$$

Since $s = 1/2$, the allowed values for *j* are $j = l + 1/2$ and $j = l - 1/2$; this second value is not allowed if $l = 0$. As before, the allowed values for m_j are

$$m_j = -j, -j+1, -j+2, \ldots, j-2, j-1, j. \tag{14.76}$$

Before calculating the energy corrections, a few comments about angular momentum are in order. $\hat{L}_z$ and $\hat{S}_z$ commute with each other, and $\hat{H}_0$, so in the absence of the spin-orbit interaction we can sensibly talk about definite values for the *z*-components of both orbital and spin angular momentum. However, $\hat{L}_z$ and $\hat{S}_z$ do not commute with $\hat{H}_{SO}$, so in the presence of the spin-orbit interaction it does not make sense to talk about definite values for their expectation values. The *z*-components of orbital and spin angular

---

[4] Relativistic quantum mechanics predicts $g_e = 2$, but quantum field theory adds corrections (on the order of 0.1%) to this.

momentum are no longer separately conserved, because they do not commute with the full Hamiltonian (see sec. 9.3). The z-component of the total angular momentum, J_z, is conserved, because $\hat{J}_z$ commutes with $\hat{H}_0$ and $\hat{H}_{SO}$.

To find the first-order energy corrections, we calculate the expectation value of eq. (14.74) using the states $\left|j,l,s,m_j\right\rangle$, in the subspace of fixed n. The result is

$$E_{SO}^{(1)}=\frac{e^2\hbar^2}{16\pi\varepsilon_0 m^2c^2}\left[j(j+1)-l(l+1)-s(s+1)\right]\left\langle\frac{1}{r^3}\right\rangle. \tag{14.77}$$

In problem 14.12 you'll show that

$$\left\langle\frac{1}{r^3}\right\rangle=\frac{1}{a_0^3n^3l(l+1)(l+1/2)}. \tag{14.78}$$

Combining the last two equations, and doing some algebra, yields

$$E_{SO}^{(1)}=\frac{mc^2\alpha^4}{4}\left[\frac{j(j+1)-l(l+1)-3/4}{n^3l(l+1)(l+1/2)}\right], \tag{14.79}$$

where we have used the fact that for an electron $s=1/2$.

There turns out to be a problem with eq. (14.79), however, for the case of $l=0$. From the Hamiltonian of eq. (14.72), it is apparent that if the orbital angular momentum is 0, there should be no spin-orbit coupling, and we should have $E_{SO}^{(1)}=0$. However, you'll show in problem 14.13 that eq. (14.79) predicts a nonzero $E_{SO}^{(1)}$ for $l=0$. What's going on? If $l=0$, then we must have $j=s$, so eq. (14.77) predicts $E_{SO}^{(1)}=0$, as long as $\left\langle r^{-3}\right\rangle$ is finite. However, eq. (14.78) says that $\left\langle r^{-3}\right\rangle$ diverges for $l=0$. Luckily, the Dirac equation for the hydrogen atom can be solved exactly [see eq. (14.88) below], which gives us the exact relativistic energies. The low-energy limit of this solution tells us that the correct answer is $E_{SO}^{(1)}=0$ for $l=0$. Our final answer for the first-order energy corrections, due to the spin-orbit interaction, is thus

$$E_{SO}^{(1)}=\begin{cases}0 & l=0\\ \dfrac{mc^2\alpha^4}{4}\left[\dfrac{j(j+1)-l(l+1)-3/4}{n^3l(l+1)(l+1/2)}\right] & l\neq 0\end{cases}. \tag{14.80}$$

14.3.3 The Darwin Term

In relativistic quantum mechanics a particle cannot be localized to a length scale smaller than its Compton wavelength, $\lambda\!\!\!^{-}_c=\hbar/mc$. Accounting for this, the Coulomb potential energy of the electron is not just $V(\boldsymbol{r})$, but has contributions from points within a volume of approximately $\lambda\!\!\!^{-}_c{}^3$ about $\boldsymbol{r}$. To see what effect this has, expand $V(\boldsymbol{r}+\boldsymbol{r}')$ (the potential at $\boldsymbol{r}+\boldsymbol{r}'$, which is near to $\boldsymbol{r}$) in a Taylor series about $\boldsymbol{r}$. Letting $x=r_1$, etc., we find

$$V(\boldsymbol{r}+\boldsymbol{r}')=V(\boldsymbol{r})+\sum_{i=1}^{3}\left(\frac{\partial V}{\partial r_i}\right)r_i'+\frac{1}{2}\sum_{i,j=1}^{3}\left(\frac{\partial^2 V}{\partial r_i\partial r_j}\right)r_i'r_j'+\ldots. \tag{14.81}$$

Note that the derivatives in parentheses must be evaluated at $\boldsymbol{r}$, and are thus constants.

If we average $V(\boldsymbol{r}+\boldsymbol{r}')$ over a sphere of radius $\lambda\!\!\!^-_c$, centered on $\boldsymbol{r}$, we find (see problem 14.14)

$$\begin{aligned}\langle V(\boldsymbol{r}+\boldsymbol{r}')\rangle &= V(\boldsymbol{r})+\frac{1}{2}\left(\frac{1}{5}\lambda\!\!\!^-{}_c^2\right)\nabla^2 V(\boldsymbol{r})+\ldots \\ &= V(\boldsymbol{r})+\frac{\hbar^2}{10m^2c^2}\nabla^2 V(\boldsymbol{r})+\ldots.\end{aligned} \tag{14.82}$$

The first term is just the Coulomb potential energy, and the second is the correction due to the unlocalized nature of the electron.

Here we have motivated the existence of a correction which accounts for the electron being unlocalized over length scales of its Compton wavelength. To get the correct answer, however, one must perform an expansion of the Dirac equation (see ref. [14.1], sec. 20.2). This results in a correction to the Coulomb Hamiltonian known as the Darwin term. The Darwin Hamiltonian is

$$H_D = \frac{\hbar^2}{8m^2c^2}\nabla^2 V(\boldsymbol{r}), \tag{14.83}$$

which is remarkably close to the correction we obtained in eq. (14.82). Using the Coulomb potential energy, we find that

$$H_D = \frac{\hbar^2}{8m^2c^2}\left[-\frac{e^2}{4\pi\varepsilon_0}\nabla^2\left(\frac{1}{r}\right)\right] = \frac{\hbar^2 e^2}{8m^2c^2\varepsilon_0}\delta^3(\boldsymbol{r}). \tag{14.84}$$

Here we've used the fact that $\nabla^2\left(r^{-1}\right) = -4\pi\delta^3(\boldsymbol{r})$, where $\delta^3(\boldsymbol{r})$ is the delta function in three dimensions [14.3].

The energy corrections due to the Darwin term are computed by evaluating the expectation value of H_D with the unperturbed states $|n,l,m_l\rangle$ (there's no spin dependence in H_D). We can calculate this expectation value using integration, which yields

$$E_D^{(1)} = \langle \hat{H}_D \rangle = \frac{\hbar^2 e^2}{8m^2c^2\varepsilon_0}\left|\psi_{nlm_l}(0)\right|^2. \tag{14.85}$$

Here the $\psi_{nlm_l}(0)$'s are the unperturbed hydrogen wave functions, evaluated at the origin. The only wave functions that are not 0 at the origin are the $l=0$ wave functions, and you'll show in problem 14.15 that eq. (14.85) can be rewritten as

$$E_D^{(1)} = \begin{cases} \dfrac{mc^2\alpha^4}{2n^3} & l=0 \\ 0 & l\neq 0 \end{cases}. \tag{14.86}$$

14.3.4 Complete Fine Structure

The magnitudes of the three corrections discussed above are comparable, and the complete first-order, fine-structure energy corrections are given by the sum of the individual energy corrections. You'll show in problem 14.16 that the final expression for the energies of the hydrogen atom, including fine structure, is

$$E_{nj} = E_n^{(0)} + E_F^{(1)} = E_n^{(0)}\left[1+\frac{\alpha^2}{n^2}\left(\frac{n}{j+1/2}-\frac{3}{4}\right)\right]. \tag{14.87}$$

Although the relativistic and spin-orbit corrections depend on l, the total energies depend only on j (not on l directly). For each value of the quantum number n, there are n allowed values for j, so each of the unperturbed hydrogen levels splits into n levels when fine structure is accounted for; this does not completely break the degeneracy of the energy levels. Because the energy levels depend on the quantum number j, we add a subscript indicating the j-value in our spectroscopic notation when including fine structure. For example, the $3p_{3/2}$ energy level corresponds to $n=3,\ l=1,\ j=3/2$. Figure 14.4, shows how the unperturbed hydrogen energy levels are split by the fine structure corrections.

The Dirac equation can be solved for the Coulomb potential, and it yields the exact relativistic energies for the hydrogen atom. The Dirac energies, after subtracting off the rest energy of the electron, are (ref. [14.4]):

$$E_{nj} = mc^2\left\{\left[1+\left(\frac{\alpha}{n-\left(j+1/2\right)+\sqrt{\left(j+1/2\right)^2-\alpha^2}}\right)^2\right]^{-1/2}-1\right\}. \tag{14.88}$$

The energies in eq. (14.87) are consistent with the Dirac energies, to terms of order α^4.

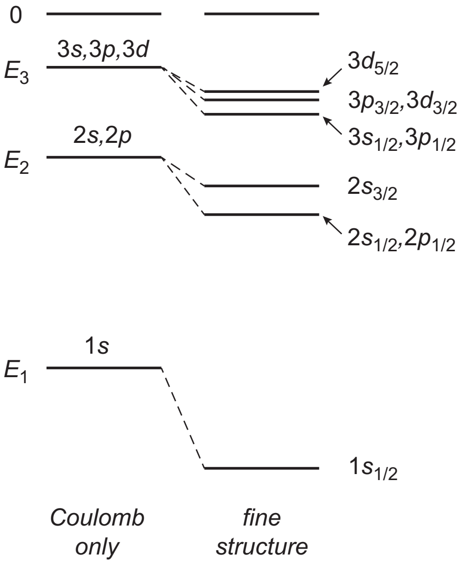


Fig 14.4 The fine structure of hydrogen (not to scale). For the Coulomb potential the energies depend only on *n*, while the fine structure energies depend on *n* and *j*.

14.4 HYPERFINE STRUCTURE OF HYDROGEN

Until now we've been treating the proton as simply a point charge. However, like the electron, it also has a magnetic dipole moment that's proportional to its spin. The proton thus produces a magnetic dipole field, which is felt by the electron. The Hamiltonian of this interaction is given by (ref. [14.5], complement A.XII)[5]

$$\hat{H}_{HF} = \frac{g_p e^2}{8\pi\varepsilon_0 m_e c^2 m_p}\left[\frac{\hat{\boldsymbol{L}}\boldsymbol{\cdot}\hat{\boldsymbol{S}}_p}{r^3} + \frac{3\left(\hat{\boldsymbol{S}}_e\boldsymbol{\cdot}\boldsymbol{u}_r\right)\left(\hat{\boldsymbol{S}}_p\boldsymbol{\cdot}\boldsymbol{u}_r\right) - \hat{\boldsymbol{S}}_e\boldsymbol{\cdot}\hat{\boldsymbol{S}}_p}{r^3} + \frac{8\pi}{3}\left(\hat{\boldsymbol{S}}_e\boldsymbol{\cdot}\hat{\boldsymbol{S}}_p\right)\delta^3\left(\hat{\boldsymbol{r}}\right)\right]. \tag{14.89}$$

Here g_p and m_p are the g factor and the mass of the proton ($g_p = 5.59$, and we've assumed $g_e \cong 2$), and $\hat{\boldsymbol{S}}_e$ and $\hat{\boldsymbol{S}}_p$ are the spin operators for the electron and the proton. This interaction is called the hyperfine interaction, because it leads to energy splittings that are smaller than the fine structure splittings.

The first term in eq. (14.89) represents the interaction of the orbital motion of the electron with the dipole field of the proton. We'll simplify our discussion here to $l = 0$ states, so this term is 0.[6] The second term represents the dipole-dipole interaction of the proton and the electron; you'll show in problem 14.18 that this term is also 0 for $l = 0$ states. Thus, the only term that contributes to the hyperfine interaction for $l = 0$ is the third and final term, which is known as the contact term.

To find the hyperfine energies, we need to find the eigenstates of $\hat{\boldsymbol{S}}_e\boldsymbol{\cdot}\hat{\boldsymbol{S}}_p$, within a given fine-structure level. Since the electron and the proton are both spin-1/2 particles, the eigenstates we seek are exactly the ones that we found in complement 8.C, which are the eigenstates of the total spin $\hat{\boldsymbol{S}} = \hat{\boldsymbol{S}}_e + \hat{\boldsymbol{S}}_p$. These eigenstates are $\left|s, m_s\right\rangle$, with $s = 0,1$ being the total spin quantum number. The square of the total spin is

$$\hat{S}^2 = \left(\hat{\boldsymbol{S}}_e + \hat{\boldsymbol{S}}_p\right)\boldsymbol{\cdot}\left(\hat{\boldsymbol{S}}_e + \hat{\boldsymbol{S}}_p\right) = \hat{S}_e^2 + \hat{S}_p^2 + 2\left(\hat{\boldsymbol{S}}_e\boldsymbol{\cdot}\hat{\boldsymbol{S}}_p\right), \tag{14.90}$$

so the hyperfine Hamiltonian for an $l = 0$ state can be written as

$$\hat{H}_{HF} = \frac{g_p e^2}{6\varepsilon_0 m_e c^2 m_p}\left(\hat{S}^2 - \hat{S}_e^2 - \hat{S}_p^2\right)\delta^3\left(\hat{\boldsymbol{r}}\right). \tag{14.91}$$

Calculating the expectation value of this Hamiltonian for the hydrogen ground state, we see that the hyperfine energies are

$$\begin{aligned} E_{HF}^{(1)} &= \frac{g_p e^2 \hbar^2}{6\varepsilon_0 m_e c^2 m_p}\left[s\left(s+1\right) - \frac{1}{2}\left(\frac{1}{2}+1\right) - \frac{1}{2}\left(\frac{1}{2}+1\right)\right]\left|\psi_{100}\left(0\right)\right|^2 \\ &= \frac{2 g_p m_e^2 c^2 \alpha^4}{3 m_p}\left[s\left(s+1\right) - \frac{6}{4}\right]. \end{aligned} \tag{14.92}$$

---

[5] You will frequently see this Hamiltonian expressed using a different notation. The operator for the spin of the nucleus is often written as $\hat{\boldsymbol{I}}$. Furthermore, you will often see the operator $\hat{\boldsymbol{F}} = \hat{\boldsymbol{J}} + \hat{\boldsymbol{I}} = \hat{\boldsymbol{L}} + \hat{\boldsymbol{S}} + \hat{\boldsymbol{I}}$, which is the total angular momentum, including that of the nucleus.

[6] For a discussion of the hyperfine interaction for $l \neq 0$, See ref. [14.6], sec. 22.

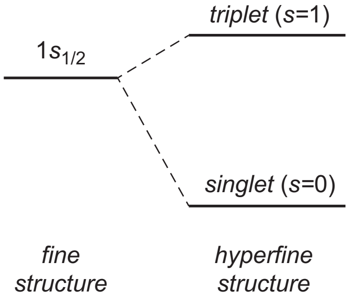


Fig 14.5 The hyperfine splitting of the hydrogen ground state.

This energy shift is smaller than the fine structure energies by a factor of approximately m_e / m_p.

The hyperfine structure of the hydrogen ground state is shown in fig. 14.5. The ground state splits into two levels, corresponding to $s = 0,1$. The true ground state is nondegenerate and is called the singlet state, while the upper level is three-fold degenerate ($m_s = 0, \pm 1$) and called the triplet state. The transition between these states is used in the hydrogen maser, and it is also very important in astronomy. The frequency of this transition has been measured to 13 decimal places, making it one of the most accurately known physical quantities.

14.5 THE ZEEMAN EFFECT

If we apply an external magnetic field $\mathcal{B}$ to a hydrogen atom, it will interact with the magnetic dipole moment of the atom; this is known as the Zeeman effect. The atomic magnetic dipole moment has contributions from both the orbital and the spin angular momenta of the electron,[7] so the Zeeman Hamiltonian is

$$\hat{H}_{Ze} = -\hat{\mu} \cdot \boldsymbol{\mathcal{B}} = -\left(\frac{-e}{2m} \hat{\boldsymbol{L}} + \frac{-g_e e}{2m} \hat{\boldsymbol{S}} \right) \cdot \boldsymbol{\mathcal{B}}. \tag{14.93}$$

The g factor of the electron affects the spin contribution to the dipole moment, whereas the orbital angular momentum contribution is given by the classical value (problem 6.2). Since there's no preferred direction in space for an atom, we are free to choose any direction we'd like for $\boldsymbol{\mathcal{B}}$. We'll choose $\boldsymbol{\mathcal{B}} = \mathcal{B}\boldsymbol{u}_z$, and use $g_e \cong 2$, which makes the Zeeman Hamiltonian

$$\hat{H}_{Ze} = \frac{e\mathcal{B}}{2m}\left(\hat{L}_z + 2\hat{S}_z \right) = \frac{e\mathcal{B}}{2m}\left(\hat{J}_z + \hat{S}_z \right). \tag{14.94}$$

If the external magnetic field is small in comparison to the internal magnetic field of the spin-orbit interaction, then we are in the limit of the weak-field Zeeman effect. We'll treat the weak-field Zeeman effect here, and you'll treat the strong-field effect in

---

[7] We're ignoring the magnetic dipole moment of the nucleus, which is small compared to that of the electron because of its much larger mass. We're consequently ignoring the hyperfine structure.

problem 14.20. In the weak-field limit, the external field splits the fine structure levels we determined in sec. 14.3. So, the unperturbed energies are the fine structure energies of eq. (14.87), and the unperturbed states are the eigenstates of total angular momentum, $\left|j,l,s,m_j\right\rangle_j$ (the subscript *j* is to distinguish these states from the states $\left|l,m_l,s,m_s\right\rangle$, which we will need below).

We need to choose the proper states for degenerate perturbation theory. The energies are degenerate for given values of *n* and *j*, so we need to diagonalize $\hat{H}_{Ze}$ in a subspace of fixed *n* and *j*. We can see that $\hat{L}^2$, $\hat{S}^2$ and $\hat{J}_z$ commute with $\hat{H}_{Ze}$, and they also commute with the unperturbed Hamiltonian, $\hat{H}_0+\hat{H}_F$. The eigenstates of $\hat{L}^2$, $\hat{S}^2$ and $\hat{J}_z$, $\left|j,l,s,m_j\right\rangle_j$ (for fixed *n* and *j*), are nondegenerate in the subspace we're interested in, so we can use these states for the perturbation calculation, without having to do any matrix diagonalization.

To first order, the energy corrections are given by the expectation values of the perturbing Hamiltonian, $\hat{H}_{Ze}$ [eq. (14.94)]. The expectation value of $\hat{J}_z$ for the state $\left|j,l,s,m_j\right\rangle_j$ is $m_j\hbar$, but to calculate $\left\langle\hat{S}_z\right\rangle$ we need to express $\left|j,l,s,m_j\right\rangle_j$ as a linear combination of the states $\left|l,m_l,s,m_s\right\rangle$. There is only one allowed value for *s*, 1/2, which simplifies things, but we won't determine these linear combinations here. We'll simply state the result, which is[8]

$$\begin{aligned}\left|l\pm1/2,l,\frac{1}{2},m_j\right\rangle_j &= \frac{1}{\sqrt{2l+1}}\left(\sqrt{l\mp m_j+\frac{1}{2}}\left|l,m_j+\frac{1}{2},\frac{1}{2},-\frac{1}{2}\right\rangle\right.\\ &\left.\pm\sqrt{l\pm m_j+\frac{1}{2}}\left|l,m_j-\frac{1}{2},\frac{1}{2},\frac{1}{2}\right\rangle\right).\end{aligned} \tag{14.95}$$

Using these states and eq. (14.94), the determination of the weak-field Zeeman energy levels is straightforward, and the result is

$$\begin{aligned}E_{Ze}^{(1)} &= \left\langle\hat{H}_{Ze}\right\rangle\\ &= \frac{e\mathcal{B}\hbar}{2m}\left[m_j+\frac{l\mp m_j+\frac{1}{2}}{2l+1}\left(-\frac{1}{2}\right)+\frac{l\pm m_j+\frac{1}{2}}{2l+1}\left(\frac{1}{2}\right)\right]\\ &= \frac{e\mathcal{B}\hbar}{2m}m_j\left[1\pm\frac{1}{(2l+1)}\right].\end{aligned} \tag{14.96}$$

The choice of ± depends on the value of *j*: $j=l\pm1/2$. The term in brackets is called the Landé *g* factor, as it plays the role of an effective *g* parameter for the electron. Thus, the energies are often written as

---

[8] This calculation is performed in ref. [14.5], complement A.X. The procedure mimics that of complement 8.C in this text. One of the linear combinations can be found fairly easily, and the raising and lowering operators can be used to generate the rest.

$$E_{Ze}^{(1)} = \frac{ge\mathcal{B}\hbar}{2m} m_j . \tag{14.97}$$

A weak magnetic field breaks the degeneracy of the fine structure levels, as the Zeeman energies also depend on the quantum number m_j.

14.6 References

[14.1] R. Shankar, *Principles of Quantum Mechanics, 2nd ed.* (Springer, New York, 1994).

[14.2] J.D. Jackson, *Classical Electrodynamics, 3rd ed.* (John Wiley and Sons, New York, 1999), Sec. 11.8.

[14.3] D.J. Griffiths, *Introduction to Electrodynamics, 3rd ed.* (Prentice Hall, Upper Saddle River, NJ, 1999), p. 50.

[14.4] J.J. Sakurai and J. Napolitano, *Modern Quantum Mechanics, 2nd ed.* (Addison-Wesley, Boston, MA, 2011), Sec. 8.4.

[14.5] C. Cohen-Tannoudji, B. Diu, and Franck Laloë, *Quantum Mechanics* (John Wiley and Sons, New York, 1977).

[14.6] H.A. Bethe and E.E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer-Verlag, Berlin, 1957).

14.7 PROBLEMS

14.1 A constant perturbation of the form $\hat{H}_p = V_0$ is applied to an arbitrary unperturbed Hamiltonian $\hat{H}_0$. Calculate the energy corrections to second-order, and the wave function corrections to first order. Assume nondegenerate energy levels.

14.2 An infinite potential well extends from $x = 0$ to $x = L$, and it has a small bump in the middle. Assume the perturbation is of the form

$$H_p = \begin{cases} V_0 & L/2 - a/2 < x < L/2 + a/2 \\ 0 & elsewhere \end{cases} , \tag{14.98}$$

where $a < L$. (a) Calculate the first-order corrections to the energies. (b) Approximate the energies for the case that $a << L$.

14.3 An infinite potential well extends from $x = 0$ to $x = L$, and it has a corrugated bottom. Assume the perturbation is of the form

$$H_p = V_0 \sqrt{\frac{2}{L}} \sin\left(\frac{M\pi}{L} x\right), \tag{14.99}$$

where M is an integer ($M > 0$). Calculate the first-order corrections to the energies. Does your answer seem reasonable?

14.4 Show that the corrected eigenstates [eq. (14.6)] are normalized to first-order (i.e., if terms of order λ^2 and higher are ignored).

14.5 Find the first-order corrections to the wave functions in example 14.1.

14.6* The translation operator $\hat{T}(D) = e^{-i\hat{p}D/\hbar}$ translates states in space (see sec. 10.2). Expand this operator to first order in D, and apply it to the Fock states $|n\rangle$. Is your answer consistent with that from example 14.2?

14.7 A 3-D, symmetric harmonic oscillator (problem 13.2) experiences a perturbation, so that it is no longer symmetric. Treat the perturbing Hamiltonian as $\hat{H}_p = (1/2)\kappa\hat{x}^2$. To first order, what affect does this have on the energies of the ground and first-excited states of the oscillator?

14.8 Use the virial theorem [eq. (13.121)] to calculate $\langle \hat{r}^{-1} \rangle$ for energy eigenstates of the hydrogen atom.

14.9 It can be shown that $\langle \hat{r}^{v} \rangle$ (where v is an integer) for a hydrogen atom in the state $|n,l,m_l\rangle$ is:

$$\langle \hat{r}^{v} \rangle = \frac{(-1)^{v+1}}{2n}\left(\frac{na_0}{2}\right)^{v}(v+1)!\sum_{k=0}^{v+1}(-1)^{k}\binom{v+1}{k}\binom{n+l+k}{v+1}\binom{n-l-1+k}{v+1}, \tag{14.100}$$

if $v \geq -1$, and

$$\langle \hat{r}^{v} \rangle = \left(\frac{na_0}{2}\right)^{v}\frac{1}{2n(\beta+1)!}\sum_{k=0}^{\beta}(-1)^{\beta-k}\frac{\binom{\beta}{k}\binom{n-l-1+k}{\beta}}{\binom{2l+\beta+1-k}{\beta+1}}, \tag{14.101}$$

if $v \leq -2$. In these expressions the symbol $\binom{j}{k}$ is the binomial coefficient, and $\beta = -(v+2)$.[9] Use these expressions to calculate $\langle \hat{r}^{-1} \rangle$ and $\langle \hat{r}^{-2} \rangle$.

14.10 Show that $\hat{H}_{SO}$ commutes with $\hat{L}^2$ and $\hat{S}^2$, but not $\hat{L}_z$ and $\hat{S}_z$.

14.11* Show that $\hat{J}^2$ and $\hat{J}_z$ commute with $\hat{H}_0$ (the unperturbed hydrogen atom Hamiltonian) and $\hat{H}_{SO}$.

14.12 Use eq. (14.101) to calculate $\langle \hat{r}^{-3} \rangle$.

14.13* Show that the expression for $E_{SO}^{(1)}$ in eq. (14.79) yields

$$E_{SO}^{(1)} = \frac{mc^2\alpha^4}{2n^3} \tag{14.102}$$

for $l = 0$. (Hint: for $l = 0$ there is only one possible value for j: $j = l + 1/2$.)

14.14 To second order in $\boldsymbol{r}'$, compute the average of $V(\boldsymbol{r}+\boldsymbol{r}')$ over a sphere of radius $\bar{\lambda}_c$, centered on $\boldsymbol{r}$.

---

[9] These formulas are adapted from ref. [14.6], sec. 3. Expressions for some specific values of v may be found there as well.

14.15 Verify eq. (14.86). You may find the following relation useful:

$$L_j^k(0)=\frac{(j+k)!}{j!k!}. \tag{14.103}$$

14.16* Add together the individual contributions to the fine structure of hydrogen to obtain eq. (14.87). (Hint: There are three cases: $l=0$, and $l\neq 0$ with two possible values for j.)

14.17 Calculate the fine structure shifts in the $n=2$ level of hydrogen. Compare them to the energy separation of the $n=2$ and $n=3$ levels.

14.18* Show that for spherically symmetric wave functions, the dipole-dipole term of the hyperfine Hamiltonian [eq. (14.89)] does not contribute to the hyperfine energies.

14.19* Calculate the frequency and wavelength of the radiation emitted when a hydrogen atom makes a transition between the two hyperfine ground states.

14.20 The strong-field Zeeman effect occurs when the external magnetic field is much larger than the internal field responsible for the spin-orbit effect. Show that in this limit, the states $\left|l,m_l,s,m_s\right\rangle$ are the appropriate ones to use in calculating the first-order energy corrections. Calculate these corrections.

14.21 If a constant, uniform electric field $\boldsymbol{\mathcal{E}}=\mathcal{E}\boldsymbol{u}_z$ is applied to an atom, we observe what is known as the Stark effect (or the DC Stark effect). (a) Show that, to first order, there is no Stark shift in the energy of the $n=1$ level of hydrogen. (b) To first order, calculate the energy splitting of the $n=2$ level. (Hint: Look at the angular integrals first.)

14.22 $\hat{A}$ commutes with $\hat{H}_0$ and $\hat{H}_p$, and it corresponds to an observable. Assume that the states $\left|\psi_{n,j}^{(0)}\right\rangle$, which are the degenerate eigenstates of $\hat{H}_0$ within a subspace of fixed n, are simultaneously nondegenerate eigenstates of $\hat{A}$. Show that the matrix representation of $\hat{H}_p$ in the basis of these states is diagonal, which means that they are the eigenstates we seek for use in degenerate perturbation theory [the eigenstates determined by eq. (14.40)].

CHAPTER 15

Time-Dependent Perturbation Theory

In chapter 9 we developed a theory of the time evolution of states. We found that if the Hamiltonian of a system is time-independent, the time evolution of the state is relatively simple. But what if the Hamiltonian does depend on time? So far we have only treated one example of a system whose Hamiltonian was time-dependent, that of magnetic resonance in complement 9.A. In this chapter we will apply perturbation theory to problems with time-dependent Hamiltonians.

15.1 TIME EVOLUTION OF THE STATE

Assume that we have a Hamiltonian of the form

$$\hat{H}(t) = \hat{H}_0 + \lambda \hat{H}_p(t). \tag{15.1}$$

Here $\hat{H}_0$ is a time-independent Hamiltonian, for which we know the exact eigenstates and eigenvalues:

$$\hat{H}_0 \left| \psi_n^{(0)} \right\rangle = E_n^{(0)} \left| \psi_n^{(0)} \right\rangle. \tag{15.2}$$

$\hat{H}_p(t)$ is a time-dependent perturbation, and λ is once again a parameter whose values can range from 0 to 1. We'll use λ as a bookkeeping parameter to allow us to keep track of the order of perturbation, just as we did in chapter 14.

We'd like to determine time dependence of an arbitrary state, so we need to use the Schrödinger equation, which tells us that [eq. (9.11)]:

$$\frac{d}{dt}\left|\psi(t)\right\rangle = -\frac{i}{\hbar}\hat{H}(t)\left|\psi(t)\right\rangle. \tag{15.3}$$

Let's write $\left|\psi(t)\right\rangle$ in the basis of the states $\left|\psi_n^{(0)}\right\rangle$ as

$$\left|\psi(t)\right\rangle = \sum_j c_j(t) e^{-i\omega_j t}\left|\psi_j^{(0)}\right\rangle. \tag{15.4}$$

The key difference between the solution for a time-independent Hamiltonian [eq. (11.16)] and a time-dependent Hamiltonian [eq. (15.4)] is that for a time-dependent Hamiltonian the coefficients $c_j(t)$ are time-dependent.

Substitute eq. (15.4) into eq. (15.3), which yields

$$\frac{d}{dt}\sum_j c_j(t) e^{-i\omega_j t}\left|\psi_j^{(0)}\right\rangle = -\frac{i}{\hbar}\sum_j c_j(t) e^{-i\omega_j t}\left(\hat{H}_0 + \lambda \hat{H}_p(t)\right)\left|\psi_j^{(0)}\right\rangle. \tag{15.5}$$

If we apply $\left\langle \psi_n^{(0)}\right|$, and use the orthogonality of the states, we find

$$\frac{d}{dt}c_n(t) e^{-i\omega_n t} = -i\frac{E_n^{(0)}}{\hbar}c_n(t)e^{-i\omega_n t} - \frac{i}{\hbar}\lambda\sum_j c_j(t)e^{-i\omega_j t}\left\langle \psi_n^{(0)}\right|\hat{H}_p(t)\left|\psi_j^{(0)}\right\rangle. \tag{15.6}$$

If we use the product rule on the derivative, and the Planck relation $E_n^{(0)} = \hbar\omega_n$, we find that we can cancel terms, yielding

$$e^{-i\omega_n t}\frac{d}{dt}c_n(t) = -\frac{i}{\hbar}\lambda\sum_j e^{-i\omega_j t}\left\langle \psi_n^{(0)}\right|\hat{H}_p(t)\left|\psi_j^{(0)}\right\rangle c_j(t). \tag{15.7}$$

If we define $\omega_{nj} \equiv \omega_n - \omega_j$, and

$$H_{p\,nj}(t) \equiv \left\langle \psi_n^{(0)}\right|\hat{H}_p(t)\left|\psi_j^{(0)}\right\rangle, \tag{15.8}$$

we can rewrite this as

$$\frac{d}{dt}c_n(t) = -\frac{i}{\hbar}\lambda\sum_j e^{i\omega_{nj}t}H_{p\,nj}(t)c_j(t). \tag{15.9}$$

This is an exact equation for the $c_n(t)$'s, given the Hamiltonian of eq. (15.1).

15.1.1 Perturbation Solution

Now let's perform a perturbation expansion of the coefficients $c_n(t)$:

$$c_n(t) = c_n^{(0)}(t) + \lambda c_n^{(1)}(t) + \lambda^2 c_n^{(2)}(t) + \ldots. \tag{15.10}$$

In order for perturbation theory to be valid when $\lambda = 1$, the corrections $c_n^{(k)}(t)$ must get smaller as the order of perturbation (k) increases. For small times this is almost always the case. If we substitute eq. (15.10) into eq. (15.9), and equate terms with equal powers of λ, we see that the zeroth-order (λ^0) terms yield

$$\frac{d}{dt}c_n^{(0)}(t)=0, \tag{15.11}$$

and the first-order terms yield

$$\frac{d}{dt}c_n^{(1)}(t)=-\frac{i}{\hbar}\sum_j e^{i\omega_{nj}t}H_{p\,nj}(t)c_j^{(0)}(t). \tag{15.12}$$

In general, the differential equation for the k^{th}-order term is

$$\frac{d}{dt}c_n^{(k)}(t)=-\frac{i}{\hbar}\sum_j e^{i\omega_{nj}t}H_{p\,nj}(t)c_j^{(k-1)}(t) \quad k=1,2,\ldots, \tag{15.13}$$

which allows us to express any order correction in terms of corrections that are one order lower.

Now we need to specify the initial conditions. We'll assume that the perturbation "turns on" at $t = 0$, so that $\hat{H}_p(t)=0$ for $t \leq 0$. At $t = 0$ eq. (15.4) becomes

$$|\psi(0)\rangle=\sum_j c_j(0)\left|\psi_j^{(0)}\right\rangle. \tag{15.14}$$

so the $c_j(0)$'s uniquely specify the initial conditions. Since there is no perturbation for $t \leq 0$, the solution at $t = 0$ is a zeroth-order solution (equivalent to $\lambda = 0$), so examining eq. (15.10) we find

$$c_n(0)=c_n^{(0)}(0), \tag{15.15}$$

$$c_n^{(k)}(0)=0, \quad k=1,2,\ldots. \tag{15.16}$$

The differential equation for the zeroth-order terms [eq. (15.11)] tells us that the $c_n^{(0)}(t)$'s are constant, and the initial conditions of eq.(15.15) then yield

$$c_n^{(0)}(t)=c_n^{(0)}(0)=c_n(0). \tag{15.17}$$

Substituting this into the differential equation for the first-order terms [eq. (15.12)], we see that

$$\frac{d}{dt}c_n^{(1)}(t)=-\frac{i}{\hbar}\sum_j e^{i\omega_{nj}t}H_{p\,nj}(t)c_j(0). \tag{15.18}$$

We can now integrate this equation; using the initial conditions of eq. (15.16), we find

$$c_n^{(1)}(t)=-\frac{i}{\hbar}\sum_j c_j(0)\int_0^t dt'\,e^{i\omega_{nj}t'}H_{p\,nj}(t'). \tag{15.19}$$

Substituting eqs. (15.17) and (15.19) into eq. (15.10) (with $\lambda = 1$), we see that to first order the time-dependent state is determined by the coefficients

$$c_n(t) = c_n(0) - \frac{i}{\hbar}\sum_j c_j(0)\int_0^t dt'\, e^{i\omega_{nj}t'} H_{p\,nj}(t'). \tag{15.20}$$

Equation (15.20) is the main result of time-dependent perturbation theory. Given the matrix elements of the perturbing Hamiltonian [eq. (15.8)] and the initial conditions, we can calculate the first-order corrections to the coefficients which determine the time-dependent state. If higher-order corrections are desired, we can substitute the lower-order solutions into eq. (15.13) and integrate.

15.1.2 Transition Probabilities

One thing that we can calculate fairly easily using time-dependent perturbation theory is the probability that a system will make the transition from one state to another. For example, suppose that at $t = 0$ system is in the initial state $\left|\psi_i^{(0)}\right\rangle$. This means that our initial conditions are

$$c_n(0) = \delta_{ni}. \tag{15.21}$$

If there is no perturbation the system is in an eigenstate of the Hamiltonian, and will stay that way forever. However, we'll assume that we turn on the perturbation at $t = 0$. We'd like to know the probability that the perturbation causes the system to make a transition to some final state $\left|\psi_f^{(0)}\right\rangle$. This time-dependent transition probability is given by

$$\begin{aligned} P_{if}(t) &= \left|\left\langle \psi_f^{(0)} \middle| \psi(t)\right\rangle\right|^2 \\ &= \left|c_f(t)\right|^2. \end{aligned} \tag{15.22}$$

Since we're talking about a transition to a final state that is different from the initial state, we know that $f \neq i$, and eq. (15.21) tells us that $c_f(0) = 0$. Using this, and also using eq. (15.21) in eq. (15.20), we see that

$$c_f(t) = -\frac{i}{\hbar}\int_0^t dt'\, e^{i\omega_{fi}t'} H_{p\,fi}(t'). \tag{15.23}$$

Equations (15.22) and (15.23) determine the probability that the system will make a transition from the initial to the final state, to first order.

15.2 SINUSOIDAL PERTURBATIONS

Now that we've developed a general theory of time-dependent perturbations, let's consider a special case, that of the two-level system shown in fig. 15.1. We'll assume that the system starts in state $|1\rangle$ at $t = 0$, so we have $c_1(0) = 1$ and $c_2(0) = 0$. Since there's only one transition, we'll define $\omega_0 \equiv \omega_{21} = (E_2 - E_1)/\hbar$ to be the resonance frequency associated with it.

From eq. (15.22), the probability that the system will make a transition from $|1\rangle$ to $|2\rangle$ is $P_{12}(t) = |c_2(t)|^2$, where $c_2(t)$ is given by eq. (15.23):

$$c_2(t) = -\frac{i}{\hbar}\int_0^t dt'\, e^{i\omega_0 t'} H_{p21}(t'). \tag{15.24}$$

We'll assume that the perturbation varies sinusoidally in time, so the perturbing Hamiltonian is

$$\hat{H}_p(t) = \hat{V}\cos\omega t. \tag{15.25}$$

The matrix element we need in eq. (15.24) is

$$H_{p21}(t') = \langle 2|\hat{V}|1\rangle \cos\omega t' = V_{21}\cos\omega t'. \tag{15.26}$$

Substituting this into eq. (15.24) yields

$$\begin{aligned} c_2(t) &= -\frac{i}{\hbar}V_{21}\int_0^t dt'\, e^{i\omega_0 t'}\cos\omega t' \\ &= -\frac{iV_{21}}{2\hbar}\int_0^t dt'\left(e^{-i(\omega-\omega_0)t'} + e^{i(\omega+\omega_0)t'}\right) \\ &= \frac{V_{21}}{2\hbar}\left(\frac{e^{-i(\omega-\omega_0)t} - 1}{(\omega-\omega_0)} - \frac{e^{i(\omega+\omega_0)t} - 1}{(\omega+\omega_0)}\right). \end{aligned} \tag{15.27}$$

The transition probability is then the square magnitude of this:

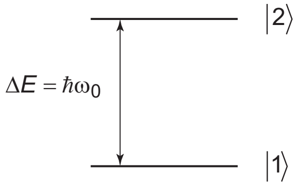


Fig 15.1 A two-level system, with a resonance frequency of ω_0.

$$P_{12}(t)=\frac{|V_{21}|^2}{4\hbar^2}\left|\frac{e^{-i(\omega-\omega_0)t}-1}{(\omega-\omega_0)}-\frac{e^{i(\omega+\omega_0)t}-1}{(\omega+\omega_0)}\right|^2. \tag{15.28}$$

This expression is valid for arbitrary drive frequencies, as long as the probability remains small. We can simplify things, however, if we assume that the drive frequency is close to resonance; this is the interesting case anyway, as it's the only time that there's significant probability of a transition. We'll assume that $\omega \simeq \omega_0$, or more specifically that $\omega-\omega_0 << \omega+\omega_0$. With this approximation we find[1]

$$\begin{aligned}P_{12}(t)&\simeq\frac{|V_{21}|^2}{4\hbar^2}\left|\frac{e^{-i(\omega-\omega_0)t}-1}{(\omega-\omega_0)}\right|^2\\&=\frac{|V_{21}|^2}{4\hbar^2}\frac{\{2-2\cos[(\omega-\omega_0)t]\}}{(\omega-\omega_0)^2}\\&=\frac{|V_{21}|^2}{\hbar^2}\frac{\sin^2[(\omega-\omega_0)t/2]}{(\omega-\omega_0)^2}.\end{aligned} \tag{15.29}$$

This probability oscillates sinusoidally in time, with a frequency of $\omega-\omega_0$. The maximum probability is

$$P_{12\max}=\frac{|V_{21}|^2}{\hbar^2(\omega-\omega_0)^2}. \tag{15.30}$$

Why does the probability oscillates in time, and not simply grow? The answer is that as soon as the probability of being in $|2\rangle$ is nonzero, the perturbation can drive the system from $|2\rangle$ back to $|1\rangle$. Indeed, you'll show in problem 15.5 that the probability of making a transition from $|2\rangle$ to $|1\rangle$ is the same as the probability of making the transition from $|1\rangle$ to $|2\rangle$, that is, $P_{21}(t)=P_{12}(t)$. Since the perturbation can drive the system in either direction, it oscillates between the two states.

What happens if we drive the system perfectly on resonance, that is, $\omega=\omega_0$? The denominator in eq. (15.29) diverges, but if we multiply the numerator and denominator by $(t/2)^2$, we find

$$\begin{aligned}P_{12}(\omega,t)&\simeq\frac{|V_{21}|^2t^2}{4\hbar^2}\frac{\sin^2[(\omega-\omega_0)t/2]}{[(\omega-\omega_0)t/2]^2}\\&=\frac{|V_{21}|^2t^2}{4\hbar^2}\operatorname{sinc}^2[(\omega-\omega_0)t/2],\end{aligned} \tag{15.31}$$

---

[1] For a further discussion of validity of this approximation, see ref. [15.1], sec. XIII.C.2.

where $\text{sinc}(x) \equiv \sin(x)/x$. This transition probability is plotted as a function of the drive frequency in fig. 15.2. It can be seen that the probability peaks at $\omega = \omega_0$, as we'd expect. Furthermore, the peak gets taller and narrower as time increases.

One way to explain this behavior is to realize that the time and frequency aspects of the perturbation oscillations are related by a Fourier transform. The longer the oscillations go on in time, the more well-defined their frequency becomes; the drive frequency must then be closer to the resonance frequency in order to drive a transition. If an oscillating signal at frequency ω is on for a time period of Δt, the width of its spectral distribution is $\Delta\omega \sim 1/\Delta t$ [see also eq. (2.A.13)], which is consistent with the width of the distribution in fig. 15.2.[2]

Perfectly on resonance, eq. (15.31) predicts that the transition probability will grow quadratically in time. Clearly at some point in time this will violate our perturbation assumption that the probability is small. How can we deal with this problem? One way is to go to a higher order of perturbation theory. Another way is to go back and try to solve the problem more exactly. Notice that we have made 2 assumptions in deriving eq. (15.31). First, we assumed the validity of first-order perturbation theory, essentially assuming that $P_{12}(t)$ is small. Second, we assumed that we were driving the system near resonance ($\omega \simeq \omega_0$). What happens if we make the near resonance approximation at the beginning? Can we avoid the perturbation assumption in that case? If you recall the discussion in complement 9.A, where we discussed magnetic resonance, you'll realize that the answer is yes. You'll explore this in problem 15.8.

15.3 ATOMS AND FIELDS

Let's apply perturbation theory to the very important example of an atom in an electromagnetic field. A hydrogen atom, for example, consists of an electron "orbiting" a proton. In general this atom will have an electric dipole moment, or have one induced by the field,

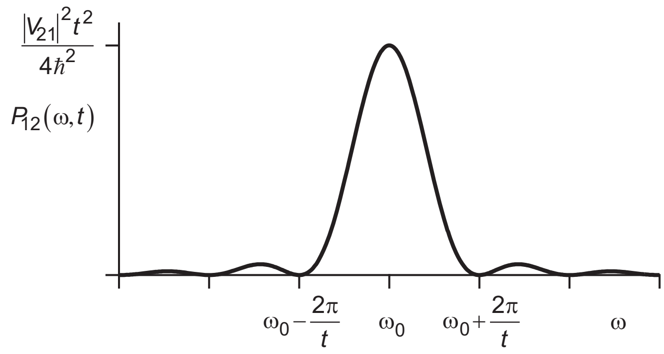


Fig 15.2 The transition probability $P_{12}(\omega,t)$ is plotted as a function of the drive frequency ω.

---

[2] Since energy and frequency are related by the Planck relation $E = \hbar\omega$, this is also consistent with the energy-time indeterminacy relation: $\Delta E \Delta t \geq \hbar/2$.

and this dipole will in turn interact with the field. If the atom is smaller than the wavelength of the field, this interaction is described by the electric-dipole Hamiltonian, which is

$$\hat{H}_{ED}(t) = -\hat{\mathbf{d}}\cdot\boldsymbol{\mathcal{E}}(t). \tag{15.32}$$

Here $\hat{\mathbf{d}}$ is the electric dipole moment operator for the atom, and the field is evaluated at the position of the atom.[3] For hydrogen $\hat{\mathbf{d}}$ is given by $\hat{\mathbf{d}} = -e\hat{\boldsymbol{r}}$, with $\hat{\boldsymbol{r}}$ being the position operator for the electron. If we assume that the field is linearly polarized along the z-direction, the electric-dipole Hamiltonian becomes

$$\hat{H}_{ED}(t) = -\hat{\mathrm{d}}_z \mathcal{E}\cos\omega t, \tag{15.33}$$

where ω is the angular frequency of the field. The theory of the interaction between the atom and the field we are presenting here is referred to as semiclassical, because the field is being treated as a classical electromagnetic wave, while the atom is being treated quantum mechanically.

The electric-dipole Hamiltonian varies sinusoidally, like the Hamiltonian in eq. (15.25). Assuming that there are only two atomic states whose resonance frequency is close to that of the drive frequency, we can then use eqs. (15.29) and (15.31) to write the probability of the atom making a transition from one state to another as

$$\begin{aligned} P_{12}(\omega,t) &= \frac{|\mathrm{d}_z|^2 \mathcal{E}^2}{\hbar^2(\omega-\omega_0)^2}\sin^2\left[(\omega-\omega_0)t/2\right] \\ &= \frac{|\mathrm{d}_z|^2 \mathcal{E}^2 t^2}{4\hbar^2}\operatorname{sinc}^2\left[(\omega-\omega_0)t/2\right]. \end{aligned} \tag{15.34}$$

The probability of the atom making a transition to the excited state varies sinusoidally in time, while its frequency behavior is illustrated in fig. 15.2.

Assume that we're talking about a hydrogen atom initially in the ground state. The dipole matrix element in eq. (15.34) is then given by

$$\mathrm{d}_z = \langle 2|\hat{\mathrm{d}}_z|1\rangle = -e\langle n,l,m_l|\hat{z}|1,0,0\rangle = -e\iiint\limits_{\substack{all\\space}} dV\,\psi^*_{nlm_l}(\boldsymbol{r})\,z\,\psi_{100}(\boldsymbol{r}). \tag{15.35}$$

In spherical coordinates we can write $z = r\cos\theta = r\sqrt{4\pi/3}\,Y_1^0(\theta,\phi)$. With this substitution, the angular part of the integral in eq. (15.35) is

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta\, Y_l^{m_l *}(\theta,\phi)\left[\sqrt{\frac{4\pi}{3}}Y_1^0(\theta,\phi)\right]\left(\frac{1}{\sqrt{4\pi}}\right) = \sqrt{\frac{1}{3}}\,\delta_{l1}\delta_{m_l 0}, \tag{15.36}$$

---

[3] This is referred to as the electric-dipole approximation. The next higher-order terms in the interaction Hamiltonian are the electric-quadrupole and magnetic-dipole terms. These terms are smaller than the electric-dipole term by a factor of the order of the fine-structure constant α, and are typically only important if the electric-dipole term is 0 [15.2].

because the spherical harmonics are orthogonal. If $\mathrm{d}_z = 0$ the probability of a transition between the states is 0, so, for our *z*-polarized electric field the only states that a ground-state hydrogen atom can make a transition to are states with $l = 1$, and $m_l = 0$. More generally, it can be shown that for an atom to make an electric-dipole transition from one state to another, the following selection rules must be satisfied: $l_2 - l_1 = \Delta l = \pm 1$, $m_{l_2} - m_{l_1} = \Delta m_l = 0, \pm 1$.[4]

15.3.1 Fields with a Broad Spectrum

To this point we have been considering monochromatic fields, but we'll find it useful to generalize to a field with a broad spectrum. We'll also find it useful to express our transition probability not in terms of the electric field, but in terms of the electromagnetic energy density (energy/volume). The time-averaged energy density *U* of an electromagnetic wave is related to the amplitude of the electric field by (ref. [15.3]):

$$U = \frac{1}{2}\varepsilon_0 \mathcal{E}^2 , \tag{15.37}$$

so eq. (15.34) becomes

$$P_{12}(\omega, t) = \frac{|\mathrm{d}_z|^2 U t^2}{2\varepsilon_0 \hbar^2} \operatorname{sinc}^2\left[(\omega - \omega_0) t/2\right]. \tag{15.38}$$

If the field has a broad spectrum, the energy density in the frequency interval between ω and $\omega + d\omega$ is

$$U = u(\omega)\, d\omega, \tag{15.39}$$

where $u(\omega)$ is the spectral energy density. Each piece of the spectrum contributes to the transition probability, so to find the total transition probability we need to integrate over all the spectral contributions. Combining eqs. (15.38) and (15.39) and integrating, we find that

$$P_{12}(t) = \frac{|\mathrm{d}_z|^2 t^2}{2\varepsilon_0 \hbar^2} \int_0^\infty d\omega\, u(\omega) \operatorname{sinc}^2\left[(\omega - \omega_0) t/2\right]. \tag{15.40}$$

There are two functions inside the integral, $u(\omega)$ and $\operatorname{sinc}^2[\ldots]$. Let's assume that the frequency spread of $u(\omega)$ is much broader than the spread of $\operatorname{sinc}^2[\ldots]$, as shown in fig. 15.3 (this is valid for times that are longer than several field oscillation periods). If this is the case, then $u(\omega)$ is approximately constant over the width of $\operatorname{sinc}^2[\ldots]$, so we can evaluate $u(\omega)$ at the peak of $\operatorname{sinc}^2[\ldots]$ (at $\omega = \omega_0$) and factor it out of the integral. Equation (15.40) then becomes

---

[4] See, for example, ref. [15.1], complement A.XIII.

$$P_{12}(t) = \frac{|\mathrm{d}_z|^2 t^2}{2\varepsilon_0 \hbar^2} u(\omega_0) \int_0^\infty d\omega \,\mathrm{sinc}^2\left[(\omega - \omega_0)t/2\right]. \tag{15.41}$$

Under the conditions we're assuming here $\mathrm{sinc}^2[\ldots]$ is narrow [fig. 15.3], so we can replace the limits of integration by $\pm\infty$. The integral then equals $2\pi/t$, so

$$P_{12}(t) = \frac{\pi |\mathrm{d}_z|^2 t}{\varepsilon_0 \hbar^2} u(\omega_0). \tag{15.42}$$

Changing from monochromatic to broad-bandwidth illumination causes the time dependence of the transition probability to go from sinusoidal in time to linear in time. This result is valid for times small enough that $P_{12}(t) << 1$.

15.3.2 Many Atoms

Suppose that instead of a single atom, we have a large collection of atoms. Unless these atoms are part of a crystal lattice, which has a preferred direction in space, their dipole moments will be randomly oriented with respect to our electric field polarization. To calculate the average probability that one of the atoms will make a transition, we need to average over the orientations of the dipole moments. This averaging will replace the factor of $|\mathrm{d}_z|^2 \mathcal{E}^2$ in eq. (15.34) with $\overline{|\mathbf{d}\cdot\boldsymbol{\mathcal{E}}(t)|^2}$ (the overbar denotes an average with respect to orientation).

We begin with

$$|\mathbf{d}\cdot\boldsymbol{\mathcal{E}}(t)|^2 = |\mathrm{d}|^2 \mathcal{E}^2 \cos^2\theta, \tag{15.43}$$

where θ is the angle between $\mathbf{d}$ and $\boldsymbol{\mathcal{E}}$. Averaging over all possible angles, we find

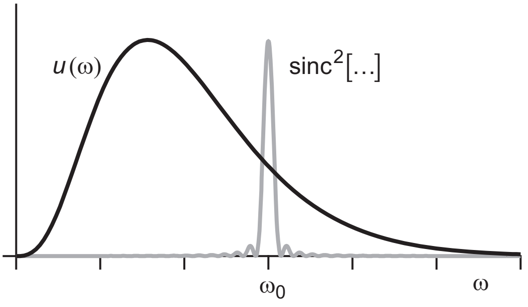


Fig 15.3 The spectral energy density, $u(\omega)$, and $\mathrm{sinc}^2\left[(\omega-\omega_0)t/2\right]$ are plotted as functions of ω.

$$\overline{\cos^2\theta} = \frac{\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta \cos^2\theta}{\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta} = \frac{\int_0^{\pi} d\theta \sin\theta \cos^2\theta}{\int_0^{\pi} d\theta \sin\theta} = \frac{1}{3}, \tag{15.44}$$

which means

$$\overline{\left|\mathbf{d}\bullet\mathcal{E}(t)\right|^2} = \frac{1}{3}\left|\mathrm{d}\right|^2 \mathcal{E}^2.$$

Thus, we can obtain the average transition probability for a randomly oriented dipole by making the substitution $\left|\mathrm{d}_z\right|^2 \to (1/3)\left|\mathrm{d}\right|^2$ in eq. (15.34). This substitution also applies to eq. (15.42), yielding

$$P_{12}(t) = \frac{\pi\left|\mathrm{d}\right|^2 t}{3\varepsilon_0\hbar^2} u(\omega_0). \tag{15.45}$$

The transition rate (transition probability per time) is simply the derivative of this, so it is

$$R_{12} = \frac{d}{dt} P_{12}(t) = \frac{\pi\left|\mathrm{d}\right|^2}{3\varepsilon_0\hbar^2} u(\omega_0). \tag{15.46}$$

The transition rate is constant, a result known as Fermi's Golden Rule.

15.3.3 Stimulated and Spontaneous Emission

In 1917 Einstein considered a collection of atoms in thermal equilibrium at temperature T with a black-body radiation field. He considered two-level atoms, with an energy difference of $\hbar\omega_0$, and the energy transfer between the levels was described by the processes shown in fig. 15.4. We'll replicate Einstein's arguments in complement 15.A; here we'll simply state the final result.

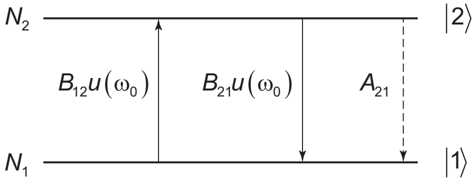


Fig 15.4 Einstein's model of a collection of two-level atoms in thermal equilibrium with a field. There are N_1 atoms in state $|1\rangle$ and N_2 atoms in state $|2\rangle$, and the transition rates are indicated.

As seen in fig. 15.4, there are three possible ways for the atoms to make transitions between energy levels. The atoms can go from the ground state to the excited state by absorbing energy from the field; the rate at which this process occurs is $B_{12}u(\omega_0)$, where B_{12} is a constant and $u(\omega_0)$ is the spectral energy density of the field at the atomic resonance frequency. The atoms can go from the excited state the ground state in two ways. The field can cause transitions at the rate $B_{21}u(\omega_0)$; this process is called stimulated emission. The atoms can also decay by spontaneous emission at the rate A_{21}.The atoms and field are assumed to be in thermal equilibrium, so $u(\omega)$ is given by the Planck black-body spectrum. Einstein showed that the constants must be related by:

$$B_{21} = B_{12}, \tag{15.47}$$

$$A_{21} = \frac{\hbar\omega_0^3}{\pi^2 c^3} B_{12}. \tag{15.48}$$

If one of the constants is known, the other two can be determined. However, quantum mechanics was not yet available to Einstein, so he had no way to go any further.

However, we just calculated the rate at which atoms will absorb energy from a field to make transitions to an excited state. Setting Einstein's absorption rate, $B_{12}u(\omega_0)$, equal to the rate we calculated in eq. (15.46), we find

$$B_{12} = \frac{\pi |\mathrm{d}|^2}{3\varepsilon_0 \hbar^2}. \tag{15.49}$$

Furthermore, in problem 15.5 you'll show that perturbation theory predicts $P_{21}(t) = P_{12}(t)$, which means $R_{21} = R_{12}$, which is in agreement with Einstein's prediction of eq. (15.47). Using eq. (15.48), the spontaneous emission rate is then

$$A_{21} = \left(\frac{\hbar\omega_0^3}{\pi^2 c^3}\right)\left(\frac{\pi |\mathrm{d}|^2}{3\varepsilon_0 \hbar^2}\right) = \frac{|\mathrm{d}|^2 \omega_0^3}{3\pi\varepsilon_0 c^3 \hbar}. \tag{15.50}$$

Since A_{21} is the rate at which atoms will decay from the upper state in the absence of an applied field, the upper state lifetime is given by $\tau = 1/A_{21}$.

15.3.4 Discussion

Using our semiclassical theory, we are able to directly calculate the values of the absorption and stimulated emission rates (Einstein *B* coefficients). However, to determine the spontaneous emission rate *A* we need to resort to Einstein's thermodynamic model and eq. (15.48). Can we calculate *A* directly? The answer is yes, but not by using the semiclassical theory we have developed here. Assume that the atom starts in the excited state, $c_2(0) = 1$, and there is no external field, $u(\omega) = 0$. Semiclassical theory predicts that the atom will stay in the excited state for all time, as it is an eigenstate of the

Hamiltonian in the absence of an external field. For the atom to make a transition to the ground state there must be a perturbation, which requires the external field.

In chap. 16 we will develop a theory that can calculate A directly.

15.4 THE PHOTOELECTRIC EFFECT

We now have all the tools necessary to give a semiclassical explanation of the photoelectric effect. In the photoelectric effect an applied electric field excites an electron that is initially in a bound state to a free state. We can then detect the presence of this free electron. In this section we will model our system as an atom, and the free states are then the positive energy states that represent the atom being ionized; there is a continuum of such states.[5] The model we will adopt is shown in fig. 15.5. As seen in this figure, the applied field oscillates at frequency ω, and this field couples the ground state to the continuum levels.

We'll choose the energy of the system to be 0 at the bottom of the continuum level. The energy of the ground state is E_1, which is negative. The ionization energy E_I is the minimum energy required to excite an electron from the ground state to the continuum levels, and as such $E_I = -E_1$. Earlier in this chapter we used the parameter ω_0 to represent the resonance frequency between two energy levels. Here, the resonance frequency associated with the ground state and a continuum level with energy E is $\omega_0 = \left(E + E_I\right)/\hbar$ [see fig. 15.5]. We can use eq. (15.34) to write the probability that the system will make a transition from the ground state to a state with energy E as

$$\begin{aligned} P(t) &= \frac{\left|\mathrm{d}_z(E)\right|^2 \mathcal{E}^2 t^2}{4\hbar^2} \operatorname{sinc}^2\left\{\left[\omega - \left(E + E_I\right)/\hbar\right]t/2\right\} \\ &= \frac{\left|\mathrm{d}_z(E)\right|^2 \mathcal{E}^2 t^2}{4\hbar^2} \operatorname{sinc}^2\left\{\left[E - \left(\hbar\omega - E_I\right)\right]t/\left(2\hbar\right)\right\}. \end{aligned} \tag{15.51}$$

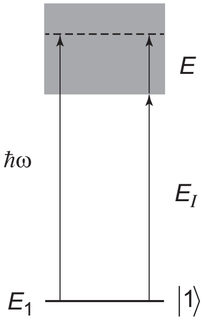


Fig 15.5 A classical electromagnetic field of frequency ω couples state $|1\rangle$ to a continuum of free states. E_I is the ionization energy, and E is the energy in the continuum ($E = 0$ at the bottom of the continuum).

---

[5] Modern photodetectors typically use semiconducting materials, and the free states are then conduction band states, where the electron is free to move.

Here we have also made it explicit that the dipole matrix element depends on the final state: $\mathrm{d}_z(E) = \langle E|\hat{\mathrm{d}}_z|1\rangle$.[6]

There is a continuum of ionized states, and we are interested in the probability that we detect a free electron. Since any of the positive energy states represent a free electron, the detection probability is given by the integral over all positive energies. In general the density of states $n(E)$ in the continuum is not uniform, and we need to multiply the probability of a transition to a particular energy state by the number of states at that energy. The number of states between energies E and $E+dE$ is $n(E)\,dE$, and integrating over all energies yields the total probability that we detect a photoelectron:

$$P(t) = \frac{\mathcal{E}^2 t^2}{4\hbar^2}\int_0^\infty dE\; n(E)\left|\mathrm{d}_z(E)\right|^2 \operatorname{sinc}^2\left\{\left[E-(\hbar\omega - E_I)\right]t/(2\hbar)\right\}. \tag{15.52}$$

The function $\operatorname{sinc}^2\{\ldots\}$ in Eq. (15.52) is a function of E, and has a peak at $E = \hbar\omega - E_I$. As time increases, this function becomes narrower in its energy spread [see the discussion following eq. (15.31)]; we'll assume that the time is long enough that $\operatorname{sinc}^2\{\ldots\}$ is narrow. In order for the integral to be nonzero, the peak must then be located at positive energies ($\hbar\omega > E_I$). The spread of $\operatorname{sinc}^2\{\ldots\}$ is also much narrower than the spread of $n(E)\left|\mathrm{d}_z(E)\right|^2$. Thus, $n(E)\left|\mathrm{d}_z(E)\right|^2$ is essentially constant over the width of $\operatorname{sinc}^2\{\ldots\}$, so it can be evaluated at the peak of $\operatorname{sinc}^2\{\ldots\}$ and factored outside of the integral, yielding[7]

$$P(t) = \frac{\mathcal{E}^2 t^2}{4\hbar^2} n(\hbar\omega - E_I)\left|\mathrm{d}_z(\hbar\omega - E_I)\right|^2 \int_0^\infty dE\; \operatorname{sinc}^2\left\{\left[E-(\hbar\omega - E_I)\right]t/(2\hbar)\right\}. \tag{15.53}$$

Once again we can approximate the integral by extending the limits to ±∞. In this case the integral is $2\pi\hbar/t$, so the probability of detecting the photoelectron is

$$P(t) = \frac{\pi}{2\hbar} n(\hbar\omega - E_I)\left|\mathrm{d}_z(\hbar\omega - E_I)\right|^2 \mathcal{E}^2 t. \tag{15.54}$$

In chap. 2 we defined the intensity I to be equal to the square magnitude of the field, and with this definition we have

$$P(t) = \eta I\, t. \tag{15.55}$$

---

[6] The final state $|E\rangle$ is the state of a free electron, so $|E\rangle$ is the 3-D analog of a state of definite momentum, as discussed in chapter 10.

[7] This situation is analogous to that described in sec. 15.3.1. There we had a broad distribution of field frequencies and a narrow atomic resonance. Here we have a well-defined field frequency and a broad distribution of atomic energies. The end result is essentially the same.

If we generalize from a single atom to a real detector, we can view η as a measure of the efficiency of the detector.[8] Remember that everything we have done here is based on perturbation theory, so eq. (15.55) is valid in the limit of small times, so that the probability $P(t)$ is small.

Equation (15.55) is very simple, but it is an important result. It says that the photoelectron detection probability is linearly proportional to the intensity striking the detector and linearly proportional to the time. Recalling that the detection rate is the time derivative of the probability, this means that the rate is constant in time.

Our model explains the following features of the photoelectric effect:

1. There is a threshold frequency, $\omega = E_I / \hbar$, for the applied field. For frequencies below the threshold no photoelectrons will be observed, while we do observe photoelectrons for higher frequencies.
2. Above the threshold frequency, the energy of the emitted electrons is linearly proportional to the frequency of the applied field: $E = \hbar\omega - E_I$.
3. For any arbitrarily small time after the field strikes the detector, there is a finite probability of observing a photoelectron.

Note that in 1905 Einstein proposed a model of the photoelectric effect that included these three features.[9] The main difference between our model and Einstein's is that ours treats the field as a classical electromagnetic wave, whereas Einstein assumed that the field consists of photons. Because Einstein's model was so successful, it was originally believed that photons were necessary to explain the photoelectric effect. Since there are no photons in our model, we have just shown that that is not the case; the photoelectric effect does not serve as proof that electromagnetic waves contain photons.

An experiment demonstrating that light truly is made of photons is described in lab 2. There we are interested in the probability of a photoelectric detection at a time t, within a small time window Δt, for a fluctuating field. In this situation eq. (15.55) generalizes to

$$P(t, \Delta t) = \eta \left\langle I(t) \right\rangle \Delta t. \tag{15.56}$$

Here the brackets indicate an average of the fluctuating intensity.

15.5 References

[15.1] C. Cohen-Tannoudji, B. Diu, and Franck Laloë, *Quantum Mechanics* (John Wiley and Sons, New York, 1977).

[15.2] R. Loudon, *The Quantum Theory of Light, 3rd ed.* (Oxford University Press, Oxford, 2000), Sec. 4.8.

---

[8] As noted in chap. 2, other texts define the intensity to be equal to the power per unit area of an electromagnetic wave. Since these two intensities are proportional to each other (note 3, page 23), and the proportionality constant can be absorbed in η, eq. (15.55) remains correct for either definition.

[9] See, for example, ref. [15.4].

[15.3] D.J. Griffiths, *Introduction to Electrodynamics, 3rd ed*. (Prentice Hall, Upper Saddle River, NJ, 1999), Sec. 9.2.3.

[15.4] P.A. Tipler and R.A Llewellyn, *Modern Physics 5th Ed*., (W.H. Freeman, New York, 2008), Sec. 3.3.

15.6 PROBLEMS

15.1 Determine the transition probability $P_{12}\left(t\right)$ for the case of a constant perturbation that turns on at $t=0$: $\hat{H}_p\left(t\right)=\hat{V}$. Use first-order perturbation theory.

15.2 A two-level system experiences a pulsed perturbation:

$$\hat{H}_p\left(t\right)=\begin{cases}\hat{V} & 0<t<T \\ 0 & t>T\end{cases}. \tag{15.57}$$

(a) Use first order perturbation theory to calculate the transition probability $P_{12}\left(t\right)$ for $t>T$. (b) Let $V=\gamma/T$, so that the "area" of the perturbation is constant, and let $T\rightarrow 0$; what is $P_{12}\left(t\right)$ in this case?

15.3 A particle with a charge q is in the ground state of an infinite square well of width L. An electric field $\boldsymbol{\mathcal{E}}=\mathcal{E}\boldsymbol{u}_x$ is turned on at $t=0$, and turned off at $t=T$. What is the probability that the charge ends up in the first-excited state after the field is turned off?

15.4* If the Hamiltonian of a system changes "suddenly" (i.e., on a timescale much faster than any timescale that the system can respond to), the state of the system before and after the change is the same; the system simply doesn't have time to change states. However, since the Hamiltonian has changed, so have its eigenstates. If the system was in an eigenstate of the Hamiltonian before the change, it will in general be in a superposition of eigenstates after.

If a 1-D infinite square well has its width suddenly increased by a factor of 10%, what is the probability that a particle in the ground state of the initial well will be found to be in the first excited state of the new well?

15.5* Calculate the probability that a system will make a transition from level 2 to level 1 $P_{21}\left(t\right)$ for a sinusoidal perturbation, assuming that the system is driven near resonance. Compare your result to $P_{12}\left(t\right)$ [eq. (15.29)].

15.6 An electric field of the form

$$\mathcal{E}\left(t\right)=\begin{cases}0 & t<0 \\ \mathcal{E}\left(1-e^{-t/\tau}\right) & 0<t\end{cases} \tag{15.58}$$

is applied to a two-level atom whose energy levels differ by $\Delta E=\hbar\omega_0$. Calculate the probability $P_{12}\left(t\right)$ that the atom will make a transition from the lower to the upper state in two limits: (a) $\tau>>1/\omega_0$, (b) $1/\omega_0>>\tau$.

15.7 A collection of hydrogen atoms are all in the $2p\ m = 0$ state. What is the lifetime of this state?

15.8* Solve the exact equations (15.9), using the electric-dipole Hamiltonian of eq. (15.33), and making the near-resonance approximation. Assume a hydrogen atom starts in its ground state, and consider only one excited state that is being driven near resonance. Show that the transition probability is given by

$$P_{12}(t) = \frac{|\mathrm{d}_z|^2 \mathcal{E}^2}{\hbar^2 \Omega_R^2} \sin^2(\Omega_R t/2), \tag{15.59}$$

where the generalized Rabi frequency Ω_R is[10]

$$\Omega_R = \sqrt{(\omega - \omega_0)^2 + \frac{|\mathrm{d}_z|^2 \mathcal{E}^2}{\hbar^2}}. \tag{15.60}$$

Show that this result reduces to the perturbative result in the appropriate limit. [Hints: (1) Use selection rules to simplify the initial differential equations. (2) If you get stuck, refer to complement 9.A.]

---

[10] The Rabi frequency is equal to the generalized Rabi frequency on resonance ($\omega = \omega_0$).

COMPLEMENT 15.A

Einstein's *A* and *B* Coefficients

Here we present Einstein's model of a collection of two-level atoms in thermal equilibrium with a black-body radiation field [15.A.1].

The basic model is depicted in fig. 15.4. A collection of two-level atoms have states whose energies are separated by $\hbar\omega_0$. There are N_1 atoms in level 1, N_2 atoms in level 2, and the atoms are in thermal equilibrium with a field whose spectral energy density is $u(\omega)$. There are three ways that atoms can make transitions between the two levels: absorption at the rate $B_{12}u(\omega_0)$, stimulated emission at the rate $B_{21}u(\omega_0)$, and spontaneous emission at the rate A_{21}.

We can write down differential equations for the numbers of atoms in the two levels; these equations are known as rate equations. Using fig. 15.4 as a guide, we see that these equations are

$$\frac{d}{dt}N_1 = -B_{12}u(\omega_0)N_1 + B_{21}u(\omega_0)N_2 + A_{21}N_2, \tag{15.A.1}$$

$$\frac{d}{dt}N_2 = B_{12}u(\omega_0)N_1 - B_{21}u(\omega_0)N_2 - A_{21}N_2 = -\frac{d}{dt}N_1. \tag{15.A.2}$$

If the atoms and the field are in thermal equilibrium, the time derivatives must be 0. Setting eq. (15.A.1) equal to 0, we see that

$$B_{12}u(\omega_0)N_1 = B_{21}u(\omega_0)N_2 + A_{21}N_2. \tag{15.A.3}$$

Solving for the spectral energy density, we find

$$u(\omega_0) = \frac{A_{21}N_2}{B_{12}N_1 - B_{21}N_2} = \frac{A_{21}}{B_{12}\left(\frac{N_1}{N_2}\right) - B_{21}} = \left(\frac{A_{21}}{B_{21}}\right)\frac{1}{\left(\frac{B_{12}}{B_{21}}\right)\left(\frac{N_1}{N_2}\right) - 1}. \tag{15.A.4}$$

The ratio of the numbers of atoms in the two energy states at temperature T (in Kelvin) are given by Boltzmann statistics:

$$\frac{N_1}{N_2} = \frac{e^{-E_1/k_BT}}{e^{-E_2/k_BT}} = e^{\left(E_2-E_1\right)/k_BT} = e^{\hbar\omega_0/k_BT}, \tag{15.A.5}$$

where $k_B = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant. Furthermore, in thermal equilibrium the field is that of a black-body, and the spectral energy density is given by Planck's blackbody radiation formula:

$$u\left(\omega_0\right) = \frac{\hbar\omega_0^3}{\pi^2 c^3}\frac{1}{e^{\hbar\omega_0/k_BT} - 1}. \tag{15.A.6}$$

Combining eqs. (15.A.4)–(15.A.6), we see that

$$\frac{\hbar\omega_0^3}{\pi^2 c^3}\frac{1}{e^{\hbar\omega_0/k_BT} - 1} = \left(\frac{A_{21}}{B_{21}}\right)\frac{1}{\left(\frac{B_{12}}{B_{21}}\right)e^{\hbar\omega_0/k_BT} - 1}. \tag{15.A.7}$$

For this equation to hold, clearly we must have

$$\frac{B_{12}}{B_{21}} = 1, \tag{15.A.8}$$

$$\frac{A_{21}}{B_{21}} = \frac{\hbar\omega_0^3}{\pi^2 c^3}, \tag{15.A.9}$$

which are equivalent to eqs. (15.47) and (15.48).

15.A.1 References

[15.A.1] A. Einstein, "On the quantum theory of radiation," Phys. Z. **18**, 121 (1917). An English translation is available in: D. ter Haar, *The Old Quantum Theory* (Pergamon Press, Oxford, 1967), p. 167.

15.A.2 PROBLEMS

15.A.1 An atom in an excited state can transition to the ground state via mechanisms other than just stimulated and spontaneous emission; collisions with other atoms, for example, can cause a decay. If the lifetime due to spontaneous emission is τ_{sp}, and the lifetime due to collisions is τ_c, what is the effective lifetime of the excited state? Recall that the lifetime is the inverse of the decay rate. You may ignore absorption and stimulated emission.

CHAPTER 16

Quantum Fields

In earlier chapters we talked about the polarization states of one- and two-photon fields, and we've alluded to a more sophisticated treatment of the field. Now it's time to provide that treatment. In this chapter we will treat the electromagnetic field as a fully quantum mechanical object. But before discussing fields, we'll take a necessary digression to discuss two different ways of treating time dependence in quantum mechanics: the Schrödinger and Heisenberg pictures.

16.1 THE SCHRÖDINGER AND HEISENBERG PICTURES OF QUANTUM MECHANICS

Previously we have distinguished between two types of time dependence: implicit and explicit. Implicit time dependence is the natural time evolution of a system; it occurs even in systems for which the Hamiltonian is time independent. Explicit time dependence typically means that there is a time-dependent external force acting on a system. We have been using the unitary time-evolution operator $\hat{U}(t)$ to propagate the state of a system forward in time:

$$\left|\psi(t)\right\rangle = \hat{U}(t)\left|\psi(0)\right\rangle . \tag{16.1}$$

By placing the time dependence in the state, we have been working in the Schrödinger picture of quantum mechanics.

Consider the expectation value of the observable A, which corresponds to the operator $\hat{A}_S$ (the subscript indicates that this is a Schrödinger-picture operator). This expectation value is

$$\begin{aligned}\langle A\rangle(t) &= \left\langle\psi(t)\right|\hat{A}_S\left|\psi(t)\right\rangle \\ &= \left\langle\psi(0)\right|\hat{U}^{\dagger}(t)\,\hat{A}_S\hat{U}(t)\left|\psi(0)\right\rangle \\ &= \left\langle\psi(0)\right|\hat{A}_H(t)\left|\psi(0)\right\rangle ,\end{aligned} \tag{16.2}$$

where we have defined

$$\hat{A}_H(t) \equiv \hat{U}^\dagger(t)\hat{A}_S\hat{U}(t). \tag{16.3}$$

Equation (16.2) tells us that we can calculate $\langle A\rangle(t)$ in two ways. The first is to use the time-independent operator $\hat{A}_S$ and the time-dependent state $|\psi(t)\rangle$; this is the Schrödinger picture. The second is to use the time-dependent operator $\hat{A}_H(t)$ and the time-independent state $|\psi(0)\rangle$; this is known as the Heisenberg picture, hence the subscript indicating that $\hat{A}_H(t)$ is a Heisenberg picture operator. Note that $\hat{U}(0)$ is the identity operator, so eq. (16.3) tells us that $\hat{A}_H(0)=\hat{A}_S$—the Schrödinger and Heisenberg-picture operators are the same at $t=0$. A Heisenberg picture operator will always commute with itself *at the same time*, but will not in general commute with itself at different times.

We obtain the equation of motion for a Heisenberg picture operator by differentiating eq. (16.3):

$$\frac{d}{dt}\hat{A}_H(t) \equiv \left[\frac{d}{dt}\hat{U}^\dagger(t)\right]\hat{A}_S\hat{U}(t)+\hat{U}^\dagger(t)\hat{A}_S\left[\frac{d}{dt}\hat{U}(t)\right]. \tag{16.4}$$

Here we have assumed that $\hat{A}_S$ has no explicit time dependence. Using eq. (9.9), and the fact that $\hat{U}(t)\hat{U}^\dagger(t)=\hat{1}$, this becomes

$$\begin{aligned}\frac{d}{dt}\hat{A}_H(t) &= \frac{i}{\hbar}\left[\hat{U}^\dagger(t)\hat{H}\hat{A}_S\hat{U}(t)-\hat{U}^\dagger(t)\hat{A}_S\hat{H}\hat{U}(t)\right]\\ &= \frac{i}{\hbar}\left[\hat{U}^\dagger(t)\hat{H}\hat{U}(t)\hat{U}^\dagger(t)\hat{A}_S\hat{U}(t)-\hat{U}^\dagger(t)\hat{A}_S\hat{U}(t)\hat{U}^\dagger(t)\hat{H}\hat{U}(t)\right]\\ &= \frac{i}{\hbar}\left[\hat{H}_H(t)\hat{A}_H(t)-\hat{A}_H(t)\hat{H}_H(t)\right]\\ &= \frac{i}{\hbar}\left[\hat{H}_H(t),\hat{A}_H(t)\right].\end{aligned} \tag{16.5}$$

This equation is known as the Heisenberg equation of motion. If there is no explicit time dependence to the Hamiltonian then $\hat{H}_H(t)=\hat{H}_S=\hat{H}$ (see problem 16.5), and the Heisenberg equation of motion becomes

$$\frac{d}{dt}\hat{A}_H(t)=\frac{i}{\hbar}\left[\hat{H},\hat{A}_H(t)\right]. \tag{16.6}$$

Let's look at an example.

EXAMPLE 16.1

Write down the Heisenberg equation of motion for the position operator of a harmonic oscillator.

The Hamiltonian for a harmonic oscillator is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2, \tag{16.7}$$

and it has no explicit time dependence. The Heisenberg equation for $\hat{x}$ is thus

$$\begin{aligned}\frac{d}{dt}\hat{x} &= \frac{i}{\hbar}\left[\frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2, \hat{x}\right] \\ &= \frac{i}{2m\hbar}\left[\hat{p}^2, \hat{x}\right] \\ &= \frac{i}{2m\hbar}\left(\hat{p}\left[\hat{p},\hat{x}\right] + \left[\hat{p},\hat{x}\right]\hat{p}\right) \\ &= \frac{1}{m}\hat{p},\end{aligned} \tag{16.8}$$

which looks like the classical equation of motion.

You may be asking yourself, "Why are we having a discussion of the Heisenberg picture now?" The answer is that the Heisenberg picture is the natural picture to use when discussing quantum fields, as we'll soon see.

16.2 THE FIELD HAMILTONIAN

Let's start with a discussion of classical free fields. Free fields exist in empty space, away from charges and currents. As such, Maxwell's equations for free fields are (ref. [16.1]):

$$\nabla \cdot \boldsymbol{\mathcal{E}} = 0, \tag{16.9}$$

$$\nabla \cdot \boldsymbol{\mathcal{B}} = 0, \tag{16.10}$$

$$\nabla \times \boldsymbol{\mathcal{E}} = -\frac{\partial}{\partial t}\boldsymbol{\mathcal{B}}, \tag{16.11}$$

$$\nabla \times \boldsymbol{\mathcal{B}} = \frac{1}{c^2}\frac{\partial}{\partial t}\boldsymbol{\mathcal{E}}. \tag{16.12}$$

If we take the curl of eq. (16.11), and then use a vector derivative identity and eq. (16.12), we see that

$$\begin{aligned}\nabla \times \nabla \times \boldsymbol{\mathcal{E}} &= -\frac{\partial}{\partial t}\nabla \times \boldsymbol{\mathcal{B}}, \\ \nabla\left(\nabla \cdot \boldsymbol{\mathcal{E}}\right) - \nabla^2\boldsymbol{\mathcal{E}} &= -\frac{\partial}{\partial t}\left(\frac{1}{c^2}\frac{\partial}{\partial t}\boldsymbol{\mathcal{E}}\right).\end{aligned} \tag{16.13}$$

Now we use eq. (16.9), and find that

$$\nabla^2\boldsymbol{\mathcal{E}}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\boldsymbol{\mathcal{E}}=0\,, \tag{16.14}$$

which is the wave equation for the electric field. By taking the curl of eq. (16.12), it is straightforward to prove that the magnetic field satisfies the same wave equation.

Consider the field inside a cubic cavity, with sides of length L. The purpose of the cavity is to allow us to use a mode expansion, and when we're all done we can let L become arbitrarily large if needed. We'll use periodic boundary conditions for the cavity, which means that the field on one side of the cavity is equal to the field on the opposite side. Here we'll consider just a single mode, and we'll write the electric field of this mode as

$$\boldsymbol{\mathcal{E}}(\boldsymbol{r},t)=\left[\mathcal{E}(t)e^{ikx}+\mathcal{E}^*(t)e^{-ikx}\right]\boldsymbol{u}_z\,. \tag{16.15}$$

To satisfy the periodic boundary conditions we must have

$$e^{ik(x+L)}=e^{ikx}\,, \tag{16.16}$$

so $k=m2\pi/L$ and m is an integer. Substituting eq. (16.15) into the wave equation [eq. (16.14)], we obtain the following equation for $\mathcal{E}(t)$:

$$\frac{\partial^2}{\partial t^2}\mathcal{E}(t)+c^2k^2\mathcal{E}(t)=0\,. \tag{16.17}$$

This is the equation obeyed by a simple harmonic oscillator, so the amplitude of a single-mode electric field behaves as a harmonic oscillator. The solution to eq. (16.17) is

$$\mathcal{E}(t)=\mathcal{E}(0)e^{-i\omega t}\,, \tag{16.18}$$

where $\omega=ck$. Given our expression for the electric field, we can use eq. (16.11) to determine the magnetic field, with the result being

$$\boldsymbol{\mathcal{B}}(\boldsymbol{r},t)=\frac{1}{c}\left[\mathcal{E}(t)e^{ikx}+\mathcal{E}^*(t)e^{-ikx}\right](-\boldsymbol{u}_y)\,. \tag{16.19}$$

The Hamiltonian determines the energy, and as such is given by

$$H=\frac{1}{2}\iiint\limits_V dV\left[\varepsilon_0\mathcal{E}^2(\boldsymbol{r},t)+\frac{1}{\mu_0}\mathcal{B}^2(\boldsymbol{r},t)\right], \tag{16.20}$$

where the integral extends over the inside of the cavity, and $V=L^3$ is the cavity volume (ref. [16.1], sec. 8.1). The square of the electric field is

$$\begin{aligned}\mathcal{E}^2&=\boldsymbol{\mathcal{E}}\bullet\boldsymbol{\mathcal{E}}\\&=\left[\mathcal{E}(t)e^{ikx}+\mathcal{E}^*(t)e^{-ikx}\right]\left[\mathcal{E}(t)e^{ikx}+\mathcal{E}^*(t)e^{-ikx}\right]\\&=\mathcal{E}^2(t)e^{i2kx}+\mathcal{E}^{*2}(t)e^{-i2kx}+\mathcal{E}^*(t)\mathcal{E}(t)+\mathcal{E}(t)\mathcal{E}^*(t)\,.\end{aligned} \tag{16.21}$$

We're dealing with classical fields, and the last two terms can be combined, but we've intentionally preserved the ordering of $\mathcal{E}(t)$ and $\mathcal{E}^*(t)$ for reasons that will become more obvious soon. Integrating this over the cavity volume, we find

$$\frac{1}{2}\iiint\limits_V dV\varepsilon_0\mathcal{E}^2(\boldsymbol{r},t)=\frac{1}{2}\varepsilon_0 V\left[\mathcal{E}^*(t)\mathcal{E}(t)+\mathcal{E}(t)\mathcal{E}^*(t)\right], \tag{16.22}$$

where we've used the fact that the integrals over the exponentials are 0. Similarly, the integral of the square of the magnetic field yields

$$\begin{aligned}\frac{1}{2}\iiint\limits_V dV\frac{1}{\mu_0}\mathcal{B}^2(\boldsymbol{r},t)&=\frac{1}{2}\frac{1}{\mu_0 c^2}V\left[\mathcal{E}^*(t)\mathcal{E}(t)+\mathcal{E}(t)\mathcal{E}^*(t)\right]\\&=\frac{1}{2}\varepsilon_0 V\left[\mathcal{E}^*(t)\mathcal{E}(t)+\mathcal{E}(t)\mathcal{E}^*(t)\right],\end{aligned} \tag{16.23}$$

where we've used the fact that $1/\mu_0\varepsilon_0=c^2$. The Hamiltonian of the field is thus

$$H=\varepsilon_0 V\left[\mathcal{E}^*(t)\mathcal{E}(t)+\mathcal{E}(t)\mathcal{E}^*(t)\right]. \tag{16.24}$$

Let's define the real quantities $\mathcal{X}$ and $\mathcal{P}$, where

$$\mathcal{E}(t)=\frac{1}{2\sqrt{\varepsilon_0 V}}\left[\omega\mathcal{X}(t)+i\mathcal{P}(t)\right]. \tag{16.25}$$

This means that

$$\mathcal{E}^*(t)\mathcal{E}(t)=\mathcal{E}(t)\mathcal{E}^*(t)=\frac{1}{4\varepsilon_0 V}\left[\omega^2\mathcal{X}^2(t)+\mathcal{P}^2(t)\right], \tag{16.26}$$

and the Hamiltonian is then

$$H=\frac{\mathcal{P}^2(t)}{2}+\frac{1}{2}\omega^2\mathcal{X}^2(t). \tag{16.27}$$

This Hamiltonian looks like the Hamiltonian for a harmonic oscillator with unit mass. It is formally equivalent to the Hamiltonian of a harmonic oscillator, as long as $\mathcal{X}$ and $\mathcal{P}$ are related as we would expect for an oscillator with unit mass:

$$\frac{d}{dt}\mathcal{X}=\mathcal{P}. \tag{16.28}$$

You'll show that this relationship does indeed hold in problem 16.9.

Thus, we've shown that the Hamiltonian for the amplitude of a single mode of the electromagnetic field is formally equivalent to that of a simple harmonic oscillator.

16.3 FIELD OPERATORS

To treat the field quantum mechanically, we assume that the behavior of each mode of the field is equivalent to that of a quantum-mechanical harmonic oscillator. We assume that the amplitude of the electric field is an operator, so that the classical Hamiltonian [eq. (16.24)] becomes the Hamiltonian operator:

$$\hat{H} = \varepsilon_0 V\left[\hat{\mathcal{E}}^\dagger(t)\hat{\mathcal{E}}(t) + \hat{\mathcal{E}}(t)\hat{\mathcal{E}}^\dagger(t)\right]. \tag{16.29}$$

We were careful to preserve the ordering of the field amplitudes when deriving eq. (16.24) so that we have the correct ordering here. The field operator $\hat{\mathcal{E}}(t)$ is a Heisenberg picture operator.

The Hamiltonian of a harmonic oscillator is given by eq. (12.10):

$$\hat{H} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right). \tag{16.30}$$

Using the commutation relation

$$\left[\hat{a}, \hat{a}^\dagger\right] = \hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} = 1, \tag{16.31}$$

this can be rewritten as

$$\hat{H} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\left[\hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a}\right]\right) = \frac{\hbar\omega}{2}\left(\hat{a}^\dagger\hat{a} + \hat{a}\hat{a}^\dagger\right). \tag{16.32}$$

To establish the equivalence between the field mode and the harmonic oscillator, we assume that the Hamiltonians in eqs. (16.29) and (16.32) are the same. For this to be the case we must have

$$\hat{\mathcal{E}}(t) = \sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\,\hat{a}(t). \tag{16.33}$$

The quantum-mechanical equivalent of the single-mode electric field in eq. (16.15) is thus

$$\begin{aligned}\hat{\boldsymbol{\mathcal{E}}}(\boldsymbol{r},t) &= \sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\left[\hat{a}(t)e^{ikx} + \hat{a}^\dagger(t)e^{-ikx}\right]\boldsymbol{u}_z \\ &= \sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\left[\hat{a}e^{i(kx-\omega t)} + \hat{a}^\dagger e^{-i(kx-\omega t)}\right]\boldsymbol{u}_z,\end{aligned} \tag{16.34}$$

where we have used the result of problem 16.4 [eq. (16.118)] for the time dependence. The operator for the magnetic field is the quantum-mechanical equivalent of eq. (16.19):

$$\hat{\boldsymbol{\mathcal{B}}}(\boldsymbol{r},t) = \frac{1}{c}\sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\left[\hat{a}(t)e^{ikx} + \hat{a}^\dagger(t)e^{-ikx}\right]\left(-\boldsymbol{u}_y\right), \tag{16.35}$$

In chapter 12 we alluded to the fact that a single mode of the electromagnetic field could be treated as a harmonic oscillator, and here we see the proof of that.

The operators $\hat{a}^\dagger$ and $\hat{a}$ are creation and annihilation operators for the field, and $\hat{n} = \hat{a}^\dagger \hat{a}$ is the corresponding photon number operator. We can rewrite the Hamiltonian of eq. (16.30) in terms of the number operator as

$$\hat{H} = \hbar\omega\left(\hat{n} + \frac{1}{2}\right). \tag{16.36}$$

One last set of operators that we will find useful are the field-quadrature-amplitude operators $\hat{X}_\theta$:

$$\hat{X}_\theta \equiv \frac{1}{2}\left(\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}\right). \tag{16.37}$$

Each value of θ yields a different quadrature amplitude. The quadrature amplitudes are observables.

16.3.1 Multimode Fields

The modes of a field are described by their wave vectors $\boldsymbol{k}$ and polarizations $\boldsymbol{\varepsilon}_{\boldsymbol{k}}$; for each wave vector there are two orthogonal polarizations, so we label the polarization vectors of a multimode field as $\boldsymbol{\varepsilon}_{\boldsymbol{k}s}$, where $s = 1,2$. The wave vector and the polarization must be orthogonal: $\boldsymbol{k} \bullet \boldsymbol{\varepsilon}_{\boldsymbol{k}s} = 0$ (see problem 16.6). The two polarizations are orthogonal as well: $\boldsymbol{\varepsilon}^*_{\boldsymbol{k}s} \bullet \boldsymbol{\varepsilon}_{\boldsymbol{k}s'} = \delta_{ss'}$. We are assuming periodic boundary conditions in a cubic cavity, so the components of the wave vector satisfy

$$k_j = m_j \frac{2\pi}{L}, \tag{16.38}$$

where $j = x, y, z$, and m_j is an integer. For each wave vector the frequency of the field is determined by the dispersion relation $\omega_{\boldsymbol{k}} = |\boldsymbol{k}|c$; note that $\omega_{\boldsymbol{k}} = \omega_{-\boldsymbol{k}}$.

The Hamiltonian of a multimode field is that of a collection of independent harmonic oscillators (ref. [16.2], ch. 4). The total field is then given by summing over the single-mode results obtained in eq. (16.34):

$$\begin{aligned}\hat{\boldsymbol{\mathcal{E}}}(\boldsymbol{r},t) &= \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\boldsymbol{k},s} \sqrt{\omega_{\boldsymbol{k}}} \left[\hat{a}_{\boldsymbol{k}s}(t) e^{i\boldsymbol{k}\bullet\boldsymbol{r}} \boldsymbol{\varepsilon}_{\boldsymbol{k}s} + \hat{a}^\dagger_{\boldsymbol{k}s}(t) e^{-i\boldsymbol{k}\bullet\boldsymbol{r}} \boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right] \\ &= \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\boldsymbol{k},s} \sqrt{\omega_{\boldsymbol{k}}} \left[\hat{a}_{\boldsymbol{k}s} e^{i(\boldsymbol{k}\bullet\boldsymbol{r} - \omega_k t)} \boldsymbol{\varepsilon}_{\boldsymbol{k}s} + \hat{a}^\dagger_{\boldsymbol{k}s} e^{-i(\boldsymbol{k}\bullet\boldsymbol{r} - \omega_k t)} \boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right].\end{aligned} \tag{16.39}$$

The corresponding magnetic field is

$$\hat{\boldsymbol{\mathcal{B}}}(\boldsymbol{r},t) = \frac{1}{c}\sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\boldsymbol{k},s} \sqrt{\omega_{\boldsymbol{k}}} \left[\hat{a}_{\boldsymbol{k}s}(t) e^{i\boldsymbol{k}\bullet\boldsymbol{r}} \left(\boldsymbol{u}_{\boldsymbol{k}} \times \boldsymbol{\varepsilon}_{\boldsymbol{k}s}\right) + \hat{a}^\dagger_{\boldsymbol{k}s}(t) e^{-i\boldsymbol{k}\bullet\boldsymbol{r}} \left(\boldsymbol{u}_{\boldsymbol{k}} \times \boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right)\right]. \tag{16.40}$$

where $\boldsymbol{u}_{\boldsymbol{k}}$ is the unit vector that points along the direction of $\boldsymbol{k}$. Because the field modes are independent, the operators corresponding to different modes commute:

$$\left[\hat{a}_{\boldsymbol{k}s}(t), \hat{a}^\dagger_{\boldsymbol{k}'s'}(t)\right] = \delta_{\boldsymbol{k}\boldsymbol{k}'}\delta_{ss'}. \tag{16.41}$$

16.4 FIELD STATES

The electromagnetic field is equivalent to a collection of harmonic oscillators, so the states of the field are harmonic oscillator states. The eigenstates of the Hamiltonian are the eigenstates of the number operator, which are the number states or the Fock states:

$$\hat{n}\left|n\right\rangle = n\left|n\right\rangle,\ n = 0,1,2,\ldots. \tag{16.42}$$

A field in the Fock state $\left|n\right\rangle$ contains exactly n photons. The creation and annihilation operators perform the following operations on the Fock states [eqs (12.28) and (12.29)]:

$$\hat{a}^{\dagger}\left|n\right\rangle = \sqrt{n+1}\left|n+1\right\rangle, \tag{16.43}$$

$$\hat{a}\left|n\right\rangle = \sqrt{n}\left|n-1\right\rangle. \tag{16.44}$$

The ground state of the field is $\left|0\right\rangle$, which contains no photons. This state is known as the vacuum.

When describing multimode fields, we need to specify the state of each mode. We will use the following conventions for doing so. The vacuum field $\left|0\right\rangle$ contains 0 photons in every mode:

$$\left|0\right\rangle = \prod_{\mathbf{k},s} \left|0\right\rangle_{\mathbf{k}s}. \tag{16.45}$$

Modes for which the state is not indicated are assumed to be in the vacuum. For example, the state $\left|n_{\mathbf{k}'s'}\right\rangle$ has n photons in mode $\mathbf{k}'s'$, and no photons in any other mode:

$$\left|n_{\mathbf{k}'s'}\right\rangle = \left|n\right\rangle_{\mathbf{k}'s'} \otimes \prod_{\substack{\mathbf{k},s\\ \neq \mathbf{k}'s'}} \left|0\right\rangle_{\mathbf{k}s}. \tag{16.46}$$

Once we know the state of the field we can calculate probabilities of measurements or expectation values, as the following example indicates.

EXAMPLE 16.2

Calculate the expectation value of $\hat{H}$ for the vacuum.

We can write the Hamiltonian for a multimode field as

$$\hat{H} = \sum_{\mathbf{k},s} \hbar\omega_{\mathbf{k}}\left(\hat{n}_{\mathbf{k}s} + \frac{1}{2}\right). \tag{16.47}$$

Taking the expectation value of this expression in a vacuum, we find

$$\begin{aligned}\left\langle \hat{H}\right\rangle &= \sum_{\mathbf{k},s} \hbar\omega_{\mathbf{k}}\left(\left\langle 0\right|\hat{n}_{\mathbf{k}s}\left|0\right\rangle + \frac{1}{2}\right)\\ &= \sum_{\mathbf{k},s} \frac{\hbar\omega_{\mathbf{k}}}{2}.\end{aligned} \tag{16.48}$$

The sum over modes contains, at least in principle, modes of infinite frequency. As such, the sum in eq. (16.48) diverges. This infinite energy is known as the zero-point energy of the field. As you can see, it's related to the zero-point energy of the harmonic oscillator. While a number of divergences in quantum electrodynamics can be explained using a procedure known as renormalization, this one is frankly embarrassing. It may be plausible to introduce a high-frequency cut-off that keeps the sum in eq. (16.48) from diverging, but I know of no completely satisfactory way around the zero-point divergence. However, despite this, measurements performed on the field are not affected by this infinite "background" energy, as we'll see in sec. 16.6.

The following example illuminates another interesting feature of the vacuum.

EXAMPLE 16.3

Calculate the expectation value and variance of the electric field of the vacuum. Consider just a single mode of the field.

Using a single mode of the field operator of eq. (16.39), we have for the expectation value:

$$\begin{aligned}\left\langle \hat{\mathcal{E}}(\boldsymbol{r},t)\right\rangle &= \sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\left[\langle 0|\hat{a}|0\rangle e^{i(\boldsymbol{k}\bullet\boldsymbol{r}-\omega t)}\boldsymbol{\varepsilon}+\langle 0|\hat{a}^\dagger|0\rangle e^{-i(\boldsymbol{k}\bullet\boldsymbol{r}-\omega t)}\boldsymbol{\varepsilon}^*\right]\\ &= \sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\left[0+0\right]\\ &=0\,.\end{aligned} \tag{16.49}$$

It shouldn't be too surprising that the expectation value of the field of a vacuum state is 0.

The variance is given by

$$\begin{aligned}(\Delta\mathcal{E})^2 &= \left\langle \hat{\mathcal{E}}^2(\boldsymbol{r},t)\right\rangle - \underset{=0}{\cancel{\left\langle \mathcal{E}(\boldsymbol{r},t)\right\rangle^2}}\\ &= \left\langle \hat{\mathcal{E}}(\boldsymbol{r},t)\bullet\hat{\mathcal{E}}(\boldsymbol{r},t)\right\rangle\\ &= \frac{\hbar\omega}{2\varepsilon_0 V}\left\langle\left[\hat{a}^2 e^{i2(\boldsymbol{k}\bullet\boldsymbol{r}-\omega t)}(\boldsymbol{\varepsilon}\bullet\boldsymbol{\varepsilon})+\left(\hat{a}^\dagger\right)^2 e^{-i2(\boldsymbol{k}\bullet\boldsymbol{r}-\omega t)}\left(\boldsymbol{\varepsilon}^*\bullet\boldsymbol{\varepsilon}^*\right)+\hat{a}\hat{a}^\dagger+\hat{a}^\dagger\hat{a}\right]\right\rangle\\ &= \frac{\hbar\omega}{2\varepsilon_0 V}\Big[\langle 0|\hat{a}^2|0\rangle e^{i2(\boldsymbol{k}\bullet\boldsymbol{r}-\omega t)}(\boldsymbol{\varepsilon}\bullet\boldsymbol{\varepsilon})+\langle 0|\left(\hat{a}^\dagger\right)^2|0\rangle e^{-i2(\boldsymbol{k}\bullet\boldsymbol{r}-\omega t)}\left(\boldsymbol{\varepsilon}^*\bullet\boldsymbol{\varepsilon}^*\right)\\ &\quad+\langle 0|\hat{a}\hat{a}^\dagger|0\rangle+\langle 0|\hat{a}^\dagger\hat{a}|0\rangle\Big]\\ &= \frac{\hbar\omega}{2\varepsilon_0 V}\left[0+0+1+0\right]\\ &= \frac{\hbar\omega}{2\varepsilon_0 V}\,.\end{aligned} \tag{16.50}$$

Note that the variance of the vacuum field is nonzero! The vacuum field has a mean of zero, but nonzero fluctuations. This has important consequences, as we'll see when discussing spontaneous emission. Other interesting effects that can be traced to the vacuum include the Lamb shift and the Casimir force (refs. [16.3] and [16.4]).

16.4.1 Coherent States

The coherent states $|\alpha\rangle$, which we discussed in sec. 12.5, are also very important field states. The coherent states are the eigenstates of the annihilation operator

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle. \tag{16.51}$$

Since the annihilation operator is not Hermitian, α can be any complex number. The coherent states can be written in terms of the Fock states as [eq. (12.59)]:

$$|\alpha\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle. \tag{16.52}$$

The coherent states are the closest quantum mechanical equivalent to a classical field with a constant amplitude. The field emitted by a stable laser is reasonably well described by a coherent state.

The probability of measuring the field to contain n photons is $P(n)$, and for a field in a coherent state $|\alpha\rangle$ this probability is given by

$$\begin{aligned} P(n) &= |\langle n|\alpha\rangle|^2 \\ &= \left|e^{-|\alpha|^2/2}\frac{\alpha^n}{\sqrt{n!}}\right|^2 \\ &= \frac{|\alpha|^{2n}}{n!}e^{-|\alpha|^2} \,. \end{aligned} \tag{16.53}$$

This equation represents a Poisson distribution of photon numbers. In the problems you'll show that this distribution has a mean of $\langle n\rangle = |\alpha|^2$. Figure 16.1 displays $P(n)$ for coherent state fields with different mean numbers of photons. If $\langle n\rangle < 1$ the distribution peaks at $n = 0$, otherwise the peak is located at $n \simeq \langle n\rangle$. If $\langle n\rangle$ is large, the distribution is approximately Gaussian.

The standard deviation of the photon number for a coherent state field is $\Delta n = \langle n\rangle^{1/2}$. The fluctuations in the number of photons get larger as the average number increases, as would be expected. However, if we normalize the standard deviation by the mean, we find that

$$\frac{\Delta n}{\langle n\rangle} = \frac{1}{\langle n\rangle^{1/2}}, \tag{16.54}$$

and the relative fluctuations decrease as the mean increases. The "signal-to-noise" ratio is the inverse of eq. (16.54), and it improves with increasing average photon number.

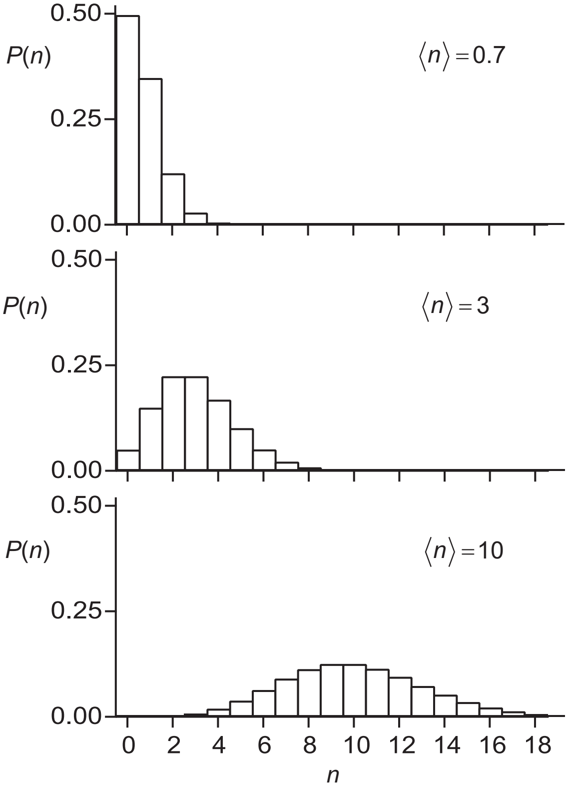


Fig 16.1 Photon number distributions, corresponding to fields in coherent states with different mean numbers of photons.

In the problems you'll show that for a coherent state, the standard deviation of the quadrature amplitudes X_θ are 1/2, independent of θ. This is consistent with the results in eqs. (12.68) and (12.70), as it should be, because the operators $\hat{X} \equiv \hat{X}_0$ and $\hat{P} \equiv \hat{X}_{\pi/2}$ defined in sec. 12.5 are merely two particular quadrature amplitudes. It is possible to measure quadrature amplitudes using a balanced homodyne detector (problem 16.26). Indeed, the measurements of the electric field amplitudes E_θ shown in fig. 12.7 are precisely such measurements: $E_\theta = X_\theta$.

16.4.2 Squeezed States

For a coherent state, the standard deviation of the quadrature amplitudes X and P are both 1/2. In the dimensionless units we're using here, their uncertainty product of 1/4 is the smallest allowed by the indeterminacy relation (problem 12.17). It can be proved that no classical state can have a quadrature-amplitude uncertainty of less than 1/2, while certain nonclassical states, called squeezed states, can. But of course the indeterminacy relation still applies, so if the uncertainty in X is less than 1/2, the uncertainty in

P must be greater. These states are called squeezed, because in the X-P plane the uncertainty in X and P for a coherent state is a circle (fig. 12.6), whereas for a squeezed state the circle is squeezed into an ellipse. Mathematically, squeezed states $\left|r\right\rangle$ are generated by applying the unitary squeezing operator $\hat{S}\left(r\right)$ to the vacuum state $\left|0\right\rangle$:

$$\left|r\right\rangle = \hat{S}\left(r\right)\left|0\right\rangle . \tag{16.55}$$

The squeezing operator is

$$\hat{S}\left(r\right) = \exp\left[\frac{1}{2}r\left(\hat{a}^2 - \hat{a}^{\dagger 2}\right)\right], \tag{16.56}$$

and r is referred to as the squeezing parameter.

You’ll examine several properties of squeezed states in the problems. You’ll show that the uncertainties in the quadrature amplitudes for a squeezed state are

$$\Delta X = \frac{1}{2}e^{-r} \tag{16.57}$$

$$\Delta P = \frac{1}{2}e^{r} . \tag{16.58}$$

You’ll also show that the photon number distribution for a squeezed state contains only even numbers of photons. This suggests that a parametric downconversion source, which produces pairs of photons, could be used to generate a squeezed state. Indeed, parametric downconversion is one of the more widely used techniques to generate squeezed light.

Squeezed states are interesting because the fluctuations of one quadrature amplitude are smaller than those for any classical state. If information can be encoded in this quadrature, the information can be transmitted with lower noise using squeezed light than it can be using ordinary laser light.

16.5 FULLY QUANTUM MECHANICAL ATOM-FIELD INTERACTIONS

The Hamiltonian of an atom-field system consists of three parts: the Hamiltonian of the atom, the Hamiltonian of the field, and the Hamiltonian of the interaction between them. We have already described the Hamiltonians of atoms and fields, and now we are interested in the interaction. To describe this interaction we will use the electric-dipole Hamiltonian that we used in sec. 15.3, only now we will use a quantum field rather than a classical field. The electric-dipole Hamiltonian is thus

$$\hat{H}_{ED}\left(t\right) = -\hat{\mathbf{d}}\bullet\hat{\boldsymbol{\mathcal{E}}}\left(\boldsymbol{r} = 0, t\right), \tag{16.59}$$

where we are assuming that the atom is placed at the origin.

The field is time dependent, and we will use time-dependent perturbation theory to describe the interaction of the atom and the field. Equations (15.22) and (15.23) tell us that the probability of making a transition from some initial state at time $t = 0$ to a different final state at time t is

$$P_{if}(t) = \left|c_f(t)\right|^2, \tag{16.60}$$

where

$$c_f(t) = -\frac{i}{\hbar}\int_0^t dt'\, e^{i\omega_{fi}t'} H_{ED\,fi}(t'). \tag{16.61}$$

The basis states we will use are those of the combined atom-field system. We will assume that only two levels of the atom are nearly resonant with the field, and that the atom is initially in the excited state $|2\rangle$, and makes a transition to the lower state $|1\rangle$ (see fig. 15.1). As such $\omega_{fi} = (E_1 - E_2)/\hbar = -\omega_0$, and eq. (16.61) becomes

$$c_f(t) = -\frac{i}{\hbar}\int_0^t dt'\, e^{-i\omega_0 t'} H_{ED\,fi}(t'). \tag{16.62}$$

We will assume that the field at $t = 0$ is $|n_{\boldsymbol{k}'s'}\rangle$ (n photons in mode $\boldsymbol{k}'s'$, and vacuum in other modes). At the moment we will not specify the final state of the field, and will designate it as $|F_f\rangle$. The matrix element in eq. (16.62) is thus

$$\begin{aligned} H_{ED\,fi}(t) &= -\langle 1, F_f|\hat{\mathbf{d}}\cdot\hat{\boldsymbol{\mathcal{E}}}(0,t)|2, n_{\boldsymbol{k}'s'}\rangle \\ &= -\langle 1|\hat{\mathbf{d}}|2\rangle\cdot\langle F_f|\hat{\boldsymbol{\mathcal{E}}}(0,t)|n_{\boldsymbol{k}'s'}\rangle \\ &= -\mathbf{d}^*\cdot\langle F_f|\hat{\boldsymbol{\mathcal{E}}}(0,t)|n_{\boldsymbol{k}'s'}\rangle, \end{aligned} \tag{16.63}$$

where $\mathbf{d}^*$ is the dipole-moment matrix element.[1]

Using eqs. (16.39), (16.43), and (16.44), the field matrix element in eq. (16.63) becomes

$$\begin{aligned} \langle F_f|\hat{\boldsymbol{\mathcal{E}}}(0,t)|n_{\boldsymbol{k}'s'}\rangle &= \sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sum_{\boldsymbol{k},s}\sqrt{\omega_{\boldsymbol{k}}}\,\langle F_f|\left[\hat{a}_{\boldsymbol{k}s}e^{-i\omega_k t}\boldsymbol{\varepsilon}_{\boldsymbol{k}s} + \hat{a}^\dagger_{\boldsymbol{k}s}e^{i\omega_k t}\boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right]|n_{\boldsymbol{k}'s'}\rangle \\ &= \sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sqrt{n_{\boldsymbol{k}'s'}\omega_{\boldsymbol{k}'}}\,\langle F_f|n_{\boldsymbol{k}'s'}-1\rangle e^{-i\omega_{k'}t}\boldsymbol{\varepsilon}_{\boldsymbol{k}'s'} \\ &\quad + \sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sqrt{(n_{\boldsymbol{k}'s'}+1)\omega_{\boldsymbol{k}'}}\,\langle F_f|n_{\boldsymbol{k}'s'}+1\rangle e^{i\omega_{k'}t}\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'} \\ &\quad + \sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sum_{\substack{\boldsymbol{k},s \\ \neq\boldsymbol{k}'s'}}\sqrt{\omega_{\boldsymbol{k}}}\,\langle F_f|n_{\boldsymbol{k}'s'}, 1_{\boldsymbol{k}s}\rangle e^{i\omega_k t}\boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}. \end{aligned} \tag{16.64}$$

There are three terms in this equation. The first term corresponds to absorption (annihilation) of a photon in mode $\boldsymbol{k}'s'$. The second term corresponds to emission (creation) of a photon in this same mode. The final term is a sum, corresponding to the creation of a photon in one of the other modes which were initially in the vacuum. Note that there are no terms corresponding to annihilation of the vacuum, since these terms are 0.

---

[1] We're using $\mathbf{d}^* = \langle 1|\hat{\mathbf{d}}|2\rangle$, because in chapter 15 we used $\mathbf{d} = \langle 2|\hat{\mathbf{d}}|1\rangle$.

Substituting eqs. (16.63) and (16.64) into eq. (16.62), we find that

$$\begin{aligned} c_f(t) &= \frac{i}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sqrt{n_{\boldsymbol{k}'s'}\omega_{\boldsymbol{k}'}}\left\langle F_f \middle| n_{\boldsymbol{k}'s'}-1\right\rangle\left(\mathbf{d}^*\bullet\boldsymbol{\varepsilon}_{\boldsymbol{k}'s'}\right)\int_0^t dt'\, e^{-i\omega_0 t'}e^{-i\omega_{k'}t} \\ &+\frac{i}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sqrt{\left(n_{\boldsymbol{k}'s'}+1\right)\omega_{\boldsymbol{k}'}}\left\langle F_f \middle| n_{\boldsymbol{k}'s'}+1\right\rangle\left(\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right)\int_0^t dt'\, e^{-i\omega_0 t'}e^{i\omega_{k'}t} \\ &+\frac{i}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sum_{\substack{\boldsymbol{k},s\\ \neq\boldsymbol{k}'s'}}\sqrt{\omega_{\boldsymbol{k}}}\left\langle F_f \middle| n_{\boldsymbol{k}'s'},1_{\boldsymbol{k}s}\right\rangle\left(\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right)\int_0^t dt'\, e^{-i\omega_0 t'}e^{i\omega_k t}\,. \end{aligned} \tag{16.65}$$

Now we make the near resonance approximation; we assume that the field must be nearly resonant with the atomic transition, which means that $\omega_{k'} \simeq \omega_0$. With this approximation the integrand in the first term of eq. (16.65) oscillates rapidly, and the integral averages to 0. Physically this says that the atom cannot make a transition from the upper state to the lower state by absorbing a photon—this process does not conserve energy. Performing the other two integrals, we are left with

$$\begin{aligned} c_f(t) &= \frac{1}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sqrt{\left(n_{\boldsymbol{k}'s'}+1\right)\omega_{\boldsymbol{k}'}}\left\langle F_f \middle| n_{\boldsymbol{k}'s'}+1\right\rangle\left(\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right)\left[\frac{e^{i\left(\omega_{k'}-\omega_0\right)t}-1}{\left(\omega_{\boldsymbol{k}'}-\omega_0\right)}\right] \\ &+\frac{1}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sum_{\substack{\boldsymbol{k},s\\ \neq\boldsymbol{k}'s'}}\sqrt{\omega_{\boldsymbol{k}}}\left\langle F_f \middle| n_{\boldsymbol{k}'s'},1_{\boldsymbol{k}s}\right\rangle\left(\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right)\left[\frac{e^{i\left(\omega_k-\omega_0\right)t}-1}{\left(\omega_{\boldsymbol{k}}-\omega_0\right)}\right]. \end{aligned} \tag{16.66}$$

16.5.1 Stimulated Emission

There are two important cases to examine. For stimulated emission, a photon is added to the mode already containing photons, so $\left|F_f\right\rangle = \left|n_{k's'}+1\right\rangle$, which yields

$$c_f(t) = \frac{1}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sqrt{\left(n_{\boldsymbol{k}'s'}+1\right)\omega_{\boldsymbol{k}'}}\left(\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right)\left[\frac{e^{i\left(\omega_{k'}-\omega_0\right)t}-1}{\left(\omega_{\boldsymbol{k}'}-\omega_0\right)}\right]. \tag{16.67}$$

The transition probability is then

$$\begin{aligned} P_{if}(t) &= \left|c_f(t)\right|^2 \\ &= \frac{\left(n_{\boldsymbol{k}'s'}+1\right)\hbar\omega_{\boldsymbol{k}'}}{2\varepsilon_0\hbar^2 V}\left|\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right|^2\left|\frac{e^{i\left(\omega_{k'}-\omega_0\right)t}-1}{\left(\omega_{\boldsymbol{k}'}-\omega_0\right)}\right|^2 \\ &= \frac{2\left(n_{\boldsymbol{k}'s'}+1\right)\hbar\omega_{\boldsymbol{k}'}}{\varepsilon_0\hbar^2 V}\left|\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right|^2\frac{\sin^2\left[\left(\omega_{\boldsymbol{k}'}-\omega_0\right)t/2\right]}{\left(\omega_{\boldsymbol{k}'}-\omega_0\right)^2} \\ &= \frac{\left(n_{\boldsymbol{k}'s'}+1\right)\hbar\omega_{\boldsymbol{k}'}t^2}{2\varepsilon_0\hbar^2 V}\left|\mathbf{d}^*\bullet\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right|^2\operatorname{sinc}^2\left[\left(\omega_{\boldsymbol{k}'}-\omega_0\right)t/2\right]. \end{aligned} \tag{16.68}$$

For simplicity we'll assume that mode $\boldsymbol{k}'s'$ is linearly polarized, and we'll choose our z-axis to point along its polarization direction, so $\boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'} = \boldsymbol{u}_z$, and

$$P_{if}(t) = \frac{\left(n_{\boldsymbol{k}'s'}+1\right)\hbar\omega_{\boldsymbol{k}'}t^2}{2\varepsilon_0\hbar^2 V}\left|\mathrm{d}_z\right|^2 \mathrm{sinc}^2\left[\left(\omega_{\boldsymbol{k}'}-\omega_0\right)t/2\right]. \tag{16.69}$$

To compare the quantum result we have just obtained with the semiclassical result of sec. 15.3, we note that since mode $\boldsymbol{k}'s'$ contains $n_{\boldsymbol{k}'s'}+1$ photons, its energy density is

$$U = \left(n_{\boldsymbol{k}'s'}+1\right)\hbar\omega_{\boldsymbol{k}'}/V. \tag{16.70}$$

Substituting this into eq. (16.69) tells us that

$$P_{if}(t) = \frac{\left|\mathrm{d}_z\right|^2 U t^2}{2\varepsilon_0\hbar^2}\mathrm{sinc}^2\left[\left(\omega_{\boldsymbol{k}'}-\omega_0\right)t/2\right]. \tag{16.71}$$

This expression is identical to that in eq. (15.38), which represents the semiclassical stimulated emission probability for a single atom in a monochromatic (i.e., single-mode) field.[2]

From Eq. (16.71), the quantum calculation of the Einstein B coefficient (the stimulated emission rate) proceeds in exactly the same manner as the semiclassical calculation in Sec. 15.3. The final expression for the Einstein B coefficient is the same as that in Eq. (15.49). Thus, the fully quantum-mechanical treatment of stimulated emission agrees with the semiclassical treatment.

16.5.2 Spontaneous Emission

While the semiclassical model works fine for stimulated emission, recall that this model cannot predict the spontaneous emission rate; it predicts that there is no spontaneous emission. In Sec. 15.3 we obtained the spontaneous emission rate with help from Einstein's thermodynamic model, not directly from the semiclassical model. As we'll now see, a fully quantum-mechanical theory can directly determine the spontaneous emission rate.

Start with Eq. (16.66) and assume that $n_{\boldsymbol{k}'s'} = 0$. In other words, none of the modes initially contain photons, and the field is in a vacuum state $\left|0\right\rangle$. All of the modes are now on an equal footing, and Eq. (16.66) becomes

$$c_f(t) = \frac{1}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sum_{\boldsymbol{k},s}\sqrt{\omega_{\boldsymbol{k}}}\left\langle F_f \middle| 1_{\boldsymbol{k}s}\right\rangle\left(\mathbf{d}^*\boldsymbol{\cdot}\boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right)\left[\frac{e^{i(\omega_k-\omega_0)t}-1}{\left(\omega_{\boldsymbol{k}}-\omega_0\right)}\right]. \tag{16.72}$$

This is 0 unless the final field state $\left|F_f\right\rangle = \left|1_{\boldsymbol{k}'s'}\right\rangle$ (i.e., the atom spontaneously emits a single photon into mode $\boldsymbol{k}'s'$), in which case

---

[2] Actually, eq. (15.38) represents the semiclassical absorption probability, but in problem 15.5 you showed that the probability of absorption and stimulated emission are the same.

$$c_f(t) = \frac{1}{\hbar}\sqrt{\frac{\hbar}{2\varepsilon_0 V}}\sqrt{\omega_{\boldsymbol{k}'}}\left(\mathbf{d}^* \bullet \boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right)\left[\frac{e^{i(\omega_{\boldsymbol{k}'}-\omega_0)t}-1}{\left(\omega_{\boldsymbol{k}'}-\omega_0\right)}\right]. \tag{16.73}$$

This is the same as eq (16.67), only with $n_{\boldsymbol{k}'s'} = 0$. Thus, the probability for the atom to spontaneously emit a photon into mode $\boldsymbol{k}'s'$ is

$$P_{if}(t) = \left|c_f(t)\right|^2 = \frac{t^2}{2\varepsilon_0 \hbar V}\omega_{\boldsymbol{k}'}\left|\mathbf{d}^* \bullet \boldsymbol{\varepsilon}^*_{\boldsymbol{k}'s'}\right|^2 \operatorname{sinc}^2\left[\left(\omega_{\boldsymbol{k}'}-\omega_0\right)t/2\right]. \tag{16.74}$$

Before going on to calculate the spontaneous emission rate, I want you to note one important thing. Equation (16.74) says that there is a nonzero probability that an atom will make a transition from the upper level to the lower level, in the absence of an applied field. This result is different from that of the semiclassical model of sec. 15.3. The reason for the difference is that in the semiclassical model, if there is no applied field, there is no mechanism by which the atom can decay—the Hamiltonian describing the interaction between the atom and the field is 0. In the quantum calculation the interaction Hamiltonian is nonzero, even though the field is in a vacuum state. In some sense the vacuum stimulates the emission of a photon (see also example 16.3).

Equation (16.74) gives us the probability that the atom will emit a photon into a particular mode. However, to find the total transition probability we must sum over all final field states. The only states for which the probability is nonzero are those containing a single photon in one mode: $\left|F_f\right\rangle = \left|1_{\boldsymbol{k}'s'}\right\rangle$. Summing eq. (16.74) over these states yields

$$P_{if}(t) = \frac{t^2}{2\varepsilon_0 \hbar V}\sum_{\boldsymbol{k}s}\omega_{\boldsymbol{k}}\left|\mathbf{d}^* \bullet \boldsymbol{\varepsilon}^*_{\boldsymbol{k}s}\right|^2 \operatorname{sinc}^2\left[\left(\omega_{\boldsymbol{k}}-\omega_0\right)t/2\right], \tag{16.75}$$

where we have made the substitution $\boldsymbol{k}'s' \rightarrow \boldsymbol{k}s$, because we no longer need the primes.

Now we will let the cavity become large, and compute the probability in eq. (16.75) by converting the sum over $\boldsymbol{k}$ to an integral. The wave vector $\boldsymbol{k}$ is 3-dimensional, so the integral is over $d^3\boldsymbol{k}$, and it needs to be weighted by the density of states in $\boldsymbol{k}$-space, $n(\boldsymbol{k})$. The components of $\boldsymbol{k}$ satisfy eq. (16.38), and the allowed values for $\boldsymbol{k}$ are displayed graphically in fig. 16.2. It can be seen that in $\boldsymbol{k}$-space the allowed modes take the form of a cubic lattice, with a unit-cell whose volume is $(2\pi/L)^3$. Each unit cell contains one state, so the density of states is

$$n(\boldsymbol{k}) = \frac{1}{(2\pi/L)^3} = \frac{V}{(2\pi)^3}. \tag{16.76}$$

In the limit that the cavity volume becomes large, the states in $\boldsymbol{k}$-space are close enough together to be considered continuously distributed, and the sum over $\boldsymbol{k}$ becomes an integral:

$$\sum_{\boldsymbol{k}} \rightarrow \int d^3\boldsymbol{k}\, n(\boldsymbol{k}) = \frac{V}{(2\pi)^3}\int d^3\boldsymbol{k}. \tag{16.77}$$

The integral over $d^3\boldsymbol{k}$ will be done using spherical coordinates. In these coordinates, the unit vector along $\boldsymbol{k}$ is

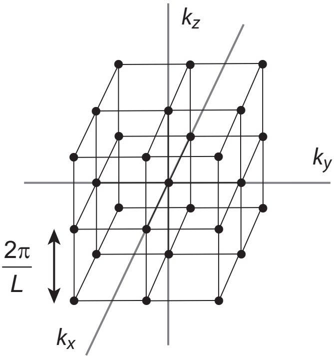


Fig 16.2 Each dot represents an allowed mode for a cubic cavity with periodic boundary conditions [eq. (16.38)].

$$\boldsymbol{u_k} = \sin\theta\cos\phi\,\boldsymbol{u}_x + \sin\theta\sin\phi\,\boldsymbol{u}_y + \cos\theta\boldsymbol{u}_z\,. \tag{16.78}$$

The two polarization vectors must be orthogonal to $\boldsymbol{u_k}$ (see problem 16.6) and to each other. For simplicity we will assume that the polarization vectors are linear, and suitable polarization vectors are (problem 16.17):

$$\varepsilon_{\boldsymbol{k}1} = \cos\theta\cos\phi\,\boldsymbol{u}_x + \cos\theta\sin\phi\,\boldsymbol{u}_y - \sin\theta\boldsymbol{u}_z\,, \tag{16.79}$$

$$\varepsilon_{\boldsymbol{k}2} = -\sin\phi\,\boldsymbol{u}_x + \cos\phi\boldsymbol{u}_y\,. \tag{16.80}$$

Since the direction of **d** is arbitrary, we can choose $\mathbf{d} = \mathrm{d}\boldsymbol{u}_z$. Using eqs. (16.79) and (16.80), we can explicitly perform the sum over polarization in eq. (16.75), with the result that

$$\sum_s \left|\mathbf{d}^*\bullet\varepsilon^*_{\boldsymbol{k}s}\right|^2 = |\mathrm{d}|^2\left[\left(\boldsymbol{u}_z\bullet\varepsilon_{\boldsymbol{k}1}\right)^2 + \left(\boldsymbol{u}_z\bullet\varepsilon_{\boldsymbol{k}2}\right)^2\right] = |\mathrm{d}|^2\sin^2\theta\,. \tag{16.81}$$

Combining eqs. (16.75), (16.77), and (16.81), we find

$$\begin{aligned} P_{if}(t) &= \frac{t^2}{2\varepsilon_0\hbar V}\frac{V}{(2\pi)^3}\int d^3\boldsymbol{k}\,\omega_{\boldsymbol{k}}\,|\mathrm{d}|^2\sin^2\theta\,\mathrm{sinc}^2\left[(\omega_{\boldsymbol{k}} - \omega_0)t/2\right] \\ &= \frac{|\mathrm{d}|^2 t^2}{16\pi^3\varepsilon_0\hbar}\int_0^{2\pi} d\phi\int_0^{\pi} d\theta\sin\theta\int_0^{\infty} dk\,k^2\omega_k\sin^2\theta\,\mathrm{sinc}^2\left[(\omega_k - \omega_0)t/2\right]. \end{aligned} \tag{16.82}$$

We can perform the ϕ and θ integrals in eq. (16.82). We can also use the fact that $\omega_k = ck$ to convert the integral over k to an integral over ω_k. The net result is

$$P_{if}(t) = \frac{|\mathrm{d}|^2 t^2}{6\pi^2\varepsilon_0 c^3\hbar}\int_0^{\infty} d\omega_k\omega_k^3\,\mathrm{sinc}^2\left[(\omega_k - \omega_0)t/2\right]. \tag{16.83}$$

Now we can once again use some of the techniques we used in the semiclassical calculation of sec. 15.3. For times longer than several oscillation periods of the field, the function ω_k^3 is very broad compared to the function $\mathrm{sinc}^2[\ldots]$, so ω_k^3 is approximately constant over the width of $\mathrm{sinc}^2[\ldots]$. We can thus evaluate ω_k^3 at the peak of $\mathrm{sinc}^2[\ldots]$ (at $\omega_k = \omega_0$), and factor it outside of the integral, yielding

$$P_{if}(t) \simeq \frac{|\mathrm{d}|^2 \, \omega_0^3 t^2}{6\pi^2 \varepsilon_0 c^3 \hbar} \int_0^\infty d\omega_k \, \mathrm{sinc}^2\left[(\omega_k - \omega_0)t/2\right]. \tag{16.84}$$

As in eq. (15.41) the integral is $2\pi/t$, and the transition probability is

$$P_{if}(t) \simeq \frac{|\mathrm{d}|^2 \, \omega_0^3 t^2}{6\pi^2 \varepsilon_0 c^3 \hbar} \left(\frac{2\pi}{t}\right) = \frac{|\mathrm{d}|^2 \, \omega_0^3 t}{3\pi \varepsilon_0 c^3 \hbar}. \tag{16.85}$$

The transition rate is the time derivative of this probability, which is

$$R = \frac{d}{dt} P_{if}(t) = \frac{|\mathrm{d}|^2 \, \omega_0^3}{3\pi \varepsilon_0 c^3 \hbar} = A_{21}. \tag{16.86}$$

This rate is equal to the Einstein A coefficient that we calculated previously [eq. (15.50)], as it should be. The difference is that here we have calculated it from first principles, rather than from a simplified model. The correct prediction of the spontaneous emission rate is one of the great triumphs of quantum field theory.

The spontaneous emission rate in eq. (16.86) assumes that the density of states is constant [eq. (16.76)], which is appropriate for an atom in free space (a cavity with $V \to \infty$). However, it is possible to modify the density of states by placing the atom in an appropriate cavity, typically a small cavity with high reflectivity mirrors. Such cavities can either enhance or inhibit spontaneous emission. For example, spontaneous emission from an atom occurs at frequencies near the atomic resonance. If this frequency is not supported by the cavity, spontaneous emission is inhibited, and its rate decreases (ref. [16.5], chapter 10).

16.6 QUANTUM THEORY OF PHOTOELECTRIC DETECTION

We can apply what we've learned about the interaction of a quantum field with atoms to build a fully quantum-mechanical theory of photoelectric detection. This will generalize the semiclassical theory we discussed in sec. 15.4. The complete elaboration of a quantum theory of photoelectric detection is beyond the scope of this text, but here we can discuss some of its basic features.[3]

The quantum model mimics the semiclassical model of sec. 15.4 (pictured in fig. 15.5). The atom is initially in the ground state $|1\rangle$, and the initial field state is $|F_i\rangle$. We'd like to find the probability that the atom makes transition to an unbound state with positive energy $|E\rangle$, which constitutes a photodetection; the final state of the field is $|F_f\rangle$. We will use the electric-dipole Hamiltonian to describe the interaction of the field with the detector. The matrix elements of this Hamiltonian are

---

[3] For a full theory see ref. [16.3], chapter 14.

$$\begin{aligned} H_{ED\,fi}(t) &= -\left\langle \psi_f \left| \hat{\mathbf{d}} \cdot \hat{\mathcal{E}}(\boldsymbol{r},t) \right| \psi_i \right\rangle \\ &= -\langle E|\hat{\mathbf{d}}|1\rangle \cdot \left\langle F_f \left| \hat{\mathcal{E}}(\boldsymbol{r},t) \right| F_i \right\rangle . \end{aligned} \tag{16.87}$$

The field operator can be written as

$$\hat{\mathcal{E}}(\boldsymbol{r},t) = \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) + \hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t), \tag{16.88}$$

where

$$\hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) = \sqrt{\frac{\hbar}{2\varepsilon_0 V}} \sum_{\boldsymbol{k},s} \sqrt{\omega_{\boldsymbol{k}}}\; \hat{a}_{\boldsymbol{k}s}(t) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \boldsymbol{\varepsilon}_{\boldsymbol{k}s}, \tag{16.89}$$

$$\hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t) = \left[\hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) \right]^{\dagger}. \tag{16.90}$$

$\hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t)$ and $\hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t)$ are often referred to as the positive and negative frequency parts of the field. We are interested in photodetection, during which a photon is absorbed. The part of the field that contributes to absorption is $\hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t)$, since it is the part containing annihilation operators (see problem 16.21).[4] Retaining just this relevant part of the field in eq. (16.87), we learn that

$$H_{ED\,fi}(t) \propto \left\langle F_f \left| \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) \right| F_i \right\rangle . \tag{16.91}$$

In perturbation theory, the probability $P_{if}(t)$ that the system will make a transition from the initial to the final state will be proportional to the square magnitude of the matrix element of the electric-dipole Hamiltonian:

$$P_{if}(t) \propto \left| H_{ED\,fi}(t) \right|^2 \propto \left| \left\langle F_f \left| \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) \right| F_i \right\rangle \right|^2 . \tag{16.92}$$

When performing photoelectric detection, what we actually measure is the free electron, not the field itself. To find the total detection probability $P(t)$ we must sum over all possible final states of the field, which means that

$$\begin{aligned} P(t) &\propto \sum_{F_f} \left| \left\langle F_f \left| \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) \right| F_i \right\rangle \right|^2 \\ &= \sum_{F_f} \left\langle F_i \left| \hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t) \right| F_f \right\rangle \cdot \left\langle F_f \left| \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) \right| F_i \right\rangle \\ &= \left\langle F_i \left| \hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t) \cdot \left(\sum_{F_f} \left| F_f \right\rangle \left\langle F_f \right| \right) \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) \right| F_i \right\rangle \\ &= \left\langle F_i \left| \hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t) \cdot \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) \right| F_i \right\rangle . \end{aligned} \tag{16.93}$$

---

[4] Terms involving $\hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t)$ contain creation operators and correspond to the emission of a photon. During photodetection, only terms involving the absorption of a photon conserve energy. See the related discussion following eqs. (16.64) and (16.65).

Here we have used the fact that the field states form a complete set, so the sum over states in the third line of eq. (16.93) is equal to the identity operator. We can define the intensity operator as

$$\hat{I}(\boldsymbol{r},t) \equiv \hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t) \boldsymbol{\cdot} \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t), \tag{16.94}$$

which means that

$$P(t) \propto \left\langle \hat{I}(\boldsymbol{r},t) \right\rangle. \tag{16.95}$$

Equation (16.95) tells us that the detection probability is proportional to the expectation value of the intensity. A full calculation yields the result that the detection probability is linear in time (as it was semiclassically), which means that the detection rate is constant.

In the laboratories at the end of this book we are measuring individual photons. The quantity we are frequently interested in is the probability that a single photon will be measured in the time interval between t and $t+\Delta t$, $P(t,\Delta t)$, for small Δt. Since the rate of detection is constant, this probability is linearly proportional to Δt. Thus, we can write

$$P(t,\Delta t) = \eta \left\langle \hat{I}(\boldsymbol{r},t) \right\rangle \Delta t, \tag{16.96}$$

where η is a measure of efficiency of the detector. Compare this to the semiclassical result in eq. (15.56).

For a single-mode field we know that

$$\hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) = \sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\; \hat{a}(t) e^{i\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{r}} \boldsymbol{\varepsilon} = \sqrt{\frac{\hbar\omega}{2\varepsilon_0 V}}\; \hat{a} e^{i(\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{r}-\omega t)} \boldsymbol{\varepsilon}. \tag{16.97}$$

and

$$\hat{I}(\boldsymbol{r},t) = \hat{\mathcal{E}}^{(-)}(\boldsymbol{r},t) \boldsymbol{\cdot} \hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t) = \frac{\hbar\omega}{2\varepsilon_0 V} \hat{a}^\dagger \hat{a} \propto \hat{n}. \tag{16.98}$$

For a single mode field, the intensity operator is proportional to the number operator for the field mode. The detection probability is then proportional to the expectation value of the photon number, which makes intuitive sense. Note that the intensity operator does NOT contain the zero-point contribution that the field Hamiltonian does. Thus, photoelectric measurements of the field are not sensitive to the zero-point energy.

16.7 BEAM SPLITTERS

Beam splitters are important optical elements, so we need to describe how fields behave when they encounter a beam splitter. Consider the experimental arrangement shown in fig. 16.3, where two fields are incident on the input ports of a beam splitter, and two fields emerge from the output ports. We'll assume that all fields are linearly polarized in the same direction. We know that the operators corresponding to the amplitude of the individual field modes satisfy the commutation relations

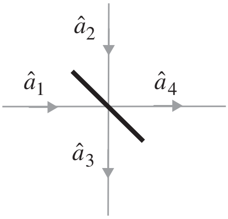


Fig 16.3 Input and output modes of a beam splitter.

$$\left[\hat{a}_1,\hat{a}_1^\dagger\right]=\left[\hat{a}_2,\hat{a}_2^\dagger\right]=\left[\hat{a}_3,\hat{a}_3^\dagger\right]=\left[\hat{a}_4,\hat{a}_4^\dagger\right]=1. \tag{16.99}$$

Furthermore, since the input modes and the output modes are separately measurable and independent, we know that their field operators must commute:

$$\left[\hat{a}_1,\hat{a}_2^\dagger\right]=\left[\hat{a}_3,\hat{a}_4^\dagger\right]=0. \tag{16.100}$$

The field reflection and transmission coefficients for mode 1 are r_1 and t_1, while those of mode 2 are r_2 and t_2. Thus, the transformation from the input to the output modes is

$$\hat{a}_3=r_1\hat{a}_1+t_2\hat{a}_2, \tag{16.101}$$

$$\hat{a}_4=t_1\hat{a}_1+r_2\hat{a}_2. \tag{16.102}$$

Assuming that the reflection and transmission coefficients are real, the commutation relations yield the following:

$$\left[\hat{a}_3,\hat{a}_3^\dagger\right]=\left[r_1\hat{a}_1+t_2\hat{a}_2,r_1\hat{a}_1^\dagger+t_2\hat{a}_2^\dagger\right]=r_1^2+t_2^2=1 \tag{16.103}$$

$$\left[\hat{a}_4,\hat{a}_4^\dagger\right]=\left[t_1\hat{a}_1+r_2\hat{a}_2,t_1\hat{a}_1^\dagger+r_2\hat{a}_2^\dagger\right]=t_1^2+r_2^2=1 \tag{16.104}$$

$$\left[\hat{a}_3,\hat{a}_4^\dagger\right]=\left[r_1\hat{a}_1+t_2\hat{a}_2,t_1\hat{a}_1^\dagger+r_2\hat{a}_2^\dagger\right]=r_1t_1+r_2t_2=0. \tag{16.105}$$

These relationships are sometimes referred to as the reciprocity relations, and you will verify in the problems that they are satisfied if

$$r_1=-r_2=r,\quad t_1=t_2=t,\quad r^2+t^2=1. \tag{16.106}$$

You can compare these relationships to those we determined classically using energy conservation [eq. (2.A.4)]. Combining eqs. (16.101), (16.102), and (16.106), the annihilation operators for the output field modes can be written as

$$\hat{a}_3=r\hat{a}_1+t\hat{a}_2, \tag{16.107}$$

$$\hat{a}_4 = t\hat{a}_1 - r\hat{a}_2. \tag{16.108}$$

This transformation for the field operators allows us to calculate the expectation values of quantities corresponding to the output fields, given the state of the input fields, as seen in the following example.

EXAMPLE 16.4

A beam splitter as reflection and transmission coefficients $r = \sqrt{0.7}$, $t = \sqrt{0.3}$, and the state of the fields incident on it is $|\psi\rangle = |10\rangle_1 |0\rangle_2$. Calculate the expectation value of the number of photons in mode 3.

$$\begin{aligned}\langle n_3 \rangle &= \left\langle \hat{a}_3^\dagger \hat{a}_3 \right\rangle \\ &= \left\langle \left(r\hat{a}_1^\dagger + t\hat{a}_2^\dagger \right)\left(r\hat{a}_1 + t\hat{a}_2 \right) \right\rangle \\ &= {}_1\langle 10|\,{}_2\langle 0| \left(r^2 \hat{n}_1 + t^2 \hat{n}_2 + rt\hat{a}_2^\dagger \hat{a}_1 + rt\hat{a}_1^\dagger \hat{a}_2 \right) |10\rangle_1 |0\rangle_2 \\ &= (0.7)(10) + 0 + 0 + 0 \\ &= 7\,. \end{aligned} \tag{16.109}$$

Since there were 10 photons incident on a beam splitter with a reflectivity of 70%, it's not surprising that an average of seven photons would be reflected. Remember, however, that photons are randomly reflected or transmitted by the beam splitter. What we have calculated here is the expectation value, not the number that will result on every trial.

The calculation in example 16.4 was performed in the spirit of the Heisenberg picture: The field operators were transformed, and the expectation value was calculated using the initial state of the system. Alternatively, we can calculate how the states transform on the beam splitter, and thus determine the output states given the input states. For general states this can be a difficult proposition, but for some simple cases it is straightforward.

We begin by inverting eqs. (16.107) and (16.108) to obtain

$$\hat{a}_1 = r\hat{a}_3 + t\hat{a}_4, \tag{16.110}$$

$$\hat{a}_2 = t\hat{a}_3 - r\hat{a}_4. \tag{16.111}$$

We can also write this as:

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} = \begin{pmatrix} r & t \\ t & -r \end{pmatrix} \begin{pmatrix} \hat{a}_3 \\ \hat{a}_4 \end{pmatrix}, \tag{16.112}$$

which is the transformation from the outputs to the inputs. We can create Fock states on the input modes by applying $\hat{a}_1^\dagger$ and $\hat{a}_2^\dagger$ to the vacuum, and we can see how these modes are transformed by using eq. (16.112). For example, if we want to know how a field containing a single photon in mode 1 is transformed by the beam splitter, we apply $\hat{a}_1^\dagger$ to the vacuum and use eq. (16.110) [or eq. (16.112)] to find that

$$\begin{aligned} \hat{a}_1^\dagger\left|0\right\rangle &= \left(r\hat{a}_3^\dagger + t\hat{a}_4^\dagger\right)\left|0\right\rangle , \\ \left|1\right\rangle_1\left|0\right\rangle_2 &= r\left|1\right\rangle_3\left|0\right\rangle_4 + t\left|0\right\rangle_3\left|1\right\rangle_4 . \end{aligned} \tag{16.113}$$

This says that if a photon in mode 1 is incident on a beam splitter, the system is transformed into a superposition state of the photon being in modes 3 and 4.

Referring back to fig. 16.3, we might be tempted to completely ignore one of the input modes if it contains vacuum. However, as you'll prove in problem 16.23, it is necessary to include all of the field operators in the beam splitter transformation [eqs. (16.107) and (16.108)]. If vacuum is present in one of the modes, that is accounted for by the state of the field.

16.7.1 The Mach-Zehnder Interferometer

A Mach-Zehnder interferometer, with a single photon incident, is shown in fig. 16.4. We can use the techniques described above to calculate the probabilities that the photon will be detected at either of the interferometer output ports (5 or 6).

Assume that both beam splitters are 50/50, so $r = t = 1/\sqrt{2}$. In fig. 16.4 the field of mode 4 is phase shifted relative to that of mode 3. Since the field operator is proportional to the annihilation operator, phase shifting the field is equivalent to phase shifting the annihilation operator: $\hat{a}_4 \rightarrow \hat{a}_4 e^{i\phi}$. The transformation corresponding to the interferometer consists of a beam splitter, a phase shift on one beam, then another beam splitter; using eq. (16.112) this is

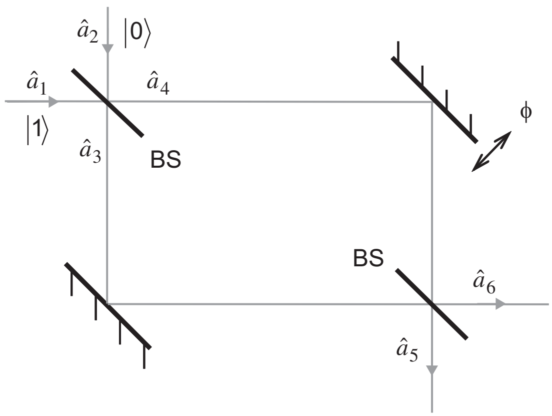


Fig 16.4 A Mach-Zehnder interferometer.

$$\begin{aligned}\begin{pmatrix}\hat{a}_1\\ \hat{a}_2\end{pmatrix} &= \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & e^{i\phi}\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}\hat{a}_5\\ \hat{a}_6\end{pmatrix}\\ &= \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}1 & 1\\ e^{i\phi} & -e^{i\phi}\end{pmatrix}\begin{pmatrix}\hat{a}_5\\ \hat{a}_6\end{pmatrix}\\ &= \frac{1}{2}\begin{pmatrix}1+e^{i\phi} & 1-e^{i\phi}\\ 1-e^{i\phi} & 1+e^{i\phi}\end{pmatrix}\begin{pmatrix}\hat{a}_5\\ \hat{a}_6\end{pmatrix},\end{aligned} \qquad (16.114)$$

so

$$\hat{a}_1 = \frac{1}{2}\left[\hat{a}_5\left(1+e^{i\phi}\right)+\hat{a}_6\left(1-e^{i\phi}\right)\right]. \qquad (16.115)$$

Assuming that a single photon is incident in mode 1, the state of the fields on the output can be calculated as:

$$\begin{aligned}\hat{a}_1^\dagger\left|0\right\rangle &= \frac{1}{2}\left[\hat{a}_5^\dagger\left(1+e^{-i\phi}\right)+\hat{a}_6^\dagger\left(1-e^{-i\phi}\right)\right]\left|0\right\rangle\\ \left|1\right\rangle_1\left|0\right\rangle_2 &= \frac{1}{2}\left[\left(1+e^{-i\phi}\right)\left|1\right\rangle_5\left|0\right\rangle_6+\left(1-e^{-i\phi}\right)\left|0\right\rangle_5\left|1\right\rangle_6\right].\end{aligned} \qquad (16.116)$$

The probability that the photon leaves the interferometer through mode 5 is then

$$P\left(1_5\right) = \left|\frac{1}{2}\left(1+e^{-i\phi}\right)\right|^2 = \frac{1}{4}\left(2+2\cos\phi\right) = \frac{1}{2}\left(1+\cos\phi\right). \qquad (16.117)$$

This compares favorably with the classical intensity leaving this port of the interferometer [eq. (2.A.9)].

By adjusting the relative phase of the arms of the interferometer, the photon can be made to emerge in mode 5, or in mode 6, or in a superposition state of both modes. Detecting the photon collapses the state, and we find that the photon is in either mode 5 or mode 6, but not both.

16.8 References

[16.1] D.J. Griffiths, *Introduction to Electrodynamics, 3rd ed.* (Prentice Hall, Upper Saddle River, NJ, 1999).

[16.2] R. Loudon, *The Quantum Theory of Light, 3rd ed.* (Oxford University Press, Oxford, 2000).

[16.3] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995).

[16.4] C.C. Gerry and P.L. Knight, *Introductory Quantum Optics* (Cambridge University Press, Cambridge, 2005).

[16.5] M. Fox, *Quantum Optics: An Introduction* (Oxford University Press, Oxford, 2006).

[16.6] C. K. Hong, Z. Y. Ou, and L. Mandel, “Measurement of subpicosecond time intervals between two photons by interference,” Phys. Rev. Lett. **59**, 2044 (1987).

16.9 PROBLEMS

16.1 Find the Heisenberg equation of motion for the momentum of a harmonic oscillator. Compare this equation to its classical equivalent.

16.2 Find the Heisenberg equations of motion for the position and momentum of a free particle. Compare these equations to their classical equivalents.

16.3 A spin-1/2 particle is placed in a uniform magnetic field that points in the z-direction. At $t = 0$ the spin of the particle points along the positive x-axis. Use the Heisenberg equations of motion to find $\langle S_x \rangle(t), \langle S_y \rangle(t)$ and $\langle S_z \rangle(t)$. Compare your results to those obtained in sec. 9.4.

16.4* Use the Heisenberg equation of motion to show that the time dependence of the annihilation operator is

$$\hat{a}(t) = \hat{a}(0)e^{-i\omega t} = \hat{a}\,e^{-i\omega t}. \tag{16.118}$$

16.5 Show that if there is no explicit time dependence to the Hamiltonian, then $\hat{H}_H(t) = \hat{H}_S = \hat{H}$.

16.6* Show that if $\boldsymbol{A}(\boldsymbol{r},t) = \boldsymbol{A}_0 e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, where $\boldsymbol{A}_0$ is independent of $\boldsymbol{r}$, then $\nabla\cdot\boldsymbol{A} = i\boldsymbol{k}\cdot\boldsymbol{A}$. Given this, what do Maxwell's equations say about the relative directions of $\boldsymbol{k}$ and $\boldsymbol{\mathcal{E}}(\boldsymbol{r},t)$? Because of this, $\boldsymbol{\mathcal{E}}(\boldsymbol{r},t)$ is said to be transverse. [Note that $\boldsymbol{\mathcal{B}}(\boldsymbol{r},t)$ is also transverse].

16.7 Prove that eqs. (16.15) and (16.18) constitute a solution to the wave equation.

16.8 Verify eq. (16.19).

16.9 Verify eq. (16.28).

16.10 Calculate the expectation value and variance of the electric field operator for a single-mode field in a Fock state.

16.11 Calculate the expectation value of the electric field operator for a single-mode field in a coherent state.

16.12 Calculate the expectation value and standard deviation of X_θ for a coherent state field. How do your answers depend on θ?

16.13 Calculate the mean and standard deviation of the photon number, for a field in a coherent state.

16.14 The Baker-Hausdorf lemma states that for any two operators $\hat{A}$ and $\hat{B}$,

$$e^{-\hat{A}}\hat{B}e^{\hat{A}} = \hat{B} + \left[\hat{B},\hat{A}\right] + \frac{1}{2!}\left[\left[\hat{B},\hat{A}\right],\hat{A}\right] + \ldots. \tag{16.119}$$

Use this lemma to show that the squeezing operator transforms the creation and annihilation operators as:

$$\hat{S}^\dagger(r)\hat{a}\hat{S}(r) = \hat{a}\cosh r - \hat{a}^\dagger \sinh r, \tag{16.120}$$

$$\hat{S}^\dagger(r)\hat{a}^\dagger\hat{S}(r) = \hat{a}^\dagger \cosh r - \hat{a}\sinh r. \tag{16.121}$$

16.15 Use the fact that $\hat{S}(r)$ is unitary, and the results of the previous problem, to verify eqs. (16.57) and (16.58).

16.16 Prove that the photon number distribution $P(n)$ for a squeezed state contains only even numbers of photons. Note: you don't need to explicitly determine $P(n)$ to accomplish this.

16.17 For a field with $\boldsymbol{u}_{\boldsymbol{k}} = \boldsymbol{u}_z$, assume that $\boldsymbol{\varepsilon}_{\boldsymbol{k}1} = \boldsymbol{u}_x$ and $\boldsymbol{\varepsilon}_{\boldsymbol{k}2} = \boldsymbol{u}_y$.

(a) Show that these assumptions are consistent with the results of problem 16.6.

(b) Rotate each of these vectors first by θ about the y-axis, and then by ϕ about the z-axis. Show that this procedure yields vectors that are consistent with eqs. (16.78)–(16.80).

16.18* Calculate the expectation value and variance of the intensity operator for a field in a vacuum state. Compare them to the expectation value and variance of the field.

16.19 Calculate the expectation value and variance of the intensity operator for a single-mode field in a Fock state.

16.20 Calculate the expectation value and variance of the intensity operator for a single-mode field in a coherent state.

16.21* Starting from eqs. (16.61) and (16.87), calculate the final state amplitude $c_f(t)$ for photoelectric detection. Show that

$$c_f(t) = \frac{i}{\hbar}\int_0^t dt'\, e^{i\omega_0 t'} \langle E|\hat{\mathbf{d}}|1\rangle \cdot \langle F_f|\hat{\mathcal{E}}^{(+)}(\boldsymbol{r},t)|F_i\rangle . \tag{16.122}$$

The important thing here is that the negative frequency part of the field does not contribute.

16.22 Verify eq. (16.106).

16.23* Suppose that mode 2 in fig. 16.3 contains vacuum. The temptation is to completely ignore it, so the beam splitter transformation of eqs. (16.107) and (16.108) becomes $\hat{a}_3 = r\hat{a}_1$ and $\hat{a}_4 = t\hat{a}_1$. Show that under this transformation the output modes do not satisfy the proper commutation relationships.

16.24 Two photons are incident on one port of a beam splitter, and vacuum is incident on the other. What is the output state?

16.25* One photon is incident on one port of a 50/50 beam splitter, and another photon of the same frequency is incident on the other port at the same time. What is the output state? What is the probability that one photon will emerge from each output port? An experiment observing this behavior is described in ref. [16.6].

16.26* In a balanced homodyne detector (fig. 16.5) signal and reference fields are incident on the input ports of a 50/50 beam splitter, and the fields emerging from the beam splitter are measured with photodetectors.[5] The detector directly measures the difference of the number of photons striking each detector $\hat{n}_{12} = \hat{n}_1 - \hat{n}_2$. Show that if the reference field is in a large-amplitude coherent state, $|\alpha\rangle_R = \left||\alpha|e^{i\theta}\right\rangle_R$, the measured difference number is proportional to the quadrature field amplitude $\hat{X}_\theta$ of the signal field. The proportionality constant

---

[5] It's referred to as homodyne detection because the signal and reference fields are of the same frequency. In heterodyne detection the fields have different frequencies.

can be determined, yielding a measurement of $\hat{X}_\theta$. Adjusting the reference field phase θ will change which quadrature amplitude is measured. (Hint: Take the expectation value of $\hat{n}_{12}$ in state $|\alpha\rangle_R$. This leaves an operator that acts only on the signal field.)

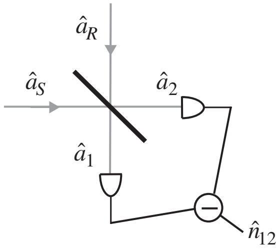


Fig 16.5 A balanced homodyne detector.

COMPLEMENT 16.A

Second-Order Coherence and the Grangier Experiment

Lab 2 describes an experiment to measure the degree of second-order coherence $g^{(2)}(0)$ for two different field states: a conditionally prepared single-photon state, and a classical state. There we show that for classical fields we must have $g^{(2)}(0) \geq 1$, while for the single-photon state we expect $g^{(2)}(0) = 0$. Here we'll derive a quantum-mechanical expression for $g^{(2)}(0)$, valid for any field state.

The basics of the experimental arrangement are shown in fig. 16.A.1. The signal field is incident on a beam splitter in mode *I*, and mode *V* contains vacuum. The transmitted and reflected fields impinge on detectors *T* and *R*. In sec. 16.6 we described the detection of a quantum field using a single detector. Here we will generalize those results to two detectors. We will assume that the fields are linearly polarized, and won't explicitly write the fields as vectors.

We are interested in the probability that a single photon is detected at *T* at time t_1, and a second photon is detected at *R* at time t_2. For the sake of definiteness we will assume that detector *T* fires first, so $t_1 < t_2$. The two-detector equivalent of eq. (16.92) is

$$P_{TR}(t_1,t_2) \propto \left| \left\langle F_f \left| \hat{\mathcal{E}}_R^{(+)}(t_2) \hat{\mathcal{E}}_T^{(+)}(t_1) \right| F_i \right\rangle \right|^2 . \tag{16.A.1}$$

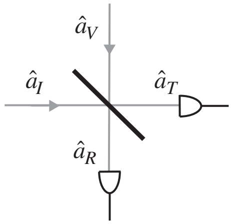


Fig 16.A.1 The experimental arrangement for measuring $g^2(0)$. The field in mode *I* (the incident field) is the signal field, while mode *V* contains vacuum.

Note the ordering of the operators: Because the detection at T occurs first, the operator for the positive-frequency part of the field at that detector, $\hat{\mathcal{E}}_T^{(+)}(t_1)$, is applied to the initial field state $|F_i\rangle$ first. Once again, we are not interested in the final state of the field, only in the state of the detectors. We can thus sum over all possible final field states, obtaining

$$\begin{aligned} P_{TR}(t_1,t_2) &\propto \sum_{F_f} \left|\left\langle F_f\left|\hat{\mathcal{E}}_R^{(+)}(t_2)\hat{\mathcal{E}}_T^{(+)}(t_1)\right|F_i\right\rangle\right|^2 \\ &= \sum_{F_f} \left\langle F_i\left|\left(\hat{\mathcal{E}}_R^{(+)}(t_2)\hat{\mathcal{E}}_T^{(+)}(t_1)\right)^\dagger\right|F_f\right\rangle\left\langle F_f\left|\hat{\mathcal{E}}_R^{(+)}(t_2)\hat{\mathcal{E}}_T^{(+)}(t_1)\right|F_i\right\rangle \\ &= \sum_{F_f} \left\langle F_i\left|\hat{\mathcal{E}}_T^{(-)}(t_1)\hat{\mathcal{E}}_R^{(-)}(t_2)\right|F_f\right\rangle\left\langle F_f\left|\hat{\mathcal{E}}_R^{(+)}(t_2)\hat{\mathcal{E}}_T^{(+)}(t_1)\right|F_i\right\rangle \\ &= \left\langle F_i\left|\hat{\mathcal{E}}_T^{(-)}(t_1)\hat{\mathcal{E}}_R^{(-)}(t_2)\left(\sum_{F_f}|F_f\rangle\langle F_f|\right)\hat{\mathcal{E}}_R^{(+)}(t_2)\hat{\mathcal{E}}_T^{(+)}(t_1)\right|F_i\right\rangle \\ &= \left\langle F_i\left|\hat{\mathcal{E}}_T^{(-)}(t_1)\hat{\mathcal{E}}_R^{(-)}(t_2)\hat{\mathcal{E}}_R^{(+)}(t_2)\hat{\mathcal{E}}_T^{(+)}(t_1)\right|F_i\right\rangle . \end{aligned} \tag{16.A.2}$$

Here we have used the fact that the field states are a complete set, so the sum over final states is the identity operator. The ordering of the operators in eq. (16.A.2) is once again important. The adjoint operation reverses the order of the operators, and as such the field operators corresponding to the first detection (at T) are next to the state of the initial field on both the left and right ends of this expression.

Recall that the positive frequency part of the field contains annihilation operators, while the negative frequency part contains creation operators. Thus, in eq. (16.A.2) all of the creation operators lie to the left of all of the annihilation operators. This ordering of the operators is called "normal ordering." Equation (16.A.2) is correct as written, but it is possible to make it look more reminiscent of the corresponding classical expression by reordering the operators. First, we use two colons, $:\cdots:$, to denote an operation in which the operators between the colons are placed in normal order, by *disregarding* commutation relations. By this we mean that $\hat{a}^\dagger\hat{a} \neq \hat{a}\hat{a}^\dagger$, but $:\hat{a}^\dagger\hat{a}:=:\hat{a}\hat{a}^\dagger:=\hat{a}^\dagger\hat{a}$. Second, we introduce the time-ordering operation $\mathcal{T}$ to indicate that creation operators (negative frequency fields) should be ordered with increasing time from left to right, while annihilation operators (positive frequency fields) should be ordered with increasing time from right to left; once again, this ordering is done by disregarding commutation relations. Using these operations, we can rewrite eq. (16.A.2) as

$$\begin{aligned} P_{TR}(t_1,t_2) &\propto \left\langle \mathcal{T}:\hat{\mathcal{E}}_R^{(-)}(t_2)\hat{\mathcal{E}}_R^{(+)}(t_2)\hat{\mathcal{E}}_T^{(-)}(t_1)\hat{\mathcal{E}}_T^{(+)}(t_1):\right\rangle \\ &= \left\langle \mathcal{T}:\hat{I}_R(t_2)\hat{I}_T(t_1):\right\rangle , \end{aligned} \tag{16.A.3}$$

where $\hat{I}(t)$ is defined in eq. (16.94). Analagous to eq. (16.96), we can write the joint probability that we will detect a photon at T between t_1 and $t_1+\Delta t_1$, and another photon at R between t_2 and $t_2+\Delta t_2$ as

$$P_{TR}\left(t_1,\Delta t_1;t_2,\Delta t_2\right)=\eta_T\eta_R\left\langle\mathcal{T}:\hat{I}_R\left(t_2\right)\hat{I}_T\left(t_1\right):\right\rangle\Delta t_1\Delta t_2\,. \tag{16.A.4}$$

Now we are ready to define the degree of second-order coherence as[6]

$$g^{(2)}\left(t_1,t_2\right)\equiv\frac{\left\langle\mathcal{T}:\hat{I}_R\left(t_2\right)\hat{I}_T\left(t_1\right):\right\rangle}{\left\langle\hat{I}_R\left(t_2\right)\right\rangle\left\langle\hat{I}_T\left(t_1\right)\right\rangle}\,. \tag{16.A.5}$$

There is no need to place symbols denoting normal ordering on the expectation values in the denominator, because the intensity operators as defined are already normally ordered. A field is said to be stationary if its statistics are independent of time. If this is the case, then the degree of second-order coherence can depend only on the time difference between the measurements, so for stationary fields

$$g^{(2)}\left(\tau\right)=\frac{\left\langle\mathcal{T}:\hat{I}_R\left(t+\tau\right)\hat{I}_T\left(t\right):\right\rangle}{\left\langle\hat{I}_R\left(t\right)\right\rangle\left\langle\hat{I}_T\left(t\right)\right\rangle}\,. \tag{16.A.6}$$

We are most interested in $\tau=0$, so all times are the same, and we can simplify our expression to

$$g^{(2)}\left(0\right)=\frac{\left\langle:\hat{I}_R\hat{I}_T:\right\rangle}{\left\langle\hat{I}_R\right\rangle\left\langle\hat{I}_T\right\rangle}=\frac{P_{TR}}{P_RP_T}\,. \tag{16.A.7}$$

Here we have used eqs. (16.96) and (16.A.4) to write $g^{(2)}\left(0\right)$ in terms of detection probabilities, and we've assumed that the time intervals $\Delta t_1=\Delta t_2=\Delta t$ are all the same. Note that the detector efficiencies cancel. In eq. (16.A.7) P_{TR} is the probability of detecting photons at both T and R (within a coincidence window of Δt), while P_T and P_R are the probabilities of detecting individual photons at the two detectors. Equation (16.A.7) agrees with our discussion in lab 2. We can use eq. (16.98) to rewrite (16.A.7) in terms of either number operators, or creation and annihilation operators, as

$$g^{(2)}\left(0\right)=\frac{\left\langle:\hat{n}_R\hat{n}_T:\right\rangle}{\left\langle\hat{n}_R\right\rangle\left\langle\hat{n}_T\right\rangle}=\frac{\left\langle\hat{a}_T^\dagger\hat{a}_R^\dagger\hat{a}_R\hat{a}_T\right\rangle}{\left\langle\hat{a}_R^\dagger\hat{a}_R\right\rangle\left\langle\hat{a}_T^\dagger\hat{a}_T\right\rangle}\,. \tag{16.A.8}$$

So far, we have written $g^{(2)}\left(0\right)$ in terms of the fields at the detectors, with no reference to the field incident on the beam splitter. We can use eqs. (16.107) and (16.108) to rewrite the field operators at the detectors in terms of the field operators incident on the beam splitter as

---

[6] See ref. [16.A.1], sec. 4.12.

$$\hat{a}_R = r\hat{a}_I + t\hat{a}_V , \tag{16.A.9}$$

$$\hat{a}_T = t\hat{a}_I - r\hat{a}_V . \tag{16.A.10}$$

Now let's use the fact that the field in mode I is in an arbitrary state $|\psi\rangle$, and the field in mode V is in a vacuum state $|0\rangle$. Using eqs. (16.A.9) and (16.A.10), the numerator of eq. (16.A.8) becomes

$$\begin{aligned}\left\langle \hat{a}_T^\dagger \hat{a}_R^\dagger \hat{a}_R \hat{a}_T \right\rangle &= \left\langle \left(t\hat{a}_I^\dagger - r\hat{a}_V^\dagger\right)\left(r\hat{a}_I^\dagger + t\hat{a}_V^\dagger\right)\left(r\hat{a}_I + t\hat{a}_V\right)\left(t\hat{a}_I - r\hat{a}_V\right)\right\rangle \\ &= {}_I\langle\psi|\,{}_V\langle 0|\left(t\hat{a}_I^\dagger - r\hat{a}_V^\dagger\right)\left(r\hat{a}_I^\dagger + t\hat{a}_V^\dagger\right)\left(r\hat{a}_I + t\hat{a}_V\right)\left(t\hat{a}_I - r\hat{a}_V\right)|\psi\rangle_I\,|0\rangle_V \\ &= {}_I\langle\psi|\left(t\hat{a}_I^\dagger\right)\left(r\hat{a}_I^\dagger\right)\left(r\hat{a}_I\right)\left(t\hat{a}_I\right)|\psi\rangle_I \\ &= t^2 r^2 \left\langle \hat{a}_I^\dagger \hat{a}_I^\dagger \hat{a}_I \hat{a}_I \right\rangle .\end{aligned} \tag{16.A.11}$$

Here the vacuum field operators have been eliminated, and the expectation value is taken using only the state of the incident field mode I. Similarly, the expectation values in the denominator of eq. (16.A.8) become

$$\left\langle \hat{a}_R^\dagger \hat{a}_R \right\rangle = r^2 \left\langle \hat{a}_I^\dagger \hat{a}_I \right\rangle , \tag{16.A.12}$$

$$\left\langle \hat{a}_T^\dagger \hat{a}_T \right\rangle = t^2 \left\langle \hat{a}_I^\dagger \hat{a}_I \right\rangle . \tag{16.A.13}$$

Using eqs. (16.A.11)–(16.A.13), the degree of second-order coherence can be written in terms of operators corresponding to the incident field as

$$g^{(2)}(0) = \frac{\left\langle \hat{a}_I^\dagger \hat{a}_I^\dagger \hat{a}_I \hat{a}_I \right\rangle}{\left\langle \hat{a}_I^\dagger \hat{a}_I \right\rangle^2} = \frac{\left\langle : \hat{n}_I^2 : \right\rangle}{\left\langle \hat{n}_I \right\rangle^2} . \tag{16.A.14}$$

Now we are in a position to calculate $g^{(2)}(0)$ for different states of the incident field. For a field in a single-photon state we have

$$g^{(2)}(0) = \frac{\langle 1|\hat{a}_I^\dagger \hat{a}_I^\dagger \hat{a}_I \hat{a}_I|1\rangle}{\langle 1|\hat{a}_I^\dagger \hat{a}_I|1\rangle^2} = \langle 1|\hat{a}_I^\dagger \hat{a}_I^\dagger \hat{a}_I \hat{a}_I|1\rangle = \langle 1|\hat{a}_I^\dagger \hat{a}_I^\dagger \hat{a}_I|0\rangle = 0 , \tag{16.A.15}$$

which is consistent with what we would expect from lab 2. The first measurement of $g^{(2)}(0)$ for a single-photon state was performed by Grangier and coworkers (ref. [16.A.2]).

The more a field fluctuates, the larger $g^{(2)}(0)$ should be. Thus, we would expect a constant-amplitude classical wave to achieve the lowest classically allowed value for $g^{(2)}(0)$. Recall that coherent states are the closest quantum analog of a classical electromagnetic waves with a constant amplitude, and for a coherent state we have

$$g^{(2)}(0) = \frac{\langle\alpha|\hat{a}_I^\dagger \hat{a}_I^\dagger \hat{a}_I \hat{a}_I|\alpha\rangle}{\langle\alpha|\hat{a}_I^\dagger \hat{a}_I|\alpha\rangle^2} = \frac{\alpha^*\alpha^*\alpha\alpha}{\left(\alpha^*\alpha\right)^2} = 1 . \tag{16.A.16}$$

This result is independent of the amplitude of the coherent state; even for weak coherent states, whose mean average photon number is much less than one, we still have $g^{(2)}(0) = 1$.[7] This is consistent with the classical inequality $g^{(2)}(0) \geq 1$, described in lab 2.

16.A.1 References

[16.A.1] R. Loudon, *The Quantum Theory of Light, 3rd ed.* (Oxford University Press, Oxford, 2000).

[16.A.2] P. Grangier, G. Roger, and A. Aspect, “Experimental evidence for a photon anticorrelation effect on a beam splitter: A new light on single-photon interferences,” Europhys. Lett. **1**, 173 (1986).

16.A.2 PROBLEMS

16.A.1 Calculate $g^{(2)}(0)$ for a field in the Fock state $|n\rangle$ ($n \geq 1$). Discuss the limits of large and small n, and compare your answer to the classical inequality $g^{(2)}(0) \geq 1$.

---

[7] A weak coherent state of this sort can be obtained by attenuating the output of a laser.

CHAPTER 17

Quantum Information

In this chapter we'll think about the information content of quantum systems. We'll explore questions such as: How can we represent information in a quantum system? How can we transmit quantum information from one place to another? How can we process quantum information? The goal here is to get you started with the basics; after that, you can continue your study of these topics in a more advanced text, such as refs. [17.1] and [17.2].

17.1 QUBITS AND EBITS

Classical information can be represented as binary bits: 0's and 1's. Standardizing the representation of information as bits makes it easier to transport and process that information. For example, all of the information in your computer is stored and processed as bits. Bits are also used to transfer information from one computer to another. Communication via bits is accomplished over a classical channel.

In quantum mechanics any two orthogonal states can be used to encode bits. For example, the polarization state $\left|H\right\rangle$ could signify 0, while $\left|V\right\rangle$ signifies 1. A bit of information stored in this manner is known as a qubit (for quantum-bit, a term coined by Benjamin Schumacher). We will refer to our qubit states as $\left|0\right\rangle$ and $\left|1\right\rangle$. This notation is generic, and can refer to a large number of different physical systems, such as polarization states, spin states, or atomic states. Communication via qubits is accomplished over a quantum channel.

At first it may seem that there is little difference between a classical bit and a qubit, but that turns out to be far from true. At any instant in time a classical bit can represent either 0 or 1, but not both. However, as we've seen many times before, quantum systems can exist in superposition states. For example, the polarization state $\left|+45\right\rangle = \left(1/\sqrt{2}\right)\left(\left|H\right\rangle + \left|V\right\rangle\right)$ is a superposition of $\left|H\right\rangle$ and $\left|V\right\rangle$ states. In our generic notation we have

$$\left|0'\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle + \left|1\right\rangle\right), \tag{17.1}$$

$$\left|1'\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle - \left|1\right\rangle\right). \tag{17.2}$$

A qubit in either of these states simultaneously signifies *both* 0 *and* 1, a property referred to as quantum parallelism. As we'll see, quantum parallelism can give quantum information processing using qubits an advantage over classical information processing using bits.

Suppose that Alice and Bob each have a qubit, and these qubits are entangled. There are four two-qubit entangled states that we will find useful, and they are known as the Bell states:

$$\left|\phi^{\pm}\right\rangle_{AB} = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_A \left|0\right\rangle_B \pm \left|1\right\rangle_A \left|1\right\rangle_B\right) \tag{17.3}$$

$$\left|\psi^{\pm}\right\rangle_{AB} = \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_A \left|1\right\rangle_B \pm \left|1\right\rangle_A \left|0\right\rangle_B\right) \tag{17.4}$$

The Bell states are orthonormal, and form a basis for a two-qubit system (see problem 8.6). A measurement performed in the Bell-basis is known as a Bell measurement.

A pair of entangled qubits, shared by separated parties, is known as an ebit. If Alice and Bob have a quantum channel, they can share an ebit, which is a resource that they can use to perform tasks that are not allowed by classical physics.

17.2 QUANTUM CRYPTOGRAPHY

In modern society it's important to be able to transmit information from one place to another securely. At present, nearly all secure communication is done using classical cryptographic protocols such as the RSA (Rivest, Shamir, and Adleman) public-key encryption scheme. The security of these schemes is based on the reasoned belief that certain computational problems are "difficult," which means that the computing resources needed for an eavesdropper to decode the message are unreasonably large. In this section we'll discuss quantum cryptography, which can be proven to be completely secure (in principle) against an eavesdropper's attack. First, however, we'll discuss some background on cryptography in general.

17.2.1 Cryptography

In cryptography we refer to three different objects: the plaintext, the key, and the ciphertext. If Alice wants to send a message to Bob, the actual message she wants to send is the plaintext. Modern communication is done electronically, so the plaintext is encoded as a series of binary bits. Alice encrypts the plaintext with the key, which is another set of binary bits. The encrypted message is known as the ciphertext, and this is what she transmits to Bob. Bob has a copy of the key, which he uses to decrypt the ciphertext, and regenerate the plaintext.

For example, suppose Alice wants to send the initials MB to Bob. The standard ASCII encoding scheme assigns one byte (eight bits) to each letter, so Alice needs to send 16 bits to Bob. She uses a key that is also 16 bits long, and produces the ciphertext by performing the XOR (exclusive-or, $\oplus$) operation on the plaintext and the key. The XOR operation yields 1 if either, but not both, of the input bits are 1, and it yields 0 otherwise. Thus, Alice performs:

$$\begin{array}{rl} \text{plaintext:} & \overbrace{01001101}^{\text{M}}\overbrace{01000010}^{\text{B}} \\ \text{key:} & \oplus\ \underline{0100010011101011} \\ \text{ciphertext:} & 0000100110101001\,. \end{array} \tag{17.5}$$

The key is assumed to be perfectly random, so the ciphertext is also perfectly random; in the absence of the key, there is no way to recover the plaintext from the ciphertext. But Bob has a copy of the key, so he performs the XOR operation on the ciphertext and the key to reconstruct the plaintext:

$$\begin{array}{rl} \text{ciphertext:} & 0000100110101001 \\ \text{key:} & \oplus\ \underline{0100010011101011} \\ \text{plaintext:} & \underbrace{01001101}_{\text{M}}\underbrace{01000010}_{\text{B}}\,. \end{array} \tag{17.6}$$

This message transmission is provably secure from attacks by an eavesdropper, Eve, under a few assumptions. The most obvious is the assumption that Eve does not have a copy of the key. The second assumption is that the key is purely random, and consists of at least as many bits as the original plaintext. The last assumption is that the key is used only once, and then discarded.

Classically, the most difficult part of this procedure is sharing the key. Alice and Bob must have copies of the key, and they need to be sure that Eve does not. Standard electronic transmission of the key is no good, because it's always possible that Eve will intercept a copy. One way to eliminate this possibility is for Alice and Bob to meet face-to-face, share the key, and then go their separate ways. They then need to keep their copies of the key secure. Once all the bits in the key have been used they must meet again to share another key, which is inconvenient.

17.2.2 Quantum Key Distribution

Quantum mechanics solves the problem of key generation and distribution. It allows Alice and Bob to remotely share a perfectly random key, and to ensure that Eve does not have a copy of it. This fact was demonstrated by Charles Bennett and Gilles Brassard in 1984, using an encryption scheme that is now known as BB84 [17.3].

First, Alice uses a quantum random number generator to create the key. This random number generator consists of a stream of individual photons incident on a 50/50 beam splitter. If a photon is transmitted it represents 0, while if it is reflected it represents 1. A key generated in this manner is perfectly random.

As shown in fig. 17.1, Alice transmits the key to Bob by sending him qubits. She uses two different bases to send the information: the 01-basis, or the $0'1'$-basis [eqs. (17.1) and (17.2)]. For each qubit she sends, Alice randomly chooses which basis to use (she uses her quantum random number generator to make the choice). At the receiving end, for each qubit Bob randomly decides which basis to perform his measurement in, and sets his measurement apparatus appropriately.

Alice and Bob can only be certain that they agree about the information contained in each qubit if they are using the same bases. For example, if Alice sends $|0'\rangle$ but Bob measures in the 01-basis, he has a 50% chance of measuring $|1\rangle$, which would result in an incorrect bit in his key. To ensure that they are using the same basis, after the measurements are completed, Bob broadcasts over a public, classical channel which basis he used for each measurement, and Alice responds over the same channel telling him whether he was correct or not. They will both be in the same basis 50% of the time, and they keep only those bits of the key; they discard the bits in which they used different bases. In order to protect the key itself, they publicly broadcast only the bases that they used, not which bit value was sent or received.

Alice and Bob now have a key that they can use to encode messages. But how can they be certain that Eve does not have a copy of it as well? Let's assume that Eve has a copy of Bob's receiving apparatus, and Alice's sending apparatus. She inserts herself between Alice and Bob, receives and measures the qubits that Alice sends, and then transmits new qubits to Bob. The qubits that Eve sends are the same as those that she determines from her measurements.

Eve doesn't know which basis Alice is using to encode any individual qubit (Alice and Bob talk about their bases only after all the measurements have been performed), so the best she can do is guess randomly. Consider a qubit which Alice sends as $|0\rangle$. Assume that Eve measures in the $0'1'$-basis, with a result of $1'$, so she sends a $|1'\rangle$ qubit to Bob. Bob is measuring in the 01-basis, so there's a 50% chance that this $|1'\rangle$ qubit will be measured by Bob to be 1, which is not the value that Alice sent.

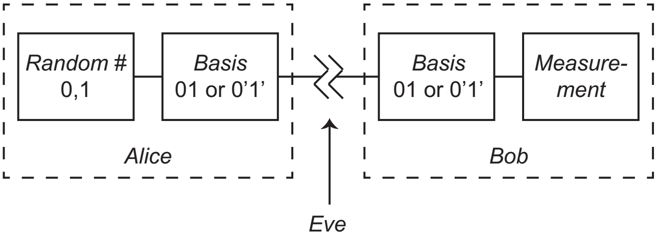


Fig 17.1 The BB84 quantum key distribution protocol. In an implementation using polarized photons, Alice uses $|H\rangle$ for $|0\rangle$ and $|V\rangle$ for $|1\rangle$. She uses a half-wave plate to determine the basis: The wave-plate axis is set at 0° for the 01-basis, and at 22.5° (45° polarization rotation) for the $0'1'$-basis. At the receiving end Bob uses a half-wave plate with the same settings to determine his measurement basis, and his measurement apparatus consists of a polarizing beam splitter with a detector at each output. Eve sits between Alice and Bob, and tries to steal a copy of their key.

Alice and Eve will be in different bases 50% of the time, so Eve will transmit a qubit to Bob in the wrong basis 50% of the time. Of those wrong-basis qubits, 50% of the time Bob will measure a value for a qubit that is not the value sent by Alice. Thus, with Eve listening in, 25% of Bob's key will be wrong. To detect the presence of Eve, Alice and Bob select some random bits from their key and compare them openly over a public, classical channel. If they find that 25% of Bob's bits are wrong, then they know that there is the possibility that Eve is listening in. If 100% of Bob's bits are correct, then they know that there cannot be an eavesdropper. Of course, the bits that they compared openly are discarded from the key.

No real system is perfect, so Alice and Bob will never agree 100% of the time. For each system it is possible to determine a threshold level of agreement (something greater than 75%) for which it is possible to distill a key in which Alice and Bob have all the same bits, and that is essentially useless to Eve. This process of distilling a secret key from a key that might be partially compromised is known as privacy amplification. During privacy amplification bits in the key are sacrificed; the larger the initial error rate, the more bits are lost. Alice and Bob sacrifice enough bits to obtain whatever level of security they desire.

There is one eavesdropping attack that we appear to have ignored: Eve makes a perfect copy of Alice's qubit. Eve measures the original qubit, and sends the copy on to Bob. If Bob gets a copy, or clone, of the original qubit, then he won't have any measurement errors, and he and Alice will have no way of detecting Eve's presence. As we'll see in the next section, however, quantum mechanics does not allow the perfect copying of an arbitrary quantum state, so this method of attack is not viable for Eve.

Finally, note that quantum cryptography has moved out of the laboratory and into commercial use. Several companies currently sell off-the-shelf quantum key distribution systems.

17.3 THE NO-CLONING THEOREM

We wish to create a cloning machine, which can copy an arbitrary state of system 1, $|\psi\rangle_1$, onto the state of system 2. System 2 starts out in a "blank" state $|B\rangle_2$, and if the cloning process is successful it ends up in state $|\psi\rangle_2$, which is a perfect copy of $|\psi\rangle_1$. We'll represent the cloning machine by the unitary operator $\check{U}_c$, and a successful cloning operation is expressed as

$$\hat{U}_c|\psi\rangle_1|B\rangle_2 = |\psi\rangle_1|\psi\rangle_2. \tag{17.7}$$

Suppose that system 1 consists of a qubit. If the qubit is in state $|0\rangle_1$ cloning operation is

$$\hat{U}_c|0\rangle_1|B\rangle_2 = |0\rangle_1|0\rangle_2, \tag{17.8}$$

while if the qubit is in state $|1\rangle_1$ the cloning operation is

$$\hat{U}_c|1\rangle_1|B\rangle_2 = |1\rangle_1|1\rangle_1. \tag{17.9}$$

If the qubit is in the superposition state $\left|\psi\right\rangle_1 = \alpha\left|0\right\rangle_1 + \beta\left|1\right\rangle_1$, eqs. (17.8) and (17.9) tell us that the cloning machine performs the operation

$$\begin{aligned}\hat{U}_c\left(\alpha\left|0\right\rangle_1 + \beta\left|1\right\rangle_1\right)\otimes\left|B\right\rangle_2 &= \alpha\hat{U}_c\left|0\right\rangle_1\left|B\right\rangle_2 + \beta\hat{U}_c\left|1\right\rangle_1\left|B\right\rangle_2 \\ &= \alpha\left|0\right\rangle_1\left|0\right\rangle_2 + \beta\left|1\right\rangle_1\left|1\right\rangle_2 .\end{aligned} \tag{17.10}$$

This is *not* the state we'd like our cloning machine to produce:

$$\hat{U}_c\left(\alpha\left|0\right\rangle_1 + \beta\left|1\right\rangle_1\right)\otimes\left|B\right\rangle_2 \neq \left(\alpha\left|0\right\rangle_1 + \beta\left|1\right\rangle_1\right)\otimes\left(\alpha\left|0\right\rangle_2 + \beta\left|1\right\rangle_2\right). \tag{17.11}$$

Thus, quantum mechanics does not allow for the existence of a device that can perfectly copy an arbitrary state of one system onto another. This is known as the no-cloning theorem, and it was proved in 1982 by Wooters and Zurek [17.4], and independently by Dieks [17.5]. The no-cloning theorem is clearly good for cryptography: It rules out the possibility that Eve could intercept Alice and Bob's key without their knowledge.

Another thing the no-cloning theorem rules out is the possibility that we could completely determine the state of a single copy of a quantum system. As described in complement 5.A and lab 4, fully determining the state of a quantum system requires many measurements, of many different observables, performed on a quantum system repeatedly prepared in the same state. If we could perfectly clone an individual realization of the ensemble, it would be possible to perform measurements on the clones, and hence determine the state of the individual quantum system. The no-cloning theorem will not allow this, so we cannot, even in principle, determine the full quantum state of an individual member of an ensemble.

17.3.1 No Super-Luminal Communication

The no-cloning theorem helps ensure that super-luminal communication is not possible. To see this, we first need to imagine how such communication might be accomplished. Assume that Alice and Bob are on opposite sides of the solar system, and manage to share an ebit in the Bell-state $\left|\phi^+\right\rangle_{AB}$, given by eq. (17.3). If Alice performs a measurement on her qubit, she collapses the state of the ebit, instantaneously affecting the state of Bob's qubit. How might Alice and Bob use this state collapse to communicate?

Assume Alice performs a measurement on her qubit in the 01-basis.The measurement result is either 0 or 1, it doesn't matter—the important thing is that Bob's qubit is left in either the state $\left|0\right\rangle_B$ or the state $\left|1\right\rangle_B$. Now, suppose that instead Alice performs a measurement on her qubit in the $0'1'$-basis. Recall that entanglement exists in any basis; in the $0'1'$-basis the ebit state $\left|\phi^+\right\rangle_{AB}$ is

$$\left|\phi^+\right\rangle_{AB} = \frac{1}{\sqrt{2}}\left(\left|0'\right\rangle_A\left|0'\right\rangle_B + \left|1'\right\rangle_A\left|1'\right\rangle_B\right), \tag{17.12}$$

as you'll show in problem 17.1. Alice's measurement in the $0'1'$-basis will project Bob's qubit into either the state $|0'\rangle_B$ or the state $|1'\rangle_B$.

Can Alice transmit information to Bob simply by deciding whether to perform a measurement in the 01-basis or the $0'1'$-basis? Regardless of the outcomes of these measurements, a measurement in the 01-basis leaves Bob's qubit in one of the states $|0\rangle_B$ or $|1\rangle_B$, while a measurement in the $0'1'$-basis leaves Bob's qubit in one of the states $|0'\rangle_B$ or $|1'\rangle_B$. If Bob could measure the state of his qubit, he could tell which measurement Alice performed, and hence obtain information from her. However, as described above, the no-cloning theorem ensures that it is not possible for Bob to determine the state of an individual qubit. Alice and Bob cannot use this technique to communicate at super-luminal speeds.

17.3.2 Imperfect Cloning

The no-cloning theorem prohibits the perfect copying of arbitrary states, but it does leave open a few other possibilities. One is the ability to make imperfect copies of states, with limits on how accurate the copies can be. Another is the ability to perfectly clone some states, but not others. For example, a certain cloning machine might copy orthogonal basis states [eqs. (17.8) and (17.9)], but not superpositions of these states. Yet another possibility is the potential for a probabilistic cloning machine. When such a machine succeeds it creates a perfect clone, but when it fails it also destroys the original input and its associated information. For some discussion of these possibilities, see appendix F of ref. [17.2].

17.4 QUANTUM TELEPORTATION

Suppose that Alice has a qubit in some unknown quantum state in her laboratory, and she wants to send that state to Bob, who's on the other side of the solar system. It's not possible for her to send her qubit directly, but she'd like to transmit the information about the state to Bob, so that he can have a copy. If Alice and Bob can do this, we say that they have successfully teleported the state of the system. Note that this teleportation is not like *Star Trek* teleportation. Here we're somehow transmitting the information about the state of the qubit and allowing it to be recreated somewhere else; the qubit itself doesn't vanish in one place and reappear in another.

Let's use some things that we already know, to infer some of the properties of quantum teleportation. First, after teleportation Alice's original qubit can no longer be in the same state; if it was then Alice and Bob would have managed to clone the state of Alice's qubit, which is forbidden by the no-cloning theorem. Second, Alice can't simply measure the state and send the information about it to Bob using a classical channel, because we showed in sec. 17.3 that it's not possible to measure the state of an individual quantum system. Since no amount of purely classical information can teleport the state, to be successful Alice and Bob must share some quantum information: Alice and Bob must share an ebit.

At the start, Alice has a qubit in state $|\psi\rangle_a = \alpha|0\rangle_a + \beta|1\rangle_a$, and Alice and Bob have a shared ebit in state $|\phi^+\rangle_{AB}$ [eq. (17.3)]. The total state of the system is thus

$$\begin{aligned}|\Psi\rangle &= |\psi\rangle_a \otimes |\phi^+\rangle_{AB}\\ &= \frac{1}{\sqrt{2}}\Big[\big(\alpha|0\rangle_a + \beta|1\rangle_a\big)\otimes|0\rangle_A|0\rangle_B + \big(\alpha|0\rangle_a + \beta|1\rangle_a\big)\otimes|1\rangle_A|1\rangle_B\Big].\end{aligned} \tag{17.13}$$

You'll show in problem 17.2 that $|\Psi\rangle$ can be written in the Bell-basis of *Alice's two qubits* as[1]

$$\begin{aligned}|\Psi\rangle = &\frac{1}{2}|\phi^+\rangle_{aA}\otimes\big(\alpha|0\rangle_B + \beta|1\rangle_B\big)\\ &+\frac{1}{2}|\phi^-\rangle_{aA}\otimes\big(\alpha|0\rangle_B - \beta|1\rangle_B\big)\\ &+\frac{1}{2}|\psi^+\rangle_{aA}\otimes\big(\beta|0\rangle_B + \alpha|1\rangle_B\big)\\ &+\frac{1}{2}|\psi^-\rangle_{aA}\otimes\big(-\beta|0\rangle_B + \alpha|1\rangle_B\big).\end{aligned} \tag{17.14}$$

From this expression, it is fairly obvious that if Alice performs a Bell measurement that indicates her qubits are in the state $|\phi^+\rangle_{aA}$, Bob's qubit is projected into the state $|\psi\rangle_B = \alpha|0\rangle_B + \beta|1\rangle_B$, which is exactly the desired state. Thus, by simply performing a Bell measurement, Alice successfully teleports the state to Bob, with probability $(1/2)^2 = 1/4$, with no further action on Bob's part. Because of her measurement result, Alice knows when this has occurred, and can inform Bob over a classical channel. Furthermore, Alice's particles are projected into the state $|\phi^+\rangle_{aA}$, so she retains no information about her original state $|\psi\rangle_a$, which ensures that the no-cloning theorem is satisfied. Thus, Alice and Bob have managed to teleport the state of a qubit, albeit with a success probability of 25%.

Suppose, however, that Alice's Bell measurement projects her qubits into the state $|\phi^-\rangle_{aA}$, which also occurs with probability $(1/2)^2 = 1/4$. From eq. (17.14) we see that Bob's qubit is projected into the state $|\psi\rangle_B = \alpha|0\rangle_B - \beta|1\rangle_B$. Bob can transform this state into the desired state by simply applying a π phase shift to the $|1\rangle_B$ component of his qubit. The unitary transformation $\hat{U}_{\phi^-}$ that accomplishes this task has the matrix representation

$$\hat{U}_{\phi^-} \doteq \begin{pmatrix}1 & 0\\ 0 & e^{i\pi}\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \tag{17.15}$$

---

[1] Note that $|\phi^+\rangle_{aA} \neq |\phi^+\rangle_{AB}$: the subscripts tell us which qubits are in the indicated state.

For polarization qubits, Bob can accomplish this transformation by simply using a half-wave plate whose axis is oriented horizontally (see table 2.2). Thus, Alice can inform Bob that her Bell measurement resulted in $\left|\phi^-\right\rangle_{aA}$, and Bob knows to apply the transformation $\hat{U}_{\phi^-}$ to his qubit to obtain the desired state.

Equation (17.14) indicates that there are two other possible outcomes of Alice's Bell measurement. As you'll show in the problems, in either case there is a unitary transformation that Bob can apply to transform his qubit into the same state as Alice's original qubit. In all cases information about the state of Alice's qubit is erased at Alice's location. Thus, the following protocol, illustrated in fig. 17.2, will allow Alice to successfully teleport the state of her qubit to Bob:

1. Alice and Bob share an ebit.
2. Alice performs a Bell measurement on the joint state of the qubit and her portion of the ebit. There are four possible outcomes for this measurement, which requires two classical bits to encode, and she spends this information to Bob over a classical channel.
3. Bob receives the information from Alice about the result of her measurement. This tells him exactly which of four unitary transformations he needs to apply to his portion of the ebit, in order to transform its state into the state of Alice's original qubit.

Quantum teleportation was first proposed in 1993 by Charles Bennett and coworkers [17.6], and was experimentally demonstrated a few years later.

17.5 QUANTUM COMPUTING

In this section we will concentrate on the processing of quantum information, with an emphasis on how quantum information processing is different from classical information processing. In order to do that, we'll need to briefly discuss classical computation.

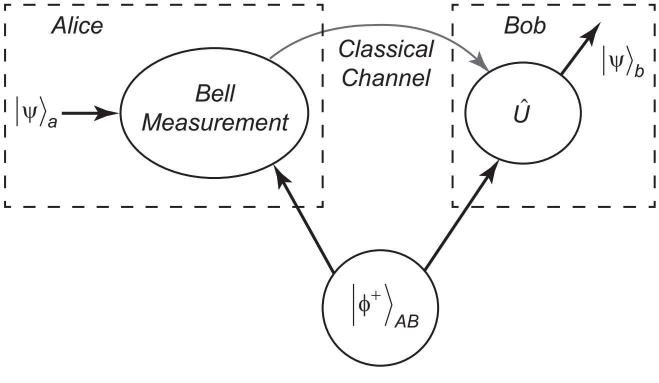


Fig 17.2 Quantum teleportation.

17.5.1 Classical Gates

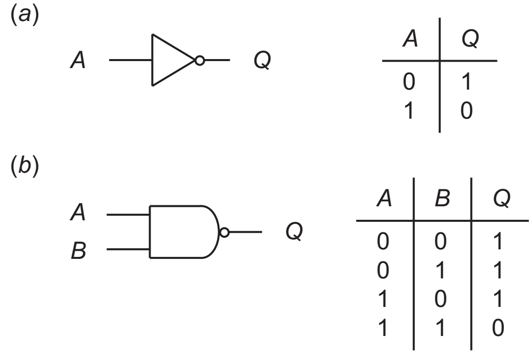
Modern computer microprocessors are phenomenally complicated objects: They contain literally billions of transistors. When thinking about microprocessors, however, we don't usually talk about individual transistors, but the logic operations that they perform. These logic operations are performed by gates that operate on bits, which are in the state 0 or 1. A logic operation can be specified by its truth table, which tells what the gate output(s) will be, given its input(s). In the circuit model of computation, gates are represented as circuit elements, and wires indicate the flow of bits from one gate to another. For example, fig. 17.3 shows the circuit elements and truth tables for a NOT gate and a NAND (NOT-AND) gate.

For classical computation the NAND gate is said to be a universal gate, because it can be used to implement any other logic operation. For example, fig. 17.4 shows how to implement the XOR operation using NAND gates. This figure also illustrates another aspect of classical logic circuits: fanout. Fanout means that it's possible to copy classical bits, simply by connecting wires together.

17.5.2 Quantum Gates

There are several important differences between classical computation and quantum computation. First, quantum computation is performed on qubits, which you recall can be in a superposition of both $|0\rangle$ and $|1\rangle$. Second, quantum computation is performed with unitary operators, which always have the same number of inputs and outputs. Thus, for example, there is no quantum analog of the NAND gate, which takes two inputs, but has only one output. Third, there is no possibility for fanout of arbitrary qubit states, as it is forbidden by the no-cloning theorem.

We'll discuss one- and two-qubit gates. The X-gate is a one-qubit gate. In the 01-basis it is the quantum equivalent of a NOT gate, as it transforms $|0\rangle$ into $|1\rangle$, and vice versa (see also problem 17.6). Its matrix representation is



A	Q
0	1
1	0

A	B	Q
0	0	1
0	1	1
1	0	1
1	1	0

Fig 17.3 (a) A NOT gate and corresponding truth table. (b) A NAND gate and corresponding truth table.

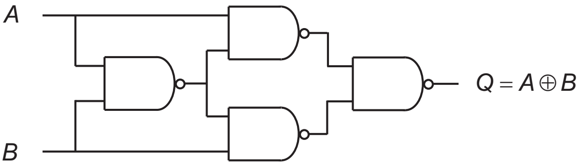
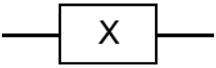
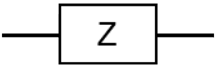
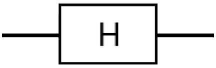
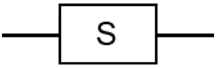


Fig 17.4 XOR using NAND gates.

Table 17.1 One-qubit quantum gates.

Name	Circuit Symbol	Matrix Representation
X (NOT)		$\hat{\mathrm{X}} \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
Y		$\hat{\mathrm{Y}} \doteq \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$
Z		$\hat{\mathrm{Z}} \doteq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
Hadamard		$\hat{\mathrm{H}} \doteq \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
Phase		$\hat{\mathrm{S}} \doteq \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$

$$\hat{\mathrm{X}} \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{17.16}$$

This gate is called the X-gate because its matrix representation is the same as the Pauli matrix $\overline{\boldsymbol{\sigma}}_x$; indeed, it is sometimes referred to as the Pauli-X gate. There is also a Y-gate and a Z-gate, which have the same matrix representations as $\overline{\boldsymbol{\sigma}}_y$ and $\overline{\boldsymbol{\sigma}}_z$. Two other important one-qubit gates are the Hadamard gate H, and the phase gate S.[2] The matrix and quantum-circuit representations of these one-qubit gates are shown in table 17.1.

Some of the most useful of the two-qubit gates are the controlled gates. The input qubits to a controlled gate are called the control (C) and the target (T); if the control qubit is $|0\rangle$ the target qubit is unchanged, while if the control is $|1\rangle$ the target undergoes some unitary transformation. For example, in a controlled-NOT (or controlled-X) gate, the NOT (Pauli-X) operation is performed on the target qubit, if the control qubit is in state $|1\rangle$. Controlled-NOT is usually referred to as CNOT. In a two-qubit space four basis vectors are $|00\rangle, |01\rangle, |10\rangle$ and $|11\rangle$, where the first number refers to the control, and the second to the target. In this basis the CNOT operation has the matrix representation

---

[2] The corresponding operators $\hat{\mathrm{H}}$ and $\hat{\mathrm{S}}$ are not italicized, to distinguish them from the Hamiltonian and spin operators.

Table 17.2 The truth table for a controlled-U operation.

C_{in}	T_{in}	C_{out}	T_{out}
$\|0\rangle$	$\|0\rangle$	$\|0\rangle$	$\|0\rangle$
$\|0\rangle$	$\|1\rangle$	$\|0\rangle$	$\|1\rangle$
$\|1\rangle$	$\|0\rangle$	$\|1\rangle$	$\hat{U}\|0\rangle$
$\|1\rangle$	$\|1\rangle$	$\|1\rangle$	$\hat{U}\|1\rangle$

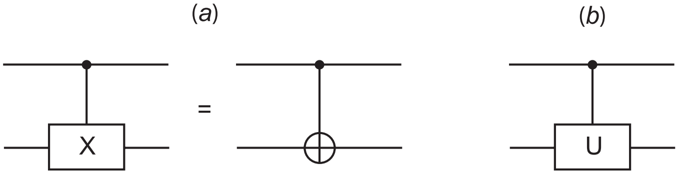


Fig 17.5 (a) Two representations of a CNOT gate. (b) A controlled-U gate. The control bit moves along the top wire, and the target bit moves along the bottom.

$$\widehat{\text{CNOT}} \doteq \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{17.17}$$

In general, a controlled-U (controlled-unitary) operation will behave as shown in table 17.2.

Because all quantum gates perform unitary operations, two-qubit gates have two inputs and two outputs. In quantum-circuit diagrams, qubits move left-to-right along horizontal lines. There are two different symbols for the CNOT operation, which are shown in fig. 17.5; also shown in this figure is the symbol for a general controlled-U operation.

Let's look at an example of a quantum circuit.

EXAMPLE 17.1

What is the output state of the quantum circuit shown in fig. 17.6?

The input state is

$$|\psi_1\rangle = |0\rangle_1 |1\rangle_2 . \tag{17.18}$$

The Hadamard gate operates on qubit 1, and after this gate the state is

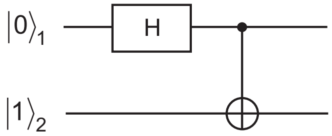


Fig 17.6 The quantum circuit considered in example 17.1.

$$\begin{aligned} |\psi_2\rangle &= \hat{\mathrm{H}}|0\rangle_1 \otimes |1\rangle_2 \\ &= \left[\frac{1}{\sqrt{2}}\left(|0\rangle_1 + |1\rangle_1\right)\right] \otimes |1\rangle_2 \\ &= \frac{1}{\sqrt{2}}\left(|0\rangle_1 |1\rangle_2 + |1\rangle_1 |1\rangle_2\right). \end{aligned} \tag{17.19}$$

After the CNOT gate we end up in one of the Bell states:

$$\begin{aligned} |\psi_3\rangle &= \widehat{\mathrm{CNOT}}|\psi_2\rangle \\ &= \frac{1}{\sqrt{2}}\left(|0\rangle_1 |1\rangle_2 + |1\rangle_1 |0\rangle_2\right) \\ &= \left|\psi^+\right\rangle_{1,2}, \end{aligned} \tag{17.20}$$

and our quantum circuit has managed to entangle the two inputs.

Any operation on N-qubits can be performed by a combination of one- and two-qubit gates. Furthermore, the CNOT gate is universal, in the sense that any two-qubit operation can be performed by a combination of CNOT's and one-qubit operations. Thus, in principle, any quantum information processor can be built with one-qubit and CNOT gates.

17.5.3 Deutsch's Algorithm

The advantage of a quantum computer is that it can perform certain tasks faster than any classical computer. To achieve this speed-up, a quantum computer uses two properties of qubits that set them apart from ordinary bits. The first is the quantum parallelism discussed in sec. 17.1. The second property that qubits display is interference; ordinary bits cannot interfere.

The first algorithm which demonstrated that a quantum computer could perform faster than a classical computer was due to David Deutsch [17.7]. The problem considered by Deutsch may seem somewhat contrived, but the fundamentals behind it serve as the basis for other, more obviously useful algorithms. Here we'll consider Deutsch's algorithm as an example of a very simple quantum computer.

Consider a function $f(x)$ which takes as its input a single bit $\{0,1\}$, and gives as an output a single bit $\{0,1\}$. There are four possible functions that operate on single bit inputs and outputs, and their behavior is shown in table 17.3. The functions $f_1(x)$ and $f_2(x)$ are known as "constant" functions, because their outputs are constant, independent of the input. The functions $f_3(x)$ and $f_4(x)$ are known as "balanced" functions; they are balanced in the sense that the outputs 0 and 1 occur equally often. The Deutsch problem is as follows: Someone gives you one of these functions, and you need to determine whether the function is constant or balanced. The key here is that you don't care which function you are given, but only whether it is constant [$f_1(x)$ or $f_2(x)$] or balanced [$f_3(x)$ or $f_4(x)$].

To solve this problem classically, how many times do you need to evaluate $f(x)$? The answer is twice. For example, suppose you input 0 to the black box, and the output is 0. Table 17.3 tells you that your function is either $f_1(x)$ or $f_3(x)$. One of these functions is constant, the other is balanced, so you must also evaluate $f(x)$ with an input of 1. If the output is again 0 you know that the function is constant. Indeed, you not only know that the function is constant, but that it is $f_1(x)$.

Now let's show that when using a quantum computer we only need to evaluate $f(x)$ once, which is half as often as is needed classically. The circuit diagram for this computer is shown in fig. 17.7. We'll describe the operation of the computer by examining the state of the qubits at the locations indicated in this figure. The input state is

$$|\psi_1\rangle = |0\rangle_1 |1\rangle_2 \tag{17.21}$$

Each of the input qubits goes through a Hadamard gate, which results in the state

$$\begin{aligned} |\psi_2\rangle &= \left[\frac{1}{\sqrt{2}}\left(|0\rangle_1 + |1\rangle_1\right)\right] \otimes \left[\frac{1}{\sqrt{2}}\left(|0\rangle_2 - |1\rangle_2\right)\right] \\ &= \frac{1}{2}\left(|0\rangle_1|0\rangle_2 - |0\rangle_1|1\rangle_2 + |1\rangle_1|0\rangle_2 - |1\rangle_1|1\rangle_2\right). \end{aligned} \tag{17.22}$$

The gate labeled U_f in fig. 17.7 works as follows. The input labeled x is passed to one of the outputs. The gate also computes $f(x)$, and then performs the XOR operation on $f(x)$ and the input y. The result of this operation, $y \oplus f(x)$, is the second output. For example, suppose the unknown function is $f_1(x)$, and the inputs are $x = 0$ and $y = 0$. From table 17.3 we know that $f_1(0) = 0$, so the outputs are $x = 0$ and $y \oplus f(x) = 0 \oplus 0 = 0$. You can verify that if the unknown function is $f_1(x)$, the state after U_f is

Table 17.3 The outputs of the four possible functions which take a single bit (x) as input, and yield a single bit of output.

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
0	0	1	0	1
1	0	1	1	0

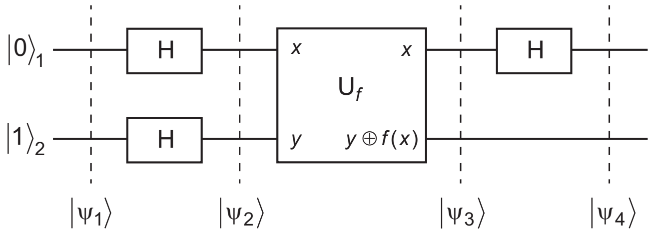


Fig 17.7 A quantum circuit implementing Deutsch's algorithm.

$$\begin{aligned}|\psi_3\rangle &= \frac{1}{2}\left(|0\rangle_1|0\rangle_2 - |0\rangle_1|1\rangle_2 + |1\rangle_1|0\rangle_2 - |1\rangle_1|1\rangle_2\right)\\ &= \left[\frac{1}{\sqrt{2}}\left(|0\rangle_1 + |1\rangle_1\right)\right]\otimes\left[\frac{1}{\sqrt{2}}\left(|0\rangle_2 - |1\rangle_2\right)\right].\end{aligned} \tag{17.23}$$

This is the same as the input state, so if the unknown function is $f_1(x)$, then this gate simply performs the identity operation: $\hat{U}_{f_1} = \hat{1}$.

In problem 17.10 you'll verify that the state after U_f depends on the unknown function, and is given by

$$|\psi_3\rangle = \begin{cases} \pm\left[\frac{1}{\sqrt{2}}\left(|0\rangle_1 + |1\rangle_1\right)\right]\otimes\left[\frac{1}{\sqrt{2}}\left(|0\rangle_2 - |1\rangle_2\right)\right] & f = f_1, f_2 \\ \pm\left[\frac{1}{\sqrt{2}}\left(|0\rangle_1 - |1\rangle_1\right)\right]\otimes\left[\frac{1}{\sqrt{2}}\left(|0\rangle_2 - |1\rangle_2\right)\right] & f = f_3, f_4 \,. \end{cases} \tag{17.24}$$

Note that apart from insignificant overall phase factors, the two constant functions yield one output state, while the two balanced functions yield a different state. Furthermore, all four functions yield the same state for the second qubit, and it is the first qubit which is in a different state depending upon the type of function. Finally, the first qubit passes through a Hadamard gate, which transforms the state to

$$|\psi_4\rangle = \begin{cases} \pm|0\rangle_1 \otimes\left[\frac{1}{\sqrt{2}}\left(|0\rangle_2 - |1\rangle_2\right)\right] & f = f_1, f_2 \\ \pm|1\rangle_1 \otimes\left[\frac{1}{\sqrt{2}}\left(|0\rangle_2 - |1\rangle_2\right)\right] & f = f_3, f_4 \,. \end{cases} \tag{17.25}$$

Now we see that it is possible to easily tell what type of function $f(x)$ is. We measure the first qubit, and if the result is 0 the function is constant, while if it is 1 the function is balanced. Note that we have made this determination by evaluating $f(x)$ only once.

Classically, it is not possible to answer the question of whether the function is constant or balanced without simultaneously determining which of the four functions is in the black box. In solving this problem classically we end up with more information than we need. The quantum computer is more economical in its solution to the Deutsch problem: It returns only the information we want, and does so with fewer function evaluations.

17.5.4 Other Quantum Algorithms

The Deutsch algorithm is important because it was the first to demonstrate an improvement in speed for a quantum computer, not because of its practical utility. This algorithm cuts the number of function evaluations in half, from 2 to 1, which isn't much of a speed improvement. However, in 1992 Deutsch and Richard Jozsa extended Deutsch's algorithm to n-bit functions, and showed that a quantum computer could solve the n-bit Deutsch problem exponentially faster than a classical computer.

There are other important quantum algorithms. Lov Grover has shown that a quantum computer can search a database faster than a classical computer. Peter Shor has shown that a quantum computer can factor large numbers faster than a classical computer, which has important implications for cryptography. The security of the RSA encryption protocol mentioned in sec. 17.2 is based on the fact that factoring large numbers (of order 1000 bits) on a classical computer is impractical. With a sufficiently large quantum computer this factorization would be feasible, making RSA encryption no longer secure. Of course, even a quantum computer would not be able to crack the quantum encryption scheme described in sec. 17.2.

17.5.5 Real Quantum Computers

In principle, there are as many different physical systems which can implement a quantum computer as there are systems which can implement qubits. These systems include photons, trapped atoms, quantum dots, superconducting circuits, and nuclear magnetic resonance. For example, it is possible to implement Deutsch's algorithm using techniques and optical elements (wave plates, beam splitters, etc.) used in the labs described at the end of this text [17.8]. For a discussion of different physical implementations of quantum computers, see refs. [17.1] and [17.9].

In addition to the model of a quantum computer using gates and circuits described above, there are other, very different models for quantum computation. In cluster-state quantum computation, the system is initialized in an entangled state of many qubits, called a cluster state, and the computation proceeds by performing a series of single-qubit measurements. In adiabatic quantum computation, the solution to a problem is represented by the ground state of a complex Hamiltonian. The system is initialized in the ground state of a simple Hamiltonian, which is then adiabatically transformed into the desired Hamiltonian. These other models for quantum computation are equivalent in power to a quantum computer implemented with gates.

One thing any real quantum computer will need to deal with is decoherence. It is necessary for the quantum states of different qubits to interfere with each other in order to perform quantum computations. Similar to the optical coherence described in complement 2.A, qubits must have stable phase relationships to interfere, and to remain entangled with each other (see sec. 8.A.4). Outside influences, such as losses or stray fields, can cause qubits to decohere. The larger the number of qubits that are involved, the more rapidly decoherence occurs. It is possible to implement quantum error correction to reduce the effects of decoherence, but decoherence invariably plays a role in limiting the performance of a quantum computer.

To date quantum computers have performed tasks such as implementing Deutsch's algorithm, and factoring the number 15 using Shor's algorithm. These tasks involve relatively small numbers of qubits, but scaling up is difficult. Entanglement among a large number of qubits is necessary for quantum computation, and as I write this the largest number of qubits that have been entangled in an experiment is 14. Quantum computing has a great deal of promise, but it will require significant work to realize that promise.

17.6 References

[17.1] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[17.2] S.M. Barnett, *Quantum Information* (Oxford University Press, Oxford, 2009).

[17.3] C.H. Bennett and G. Brassard, "Quantum cryptography: public key distribution and coin tossing," in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing*, Bangalore, India, p. 175 (1984).

[17.4] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," Nature **299**, 802 (1982).

[17.5] D. Dieks, "Communication by EPR devices," Phys. Lett. A 92, 271 (1982).

[17.6] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W.K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels," Phys. Rev. Lett. **70**, 1895 (1993).

[17.7] D. Deutsch, "Quantum-theory, the Church-Turing principle and the universal quantum computer," Proc. R. Soc. London A **400**, 97 (1985).

[17.8] P. Zhang, R.F. Liu, Y.F. Huang, H. Gao, and F.L. Li, "Demonstration of Deutsch's algorithm on a stable linear optical quantum computer," Phys. Rev. A **82**, 064302 (2010).

[17.9] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, "Quantum computers," Nature **464**, 45 (2010).

17.7 PROBLEMS

17.1* Prove that if the state $\left|\phi^{+}\right\rangle_{AB}$ in eq. (17.3) is expressed in the $0'1'$-basis, the result is given by eq. (17.12).

17.2* Verify that eq. (17.13) can be rewritten in the form given in eq. (17.14).

17.3 Determine the two transformations $\hat{U}_{\psi^{\pm}}$ that Bob must perform on his qubits in order to successfully perform teleportation when the results of Alice's Bell measurement are $\left|\psi^{\pm}\right\rangle_{aA}$.

17.4 Show that quantum teleportation can be accomplished with an ebit in the state $\left|\psi^{+}\right\rangle_{AB}$. (An ebit in any of the Bell states can be used for teleportation.)

17.5 Prove that the circuit in fig. 17.4 performs the XOR operation.

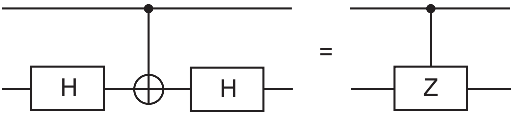


Fig 17.8 A CZ gate implemented with Hadamard and CNOT gates.

17.6 Determine the operation of an X-gate on the states $|0'\rangle$ and $|1'\rangle$. Does it behave like a NOT gate for these states?

17.7 Show that the Bell state $|\phi^+\rangle_{AB}$ can be transformed into any of the other Bell states by using one of the Pauli gates (X, Y, or Z) to operate on one of the qubits. (Any of the Bell states can be transformed into any of the other Bell states in a similar manner.)

17.8* (a) Work out the truth table (similar to that of table 17.2) for a CNOT gate in 01-basis. (b) Work out the truth table for a CNOT gate in $0'1'$-basis. (c) The quantum circuit symbol for a CNOT gate seems to indicate that the control bit is not changed by the operation of the gate. Is this always true? In the $0'1'$-basis, which qubit appears to be the control, and which qubit appears to be the target?

17.9 (a) Work out the matrix representation for a controlled-Z (CZ) gate in the 01-basis. (b) Show that the circuit of fig. 17.8 implements the CZ operation.

17.10 Verify eq. (17.24).

17.11* If the U_f-gate of fig. 17.7 is programmed with $f_1(x)$, the gate implements the unitary transformation $\hat{U}_{f_1} = \hat{1}$. (a) Work out the unitary transformations that correspond to the other three possible functions. Express your answers in terms of the one-qubit unitary operators in table 17.1 and $\widehat{\text{CNOT}}$. (b) Find quantum circuit diagrams for the four possible U_f gates used in implementing Deutsch's algorithm.

Laboratories

Previously in this text we have discussed certain physical phenomena that are counterintuitive. For example, it's difficult to understand how a single photon will interfere with itself, or how local realism can be violated. Because these phenomena defy common sense, it is important to verify them experimentally. This is what you will do in the laboratories described here.

GETTING STARTED

Suggestions on the equipment used to perform these laboratories can be found in a number of places. The original journal articles that first described many of these experiments often contain equipment lists. References to these articles are presented at the end of each individual laboratory.[1] Equipment lists are also available on several websites; my website http://www.whitman.edu/~beckmk/QM/ contains such a list.

The LabVIEW programs used to acquire the experimental data are also available from my website. The descriptions of the laboratories presented here assume the use of these programs, but the labs can be easily adapted for use with other software.

BEFORE LAB

Each lab contains a "Lab Ticket" that you should complete before attending the lab.

---

[1] Please keep in mind, however, that technology and manufacturers change. Original equipment may now be out of date or unavailable.

IMPORTANT LABORATORY SAFETY TIPS

- For your safety, **NEVER LOOK DIRECTLY INTO THE LASER.** You should **WEAR THE LASER SAFTY GOGGLES** while in the lab; they will block the intense blue laser light, but still allow you to see.
- **REMOVE YOUR WATCH OR ANY OTHER SHINY JEWLERY FROM YOUR HANDS OR WRISTS**. A reflection off of a shiny surface can be accidentally directed to someone's eye.
- For the safety of the equipment, **NEVER TURN ON THE ROOM LIGHTS WHILE THE PHOTON COUNTING MODULES ARE ON**.

LAB 1

Spontaneous Parametric Downconversion

LAB TICKET

You have a nonlinear crystal which has been cut to downconvert pump light at 405 nm into signal and idler beams at 810 nm. If the index of refraction of the crystal is 1.659 at 405 nm, and 1.661 at 810 nm, what angle do the signal and idler beams make with respect to the pump?

L1.1 INTRODUCTION

All of the experiments described here use photon pairs produced via spontaneous parametric downconversion as a light source. This physical process comes in several variations, but at its most basic level it is a process in which light of one frequency is converted into light of a different frequency. Any optical process which changes the frequency of a light beam is inherently nonlinear. Most of the other optical processes you are probably familiar with (absorption, reflection, refraction, polarization rotation, etc.) are linear processes; they may affect many properties of a light field, but linear processes can never change the frequency.

In the process of spontaneous parametric downconversion, shown schematically in fig. L1.1, a single photon of one frequency is converted into two photons of lower frequency (by approximately a factor of 2) in a nonlinear crystal. While downconversion is extremely inefficient (10s of milliwatts of input power generate output beams that must be detected using photon counting) it is much more efficient than other sources of photon pairs (for example, atomic emission of 2 photons).

The input wave is referred to as the pump (at angular frequency ω_p), while the two outputs are referred to as the signal and idler (at angular frequencies ω_s and ω_i). Spontaneous parametric downconversion is said to be "spontaneous" (as opposed to "stimulated") because there are no input signal and idler fields—they're generated spontaneously inside the crystal. The process is "parametric" because it depends on the electric fields, and not just their intensities. This means that there is a definite phase relationship between the

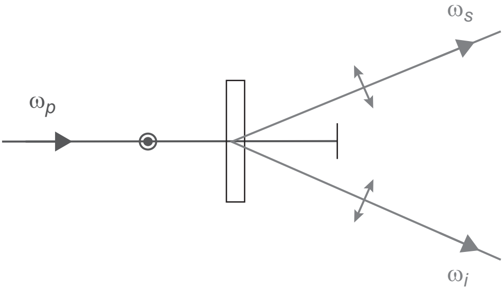


Fig L1.1 Type-I spontaneous parametric down conversion. Inside a crystal one pump photon at angular frequency ω_p is converted into signal and idler photons at angular frequencies ω_s and ω_i. The polarizations of the signal and idler photons are orthogonal to that of the pump.

input and output fields. It is called "downconversion" because the signal and idler fields are at a lower frequency than the pump field.

Energy conservation requires that the energy of the pump photon is equal to the sum of the energies of signal and idler photons:

$$\begin{aligned} \hbar\omega_p &= \hbar\omega_s + \hbar\omega_i \ , \\ \omega_p &= \omega_s + \omega_i \ . \end{aligned} \tag{L1.1}$$

We see that energy conservation implies that the frequencies of the signal and idler waves add up to the frequency of the pump. Momentum conservation is equivalent to a classical condition known as phase-matching, which requires that the wave vectors of the input and output fields satisfy:

$$\begin{aligned} \hbar\boldsymbol{k}_p &= \hbar\boldsymbol{k}_s + \hbar\boldsymbol{k}_i \ , \\ \boldsymbol{k}_p &= \boldsymbol{k}_s + \boldsymbol{k}_i \ . \end{aligned} \tag{L1.2}$$

Recall from chap. 2 that the frequencies and wave vectors are not independent of each other; they are related by a dispersion relationship. For the pump wave we have

$$k_p = \frac{n_p \omega_p}{c}, \tag{L1.3}$$

where n_p is the index of refraction of the downconversion crystal at the pump frequency. There are similar expressions for the signal and idler waves.

It is important to note that the indices of refraction depend on the frequency: $n(\omega)$. If the pump, signal, and idler waves are nearly collinear, eqs. (L1.1) and (L1.2) imply that the indices of refraction of all three waves are nearly the same. For most transparent optical materials, the index of refraction increases with frequency in the visible part of the spectrum. Since the pump is at nominally twice the frequency of the downconverted waves, it will ordinarily have a very different index of refraction from the signal and idler. Thus, we need to use a "trick" to satisfy the phase-matching condition. The trick is to use a birefringent downconversion crystal.

In type-I downconversion the polarizations of the signal and idler are determined by the crystal orientation, and are parallel to each other. For maximum efficiency the pump polarization is perpendicular to that of the signal and idler. Since the pump is polarized orthogonal to the signal and idler, in a birefringent crystal its dispersion relationship is different than that of the downconverted beams. For certain crystals it is possible to satisfy the constraints imposed in eqs. (L1.1) and (L1.2) by proper orientation of the pump beam wave vector $\boldsymbol{k}_p$ and polarization with respect to the crystal axes.

The crystal we typically use is β-Barium Borate (BBO). In these experiments the pump laser has a wavelength of around 405 nm, while the signal and idler beams are at 810 nm (twice the wavelength, half the frequency). In order to separate the signal and idler, they are chosen to make a small angle (about 3°) with the pump beam; the signal comes out a few degrees from the pump, and the idler comes out a few degrees on the other side of the pump (fig. L1.1). Since only the relative angles between the pump, signal, and idler are important, the signal and idler beams are emitted into cones surrounding the pump beam (see, e.g., ref. [L1.1]).

However, for a given crystal orientation, there is not a unique solution to the constraints imposed in eqs. (L1.1) and (L1.2). The sums of the frequencies and wave vectors are constrained, but not the individual frequencies and wave vectors. If the idler frequency is somewhat more than half the pump frequency, it is possible for energy to be conserved [eq. (L1.1)] if the signal frequency is an equal amount less. In order for momentum to be conserved [eq. (L1.2)] the idler then makes a slightly greater angle with respect to the pump, and the signal makes a slightly less angle. Thus, the light coming out of a down conversion crystal is emitted into a range of angles (up to a few degrees), and wavelengths (on the order of 10s of nm, centered about twice the pump wavelength.)

While the emitted photons are allowed to come out in many directions, and with many frequencies, they always come in signal-idler pairs, with the pairs satisfying the constraints in eqs. (L1.1) and (L1.2). It is also the case that these photon pairs are emitted at the same time (to a very high precision), and to distinguish specific pairs of photons we use this fact. We find the pairs by using a technique called coincidence counting. If two photons are detected within a narrow time interval (about 8 ns wide in these experiments) we say that they are coincident, and assume that they constitute a signal-idler pair.

In Lab 1 we'll be exploring things like the momentum conservation rule and the precision of the timing of the photon pairs. The experimental apparatus is shown in fig. L1.2. The signal and idler photons are collected with lenses, coupled into optical fibers and directed to single-photon counting modules (SPCMs) where they are detected. Not shown in this figure are colored glass filters, which are in-line with the fibers, between the collection lenses and the SPCMs. These are RG780 filters that block wavelengths shorter than 780 nm, and transmit wavelengths longer than this. Their purpose is to transmit the downconverted light, while blocking scattered blue pump light and the green safe light used to illuminate the laboratory.

The SPCMs output an electrical pulse every time they detect a photon. These pulses then go the coincidence-counting unit (CCU). The CCU takes inputs from up to four detectors, and uses a programmable logic chip (a field programmable gate array, or FPGA) to implement the coincidence logic and eight counters. In the labs described here, four of the counters count the pulses coming directly from the individual SPCMs (called singles counts), while the other four register 2-, 3-, or 4-fold coincidence counts.

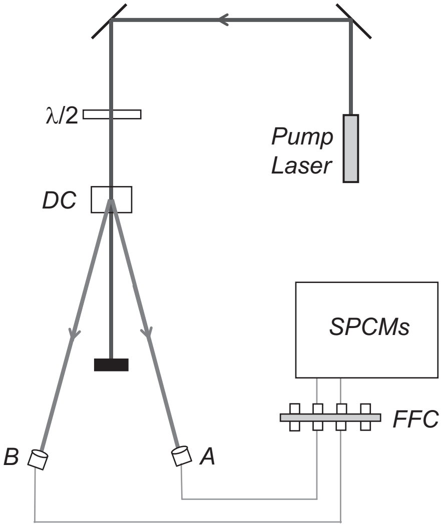


Fig L1.2 The experimental apparatus. Here λ/2 denotes a half-wave plate, DC denotes the downconversion crystal, FFC denotes fiber-to-fiber coupler, and SPCMs denotes the single photon counting modules. The signal and idler beams impinge on the collection optics A and B, are focused into fibers, and delivered to the SPCMs.

Your instructor will show you how to configure the CCU to obtain the proper coincidence signals. Data from the counters is streamed from the CCU to the host computer.

An important aspect of the experiment is properly coupling the signal and idler beams into the fibers, and maximizing the number of coincidence counts obtained between the signal and idler beams. This alignment will be your primary task in this lab.

We speak of aligning detector-**A** (for example), but really we mean aligning the lens and optical fiber that deliver the downconverted light to the detector. Figure L1.3 shows the mounts, lenses, and fibers of the signal and idler light collection optics. The bases which hold the mounts slide along a ruler, which is fastened to the table. This allows reasonably precise translation of the mounts, in order to position them at the correct angles to detect the downconverted photons.

L1.2 ALIGNING THE CRYSTAL

Note to instructors: I would suggest that the crystal alignment described in this section be performed prior to students coming into the laboratory. You can then remove the detectors so that the students can align them, beginning with the procedure presented in sec. L1.3.

The pump laser should be mounted so that its polarization is vertical (or horizontal). The beam must be aligned so that it is approximately collimated and traveling level to the table at a convenient height (4 inches or so). It is also useful if the beam is traveling directly above a row of holes in the optical table or breadboard. If your laser is not precollimated, you'll need to insert a collimation lens. The beam can be leveled and directed along a row of holes using the two mirrors shown in fig. L1.2. Once the beam

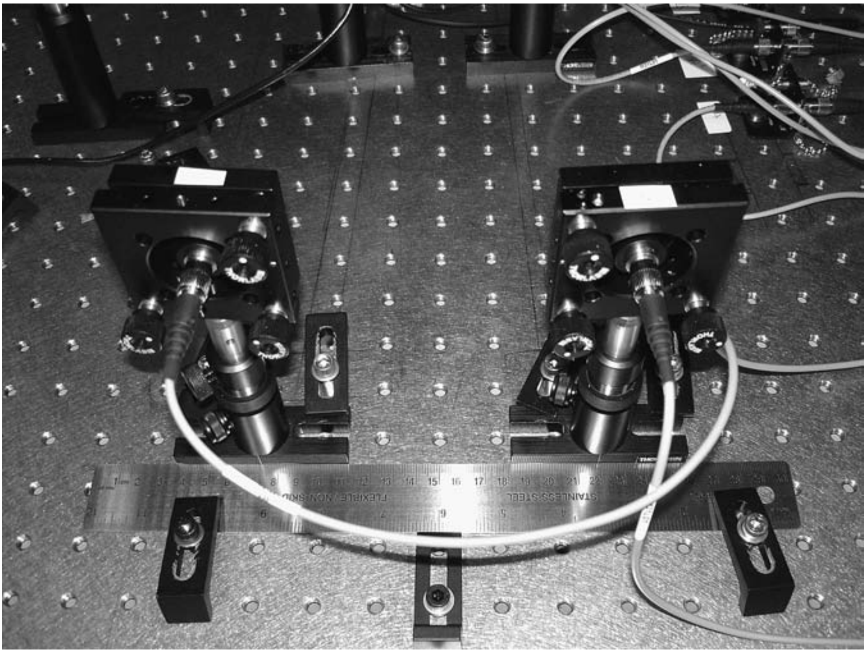


Fig L1.3 The fiber-coupling mounts.

has been leveled, it serves as a convenient reference that will allow you to place other optics, such as the fiber coupling-lenses, at this same height.

The crystal alignment proceeds as follows:

- **Double check that the detectors are turned off.**
- Insert the 405 nm half-wave plate and the downconversion crystal into the pump beam. Orient them so that they are perpendicular to the beam. This is most easily done by looking at the back reflection from the optic, and making sure that this reflection travels nearly along the direction of the incoming beam. Rotate the half-wave plate so that it is at 0°.
- Use the alignment laser to coarsely align detector-**A**, as described below in sec. L1.3. When this is done, make sure that the alignment laser is turned off, and then disconnect the **A**-fiber from the alignment laser, and connect it to the fiber leading to the **A**-detector SPCM.
- Run the LabVIEW program "Coincidence.vi" (some of the basics of this program are discussed in sec. L1.3, and more details are given in the documentation that came with the program).
- **Make sure the room lights are turned off** (it's OK for the green safe lights to be on) and turn on the detectors.
- Adjust the horizontal and vertical tilts on the mount holding the detector-**A** collection optics to maximize the count rate.

At this point you will hopefully see a large number of counts on detector-**A**, but you need to make sure that these counts are actually downconversion and not just background.

- Rotate the half-wave plate in the pump beam while observing the counts on detector-**A**. As the wave plate is rotated the counts should increase and decrease; this dependence on the pump polarization is confirmation that you are seeing downconversion. Rotate this wave plate to maximize the number of counts on detector-**A**.

If you do not see a dependence on the polarization of the pump, you are not seeing downconversion. There are two possibilities: (1) Detector-**A** is not properly aligned, so you need to go back and realign it. (2) Your crystal is possibly too far away from the correct tilt angle, so skip the next bulleted step, and then continue following this procedure. You may need to iterate back-and-forth between adjusting the pump polarization and the crystal tilt in order to see downconversion.

- Perform the fine alignment of detector-**A** described below in sec. L1.3, sliding the collection optics along the ruler to optimize the angle that the signal beam makes with the pump beam.

At this point the crystal should still be perpendicular to the pump beam; now you want to optimize its tilt angle. Your crystal can be tilted by using the adjustment screws on its mount. The downconversion efficiency will be extremely sensitive to tilt in one direction (horizontal or vertical), but very insensitive to tilt in the other. Which direction is sensitive depends on the orientation of your crystal, and the pump polarization: For a vertically polarized pump, the crystal should be sensitive to tilt in the vertical direction.

- Using the adjustment screws on the crystal mount, slowly tilt the crystal horizontally, while monitoring the counts on detector-**A**. If the count rate is sensitive to this tilt, adjust the tilt to maximize the count rate. If the count rate is insensitive to this tilt, adjust the crystal so that it is perpendicular to the pump.
- Repeat the last step while adjusting the vertical tilt.

Your downconversion crystal should now be reasonably well aligned. When you have completed the full alignment of both detectors, described below, you might want to once again carefully adjust the tilt of the crystal in order to maximize the coincidence count rate.

L1.3 ALIGNING DETECTOR A

The downconversion crystal is cut so that when properly aligned, the signal and idler beams make nominal 3° angles with respect to the pump beam. The first order of business is to place detector-**A** so that it makes a 3° angle with the pump beam, and get it facing the downconversion crystal.

- **Double check that the detectors are turned off.**
- **Insert a beam block in the blue pump beam.**

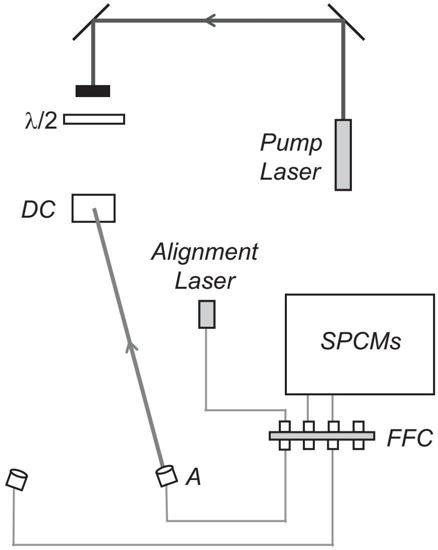


Fig L1.4 Using the alignment laser. The beam should travel backward from the collection optics, and onto the downconversion crystal.

- The fiber from the **A** collection optics leads to a fiber-fiber coupler, where it joins another fiber which leads to the filters and a SPCM. Unscrew the fiber coming from the **A** collection optics, and connect it instead to the fiber leading from the alignment laser, as shown in fig. L1.4. **Please do this carefully,** so that you don't scratch the fiber tip.
- Place the detector-**A** mount so that it makes an angle of approximately 3° from the pump laser, and that its base is pushed up against the ruler (as shown in fig. L1.3).
- Turn on the alignment laser (**your instructor will tell you the maximum current for your laser**), then adjust the vertical and horizontal tilt of the mount, using the knobs on the back, so that the beam strikes the center of the downconversion crystal. The laser light appears dim because it is at a wavelength that your eye is not very sensitive to.[1] If you stick white paper in the beam, you should be able to see it when the room lights are out. You should also be able to see it very easily using a CCD camera and a monitor.
- Gently screw the **A**-detector mount to the table, and double check that the laser is still shining on the center of the crystal.
- **Turn off the alignment laser.**

---

[1] The alignment laser puts out several milliwatts of light at about 800 nm. Your eye is not very sensitive to this light, but it IS a fairly intense beam, and you should **be careful not to look directly into it.** Also, **take care not to accidentally couple the laser light into the fiber leading to the SPCMs.** The filters will not block much light at this wavelength, and the SPCMs are VERY sensitive to it and could easily be damaged. It is wise to **make sure that the SPCMs are turned off while the alignment laser is turned on**.

- Unscrew the **A**-fiber from the alignment laser, and connect it to the fiber leading to the **A**-detector SPCM. Take care when you do this.

The **A**-detector is now coarsely aligned. The next step is to perform a fine alignment, to maximize the signal on this detector.

- Start by opening the LabVIEW program “Coincidence.vi”.
- Run the program by clicking the Run icon (the arrow in the upper left corner of the window).
- If you ever need to stop the program, do so by pushing the **STOP** button in the upper left. Do not simply close the window without stopping—if you do this the program does not exit gracefully. At the very least you’ll need to restart LabVIEW.

After a few seconds the program is running, reading the counters, and updating the screen in real time—although this may not be obvious at first because the detectors should still be turned off.

- Remove the beam block from the blue pump beam.
- **Make sure the room lights are turned off** (it’s OK for the green safe lights to be on) and turn on the detectors.

Now the indicators on the screen should be changing, and it should be more obvious that the program is running. The mode the program is currently running in is useful for “tweaking”: Adjusting the various parameters and seeing how they affect the measured count rates. Once things are adjusted as you want them to be, you press the **Take Data** button and the program switches to data record mode, in which the data is saved to a file on the disk.

- Make sure the **Experimental Setup** dial is set to **Coincidence** (if it isn’t, click on the dial and rotate it), and that **Update Period** is set to 0.1 or 0.2s (if it isn’t, highlight the value, type “0.2”, and hit <Enter>).

You should see some **A** counts, but possibly not a large number. For the moment don’t worry about the **B** counts or the **AB** coincidence counts.

- Set the full scale reading on the “thermometer” measuring the **A** counts to be about 3 times larger than the present count rate (highlight the value at the top, and replace it with a new value). The idea is that you’re going to be trying to increase the **A** count rate, so you want to have some room on the display to see the count rate increase.
- Slowly adjust the horizontal tilt of the detector-**A** mount while observing the count rate. If tilting in one direction decreases the counts, tilt the other way. Keep tilting back and forth until you have maximized the count rate. If you need to change the full-scale reading on the “thermometer” while performing this adjustment, then do so.
- Repeat the last step, but using the vertical tilt adjustment.

Now the detector is looking directly back at the downconversion crystal, but we need to optimize the angle that the detector makes with the pump beam.

- In your lab notebook, record the position of detector-**A** (using the ruler as a guide) and the count rate.
- While holding the mount in place with one hand, and keeping it pushed against the ruler, unfasten the screw that secures the mount to the table. Using one hand to hold the mount in place on the table, use your other hand to readjust the horizontal tilt so that the **A** count rate is maximized. Note that the vertical adjustment should not need to be changed, but you can adjust it if you do need to.
- Slide the detector mount 0.5 or 1 mm in either direction, and readjust the tilt to maximize the **A** detection rate. Note the position and the rate.
- Continue to move and tilt the detector (with ever-finer adjustments) until you find an alignment which maximizes the **A** count rate. Note that each time you slide the detector you should maximize the count rate by adjusting the tilt before you start to slide it again.
- Once the count rate is maximized, carefully screw the mount securely to the table, then give one final adjustment of both the vertical and horizontal tilts.
- Record in your notebook the position of the **A**-detector and the detection rate.
- **Turn off the detectors.**

Q1: What is the angle of the **A**-detector from the pump beam at this optimal position?

L1.4 ALIGNING DETECTOR B

The idea now is to place detector-**B** in the proper location to maximize the number of **AB** coincidence counts. Remember that although the downconverted light is emitted in many directions, individual pairs of photons have well-defined angles, as determined by eq. (L1.2). Since detector-**A** is now fixed, it is necessary to place detector-**B** in the correct spot to properly detect the photon pairs. Thus, we are interested in maximizing the **AB** coincidence counts, not the singles counts on the **B** detector.

- **Make sure the detectors are turned off.**
- Insert a beam block in the blue pump beam.
- Connect the fiber from the collection optics for detector-**B** to the alignment laser at the fiber-fiber coupler (similar to fig. L1.4, but with **B** connected instead of **A**).
- Place the detector-**B** mount so that the base pushes the up against the ruler, as shown in fig. L1.3. Begin by locating detector-**B** so that it makes roughly the same angle from the pump beam as detector-**A**.
- Turn on the alignment laser, and adjust the tilt of the mount so that the beam shines onto the downconversion crystal in order to coarsely align detector-**B**. Gently tighten the mount to the table.
- **Turn off the alignment laser.**
- Carefully unscrew the **B**-fiber from the alignment laser, and connect it to the fiber leading to SPCM **B**.

- Remove the beam block from the blue pump beam.
- **Make sure the room lights are turned off,** and then turn on the detectors.
- Slowly adjust the horizontal tilt of detector-**B** while observing the **AB** coincidence count rate (NOT the **B** singles rate). Keep adjusting until you have maximized the coincidence count rate. Once again, you may need to change the full-scale reading on the **AB** "thermometer" while performing this adjustment. If you do not see any coincidences, or just very few coincidences, you can start by simply maximizing the singles count rate on detector-**B**.
- Repeat the last step, but using the vertical tilt adjustment.
- In your lab notebook, record the position of detector-**B**, and the **B** and **AB** detection rates.

Now perform the fine alignment of detector-**B**.

- While holding the **B** mount in place with one hand, and keeping it pushed against the ruler, unfasten the screw that secures the mount to the table. Using one hand to hold the mount in place on the table, use your other hand to readjust the horizontal tilt so that the **AB** coincidence count rate is maximized.
- Now repeat the procedure you did before, but with detector-**B**. Slide the mount 0.5 or 1 mm in either direction, and readjust the tilt to maximize the **AB** coincidence rate.
- Continue to move and tilt the detector until you find an alignment which maximizes the coincidence count rate. Ask your instructor what count rate you should be shooting for.
- Once the count rate is maximized, carefully screw the mount securely to the table, then give one final adjustment of both the vertical and horizontal tilts.
- Record the position of detector-**B**, and the optimal **B** and **AB** count rates.

Q2: What is the angle of the **B**-detector from the pump beam at this optimal position?

L1.5 ANGULAR CORRELATIONS – MOMENTUM CONSERVATION

Having done the above, you should have the optimal alignment, and some idea of the range over which you can obtain coincidences. Now make a careful study of how the coincidence rate depends on angle. You will fix the position of one detector, and scan the other over a 1° range of angles, centered about the optimal angle. You should take at least 10 data points across this range, calculating the average and standard deviation of the coincidence rate at each point. You will use the **Take Data** feature of the program to do this.

- Place detector-**B** so that it is at the smallest angular position you will acquire data at, and adjust the tilt to optimize the coincidence rate.

- In the **Data Taking Parameters** section of the program set **Update Period (Data Run)** to 1.0 s, and **Number of Points** to 10. Now press the **Take Data** button.

A new window will automatically open as the computer switches to data taking mode. The computer will automatically take 10 measurements with 1.0 s counting windows, plot the data on the screen, calculate the mean and standard deviation of the coincidence rate, and save all the data to a file. The data file is automatically named according to the date and time.

- In your notebook record the filename, important parameters (e.g., the location of the detector), and the results.
- Once you have written down all of these parameters, you can close the window of the data recording program.
- Check the data file you just created by opening it up and looking at it (it's easiest to use a spreadsheet to examine the data, so open the file from within the spreadsheet program). The info in the file should agree with what you wrote in your notebook.
- Repeat the above process for at least 10 data points over a 1° range of angles. At the two extremes of your data, the coincidence rates should be less than 10% of the maximum rate; if this is not the case, you should take more data at larger and/or smaller angles. Also, if the peak in your data is not well resolved take more data with finer angular adjustments.
- When you're all done taking data, place your detector at the optimum position, and screw it to the table.

  After you leave the lab, create a graph of coincidence count rate versus angle, including error bars.

Q3: Approximately what is the full width at half maximum (FWHM) of the angular spread of this distribution?

L1.6 POLARIZATION

Now you're going to examine the polarization properties of your downconversion source.

- Place a linear polarizer in the blue pump beam (after the crystal) and rotate it to maximize the amount of blue light transmitted through the polarizer—just look at the blue light on a screen and maximize its brightness. You have now oriented the polarizer so that it is parallel to the polarization of the blue pump beam.
- Move the polarizer so that it is in front of the **A**-collection optics.

Q4: What happens to the count rate on the **A**-detector when you insert this polarizer? What does this say about the relative orientation of the polarization of the blue beam and the polarization of the beam hitting the **A**-detector?

- Repeat this measurement by placing the polarizer in front of the **B**-collection optics.

Q5: What happens to the count rate on the **B**-detector when you insert the polarizer? What does this say about the relative orientation of the polarization of the blue beam and the polarization of the beam hitting the **B**-detector?

- Remove the polarizer. While looking at the count rates, rotate the pump-beam half-wave plate.

Q6: How many degrees do you need to rotate the wave plate in order to minimize the count rates? By how far has the polarization of the blue beam been rotated? Does the polarization of the downconversion change as the pump-beam polarization changes? (Check this by using your polarizer.)

- Remove the polarizer and rotate the wave plate back to maximize the count rates.

L1.7 TIMING

As discussed above, the photon pairs are produced not only with well-defined angles, but also at well-defined times. Indeed, the two photons are produced at the same time. We can't say that they are produced at *exactly* the same time, because there is an indeterminacy relationship between energy (frequency) and time:

$$\begin{aligned} \Delta E \Delta t &\geq \frac{\hbar}{2}, \\ \Delta(\hbar\omega)\Delta t &\geq \frac{\hbar}{2}, \\ \Delta(2\pi f)\Delta t &\geq \frac{1}{2}. \end{aligned} \tag{L1.4}$$

The uncertainty in the time is given approximately by the inverse of the bandwidth of the downconverted light:

$$\Delta t \geq \frac{1}{4\pi\Delta f}, \tag{L1.5}$$

[see also eq. (2.A.13)]. For our experimental parameters the uncertainty in time is on the order of 10 fs—a time that is much shorter than our experimental apparatus can resolve. Here we will be able to show that this uncertainty is less than about a nanosecond.

In order to perform this measurement you'll use an instrument known as a time-to-amplitude converter (TAC), which measures the time interval between two events. In our case the two events are the detections of two photons. One detector is connected to the START input of the TAC, while the other is connected to the STOP input. The TAC outputs a voltage pulse whose height is proportional to the time interval between these events. We ensure that the STOP always comes after the START by inserting an extra length of electrical cable between the STOP detector and the TAC. We will thus measure a time interval that is offset from zero by this electrical delay; we are not interested in this offset, but rather in the width of the distribution of the arrival times.

The output pulses from the TAC are most easily analyzed with an instrument known as a multichannel analyzer (MCA). An MCA reads in the voltage pulses and displays a histogram of the pulse heights. Because we can calibrate the TAC, this is essentially a measurement of the distribution of time intervals.

- Your instructor will show you how to connect the detectors to the TAC.
- **Close the Coincidence.vi**—it cannot be in memory at the same time as the MCA.vi.
- Run the "MCA.vi" LabVIEW program.
- This program updates the screen after a certain number of pulses are measured, not after a certain time interval. So, if nothing happens after a few seconds it probably means that there are no voltage pulses present. Make sure that the lights are off, and the detectors are on.
- You should see a sharp peak appear in the distribution on your screen. Place the cursors on either side of this peak, then click **Zoom to Cursors** to zoom in on it.
- Hit the **Clear Buffer** button, and then accumulate enough data so that you can clearly see the peak. Push the **Save Data** button to save the data to a file.
  After you leave the lab, create a graph displaying this data. Don't necessarily show all the data, but show about a 10 ns window around the coincidence peak.

Q7: Approximately what is the full width at half maximum (FWHM) of this distribution?

Q8: How wide a time window would you need to count essentially all of the coincidences, but as little as possible of the background?

Q9: The coincidence circuit in your CCU has a coincidence window of about 8 ns (assume that it's centered about the peak). Given this information, estimate what fraction of the coincidences that are measured are true coincidences, and what fraction are "accidental" coincidences, due to the background. Explain how you came up with this estimate.

L1.8 References

[L1.1] D. Dehlinger and M. W. Mitchell, "Entangled photons, nonlocality, and Bell inequalities in the undergraduate laboratory," Am. J. Phys. **70**, 903 (2002).

LAB 2

"Proof" of the Existence of Photons

LAB TICKET

Explain the differences between the two-detector and three-detector measurements of $g^{(2)}(0)$. For which do we hope to measure $g^{(2)}(0)<1$, and why? What would such a measurement signify? What do we expect to obtain for the other measurement, and why?

L2.1 INTRODUCTION

First, we should define what we mean by "proving" that photons exist. Classically, light is an electromagnetic wave. Quantum mechanically, light has both wave-like and particle-like properties, and the particle-like quanta are called photons. So, to prove that light is made of photons you want to do an experiment which makes this granular, particle-like nature of the field apparent.

However, this is not enough. We never measure light directly, but always measure the current from a detector instead. It is possible for the granularity of our measurements to be caused by the discrete nature of the electrons in our detector. Assume a classical wave is incident on a detector. How do we determine whether the granularity of our measurements is due to the discreteness of electrons in the detector, or the discreteness of photons in the field? There is no easy way to do this. Observing granularity in measurements of a field is a necessary, but not a sufficient, condition for the existence of photons. Granularity motivates the existence of photons, but does not prove it.

Since photons are inherently quantum mechanical objects (classical waves aren't made of photons), an experiment which requires a quantum mechanical explanation would imply that the field contains photons. In other words, if we do an experiment that we cannot explain classically, it means there is more to the field than just classical waves. To be precise, the experiment we will do distinguishes between quantum mechanical and classical theories of the electromagnetic field. If classical waves cannot explain the results, then we take this to mean that photons exist.

You will be performing an updated version of the experiment performed originally by Grangier, Roger, and Aspect [L2.1, L2.2]. The experimental apparatus is shown in figs. L2.1 and L2.2. The basic idea is to show that if a single photon is incident on a beam splitter, it can be transmitted and detected at **B**, or reflected and detected at **B′**, but not both. We ensure that a single photon is present in the signal beam incident on the beam splitter by using the spontaneous parametric downconversion source we studied in lab 1. By conditioning the measurements on the detection of an idler photon at **A**, the signal beam is projected into a single-photon state.

The measurements are quantified using a parameter called the degree of second-order coherence, $g^{(2)}(0)$. Below we'll show that a classical wave theory of light requires $g^{(2)}(0) \geq 1$. This is an instance of a classical inequality, of which there are several in physics.[1] There are situations in which quantum mechanics can violate classical inequalities, and in these cases we say that we are observing strictly quantum mechanical phenomena. For this experiment, if we measure a value of $g^{(2)}(0) < 1$, then the field cannot be explained classically. As stated above, we will take this as proof of the existence of photons. The quantum state which yields maximum violation of this classical inequality is the single-photon state, for which the quantum prediction is $g^{(2)}(0) = 0$.

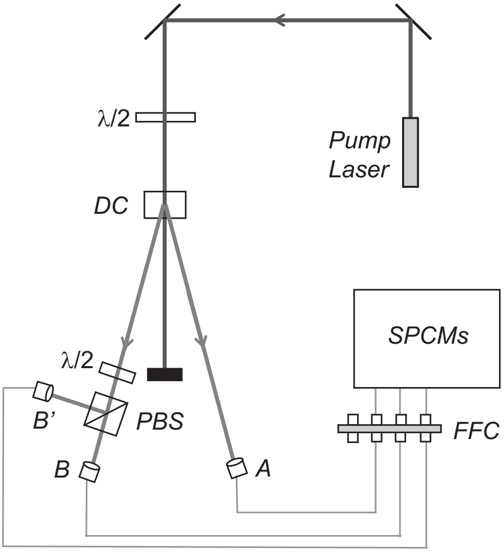


Fig L2.1 The experimental apparatus. Here λ/2 denotes a half-wave plate, DC denotes the downconversion crystal, PBS denotes a polarizing beam splitter, FFC denotes fiber-to-fiber coupler, and SPCMs denotes the single photon counting modules.

---

[1] Other classical inequalities are described in chap. 8 and lab 5.

The above described experiment uses three detectors to measure $g^{(2)}(0)$. We can think of this as measuring $g^{(2)}(0)$ for the signal beam striking the beam splitter, *conditioned* on a detection in the idler beam. The conditioning is key—it's what prepares the signal beam in a single-photon state. Without this conditioning the beam striking the beam splitter is classical. To prove this you will also perform a two-detector, unconditional measurement on the signal beam, for which you should observe $g^{(2)}(0) \geq 1$.

Measurements of $g^{(2)}(0)$ were very important historically in the development of quantum optics. For some discussion of this history see ref. [L2.2], and the references therein.

L2.2 THEORY

Here we'll discuss the classical and quantum predictions for $g^{(2)}(0)$.

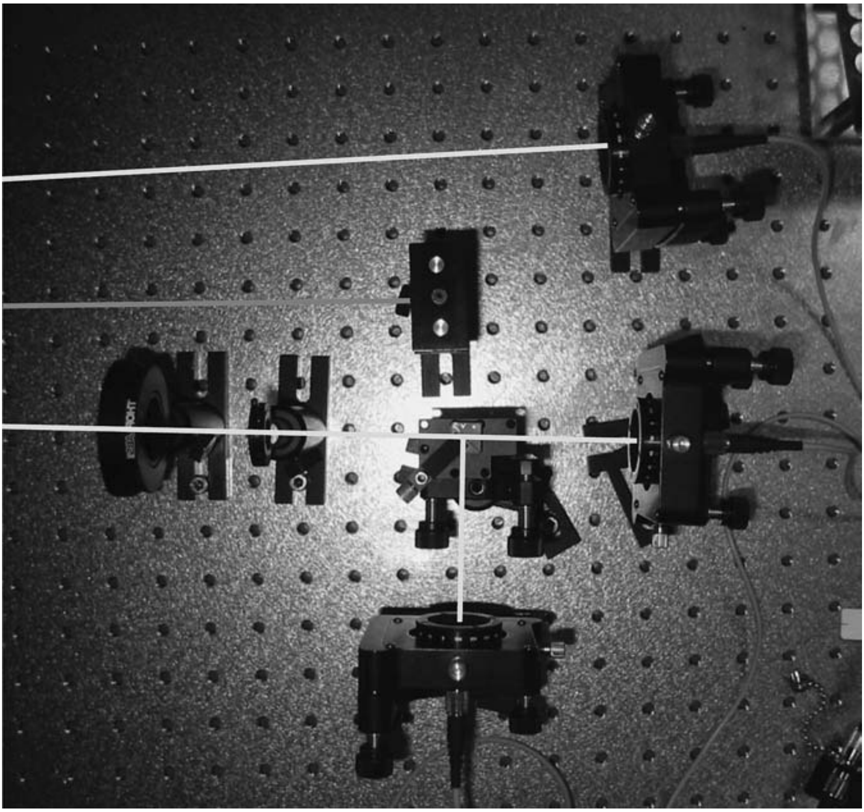


Fig L2.2 An overhead view of the wave plate, iris, PBS, and the collection optics. The beam paths are shown for clarity.

L2.2.1 Classical Fields

By a classical field we mean an electromagnetic wave that is completely described by Maxwell's equations. Referring to fig. L2.1, the intensity of the field incident on the polarizing beam splitter (PBS) is $I_I(t)$. The transmitted and reflected fields from the PBS go to detectors **B** and **B′**, and their intensities are $I_B(t)$ and $I_{B'}(t)$. The correlations between $I_B(t)$ and $I_{B'}(t)$ are given by the degree of second-order (temporal) coherence, $g^{(2)}_{B,B'}(\tau)$, which is a function of the time delay τ between the intensity measurements:

$$g^{(2)}_{B,B'}(\tau) = \frac{\langle I_B(t+\tau) I_{B'}(t)\rangle}{\langle I_B(t+\tau)\rangle\langle I_{B'}(t)\rangle}. \tag{L2.1}$$

If the light source is stationary (i.e., if its statistics do not change in time), then we can interpret the brackets as referring to ensemble averages rather than time averages. This quantity is called the degree of second-order coherence because it involves correlations between intensities, whereas the degree of first-order coherence describes correlations between fields.

Of particular importance to us here is the case of simultaneous $(\tau = 0)$ intensity correlations. If the beam splitter has an intensity transmission coefficient of $\mathcal{T}$, and reflection coefficient of $\mathcal{R}$, then the transmitted, reflected and incident intensities are related by

$$I_B(t) = \mathcal{T}\, I_I(t) \text{ and } I_{B'}(t) = \mathcal{R}\, I_I(t). \tag{L2.2}$$

Inserting these expressions into eq. (L2.1), we find that

$$g^{(2)}_{B,B'}(0) = g^{(2)}(0) = \frac{\left\langle [I_I(t)]^2 \right\rangle}{\langle I_I(t)\rangle^2}, \tag{L2.3}$$

which we note is independent of the splitting ratio of the beam splitter. From the Cauchy-Schwartz inequality, it is straightforward to prove that $\left\langle [I_I(t)]^2 \right\rangle \geq \langle I_I(t)\rangle^2$.[2] Using this, we find that

$$g^{(2)}(0) \geq 1 \text{ (classical fields)}, \tag{L2.4}$$

where we emphasize that this result has been derived using classical wave theory. In eq. (L2.4), equality with 1 is achieved if the input intensity is perfectly stable with no fluctuations, while for fluctuating intensities the second-order coherence is greater than 1.

The closest experimental realization of a stable, classical wave is laser light. Light from a stable laser yields $g^{(2)}(0) = 1$, no matter how highly attenuated it is. For "chaotic"

---

[2] A way to motivate the validity of this inequality is to note that the variance of the intensity must be a positive number. Since the variance can be written as $\Delta I^2 = \langle I^2\rangle - \langle I\rangle^2 \geq 0$, it must be true that $\langle I^2\rangle \geq \langle I\rangle^2$.

light (e.g., light from a thermal source, such as a vapor lamp) it can be shown that $g^{(2)}(0) = 2$ [L2.3].

L2.2.2 Semiclassical Theory of Photodetection

So far we have been speaking about correlations between the intensities of the fields leaving the beam splitter. In an experiment, however, one does not measure the intensity directly, but rather the photocurrent from a detector. It is then necessary to model the photodetection process. Since to this point we have been discussing classical fields, it is most appropriate to use a model that treats the field classically. The most rigorous theory of this sort is the semiclassical theory of photoelectric detection, in which the field is treated classically and the photodetector is treated quantum mechanically. We describe this theory in chapter 15, but the details are not important for our discussion here. Here we'll describe enough of the theory to explain the measurements performed in this lab

In the semiclassical theory of photoelectric detection, it is found that the conversion of continuous electromagnetic waves into discrete photoelectrons is a random process. The probability of obtaining a single photocount from a single detector (for example, detector-**B**), within a short time window Δt, is proportional to the average intensity of the field striking that detector:

$$P_B = \eta_B \left\langle I_B(t) \right\rangle \Delta t \,, \tag{L2.5}$$

where η_B is a constant that characterizes the detection efficiency of detector-**B**.[3] Equation (L2.5) tells us that the probability of detection increases with increasing intensity. Furthermore, this equation says that the detection probability is linearly proportional to the width of the time window, so the detection rate is constant if the intensity is constant. These results are reasonable.

The joint probability of getting a photocount (within a short time widow Δt) at detector-**B'**, and then after a time τ obtaining a photocount at detector-**B** (also within a time widow Δt) is given by

$$P_{BB'}(\tau) = \eta_B \eta_{B'} \left\langle I_B(t+\tau) I_{B'}(t) \right\rangle \Delta t^2 \,. \tag{L2.6}$$

Combining eqs. (L2.1), (L2.5), and (L2.6), we find that

$$g^{(2)}_{B,B'}(\tau) = \frac{P_{BB'}(\tau)}{P_B P_{B'}} \,. \tag{L2.7}$$

This tells us that if we measure the probability of joint and individual photocounts at detectors **B** and **B′**, we can determine the degree of second-order coherence.

---

[3] You can compare eq. (L2.5) to eq. (15.56).

Again, we are most interested in $g^{(2)}(0)$. Using eqs. (L2.4) and (L2.7), we find that for classical fields, the measured degree of second-order coherence must be greater than or equal to 1:

$$g^{(2)}_{B,B'}(0) = \frac{P_{BB'}(0)}{P_B P_{B'}} = g^{(2)}(0) \geq 1 \text{ (classical fields).} \tag{L2.8}$$

Here $P_{BB'}(0)$ is the probability of coincidence detections (within Δt) at **B** and **B′**. Furthermore, we see that if the joint probability factorizes, $P_{BB'}(0) = P_B P_{B'}$, which occurs when the detections at **B** and **B′** are completely uncorrelated, then $g^{(2)}(0)$ is minimized and is equal to 1.

Experimentally, we must express the probabilities in terms of measured count rates. For example, the probability of a detection at **B** in a short time interval Δt is simply given by the average rate of detections, multiplied by Δt. The average rate of detections at **B** is just the number of detections N_B divided by the counting time T that we are averaging over. The probabilities for **B** detections and **BB′** coincidences are given similarly:

$$P_B = \left(\frac{N_B}{T}\right)\Delta t,\ P_{B'} = \left(\frac{N_{B'}}{T}\right)\Delta t,\ P_{BB'} = \left(\frac{N_{BB'}}{T}\right)\Delta t\,. \tag{L2.9}$$

Here $N_{BB'}$ is the number of coincidence counts between the two detectors. Substituting eq. (L2.9) into eq. (L2.8), we find that

$$g^{(2)}(0) = \frac{N_{BB'}}{N_B N_{B'}}\left(\frac{T}{\Delta t}\right) \text{(2-detector).} \tag{L2.10}$$

Here we note that this expression corresponds to 2-detector measurements of $g^{(2)}(0)$, which must be distinguished from the 3-detector measurements we will discuss below.

Just to be perfectly clear about how the variables in eq. (L2.10) are related to experimentally measurable quantities: Δt is the coincidence time window determined by the resolution of the coincidence unit (typically about 10 ns), T is the counting interval (typically greater than 0.1 s), N_B and $N_{B'}$ are the numbers of singles counts in time T, and $N_{BB'}$ is the number of coincidence counts in that same time.

We can summarize what we have learned about classical field statistics as follows. It is possible to measure the degree of second-order coherence between the fields leaving a beam splitter, $g^{(2)}(0)$, by measuring the probability of joint and individual photocounts at detectors **B** and **B′**. The second-order coherence must satisfy the inequality $g^{(2)}(0) \geq 1$. When the photocounts at **B** and **B′** are completely uncorrelated $g^{(2)}(0) = 1$, which occurs when the input field to the beam splitter is a perfectly stable wave, such as the light from a laser. If the input field fluctuates, then $g^{(2)}(0) > 1$ indicating positive correlations between the photocounts; such fields are said to be "bunched."

Since $g^{(2)}(0)$ cannot be less than 1, we are left with the conclusion that for classical fields the measured photocounts at **B** and **B′** cannot be anticorrelated. This makes sense, because a beam splitter simply splits a classical input field into two identical copies. These output fields either fluctuate together (positive correlation) or don't fluctuate at all (no correlation). It is not possible for the intensity of one to decrease while the intensity of the other increases (anti-correlation, or "anti-bunching").

L2.2.3 Three-Detector Measurements

We are also interested in performing measurements of $g^{(2)}(0)$, conditioned on the measurement of a photocount on a third detector. In such a measurement all probabilities are conditioned upon a detection at **A**. In this case, in place of eq. (L2.7) we have (at $\tau = 0$)

$$g^{(2)}(0) = \frac{P_{ABB'}(0)}{P_{AB}(0)\,P_{AB'}(0)} \text{ (3-detector),} \tag{L2.11}$$

where $P_{ABB'}(0)$ is the probability of threefold coincidence detection, and $P_{AB}(0)$ and $P_{AB'}(0)$ are the probabilities of coincidence detection between detector-**A** and detectors **B** and **B′**, respectively.

Since we are not interested in any events unless detector **A**-fires, the number of detections at **A**, N_A, serves as the total number of trials, which we can use to normalize our probabilities:

$$P_{AB}(0) = \frac{N_{AB}}{N_A},\ P_{AB'}(0) = \frac{N_{AB'}}{N_A},\ P_{ABB'}(0) = \frac{N_{ABB'}}{N_A}. \tag{L2.12}$$

Using these probabilities, $g^{(2)}(0)$ for the three-detector measurements becomes

$$g^{(2)}(0) = \frac{N_A N_{ABB'}}{N_{AB} N_{AB'}} \text{ (3-detector).} \tag{L2.13}$$

For classical fields, it is still the case that $g^{(2)}(0)$ as expressed in eq. (L2.13) must satisfy $g^{(2)}(0) \geq 1$ [L2.1].

L2.2.4 Quantum Fields

In order to describe the degree of second-order coherence for an arbitrary quantum field, it is necessary to have a full quantum theory of fields. Such a theory is described in chapter 16, and applied to $g^{(2)}(0)$ in complement 16.A. However, for the special case of a single-photon field we can easily understand what the quantum result must be. To quote from Grangier et al. in ref. [L2.1], "a single photon can only be detected once!" Hence, if a single quantum of light is incident on the beam splitter, it should be detected at the transmission output, or at the reflection output, but not both: there should be no coincidence detections between the two outputs. In other words, for a single-photon incident on the PBS in fig. L2.1, we must have $N_{BB'} = 0$ in eq. (L2.10), and $N_{ABB'} = 0$ in eq. (L2.13). In both cases quantum mechanics predicts $g^{(2)}(0) = 0$ for this single-photon field.

There is a subtlety that we need to consider. In order to measure $g^{(2)}(0) = 0$, it is necessary to have a single photon incident on the PBS in fig. L2.1. By using our spontaneous parametric downconversion source, this is achieved conditionally upon the detection of an idler photon at detector-**A**. Thus, we must perform a three-detector measurement in order to see $g^{(2)}(0) = 0$ with this source. Using the same source without conditioning, the field incident on the PBS is *not* a single-photon field; a measurement of $g^{(2)}(0)$ on this field should yield $g^{(2)}(0) \geq 1$.[4]

Finally, in a real experiment we cannot expect to measure $g^{(2)}(0) = 0$, because there will always be some accidental three-fold coincidence counts $N_{ABB'}$. By knowing quantities such as the effective coincidence window Δt, the counting time T, and the average count rates on the detectors, we can predict the number of accidental counts, and hence determine an expected value for $g^{(2)}(0)$ given our experimental parameters. For a further discussion of how this expected value is calculated, see ref. [L2.2].

L2.3 ALIGNING THE IRISES AND THE BEAM SPLITTER

In lab 1 you examined the behavior of a spontaneous parametric downconversion source. You learned how to maximize the coincidence count rate between detectors **A** and **B**, and here we will assume that this part of the alignment has already been completed. Your task in this lab will be to insert a beam splitter in the signal beam, and to align detector **B′**. Once this is done you'll be able to measure $g^{(2)}(0)$.

- **Make sure the detectors are off.**
- Insert a beam block in the pump beam.
- The fiber from the **B** collection optics leads to a fiber-fiber connector, where it joins another fiber which leads to the filters and a SPCM. Unscrew the fiber coming from the **B** collection optics, and connect it instead to the fiber leading from the alignment laser, as shown in fig. L2.3. **Take care when you do this.**
- Turn on the alignment laser. Light from this laser will shine backward through the **B**-fiber, and emerge as a collimated beam from the fiber-coupling lens attached to the end of the fiber, see fig. L2.3.[5] The laser light appears dim because it is at a wavelength your eye is not very sensitive to. If you stick white paper in the beam you should be able to see it with the room lights out. You should be able to see it very brightly using a CCD camera and monitor. If the laser light is not shining onto the downconversion crystal, detector-**B** is not properly aligned, and you'll need to complete this alignment (see lab 1) before continuing.

---

[4] You may be surprised that a classical field having a very low photon flux, such as a highly attenuated laser, cannot give $g^{(2)}(0) = 0$. However, as described in chapters 12 and 16, the light from a laser is described quantum mechanically by a "coherent state." No matter how highly attenuated it is, even to the level of just one photon per second or less, there is always a nonzero probability for laser light to contain two or more photons. This means that $N_{BB'} \neq 0$, allowing $g^{(2)}(0) = 1$ [eq. (L2.10)].

[5] Recall that the alignment laser puts out several milliwatts of light at about 800nm. This light can be dangerous to both your eyes and the SPCMs, so use care.

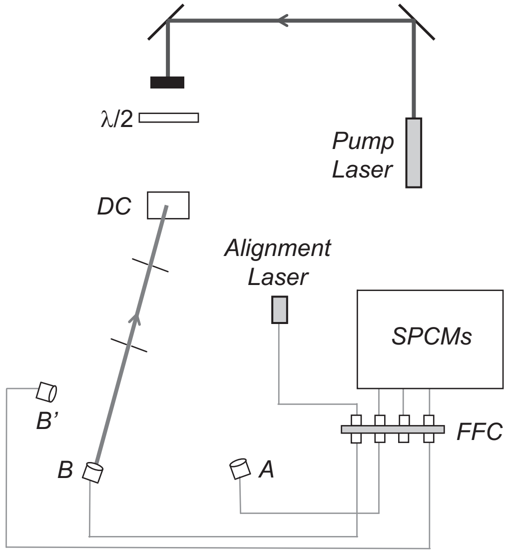


Fig L2.3 Aligning the irises.

- First you want to center two irises on the alignment beam. One should be about a foot (maybe a bit more) away from the detector, and the other about a foot away from the downconversion crystal. Either by eye, or using the CCD camera, adjust the height and position of the irises so that the alignment beam passes through their centers. You should be able to close the irises down to about 2–3 mm and still get nearly all the beam through. Once the irises are centered on the beams, lock them to the table.
- Make sure that the iris near the downconversion crystal does not block the beam going to the **A** detector. You can check this by letting the blue beam through and seeing where it hits the iris. The **A** beam should make roughly the same angle with the pump as the **B** beam, so you should have a pretty good idea of where it's at with respect to the iris.
- Insert the polarizing beam splitter (PBS) about 3–4 inches away from the **B** collection optics (fig. L2.2). Make sure it is oriented so that light coming from the downconversion crystal will be reflected away from the beam going to detector-**A**. Make sure that the beam passes through the center of the beam splitter.
- Orient the beam splitter so that its face is perpendicular to the beam. Do this by looking at the back-reflection from the beam splitter, which shines back toward the collection optics (this beam will be VERY dim, so you will probably need the CCD camera to see it). Orient the beam splitter so that this back-reflection goes straight back on top of the incident beam. Double check that the beam is still centered on the beam splitter, then screw the beam splitter mount to the table.
- Insert the half-wave plate between the PBS and the downconversion crystal, about an inch or two from the PBS. Again, center it, orient it perpendicular to the beam, and screw it to the table.

L2.4 ALIGNING THE B′ DETECTOR

You will be aligning the **B′** detector so that it collects light from the same beam as the **B** detector.

- Place the mount with the **B′** collection optics on the reflection side of the beam splitter, about an equal distance from the beam splitter as the **B** collection optics (fig. L2.2). Don't screw it down yet.
- Unscrew the **B**-fiber from the alignment laser, and reconnect it to the fiber leading to SPCM **B**. Connect the **B′**-fiber to the alignment laser, as shown in fig. L2.4. At this point the beam won't go back through the irises.
- The task now is to get the alignment laser to shine back through the irises and onto the downconversion crystal. By moving the mount with the **B′** collection optics sideways (perpendicular to the beam), rotating it, and adjusting its vertical tilt, position it as well as you can to shine the light back through the two pinholes. The CCD camera will be helpful for this. It won't be perfect, but the better job you do on this coarse alignment, the easier the fine alignment will be.
- **Alignment Hint:** Slide the mount back and forth sideways to center the beam on the first iris (closest to the beam splitter). Adjust the tilt of the mount to center the beam on the second iris. Iterate back and forth between these two adjustments.
- Once you've got it reasonably well aligned, screw it to the table.
- Adjust the vertical and horizontal tilt of the **B′** collection optics to perfectly center the beam on the iris closest to the beam splitter.

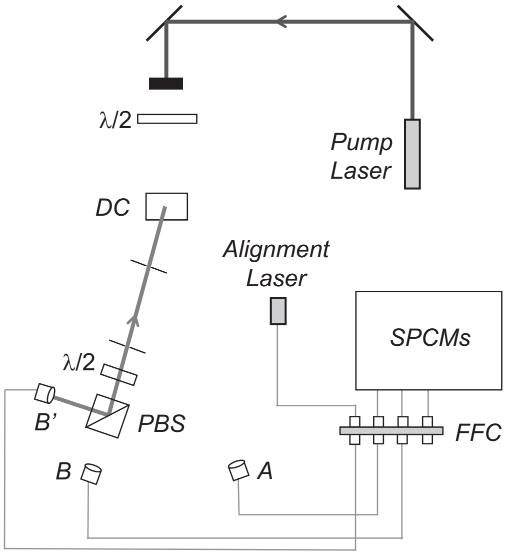


Fig L2.4 Aligning detector-**B′**.

- Adjust the vertical and horizontal tilt of the beam splitter to center the beam on the iris closest to the downconversion crystal.
- Alternate back and forth between the last two steps, always adjusting the collection optics to center the beam on the first iris, and the beam splitter to center the beam on the second. When the beam is well centered on both irises, you're done.
- **Turn off the alignment laser**.
- Unscrew the **B′**-fiber from the alignment laser, and reconnect it to the fiber leading to SPCM **B′**.
- Run the LabVIEW program "Coincidence.vi".
- Set **Experimental Setup** to **g(2) 3-det,** and **Update Period** to 0.2s. In the pane that displays the graphs, choose the **AB & AB′** tab.
- The **ABB′ Coincidence Window (ns)** parameter tells the computer the effective time window for the three-fold coincidence determination. In three-detector measurements this parameter is needed to calculate the expected value for $g^{(2)}(0)$. Ask your instructor for the precise value; it should be on the order of 5–10 ns.
- **Make sure that the lights are out**, and then turn on the detectors. Open the irises wide, and unblock the blue pump beam.
- Slowly rotate the wave plate in front of the beam splitter while monitoring the count rates. You should notice that for some wave plate angles you get lots of **B** and **AB** counts, but almost no **B′** and **AB′** counts. For other angles you get lots of **B′** and **AB′** counts, but almost no **B** and **AB** counts.
- Rotate the wave plate to maximize the **AB** counts. Adjust the tilt on the **B** mount to maximize this coincidence rate.
- Rotate the wave plate to maximize the **AB′** counts. Adjust the tilt on the **B′** mount to maximize this coincidence rate.
- The maximum **AB** and **AB′** count rates should be approximately the same (hopefully within 10–20%). If this isn't the case, there are a few possible problems. The first is that either detector-**B** or **B′** is not well aligned. You can try tweaking the alignment to improve things, but you might need to go back and realign. Other possibilities are that the fiber coupling lens is better aligned for one of the detectors than the other, or the end of one of the fibers has dirt on it. If you get drastically different maximum count rates for **AB** and **AB′**, ask your instructor what you should do.

Q1: When the wave plate is set to 0°, are the **B** or **B′** counts maximized? How far do you have to rotate the wave plate in order to maximize the other count rate? Explain why the count rates change the way they do when the wave plate is rotated.

Q2: The polarizing beam splitter reflects vertically polarized light, and transmits horizontally polarized light. What polarization is the light emerging from the downconversion crystal?

L2.5 MEASURING $g^{(2)}(0)$ FOR A SINGLE-PHOTON STATE

- Rotate the wave plate to roughly equalize the **AB** and **AB′** count rates.
- Increase **Update Period** to 1 s.

You should notice that there are very few **ABB′** three-fold coincidences.

- In the pane with the plots, click on the **g(2)(0)** tab. In the pane labeled **g(2) Measurements,** click **Clear Buffer** and get an idea of the average value of $g^{(2)}(0)$.

The $g^{(2)}(0)$ measurements will fluctuate significantly, but hopefully they should always be less than one. Now you should be ready to take data.

- In the **Data Taking Parameters** pane set **Update Period (Data Run)** to 10 s, and **Number of points** to 10. Click the **Take Data** button.

Hopefully you'll find a value of $g^{(2)}(0)$ that is several standard deviations below 1; if you don't, ask your instructor for help. Calculate how many standard deviations your value for $g^{(2)}(0)$ is below 1.

- Take several more data sets with different parameters for **Update Period (Data Run)** and **Number of points**. Always take at least 10 points in order to assure reasonable statistics.

Note that while the ideal theoretical prediction for $g^{(2)}(0)$ of a single photon is 0, experimentally you can't achieve this. If you measure $g^{(2)}(0) = 0$, then you need to increase the counting time, or align things better to get higher count rates to get a nonzero $g^{(2)}(0)$. Try and adjust the experimental parameters so that you get a data set for which the standard deviation of $g^{(2)}(0)$ is several times smaller than the average value of $g^{(2)}(0)$. In other words, try to get a measurement of $g^{(2)}(0)$ which is different from 0 by more than one standard deviation.

Q3: Explain why you shouldn't measure a value of 0 for $g^{(2)}(0)$.

Q4: Does your measured value for $g^{(2)}(0)$ agree with the "expected" value (to roughly within the error of the measurement) returned by the computer? If not, what might be wrong?

L2.6 TWO-DETECTOR MEASUREMENT OF $g^{(2)}(0)$

When you've completely finished making measurements on the single-photon state, you are ready to show that a single beam of your downconversion source behaves classically.

- Set **Experimental Setup** to **g(2) 2-det,** and **Update Period** to 1.0 s [this plays the role of T in eq. (L2.10)]. The computer uses the **BB′ Coincidence Window (ns)** parameter as Δt in this same equation; ask your instructor what value to use for this parameter (it should be on the order of 10 ns, but won't necessarily be the same as you used for 3-detector measurements). In the pane that displays the graphs, choose the **g(2)(0)** tab.
- Adjust the half-wave plate so that the **B** and **B′** counts are approximately equal.

The $g^{(2)}(0)$ measurements will fluctuate significantly, but hopefully they should always be around one, or greater.

- Click **Clear Buffer** and get an idea of the average value of $g^{(2)}(0)$.

Now you should be ready to take data.

- In the **Data Taking Parameters** pane set **Update Period (Data Run)** to 10 s, and **Number of points** to 10. Click the **Take Data** button.

Hopefully you'll find a value of $g^{(2)}(0)$ that is greater than or equal to 1. Likely you will find that $g^{(2)}(0)$ is equal to 1 to within the error of the measurement.[6]

- Take several more data sets with different parameters for **Update Period (Data Run)** and **Number of points**. Always take at least 10 points in order to assure reasonable statistics.

Hopefully in this lab you've shown that if you condition your measurements on the presence of a photon in the idler beam, you produce a single-photon state in the signal beam. This is a highly nonclassical state. If you don't condition your measurements, your signal beam behaves classically.

Q5: Why don't you obtain the same value for $g^{(2)}(0)$ when performing two- and three-detector measurements?

---

[6] In an ideal world we would expect to measure $g^{(2)}(0)=2$ for this thermal-like beam, however, we would need much higher time resolution (on the order of 10s of femtoseconds) to achieve this result. Given the nanosecond resolution that we have, the expected result is $g^{(2)}(0)=1$. See ref. [L2.3] for more details.

L2.7 References

[L2.1] P. Grangier, G. Roger, and A. Aspect, “Experimental evidence for a photon anticorrelation effect on a beam splitter: A new light on single-photon interferences,” Europhys. Lett. **1**, 173 (1986).

[L2.2] J.J. Thorn et al., “Observing the quantum behavior of light in an undergraduate laboratory,” Am. J. Phys **72**, 1210 (2004).

[L2.3] R. Loudon, *The Quantum Theory of Light, 3rd ed*. (Oxford University Press, Oxford, 2000).

LAB 3

Single-Photon Interference

LAB TICKET

Referring to fig. L3.1, if we measure the number of output photons on detector-**B**, N_B, in a given time interval, we'll find that N_B varies sinusoidally with ϕ—we'll see interference. The visibility V of this interference pattern is defined as

$$V = \frac{N_{B\,\max} - N_{B\,\min}}{N_{B\,\max} + N_{B\,\min}}, \tag{L3.1}$$

where $N_{B\,\max}$ is the maximum number of output counts (as ϕ is varied), and $N_{B\,\min}$ is the minimum number. Assuming vertically polarized input photons, calculate the visibility of the measured pattern on detector **B**, as a function of the angle of waveplate 1, θ_1.

L3.1 INTRODUCTION

In lab 2 you showed that it was possible to create a beam that consists of individual photons. In this laboratory you will show that if these individual photons are passed through an interferometer, they will interfere with themselves. Indeed, it is possible to do both of these experiments at the same time: You'll be performing an experiment which simultaneously shows both the wave-like and particle-like aspects of light.

You'll be using the polarization interferometer (PI) described in chapters 2 and 3, and depicted in fig. L3.1. You should already be familiar with the workings of this interferometer, so we will not go into detail about how it works here. A diagram of the complete experimental apparatus is shown in fig. L3.2, while a picture of the beam displacing polarizers (BDPs) is shown in fig. L3.3.

Recall that in lab 2 we were able to show that the signal beam contained individual photons, as long as we conditioned detections at **B** and **B′** on the detection of an idler photon at **A**. This was done by looking for a lack of coincidences between two detectors

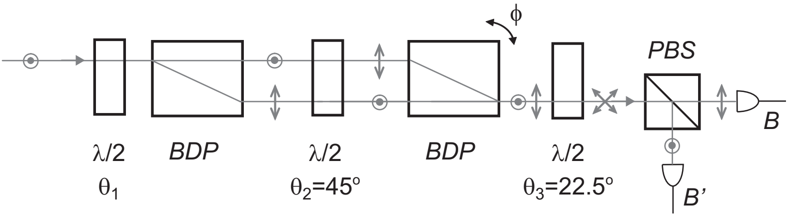


Fig L3.1 The polarization interferometer. Here λ/2 denotes a half-wave plate, BDP denotes a beam displacing polarizer, and PBS denotes a polarizing beam splitter. The phase of the interferometer ϕ is adjusted by tilting one of the BDPs.

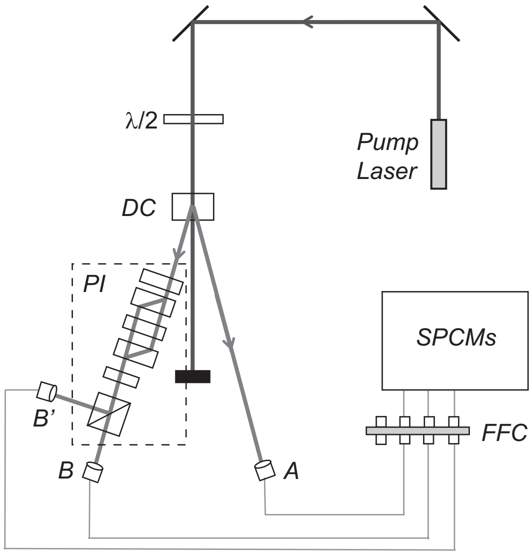


Fig L3.2 The experimental apparatus, with the polarization interferometer (PI) of fig. L3.1 in the dashed box. Here λ/2 denotes a half-wave plate, DC denotes the downconversion crystal, FFC denotes fiber-to-fiber coupler, and SPCMs denotes the single photon counting modules.

monitoring the outputs of a beam splitter. Note that in fig. L3.2, we still have two detectors at the outputs of a polarizing beam splitter (PBS), and a detector for the idler beam. The main difference from lab 2 is that the beam splitter is now part of the PI.

As far as the output detectors are concerned, the PI is a beam splitter; it takes one input beam and creates two output beams. The splitting ratio of this "beam splitter" depends on the phase of the interferometer. For some phases most of the light is transmitted, while for others most of the light is reflected. We can even adjust the interferometer phase so that the beam splits equally.

In lab 2 you measured the quantity $g^{(2)}(0)$ with a beam splitter adjusted for a 50/50 splitting ratio. As proved there, however, $g^{(2)}(0)$ is independent of the splitting ratio. You can use a 50/50 beam splitter, or an 80/20 beam splitter, and you will always measure

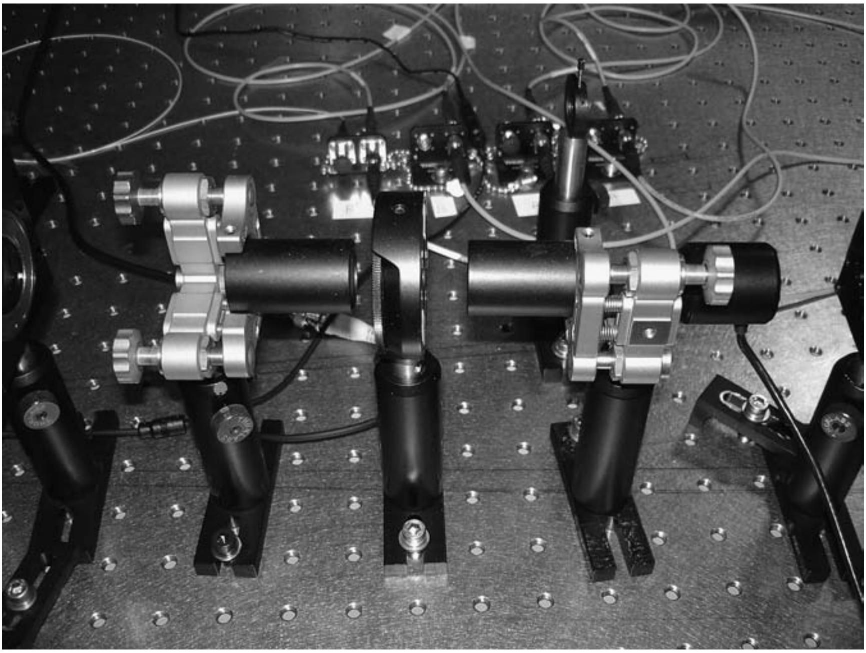


Fig L3.3 The beam displacing polarizers (BDPs) sandwiched around a half-wave plate. The large black "knob" on the mount on the right is a stepper motor. It allows the computer to tilt the mount to adjust the phase of the interferometer.

$g^{(2)}(0)$ to be the same. The only caveat is that you can't get too close to 100/0 or 0/100, because then one of the detectors doesn't get much light, both the numerator and denominator of $g^{(2)}(0)$ get small, and experimentally $g^{(2)}(0)$ is not well behaved.

The upshot of all this is that we have all the information we need to measure $g^{(2)}(0)$ at the same time that we measure the interference pattern. It's all the same data, acquired at the same time, but looked at in slightly different manners. You'll be able to see both wave-like behavior (interference), and particle-like behavior [$g^{(2)}(0)<1$] in the same experiment.

Before describing the experiment, I'll note that there are other interesting interference experiments that can be performed using a spontaneous parametric downconversion source; see ref. [L3.1] for some examples.

L3.2 ALIGNING THE POLARIZATION INTERFEROMETER

Note to instructors: I would suggest that, at the very least, before students come to the laboratory the BDPs should be properly rotated in their mounts, so that they will properly separate and recombine the beams to form an interferometer. To save lab time you could pre-align the interferometer and equalize the path lengths, and have students begin the lab starting at section L3.4.

We'll assume that the alignment of the crystal, and of detectors **A** and **B**, as described in lab 1, has already been performed.

Your PI consists of three half-wave plates, two BDPs, and a PBS, as shown in fig. L3.1. From left-to-right in that figure, we will refer to the wave plates as wave-plate 1, wave-plate 2, etc. Before beginning the alignment, arrange these components on the table, and inch or so apart, roughly as they will be placed in the experiment (see fig. L3.3). The point here is to get an idea of how much space the interferometer will occupy. Line the components up parallel to the path of the beam headed to detector-**B** (but not yet in the beam), with the idea that you could slide them into their appropriate place along the beam.

- **Make sure the detectors are off and the blue pump beam is blocked.**
- Connect the fiber leading from the **B** collection optics to the alignment laser, as shown in fig. L3.4. Turn on the alignment laser.
- Before inserting any of the polarization interferometer components, you want to align the two irises in fig. L3.4 that are closest to the downconversion crystal. They should be centered on the alignment laser beam, as described in lab 2. You want these two irises to be a foot or so apart from each other, but they also both need to be located between the downconversion crystal and the PI. Furthermore, you don't want them to block the beam headed to detector-**A**.

You'll see in fig. L3.4 that the effect of the PI is to displace the beam coming from the downconversion crystal to detector-**B**. The amount of displacement is determined by the BDPs. Ask your instructor how far this displacement is; it's typically on the order of about 4 mm. As can be seen in fig. L3.4, when looking from the perspective of detector-**B** toward the downconversion crystal, detector-**B** will need to move to the left.

- Note the location of detector-**B** with respect to the ruler it is pushed up against (see fig. L1.3). Unscrew detector-**B** from the table, slide it to the left by 4 mm (or whatever the displacement of your BDPs is), and then screw it back to the table. For the moment, do **not** adjust the tilt of the detector mount.
- Begin assembling the PI by inserting the BDP which goes closest to detector-**B**. Insert it in its proper location along the beam path, leaving space for the other components. This should be the BDP with the stepper motor on its horizontal tilt. Note from fig. L3.4 that when looking from detector-**B** toward the downconversion crystal, the alignment laser beam should enter the BDP toward its left edge. One beam (polarization) will go straight through, and the other will be deflected to the right by 4 mm, and then emerge parallel to the first beam. If the BDP does not displace this second beam horizontally to the right, you will need to rotate it in its mount so that it does (**double check with your instructor before doing this**).
- Orient this BDP so that its surfaces are perpendicular to the alignment laser (do this by looking at the back reflection from the first surface). Check that both beams emerge from the BDP without clipping its edges. Once both beams are emerging unclipped, and the faces of the BDP are roughly perpendicular to the beam, screw its mount to the table.

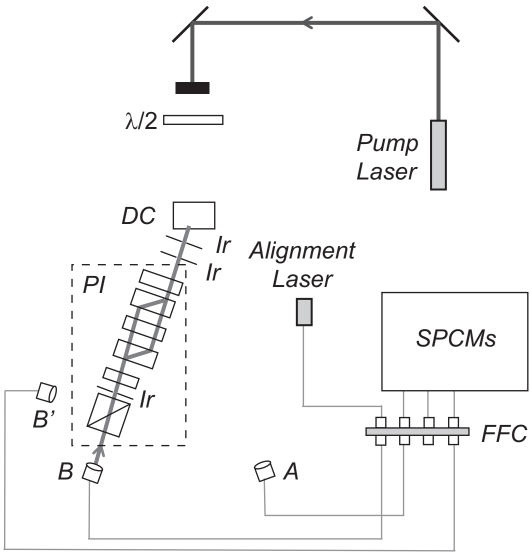


Fig L3.4 Aligning the polarization interferometer. Here Ir denotes an iris.

- There should now be two beams emerging from the BDP separated by approximately 4 mm. The beam on the right (looking from detector-**B** toward the downconversion crystal) should be going approximately through the two irises, and onto the downconversion crystal, as in fig. L3.4.
- If the alignment beam is not centered on the two irises, unscrew detector-**B** from the table, and gently slide it along the ruler to center the beam on the irises. Do not adjust the tilt of its mount in order to center the beams. Once the beam is centered on the irises, screw the mount back to the table. Double check that none of the beams entering or emerging from the BDP are clipped (slightly move the BDP if necessary).
- If simple translation of detector-**B** is not sufficient to get the alignment beam traveling back through the center of the irises, you may need to rotate the BDP in its mount (**double check with your instructor before doing this**).
- Referring to figs. L3.1 and L3.4, insert the PBS, wave-plate 3, and the third iris. Center them and orient them perpendicular to the alignment beam. Rotate the wave plate so that its axis is at 22.5°, which should equalize the intensities of the beams emerging from the BDP.
- Insert wave-plate 2, and rotate its axis to 45°.
- Insert the second BDP so that both beams enter without clipping, and orient it perpendicular to the beams. This BDP should be rotated in its mount so that both of the incoming beams are recombined into a single output beam emerging from the right side of the BDP (looking from detector-**B** toward the downconversion crystal) and passing back through the irises, as shown in fig. L3.4. Slight adjustment of the rotation angle of wave-plate 2 might help to get all of the left beam to deflect to the right. If this doesn't help, you may need to rotate the BDP in its mount (**double check with your instructor before doing this**).

- Insert wave-plate 1, and orient its axis at 22.5°.
- Insert a linear polarizer between wave-plate 1 and the downconversion crystal. Orient this polarizer to be either vertical or horizontal. Place a screen behind it so that you can see the alignment beam.
- Open the iris apertures.
- Very slowly adjust the horizontal tilt of the BDP closest to the downconversion crystal (the BDP with the manual horizontal adjustment knob), which scans the interferometer phase, while looking at the screen. You should see the beam on the screen getting alternately brighter and darker; in other words, you should observe interference.
- At this point you'd like to maximize the contrast (visibility) of the interference. Ideally you would like to see the beam get uniformly brighter and darker as the phase of the interferometer is adjusted; it should look like it's "breathing". It's easier to adjust things by looking at a dark fringe and trying to minimize its intensity. You can try slight adjustments of the polarizer or wave-plate rotation angles to improve the visibility.
- If instead of a single fringe, you see alternating bright and dark fringes that move as the phase is adjusted, somewhere along the line the interferometer beams are tilted with respect to each other. You can try adjusting the vertical tilt of the BDPs, or rotating the BDPs in their mounts to correct this (**double check with your instructor before doing this**).
- Once you've improved the visibility of the interferometer as best you can, adjust the horizontal tilt of the BDP so that it is roughly perpendicular to the beams.

If the interferometer alignment is being performed by students as part of the lab, the material in sec. L3.4 should be completed at this point.

- Turn off the alignment laser, and reconnect the **B** fiber to the **B** SPCM.
- Connect the **B′** fiber to the alignment laser, and then turn the laser back on.
- Align the **B′** detector using the procedure described in lab 2. This beam should pass through the centers of all three irises. Do your best to coarsely align this detector before adjusting the PBS. As much as possible, try to get the beam through all three irises by simply translating and tilting the **B′** detector mount.
- Turn off the alignment laser, and reconnect the **B′** fiber to the **B′** SPCM.

L3.3 EQUALIZING THE PATH LENGTHS

The interferometer is now nearly aligned. The last step is to equalize the path lengths of the two interferometer arms. This is necessary in order to see interference with the downconversion source, because it has a very short coherence length (see complement 2.A).

- Run the LabVIEW program "Coincidence.vi". Set **Experimental Setup** to **Interference**, and **Update Period** to 0.2s. In the pane with the graphs, click on the **AB** & **AB′** tab.

- You should see that the **Pause Stepper** button is grayed (“pushed-in”), and the **Current position (μsteps)** indicator is not changing.
- If the **Current position (μsteps)** indicator does not read 26000, set **Set position** to 26000, and then click the **Go to Set Position** button. The stepper motor should move to this position, and the **Current position (μsteps)** indicator should read 26000. This is the center of travel for the stepper motor, and you want to adjust the path lengths of the arms so that they are roughly equal when the stepper is at this location.
- **Make sure the room lights are off**, and turn on the detectors.
- Set the wave plate angles as follows: $\theta_1 = 0^o$, $\theta_2 = 45^o$ and $\theta_3 = 22.5^o$. You should see roughly equal numbers of **AB** and **AB′** coincidence counts. If you don’t, try to tweak the tilt of detectors **B** and **B′** to improve things. You can also slightly rotate wave-plate 3 to better equalize the count rates, but you don’t want this wave plate to get too far from $\theta_3 = 22.5^o$.
- Now rotate wave-plate 1 to $\theta_1 = 22.5^o$. At this point, doing so will probably have very little effect on the coincidence count rates.
- Slowly adjust the horizontal tilt of the BDP closest to the downconversion crystal (the BDP with the manual horizontal adjustment knob), while observing the **AB** and **AB′** coincidence counts (observing the **B** and **B′** singles counts will work as well).

If the path lengths of the two arms are off by more than the coherence length, you should see no interference; adjusting this tilt, which adjusts the interferometer phase, will have essentially no effect on the counts. If the path lengths are within a coherence length you will notice that the counts oscillate as the phase is adjusted. The **AB** and **AB′** counts should oscillate out of phase with each other: as the **AB** counts increase, the **AB′** counts will decrease. As you scan the path length, you will see that at first the amplitude of the oscillations is small, with low contrast, because the path lengths are off by slightly more than a coherence length. When the path lengths are exactly equal you should see large amplitude, high contrast oscillations. The **AB** and **AB′** counts should alternately drop to nearly 0. Continuing to scan will take you beyond the coherence length, so the oscillation contrast will once again decrease. The idea is to adjust the lengths to be equal, so that the contrast is largest.

- Slowly scan the horizontal tilt of the BDP to find maximum contrast in the oscillation of the **AB** and **AB′** counts. If tilting one way doesn’t work, go back and tilt the other way. Don’t be afraid to scan a long way.
- If large tilts in either direction do not yield interference, double check that your wave plates are adjusted to the correct angles: $\theta_1 = 22.5^o$, $\theta_2 = 45^o$ and $\theta_3 = 22.5^o$.
- If you still don’t see interference, you probably need to go back to the steps above and double check the alignment of the interferometer with the alignment laser. If the interferometer is properly aligned with the laser, the only adjustments you should use to obtain interference with the downconversion are the wave-plate angles and the horizontal tilt of the BDPs.
- Once you have adjusted the path lengths to be equal, and maximized the contrast by tilting the BDP, there are a few tweaks you can do to improve the contrast further. Slight adjustment of the wave-plate angles may improve the contrast. Also, closing one or more of the iris apertures may improve the contrast, but this will also reduce

the overall count rates. Ideally, you'd like to get a visibility of 90% or better, but even if you only manage 70%, you'll still convincingly see single-photon interference.

L3.4 THE POLARIZATION INTERFEROMETER

Note to instructors: If students are aligning the interferometer themselves, this part of the laboratory should be completed at the point in sec. L3.2 where it is noted. If the interferometer is pre-aligned before students enter the lab, they should start here.

Looking at fig. L3.1, from left-to-right we will refer to the wave plates as wave-plate 1, wave-plate 2, etc.

- **Make sure the detectors are off and the blue pump beam is blocked.**
- Connect the fiber leading from the **B** collection optics to the alignment laser, as shown in fig. L3.4. Turn on the alignment laser.
- Set the wave plate angles to $\theta_1 = 22.5^o$, $\theta_2 = 45^o$, and $\theta_3 = 22.5^o$.
- Take a piece of white paper or an index card, and carefully insert it in different locations along the beam path so you can see where the beam goes. You'll probably need to have the lights out for this, as the beam is not very bright. You should verify that the beam follows the path shown in figs. L3.1 and L3.4. In particular, you should notice two beams in between the BDPs.
- While looking at the two beams between the BDPs, slowly rotate wave-plate 3.

Q1: What happens to the beams as this wave plate is rotated? Explain.

- Rotate wave-plate 3 back to $\theta_3 = 22.5^o$.
- Insert a linear polarizer between wave-plate 1 and the downconversion crystal. Orient this polarizer to be either vertical or horizontal. Place a screen behind it so that you can see the alignment beam. This beam may be very dim and hard to see; don't worry about that for the moment.
- Run the LabVIEW program "Coincidence.vi". Set **Experimental Setup** to **Interference**, and **Update Period** to 0.2s.
- Even though the detectors are not on, and you cannot see any counts, you are using the vi because it has some controls for the stepper motor that scans the interferometer phase. You'll be using it to slowly sweep the phase, so that you can see the interference pattern. When the program is first initialized, the motor is not moving. You should see that the **Pause Stepper** button is grayed ("pushed-in"), and the **Current position (μsteps)** indicator is not changing.
- If the **Current position (μsteps)** indicator does not read 26000, set **Set position** to 26000, and then click the **Go to Set Position** button. The stepper motor should move to this position, and the **Current position (μsteps)** indicator should read 26000.
- Move the **Stepper increment** slider to 4, then click the **Pause Stepper** button to release it. The motor should now be moving, and the **Current position (μsteps)** indicator should be increasing.
- ***Note: Don't let the stepper motor scan too far***. You can stop the motor from scanning by clicking the **Pause Stepper** button. Try to keep the stepper motor in the range

25000–27000. To get back to the center of the range, simply set **Set position** to 26000, and then click the **Go to Set Position** button.

- Observe the beam behind the polarizer. You should notice it getting alternately bright and dark as the phase of the interferometer is scanned to produce constructive and destructive interference. If you don't see this, it is likely that the wave plate or linear polarizer angles are not properly set—double check them. The contrast is most easily adjusted by rotating the linear polarizer or wave-plate 1. If you still have problems seeing interference, consult with your instructor about realigning the interferometer.

Q2: As the motor scans, you should notice that the beam stays bright in front of the polarizer, but gets brighter and darker behind it. Why is this?

- Pause the stepper motor.
- Remove the screen and the linear polarizer, **turn off the alignment laser**, and reconnect the **B** fiber to the **B** SPCM.

*If you are completing the alignment of the polarization interferometer as part of your laboratory, return to sec. L3.2 to perform the alignment of the detector-**B′**. Then continue on to sec. L3.3, and equalize the path lengths of the arms of the interferometer.*

L3.5 SINGLE-PHOTON INTERFERENCE AND THE QUANTUM ERASER

By now the path lengths of the arms of your interferometer are equalized, and you are ready to observe single-photon interference.

- **Make sure the lights are out**, then turn on the detectors and unblock the blue beam.
- Set the wave plate angles to $\theta_1 = 22.5^\text{o}$, $\theta_2 = 45^\text{o}$ and $\theta_3 = 22.5^\text{o}$.
- Set **Update Period** to 0.2s, and in the pane with the graphs click on the **AB** & **AB′** tab.
- Begin with the stepper motor at 26000, set the **Stepper increment** to 1 or 2, then unpause the stepper and let it scan. Again, don't let it scan too far (keep the stepper motor in the range 25000–27000).
- You should see interference fringes in the **B**, **B′**, **AB**, and **AB′** count rates as the motor scans. The visibility should be fairly high; hopefully it will be 90% or better. If the visibility of your fringe pattern is less than about 70%, there are a few things you can do:

1. Tweak the tilt of the **B** and **B′** mounts.
2. The path lengths of the arms of your interferometer may not be equal. Try scanning in both directions (you can use a negative **Stepper increment**) until you find the point of equal path lengths. You want to perform your measurements near equal path lengths, where the visibility of the fringe pattern is maximized.
3. Slight adjustments of the rotation angles of your wave plates may improve the visibility.

4. Close down one or more of the iris apertures, although this will decrease your overall count rates.

If none of these work, please ask for help from your instructor, as part of the alignment procedure may need to be redone.

- Once you've observed high-visibility interference, set the **Data Taking Parameters** to: **Update Period (Data Run)** 1s, **Number of points** 60, **Center stepper position (μsteps)** 26000 (or whatever the center of the highest visibility portion of your scan is), **Range to scan (±)** 150, then click the **Take Data** button to acquire a data set.
- Hopefully, your data should have between 2 and 3 full oscillations, with a high visibility. If you don't have between 2 and 3 oscillations, adjust **Range to scan (±)**, then take a new scan.
- The singles counts **B** and **B′** represent ordinary interference that you'd see with any light source. The coincidence counts **AB** and **AB′** represent true, single-photon interference. This is shown by the fact that your measured values for $g^{(2)}(0)$ should be less than one. At the moment these $g^{(2)}(0)$ measurements probably fluctuate greatly; don't worry about this now, we'll come back and clean them up shortly.
- Take at least two more data sets, using $\theta_1 = 10^{\circ}$ and $\theta_1 = 0^{\circ}$.

Q3: After you leave the lab, calculate the measured visibilities of your interference patterns (use the **AB** coincidence data). Compare them to the prediction made in your lab ticket.

Q4: Compare the interference patterns you measured with $\theta_1 = 22.5^{\circ}$ and $\theta_1 = 0^{\circ}$. Which setting has higher visibility? For which setting do you know the path of the photons through the interferometer? For which setting is the path information "erased"? Explain your observations in terms of complementarity (see sec. 5.6).

- Now set your wave plate angles to $\theta_1 = 22.5^{\circ}$, $\theta_2 = 45^{\circ}$ and $\theta_3 = 0^{\circ}$, and acquire another data set.

Q5: Compare the interference patterns you measured with $\theta_3 = 22.5^{\circ}$ and $\theta_3 = 0^{\circ}$. Which setting has higher visibility? For which setting do your detectors provide you with information about the path of the photons through the interferometer? For which setting is the path information "erased"? Explain your observations in terms of complementarity (see sec. 5.6).

L3.6 "EXPERIMENT 6"

In sec. 3.7 we discussed experiment 6, which is shown in figs. 3.7 and 3.8. Now you will perform that experiment. Comparing the experiment on your table to fig. 3.8, wave-plate 1 at $\theta_1 = 22.5^{\circ}$ is used to rotate the polarization of the photons from the downconversion crystal to place them in the state $|+45\rangle$. The two BDPs (PA_{HV}'s) and wave-plate 2 at $\theta_2 = 45^o$ are already set on your table. Wave-plate 3 with $\theta_3 = 22.5^o$ is used in combination with the PBS to implement the PA_{45}. The only thing we need to do is adjust the interferometer phase so that all of the light exits from only one port of the PBS.

- Set the wave plate angles to $\theta_1 = 22.5^o$, $\theta_2 = 45^o$ and $\theta_3 = 22.5^o$.
- While observing the **AB** and **AB′** coincidence counts, let the stepper motor run slowly to scan the interferometer phase. When you are near the point of highest visibility, and the interferometer output is such that **AB** maximized and **AB′** minimized, stop the stepper.
- Your interferometer is now adjusted so that nearly all the light is exiting to detector-**B**. It will now behave equivalently to experiment 6.
- Using a piece of paper or index cards, block the beam in the interferometer that is closest to the edge of the table [as in fig. 3.7(a)].

Q6: What do you notice about the **AB** and **AB′** coincidence rates? Explain this behavior.

- Unblock the beam closest to the edge of the table, and instead block the other beam [as in fig. 3.7(b)].

Q7: What do you notice about the **AB** and **AB′** coincidence rates? Explain this behavior.

Q8: With both beams unblocked, what do you notice about the **AB** and **AB′** coincidence rates? Explain this behavior.

Remember that when we first discussed this experiment in sec. 3.7, you probably found its results to be counterintuitive.

L3.7 PARTICLES AND WAVES

Now it's time to improve the $g^{(2)}(0)$ measurement, and truly demonstrate the wave-like and particle-like behaviors of light in the same experiment.

- Set θ_1 to an angle that you predict will yield an interference pattern with $V \approx 0.9$. (As discussed in sec. L3.1, if the visibility is too close to 1, experimental measurements of $g^{(2)}(0)$ will not be well behaved at interferometer phases where one of the detectors receives almost no light due to destructive interference.)
- Take a data run that scans over approximately 2 full oscillations, has at least 30 points, and has an integration time of at least 10s per point.

Q9: What value do you measure for $g^{(2)}(0)$? Be sure to quote your error.

Your coincidence count rates should show a high visibility interference pattern. This shows conclusively that the photons are behaving as waves as they traverse the interferometer. They must take both paths through the interferometer after they leave the first BDP in order to display interference. Hopefully your measured values of $g^{(2)}(0)$ will all be less than 1, with a mean value that is significantly less than 1. As discussed in detail in lab 2, this shows that the photons are behaving as particles. They can only take one path after they leave the final beam splitter.

You've done an experiment that unambiguously shows both the wave-like, and the particle-like nature of light.

L3.8 References

[L3.1] E. J. Galvez et al., "Interference with correlated photons: Five quantum mechanics experiments for undergraduates," Am. J. Phys. **73**, 127 (2005).

LAB 4

Quantum State Measurement

LAB TICKET

A beam consisting of individual photons, in the arbitrary polarization state $|\psi\rangle$ of eq. (L4.1), passes through a quarter-wave plate (QWP), then a half-wave plate (HWP), and finally a horizontal polarizer (the transmission port of a polarizing beam splitter), as shown in fig. L4.1. Show that for the wave-plate settings in table L4.1, the probabilities of a photodetection at detector-**B** are given by eqs. (L4.2), (L4.5) and (L4.6). This proves that these wave-plate settings correspond to the desired measurements.

L4.1 INTRODUCTION

The procedure for determining the polarization state of a beam of individual photons is described in detail in complement 5.A. Here we'll briefly review that procedure.

The polarization state we wish to determine can be written as

$$|\psi\rangle = a|H\rangle + be^{i\phi}|V\rangle, \tag{L4.1}$$

so we must determine three real-valued quantities: a, b and ϕ. By performing many polarization measurements in the *HV*-basis, we can measure the probability that a photon will be horizontally polarized, given that the beam is prepared in state $|\psi\rangle$:

$$P(H\,\|\,\psi\rangle) = |\langle H|\psi\rangle|^2 = a^2. \tag{L4.2}$$

From this, and the fact that the state vector must be normalized, we can determine a and b:

$$a = \sqrt{P(H\,\|\,\psi\rangle)}, \tag{L4.3}$$

$$b=\sqrt{1-a^2}. \tag{L4.4}$$

In order to determine ϕ, we must perform polarization measurements in other bases. As shown in complement 5.A, the probabilities of measuring a +45° or a left-circularly polarized photon are

$$P\left(+45\|\psi\rangle\right)=\frac{1}{2}(1+2ab\cos\phi), \tag{L4.5}$$

$$P\left(L\|\psi\rangle\right)=\frac{1}{2}(1+2ab\sin\phi). \tag{L4.6}$$

We can combine these two equations, to find that

$$\phi=\tan^{-1}\left(\frac{\sin\phi}{\cos\phi}\right)=\tan^{-1}\left[\frac{P\left(L\|\psi\rangle\right)-\frac{1}{2}}{P\left(+45\|\psi\rangle\right)-\frac{1}{2}}\right]. \tag{L4.7}$$

ϕ can be uniquely determined on the interval $(-\pi,\pi]$ by knowing the signs of the numerator and denominator of the fraction inside the inverse tangent.[1]

The measurement procedure is then as follows. Many measurements of the polarization of identically prepared photons are performed in each of three different bases, in order to determine the probabilities $P\left(H\|\psi\rangle\right)$, $P\left(+45\|\psi\rangle\right)$, and $P\left(L\|\psi\rangle\right)$. From these measurements, eqs. (L4.3), (L4.4), and (L4.7) are used to determine the coefficients that describe the state $|\psi\rangle$ [eq. (L4.1)].

Note that we are assuming that the polarization state is pure, so that it is possible to describe the state in terms of a state vector (ket). More generally, the state can be described by a density matrix (see complement 8.A). For information about reconstructing the density matrix of a polarization state, see ref. [L4.1].

L4.1.1 Wave-Plate Angles

Measurements in the *HV*-basis are easily performed using a polarizing beam splitter (PBS). Measurements in the ±45-basis can be performed by inserting a half-wave plate in front of the PBS, while measurements in the *LR*-basis can be performed by inserting a quarter-wave plate in front of the PBS. However, experimentally it is more convenient to simply rotate a wave plate than it is to insert or remove one. With this in mind, the experimental arrangement you will use for state measurement is depicted in fig. L4.1.

---

[1] In a computer, this is done using the function $\text{atan2}(x,y)\equiv\tan^{-1}(y/x)$.

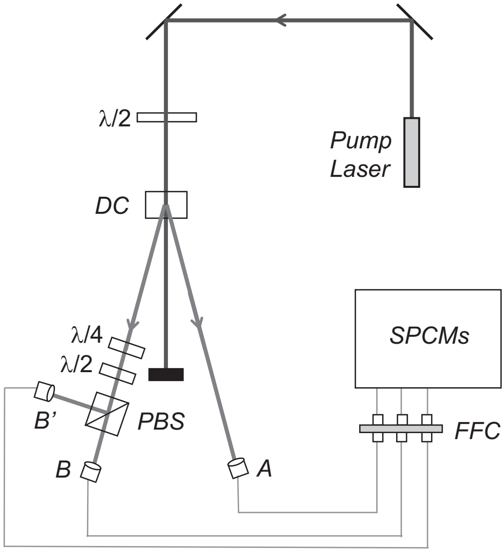


Fig L4.1 The experimental apparatus for polarization state measurement. Here λ/2 denotes a half-wave plate, λ/4 denotes a quarter-wave plate, DC denotes the downconversion crystal, PBS denotes a polarizing beam splitter, FFC denotes fiber-to-fiber coupler, and SPCMs denotes the single-photon-counting modules.

As in our previous labs, detection of an idler photon at **A** prepares the signal beam in a single-photon state. It is the polarization state of the signal beam that is being measured. The signal beam first passes through a quarter-wave plate, then a half-wave plate, and then strikes the PBS, which sends beams to detectors **B** and **B′**. The PBS transmits horizontally polarized photons to detector-**B**. The probability that a single photon is transmitted by the PBS is equal to the probability of an **AB** coincidence count, and is determined from the measured coincidence counts by

$$P(AB) = \frac{N_{AB}}{N_{AB} + N_{AB'}}. \tag{L4.8}$$

In the lab ticket you will show that the wave-plate settings in table L4.1 allow you to obtain measurements in the three needed bases.

Table L4.1 The wave-plate settings needed to perform measurements in the indicated bases.

Basis	Fast Axis of Quarter-Wave Plate	Fast Axis of Half-Wave Plate
HV	0°	0°
LR	45°	0°
±45	45°	22.5°

L4.2 ALIGNMENT

As can be seen by comparing figs. L4.1 and L2.1, the experimental apparatus for this lab is nearly identical to that of lab 2. There are only two differences: In this lab there is a quarter-wave plate (QWP) in the signal beam, and the wave plates are on motorized rotation stages, so the computer can control their rotation angles.

- Follow the procedure in lab 1 to align the crystal, and detectors **A** and **B**.
- Follow the procedure in lab 2 to align the polarizing beam splitter, the half-wave plate (HWP) and detector **B′**. Do not yet insert the QWP.
- Run the LabVIEW program “QSM.vi”.

Documentation for this program comes with the software. It starts by initializing the counters and the motors which control the wave-plate rotation stages; this takes a few seconds, and the **Status** indicator reads “Initializing.” Once everything initializes, the **Status** should switch to “Reading Counters.” The program is now reading the counters, and updating the screen in real time. When the program is done initializing, the QWP and HWP angles are set to 0 (double check that the **QWP Position** and **HWP Position** parameters read 0).

- The wave-plate angles are changed by first entering the desired rotation angle into the parameters **QWP (HWP) Desired Position** and then pressing **Move Motors**. Check that this is working properly by entering 45 for **HWP Desired Position**, then pressing **Move Motors**. The HWP should rotate, and **HWP Position** should now read 45.
- Set **HWP Desired Position** to 0, then press **Move Motors**, in order to rotate the HWP back to 0°.
- While **HWP Position** should now read 0, **HWP Motor Position** may not read 0; it will read the value given in the **HWP zero** parameter. The **HWP zero** and **QWP zero** parameters are needed because the 0 angles of the motors are not necessarily perfectly aligned with the 0 angles of the wave plates. For an accurate state measurement, we need to make sure that these parameters are properly set.
- I’ll assume that your downconversion source is aligned so that it is producing horizontally polarized signal and idler pairs. The HWP is then properly zeroed when all of the photons are transmitted through the PBS to detector-**B**.[2] While monitoring the **AB** and **AB′** coincidences, slightly adjust (in about 1° increments) the rotation angle of the HWP by entering values into **HWP Desired Position**, then pushing the **Move Motors** button. The goal is to maximize the **AB** coincidences while minimizing the **AB′** coincidences. Once you’ve done this, note the reading in **HWP Motor Position**—this is the correct 0 setting for the HWP, and you should enter it into the **HWP zero** parameter.
- Once the correct **HWP zero** value has been entered, the program must be stopped and then restarted in order to recognize the new value. Write down the correct value, because any time you quit and restart LabVIEW you may need to reenter it.

---

2. If your source produces vertically polarized pairs, you want all the photons to be reflected by the PBS and sent to detector-**B′**.

Now that the HWP is properly zeroed, you need to insert and zero the QWP.

- Insert the QWP between the downconversion source and the HWP (insert it near the HWP) and orient it perpendicular to the beam. You can use the alignment laser for this, but that's probably not necessary. Set the QWP at the same height as the HWP, and insert it into the beam. Slide the QWP back and forth perpendicular to the beam while monitoring the count rates. If the count rates decrease then the beam is being clipped; adjust the wave plate position so that it does not clip the beam.
- Set both **QWP Desired Position** and **HWP Desired Position** to 0, and press **Move Motors**.
- While monitoring the **AB** and **AB′** coincidences, slightly adjust (in about 1° increments) the rotation angle of the QWP by entering values into **QWP Desired Position**, then pushing the **Move Motors** button. The goal is to maximize the **AB** coincidences while minimizing the **AB′** coincidences.[3] Once you've done this, note the reading in **QWP Motor Position**—this is the correct 0 setting for the QWP, and you should enter it into the **QWP zero** parameter.
- Once again, after the correct **QWP zero** value has been entered, the program must be stopped and then restarted in order to recognize the new value. Write down the correct value, because any time you quit and restart LabVIEW you may need to reenter it.

L4.3 MEASUREMENT OF LINEAR POLARIZATION STATES

- Set both **QWP Desired Position** and **HWP Desired Position** to 0, and then press **Move Motors**. If your source produces horizontally polarized photons, the **AB** coincidences should be maximized, while the **AB′** coincidences should be minimized.
- In the **Data Taking Parameters** box set **Update Period (Data Run)** to 5.0s, and **No. Of Samples** to 10, then push the **Take Data** button.

Control of the computer is now switched to the data acquisition program. This program requires nothing from you; it automatically adjusts the wave plates to the correct angles (given in table L4.1), calculates probabilities, determines the parameters a, b, and ϕ, and saves the data to a file. For the parameters you just entered the data run will take approximately 2 1/2 min. There is no graceful way to exit this program while it is still running, and if you exit in the middle by closing the window, chances are you'll need to reboot the computer—better to just let it run.

- The program is done running when the **Operation** box reads "Finished." The data file is automatically named according to the date and time. In your notebook record the filename, important parameters (**Update Period**, **No. Of Samples**), and results (a, b, and ϕ, errors, probabilities, etc.).

---

[3] Again, this assumes horizontally polarized photons coming out of the downconversion crystal. For vertically polarized photons, maximize **AB′** and minimize **AB**.

Q1: Is the state determined by the program the state you expect? Explain why or why not.

Q2: If either of the parameters a or b is very small, then ϕ is not really meaningful. Explain why this is so.

Now look at some other polarization states.

- Insert a HWP in the beam between the downconversion crystal and the QWP. You'll use this HWP to rotate the polarization coming from the downconversion crystal, to generate different linear polarization states. As such, we'll refer to this HWP as the "state-generation" HWP.
- Use the state-generation HWP to generate the state $|+45\rangle$. When creating this state, first set the HWP to the angle that you think it should be at. You can fine-tune the angle by setting the measurement wave plates to 0°, and adjusting the state-generation HWP so that $P\left(H \| \psi\rangle\right)$ is what you would expect for the state. The indicator **P(AB)** is useful for this, because with the measurement wave plates set to 0°, it reads $P\left(H \| \psi\rangle\right)$.
- Collect a data set to determine the state.

Q3: For this state, calculate the expected values of the parameters a, b, and ϕ, and compare them to the measured values. Do this before proceeding, because if you find that you've created the state $|-45\rangle$ instead of $|+45\rangle$, chances are that you're rotating the state-generation HWP in the wrong direction.

- Now generate and perform a state measurement for each of the polarization states $|-45\rangle$ and $|+30\rangle$. Once again, first set the state-generation HWP to the angle that you think it should be at, then fine tune it by monitoring **P(AB)**.

Q4: For each of these states, calculate the expected values of the parameters a, b, and ϕ, and compare them to the measured values.

L4.4 MEASUREMENT OF CIRCULAR AND ELLIPTICAL POLARIZATION STATES

- Remove the state-generation HWP and replace it with a state-generation QWP. You'll use this QWP to generate different circular/elliptical states.
- Use the state-generation QWP to generate the state $|L\rangle$. When creating the state, first set the QWP to the angle that you think it should be at. You can fine-tune the angle by setting the measurement wave plates to 0°, and monitoring **P(AB)**.
- Collect a data set to determine the state.

Q5: For this state, calculate the expected values of the parameters a, b, and ϕ, and compare them to the measured values. Do this before proceeding, because if you find that you've created the state $|R\rangle$ instead of $|L\rangle$, chances are that you're rotating the QWP that is producing the state in the wrong direction.

- Now generate and perform a state measurement of the state $|R\rangle$.
- Finally, rotate your QWP axis to 15°, and perform a state measurement.

Q6: For each of the circular/elliptical polarization states above, calculate the expected values of the parameters a, b, and ϕ, and compare them to the measured values. (NOTE: the program calculates a assuming that it is real. This means you must factor out any overall phase factor in your calculations, to get your results to agree with the program.)

L4.5 References

[L4.1] J.B. Altepeter, E.R. Jeffrey, P.G. Kwiat, "Photonic state tomography," in *Advances In Atomic, Molecular, and Optical Physics*, P.R. Berman and C.C. Lin, eds. (Elsevier, Amsterdam, 2006), p. 105.

LAB 5

Testing Local Realism

LAB TICKET

Create a spreadsheet that calculates $P(-\alpha,\alpha)$, $P(\beta,-\beta)$, $P(\beta,\alpha^{\perp})$, $P(-\alpha^{\perp},-\beta)$, and H, given the state $|\psi_1\rangle$ in eq. (L5.13) below. For the fixed value $\alpha = 35^{\text{o}}$, make plots of $P(\beta,\alpha^{\perp})$, $P(-\alpha^{\perp},-\beta)$, and H as functions of β (use values of β between 0^{o} and 90^{o}). How can you maximize H, while still keeping $P(\beta,\alpha^{\perp})$, $P(-\alpha^{\perp},-\beta)$ less than 0.01?

L5.1 INTRODUCTION

In this experiment you will be testing local realism. By local we mean that measurements performed in one place cannot affect the outcomes of measurements preformed somewhere else. By realism we mean that objects have values for measurable quantities, regardless of whether or not we measure them. According to local realism, if two photons are produced by a source, their polarizations are completely defined once they leave the source. Thus, polarization measurements performed on one photon should not affect the results of polarization measurements performed on the other photon.

Local realism is common sense, and all classical systems are bound by it. However, as you will demonstrate in this lab, quantum systems are not constrained by local realism. In order to explain the results of certain experiments we must abandon either locality or reality.

You will be testing local realism using the method suggested by Lucian Hardy that we discussed in sec. 8.4 and complement 8.B.[1] You will also perform another test, one that uses a “Bell inequality” that was originally derived by Clauser, Horne, Shimony and Holt (CHSH). This is a historically important test of local realism, which has been

---

[1] See also refs. [L5.1]–[L5.5].

around longer than Hardy's test. We'll describe the basics of this test below, but for more details see ref. [L5.6], and the references therein.

The experimental apparatus is shown in fig. L5.1 (see also fig. 8.3). The downconversion source is similar to the one we've used in previous labs, but there's an important difference. There are actually 2 downconversion crystals sandwiched back-to-back, with their crystal axes rotated at 90° with respect to each other. As described in sec. 8.4, this source produces photons in the state

$$\left|\psi\right\rangle = \sqrt{a}\left|H\right\rangle_A\left|H\right\rangle_B + \sqrt{1-a}\, e^{i\varphi}\left|V\right\rangle_A\left|V\right\rangle_B. \tag{L5.1}$$

Experimentally, to change the ratio of the probability of the production of horizontally or vertically polarized pairs (the parameter a), you rotate the half-wave plate in the pump beam, which changes the pump polarization (e.g., if the pump has a larger horizontal component, then vertically polarized output photons are more likely). The birefringent plate in the pump beam is used to adjust the relative phase φ. With this experimental arrangement, we can create polarization states with any arbitrary linear combinations of the states $\left|H\right\rangle_A\left|H\right\rangle_B$ and $\left|V\right\rangle_A\left|V\right\rangle_B$.

The idler photon travels to Alice and her two detectors (**A** and **A′**), while the signal photon travels to Bob and his two detectors (**B** and **B′**). We are interested in signal-idler pairs where Alice and Bob detect photons at the same time, and the raw data collected in the experiment consists of measuring numbers of coincidence counts in a given time window ($N_{AB'}$ is the number of coincidences between detectors **A** and **B′** in a given counting interval, for example.)

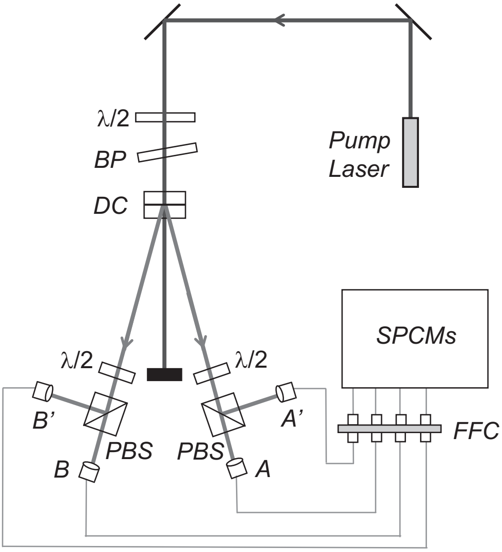


Fig L5.1 The experimental apparatus. Here λ/2 denotes a half-wave plate, BP denotes a birefringent plate, DC denotes the downconversion crystals, PBS denotes a polarizing beam splitter, FFC denotes fiber-to-fiber coupler, and SPCMs denotes the single-photon-counting modules.

Given the four measurements of coincidence counts (N_{AB}, $N_{AB'}$, $N_{A'B}$, and $N_{A'B'}$) we can determine the probability that Alice's and Bob's photons had a particular set of polarizations. For example, assume that Alice's half-wave plate is set to transmit photons polarized along the angle θ_A to detector **A**, and Bob's half-wave plate is set to transmit photons polarized along the angle θ_B to detector **B**. The joint probability that they will measure photons polarized along these directions, $P\left(\theta_A, \theta_B\right)$ is given by

$$P\left(\theta_A, \theta_B\right) = \frac{N_{AB}}{N_{AB} + N_{AB'} + N_{A'B} + N_{A'B'}}. \qquad \text{(L5.2)}$$

We will assume that in joint probabilities of the form $P\left(\theta_A, \theta_B\right)$, the first variable always refers to Alice's polarization.

L5.2 THEORY

L5.2.1 Hardy Test

Hardy's test of local realism is described in detail in sec. 8.4 and complement 8.B. In the idealized situation presented in sec. 8.4, this test involves testing the classical inequality

$$P\left(\theta_{A1}, \theta_{B1}\right) \leq P\left(\theta_{A2}, \theta_{B2}\right). \qquad \text{(L5.3)}$$

However, in complement 8.B we show that this inequality is simply a special case of the more general, experimentally testable, Bell-Clauser-Horne inequality [eq. (8.B.10)]:

$$P\left(\theta_{A1}, \theta_{B1}\right) \leq P\left(\theta_{A2}, \theta_{B2}\right) + P\left(\theta_{A1}, \theta_{B2}^{\perp}\right) + P\left(\theta_{A2}^{\perp}, \theta_{B1}\right), \qquad \text{(L5.4)}$$

where $\theta^{\perp} = \theta \pm 90^{\circ}$. In our experiment the angles of interest are two angles, α and β, and their negatives. By assigning $\theta_{A1} = \beta$, $\theta_{B1} = -\beta$, $\theta_{A2} = -\alpha$ and $\theta_{B2} = \alpha$, eq. (L5.4) becomes

$$P\left(\beta, -\beta\right) \leq P\left(-\alpha, \alpha\right) + P\left(\beta, \alpha^{\perp}\right) + P\left(-\alpha^{\perp}, -\beta\right). \qquad \text{(L5.5)}$$

It is convenient to define the quantity H, where

$$H \equiv P\left(\beta, -\beta\right) - P\left(-\alpha, \alpha\right) - P\left(\beta, \alpha^{\perp}\right) - P\left(-\alpha^{\perp}, -\beta\right). \qquad \text{(L5.6)}$$

In terms of H, the Bell-Clauser-Horne inequality is $H \leq 0$.

As described in complement 8.B, if we measure $H \leq 0$, then the data are consistent with local realism. If we measure $H > 0$, then local realism is violated, and we are forced to abandon some of our classical ideas.

L5.2.2 CHSH Test

Here we describe the CHSH inequality, but we will not prove it. For a proof of this inequality, see ref. [L5.6].

The CHSH inequality involves a particular combination of expectation values. Consider the polarization operator defined in chap. 5. The eigenstates and eigenvalues of the polarization operator for linear polarization along the angle θ are

$$\hat{\wp}_\theta\left|\theta\right\rangle=\left(+1\right)\left|\theta\right\rangle,\ \hat{\wp}_\theta\left|\theta^\perp\right\rangle=\left(-1\right)\left|\theta^\perp\right\rangle. \tag{L5.7}$$

The joint polarization operator for Alice and Bob is defined as

$$\hat{\wp}^{AB}_{\theta_A\theta_B}=\hat{\wp}^{A}_{\theta_A}\hat{\wp}^{B}_{\theta_B}. \tag{L5.8}$$

It is traditional to represent the expectation value of this joint polarization operator as

$$E\left(\theta_A,\theta_B\right)\equiv\left\langle\hat{\wp}^{AB}_{\theta_A\theta_B}\right\rangle=P\left(\theta_A,\theta_B\right)+P\left(\theta_A^\perp,\theta_B^\perp\right)-P\left(\theta_A^\perp,\theta_B\right)-P\left(\theta_A,\theta_B^\perp\right), \tag{L5.9}$$

and to define the quantity S as

$$S\equiv E\left(\theta_{A1},\theta_{B1}\right)+E\left(\theta_{A2},\theta_{B1}\right)-E\left(\theta_{A2},\theta_{B2}\right)+E\left(\theta_{A1},\theta_{B2}\right). \tag{L5.10}$$

In a universe that is consistent with local realism, S satisfies the inequality $\left|S\right|\leq 2$, for any choices of the angles.

However, assume that the photons are prepared in the Bell state:

$$\left|\phi^+\right\rangle=\frac{1}{\sqrt{2}}\left(\left|H\right\rangle_A\left|H\right\rangle_B+\left|V\right\rangle_A\left|V\right\rangle_B\right). \tag{L5.11}$$

For this state, and for proper combinations of angles, quantum mechanics predicts $S=2\sqrt{2}$, which yields a maximal violation of the CHSH inequality.

Q1: Prove that the expectation value $E\left(\theta_A,\theta_B\right)$ is given by the combination of probabilities in eq. (L5.9).

Q2: For the state $\left|\phi^+\right\rangle$, and the angles $\theta_{A1}=0^\circ$, $\theta_{B1}=22.5^\circ$, $\theta_{A2}=45^\circ$, and $\theta_{B2}=-22.5^\circ$, show that the quantum mechanical prediction yields $S=2\sqrt{2}$.

L5.3 ALIGNMENT

Note to instructors: To save time in the lab, the alignment described in this section could be done ahead of time. Students would then begin their lab work with sec. L5.4.

- Begin with the pump-beam half-wave plate set so that the pump is vertically polarized when it strikes the downconversion crystal pair. In this orientation, only one of the downconversion crystals is being pumped, and the polarization state of the downconverted photons is that of eq. (L5.1) with $a = 1$, i.e., $|\psi\rangle = |H\rangle_A |H\rangle_B$.
- Following the procedure in lab 1, align the crystal, and detectors **A** and **B**.
- Following the procedure in lab 2, align the polarizing beam splitters, wave plates, and detectors **A′** and **B′**.
- So far, we have been aligning the system by pumping only one of the downconversion crystals. This first crystal is sensitive to tilt in one direction, but not the other. For a vertically polarized pump, the crystal should be sensitive to tilt in the vertical direction, but not in the horizontal direction.
- In order to align the second crystal, rotate the pump-beam wave plate by 45°, which rotates the pump-beam polarization to horizontal. Now the second crystal is being pumped, but not the first. Adjust the horizontal tilt of the crystal pair to maximize the count rates. The second crystal should be sensitive to this tilt, but the alignment of the first crystal will not be affected because it is not sensitive to this tilt. This tilt is the only adjustment you should need to make in order to align the second crystal, and complete the alignment.

In the next section, you will adjust the pump-beam wave plate and the birefringent plate, in order to create the proper polarization-entangled state.

L5.4 CREATING THE BELL STATE

- Run the LabVIEW program “Angle_scan.vi”.

Documentation for this program comes with the software. It starts by initializing the counters and the motors which control the wave-plate rotation stages; this takes a few seconds and the **Status** indicator reads “Initializing.” Once everything initializes, the **Status** should switch to “Reading Counters.” The program is now reading the counters and updating the screen in real time.

- Make sure that **Update Period** is set to somewhere between 0.2s and 1s. Set the **Subtract Accidentals?** switch to **Yes**; check with your instructor about what coincidence time resolutions you should use.

As you learned in lab 1 (sec. L1.7), there are always some background “accidental” coincidences that are detected, even when you don’t expect to get any. This is due to the randomness of the photon emission process. By knowing the count rates and coincidence time resolution, we can calculate how may accidental coincidences we would expect to get, and subtract these accidentals from the raw count rates to correct the data. For now we’ll subtract them, but later you can explore what happens if you don’t.

When the program is done initializing, the **A** and **B** wave plate angles are set to zero (double check that the **A Position** and **B Position** parameters read 0). With these wave plates settings the **A** and **B** detectors monitor horizontally polarized photons coming from the source, and the **A′** and **B′** detectors monitor vertically polarized photons. Recall that the state of the downconverted photons is given by eq. (L5.1), with the parameters a and φ determined by the settings of the half-wave plate and the birefringent plate in the pump beam.

- Adjust the pump-beam half-wave plate to roughly equalize the **AB** coincidences and **A′B′** coincidences. Given that the source produces photons in the state of eq. (L5.1), and the wave-plate axes are set to 0°, you should notice that there are **AB** and **A′B′** coincidences, but essentially no **A′B** or **AB′** coincidences (there are always a few due to experimental imperfections).

Q3: Why is this? Calculate the probability of an **A′B** or **AB′** coincidence given the state in Eq. (L5.1).

- If you do notice significant **A′B** or **AB′** coincidences, it probably means that the **A** and/or **B** wave plates are not properly zeroed. Enter 0 for the **A Desired Position** and the **B Desired Position** parameters, then push the **Move Motors** button. The wave plates should rotate to 0: **A Position** and **B Position** should read 0. Note, however, that the **A Motor Position** and **B Motor Position** displays will not necessarily read zero; they will read the values given in the **A zero** and **B zero** parameters. These parameters are needed because the 0 angles of the motors are not necessarily perfectly aligned with the 0 angles of the wave plates.

  Slightly adjust (in about 1° increments) the rotation angles of the wave plates by entering values into **A Desired Position** and the **B Desired Position** parameters, then push the **Move Motors** button. Once you've minimized the **A′B** and **AB′** coincidences, note the readings in the **A Motor Position** and **B Motor Position**—these are the correct 0 readings, and you should enter them as the **A zero** and **B zero** parameters.
- Once the correct **A zero** and **B zero** parameters have been entered, the program must be stopped and then restarted in order to recognize the new values. Write down these correct values, because any time you quit and restart LabVIEW you may need to reenter them.
- Rotate the wave plate in the pump beam by about 10°, then wait a few seconds for the computer to catch up with readings at this new setting.

Notice that no matter how you set the polarization of the pump beam, you can change the ratio of the **AB** and **A′B′** coincidences, but you should never produce any significant **A′B** or **AB′** coincidences.

- Adjust the wave plate in the pump beam so that the **AB** and **A′B′** coincidence rates are roughly the same.

- Enter 0 for the **A Desired Position** and 45 for the **B Desired Position** parameters, then push the **Move Motors** button.

Q4: What happens to the coincidence rates? Explain this.

- You may want to adjust the **Update Period**. If it is too short the counts will fluctuate a lot, and it will be difficult to get a good reading. If it is too long you need to adjust things very slowly, and wait for the screen to catch up. Values between 0.2 and 1.0 s should work, depending on your count rates. You'll also need to adjust the full scale reading on your meters.
- Set the **A** and **B** wave plates to 0°. Adjust the pump-beam half-wave plate so that the ratio of the **AB** and **A′B′** coincidences is roughly 1:1.
- Now use the motors to set the **A** and **B** wave plates to 22.5°. Adjust the tilt of the birefringent plate in the pump beam (NOT the pump-beam half-wave plate) to minimize the **A′B** and **AB′** coincidences. You won't be able to get these coincidences to be as low as with the wave plates set to zero, but you should be able to get them fairly low.
- Iterate back-and-forth between the last two steps. With the **A** and **B** wave plates set to 0°, adjust the ratio of the **AB** and **A′B′** coincidences to be equal using the pump-beam half-wave plate; with the **A** and **B** wave plates set to 22.5°, minimize the **A′B** and **AB′** coincidences with the tilt of the birefringent plate. You should notice that even with the **A** and **B** wave plates set to 22.5°, the ratio of the **AB** and **A′B′** coincidences should still be roughly 1:1.

Now the state of your downconverted photons should be given approximately by the Bell state $\left|\phi^+\right\rangle$ of eq. (L5.11). Consider how we know this:

Q5: With the **A** and **B** wave plates set to 0°, detectors **A** and **B** are measuring horizontally polarized photons from the source, and detectors **A′** and **B′** are measuring vertically polarized photons from the source. If the **AB** and **A′B′** coincidences are equal, what do we know about the parameter a in eq. (L5.1)? Do we know anything yet about the parameter φ? Write down the state produced by the source, assuming φ to be unknown.

Measurements with the wave plates set to 0° determines the parameter a, but not φ. In order to determine φ, you need to use the results of your measurements with the wave plates set to 22.5°.

Q6: Given the state you wrote down in the last question, what must φ be in order to explain the fact that the probability of an **A′B** (or an **AB′**) coincidence is 0 with the **A** and **B** wave plates set to 22.5°? (Ignore your experimental inability to make this coincidence rate perfectly 0.) Write down the state produced by the source.

Q7: Given the state you just determined, calculate the joint probability $P(\theta_A, \theta_B)$ that Alice will measure her photon to be polarized along θ_A, and Bob will find his photon polarized along θ_B.

L5.5 EXPLORING QUANTUM CORRELATIONS—ENTANGLED STATES AND MIXED STATES

Before you actually try to test local realism, you'll first explore some of the interesting correlations that allow quantum mechanics to violate it.

Now you're ready to scan one of the wave-plate angles, and measure the joint probability $P(\theta_A, \theta_B)$. The program "Angle_scan.vi" is designed to fix θ_A, and scan θ_B over a preset range of angles. **NOTE: the computer scans (and records in a datafile) a *wave-plate* angle, whereas when we talk about $P(\theta_A, \theta_B)$, the angles θ_A and θ_B refer to the angles of a *polarizer*. Remember that polarizer angles are twice the wave plate angles, because the angle of the output polarization from a wave plate rotates twice as fast as the rotation angle of the wave plate.**

- In the **Data Taking Parameters** section, set **A** to 0 (this is the fixed wave-plate angle), then take a scan with 5 samples per point with counting times of 3–5 s per sample. This data file will automatically be saved. Note that the computer acquires data at 17 values of the **B** wave-plate angle between 0° and 90°.
- Repeat this experiment with **Angle A** set to 22.5° (corresponding to a polarizer angle of 45°).

In your lab report you should create two graphs. The first is theory and experiment for $P(\theta_A = 0^{\circ}, \theta_B)$, and the second for $P(\theta_A = 45^{\circ}, \theta_B)$. Plot the theory as a solid line and the data as points. For the theory curves, use the probabilities you obtained in Q7.

In chapter 8, we talked about the difference between an entangled state and a mixed state. Equation (L5.11) assumes an entangled state—in other words, that at any given time the photons are in both the states $|H\rangle_A |H\rangle_B$ and $|V\rangle_A |V\rangle_B$. Is this assumption correct? Can we explain our data instead assuming that the photons are in a classical mixture of either the $|H\rangle_A |H\rangle_B$ or $|V\rangle_A |V\rangle_B$ states?

Q8: Calculate the probability $P(\theta_A, \theta_B | H_A, H_B)$—the joint probability that Alice will measure her photon to be polarized along θ_A and Bob will find his photon polarized along θ_B, assuming that the photons are in the state $|H\rangle_A |H\rangle_B$.

Q9: Calculate the probability $P(\theta_A, \theta_B | V_A, V_B)$.

Q10: If you refer back to sec. 8.3, you'll see that the probability of joint polarization measurements in a mixed state is

$$P_{mix}(\theta_A,\theta_B) = P(\theta_A,\theta_B|H_A,H_B)P(H_A,H_B) + P(\theta_A,\theta_B|V_A,V_B)P(V_A,V_B), \quad \text{(L5.12)}$$

where $P(H_A,H_B)$ is the probability that the photons are produced in the state $|H\rangle_A|H\rangle_B$, and similarly for $P(V_A,V_B)$. These probabilities are both 1/2 here. Calculate $P_{mix}(\theta_A,\theta_B)$.

On your two graphs of $P(\theta_A,\theta_B)$, add graphs of $P_{mix}(\theta_A,\theta_B)$, for appropriate values of θ_A; plot $P_{mix}(\theta_A,\theta_B)$ as a dashed line.

Q11: Is it possible to explain your experimental data using this mixed state? If not with this mixed state, can you think of any mixed state that will agree with both of your data sets? By this I mean, are there any $P(H_A,H_B)$ and $P(V_A,V_B)$ that will allow $P_{mix}(\theta_A,\theta_B)$ to agree with both data sets?

Remember, if the data are consistent with an entangled state, we must conclude that although the polarizations of the two photons are perfectly correlated with each other, neither photon is in a well-defined state before a measurement.

L5.6 TESTING THE CHSH INEQUALITY

The Bell state $|\phi^+\rangle$ you've created is the ideal state to test the CHSH inequality, so let's perform this test.

- Close "Angle_scan.vi," as it cannot be in memory while running "Hardy-Bell.vi."
- Open the LabVIEW program "Hardy-Bell.vi", set the **A zero** and **B zero** parameters to the values you previously determined, then run the program.
- Set the **Experimental Setup** dial to **S**, and **Update Period** to something between 0.2 and 1.0s. Set the **Subtract Accidentals?** switch to **Yes**.
- Double check that you're in the Bell state $|\phi^+\rangle$. The **AB** and **A′B′** coincidences should be roughly equal, and these coincidences should be maximized while the **A′B** and **AB′** coincidences are minimized. This should be true with the **A** and **B** wave plates at both 0° and 22.5°. A good figure of merit for this is the "E-meter"—the big blue bar on the right. The E-meter reads the expectation value $E(\theta_A,\theta_B)$ of eq. (L5.9), which is 1 if **A′B** and **AB′** are 0. Your goal is to maximize $E(\theta_A,\theta_B)$ at both wave plate settings.
- In the **Data Taking Parameters** box set **Update Period (Data Run)** to 5 s, and **No. of Samples** to 10, then push the **Take Data** button.

Control of the computer is now switched to the data acquisition program. This program requires nothing from you; it automatically adjusts the wave plates to the correct angles (corresponding to the polarizer angles given in Q2 above), makes readings, calculates expectation values and S, and saves the data to a file. For the parameters you just entered the data run will take approximately 4 min. There is no graceful way to exit this program while it is still running, and if you exit in the middle by closing the window, chances are you'll need to reboot the computer—better to just let it run.

- The program is done running when the **Operation** box reads "Finished." The data file is automatically named according to the date and time. In your notebook record the filename, important parameters (**Update Period**, **No. of Samples**), and results (S, errors, etc.). The **Violations** parameter gives the number of standard deviations by which your result violates local realism.

Try and get a result that violates local realism by at least 10 standard deviations. If your value for S is greater than 2, but you don't have a 10 standard deviation violation, use more than 5 s per point to decrease the error (the error given is the standard deviation of the **No. of Samples** measurements of S). If $S < 2$ you probably need to tweak the state using the wave plate and the birefringent plate in the pump beam. You shouldn't have too much difficulty getting an S value of at least 2.3.

L5.7 MEASURING *H*

Now you're ready to perform Hardy's test of local realism.

As discussed above, you'll be measuring the quantity H, which depends on several probabilities, all determined by the parameters α and β. Maximum violation of local realism can be achieved using either of the states

$$\left|\psi_1\right\rangle = \sqrt{0.2}\left|H\right\rangle_A\left|H\right\rangle_B + \sqrt{0.8}\left|V\right\rangle_A\left|V\right\rangle_B, \tag{L5.13}$$

$$\left|\psi_2\right\rangle = \sqrt{0.8}\left|H\right\rangle_A\left|H\right\rangle_B + \sqrt{0.2}\left|V\right\rangle_A\left|V\right\rangle_B. \tag{L5.14}$$

We'll start with $\left|\psi_1\right\rangle$, for which the angle parameters are $\alpha = 35^{\mathrm{o}}$ and $\beta = 19^{\mathrm{o}}$. Of course, experimentally it is difficult to produce exactly the state $\left|\psi_1\right\rangle$, so optimal violation may occur for slightly different values of α and β. You'll begin by attempting to produce state $\left|\psi_1\right\rangle$, and assuming $\alpha = 35^{\mathrm{o}}$ and $\beta = 19^{\mathrm{o}}$. The magnitudes of the amplitudes of the states in eqs. (L5.13) and (L5.14) are set by monitoring the coincidence probabilities with the **A** and **B** half-wave plates set at 0°. The relative phase of the states is adjusted by attempting to ensure that $P(-\alpha,\alpha) = 0$.

- Run the LabVIEW program "Hardy-Bell.vi."
- Make sure the **Experimental Setup** dial is set to **H**, and that **Update Period** is set to between 0.2 and 1.0s. Set the **Subtract Accidentals?** switch to **Yes**.
- Make sure that **Alpha** is set to 35° and **Beta** is set to 19°. You'll notice that in the lower right hand portion of the screen, **H HWP Measurement Angles** are displayed. These are the angles that the half-wave plates will need to be set to, in order to measure the four probabilities that comprise H [eq. (L5.6)]; they are determined from the **Alpha** and **Beta** parameters entered on the left of the screen (remember that the polarization rotates through an angle 2θ when the wave plate rotates by θ).

- Set the **A** and **B** wave plates to 0°. Adjust the pump-beam wave plate so that the ratio of the **AB** and **A′B′** coincidences is roughly 1:4. This is most easily done by watching the **P Meter**, which reads the probability of an **AB** coincidence. You would like it to read 0.2.
- Set the **A** and **B** wave plates so that you are measuring $P(-\alpha,\alpha)$. Adjust the tilt of the birefringent plate to minimize this probability.
- Iterate back-and-forth between the last two steps. With the **A** and **B** wave plates set to 0° adjust the ratio of the coincidences using the pump-beam wave plate; with the wave plates set to measure $P(-\alpha,\alpha)$, minimize this probability with the tilt of the birefringent plate. When you've got everything adjusted fairly well, set **Update Period** to at least 1.0 s, to get better statistics.
- Set your wave plates to measure $P(\beta,\alpha^{\perp})$ and $P(-\alpha^{\perp},-\beta)$. These probabilities should be fairly small. Set your wave plate to measure $P(\beta,-\beta)$; this probability should be larger than the others.

By now the pump-beam wave plate and the birefringent plate should be reasonably well adjusted to produce the state $|\psi_1\rangle$. You're ready to take a data run which measures *H*.

- In the **Data Taking Parameters** box set **Update Period (Data Run)** to 10.0 s, and **No. Of Samples** to 5, then push the **Take Data** button.

Control of the computer is now switched to the data acquisition program. This program requires nothing from you, it automatically adjusts the wave plates to the correct angles, makes readings, calculates probabilities and *H*, and saves the data to a file.

- The program is done running when the **Operation** box reads "Finished." The data file is automatically named according to the date and time. In your notebook record the filename, important parameters (**Alpha**, **Beta**, **Update Period**, angle of the pump wave plate, etc.), and results (*H*, probabilities, errors, etc.).
- Once you have written down all of these parameters, you can close the window of the data recording program.

L5.8 OPTIMIZING YOUR RESULTS

How do your data look? Chances are you measured a value for *H* that was less than 0; or maybe it was greater than 0, but not by very much (the **Violations** result tells you by how many standard deviations your value of *H* exceeds 0). You'd really like to see values of 0.02 or less for the three probabilities that you expect to be 0, and you'd like to see 10 or more violations. This would be a very convincing result.

You can increase the number of violations you get by either increasing *H*, or decreasing the error. At the same time, you'd like to make sure that you stay below 0.02 or 0.03 for the probabilities you expect to be 0. For a reasonable error measurement you should be using at least 10 for the **No. of Samples** parameter (we only used 5 for the first data set because we wanted to get a quick run). Your final data runs should always include at least 10 samples. Increasing the **No. of Samples** parameter will not decrease the error, it will only make the error measurement more accurate (the error is the standard deviation of the **No. of Samples** measurements of *H*). The only good way to decrease the error of your measurement is to increase the **Update Period (Data Run)** parameter, which increases the time for a data run.

In order to increase *H*, you need to increase $P(\beta,-\beta)$, while trying to keep all the rest of the probabilities on the order of 0.02. How do you do this? Start by using the results of your lab ticket. Assuming the value for α is unchanged, how do you adjust β in order to increase *H*? Remember that you need to keep three of your probabilities down around 0.02.

- Rerun the program using a new value for **Beta**.
- Double check the alignment of the pump-beam wave plate and the birefringent plate. With the **A** and **B** wave plates set to 0° adjust the ratio of the coincidences using the pump wave plate; with the wave plates set to measure $P(-\alpha,\alpha)$, minimize this probability with the tilt of the birefringent plate.
- Take another data run.

Keep adjusting your parameters, and retaking data, until you get at least a 10 standard deviation violation of local realism, with 3 of your probabilities as low as reasonably possible (a few percent). This data run needs to use at least 10 samples.

L5.8.1 Optimization Hints

Don't stress too much about getting $P(-\alpha,\alpha)$ super low by tweaking the pump-beam wave plate and the birefringent plate—it should be down near 0.02 or 0.03, but it's been my experience that this is the most difficult of the probabilities to get very low. I've found that the main parameter to adjust is **Beta**, while adjustments of **Alpha** will help some as well.

Once you've got your value for *H* up to 0.05 or above, the best way to increase your number of violations is to decrease your error by taking longer data runs with an increased **Update Period (Data Run).**

I know it can be tedious making small adjustments to parameters, and waiting 10 minutes or so for a data run to complete. Having the computer finish taking data, and then simply spit out a value for *H* can be somewhat anti-climactic. However, try not to lose sight of the big picture. Remember the argument in chap. 8 that a value of $H > 0$ means that local realism is violated. When you're all done, you'll have proven that classical mechanics doesn't always work!

For all of the above experiments you've been subtracting the accidental coincidences. Once you've gotten a convincing result, turn off the accidentals subtraction, but leave everything else the same. How does this affect your results?

L5.9 LAST EXPERIMENT

Redo what you've done above using the state $|\psi_2\rangle$ of eq. (L5.14).

Note that this is a different state, so it will require different values for α and β. Remember: don't just change α and β—**you have to adjust the pump-beam wave plate and the birefringent plate to change the state.** Think about how you'll have to adapt the procedure described above to create the state $|\psi_2\rangle$.

Q12: For the state $|\psi_2\rangle$, what values of α and β will yield 0 for $P(-\alpha,\alpha)$, $P(\beta,\alpha^{\perp})$, and $P(-\alpha^{\perp},-\beta)$? Start by finding the value for α by looking at $P(-\alpha,\alpha)$, then find β. You may find it useful to redo the lab ticket using the state $|\psi_2\rangle$, and these new values of α and β.

L5.10 References

[L5.1] L. Hardy, "Nonlocality for two particles without inequalities for almost all entangled states," Phys. Rev. Lett. **71**, 1665 (1993).

[L5.2] N. D. Mermin, "Quantum mysteries refined," Am. J. Phys. **62**, 880 (1994).

[L5.3] A. G. White et al., "Nonmaximally entangled states: production, characterization, and utilization," Phys. Rev. Lett. **83**, 3103 (1999).

[L5.4] P. G. Kwiat and L. Hardy, "The mystery of the quantum cakes," Am. J. Phys. **68**, 33 (2000).

[L5.5] J. A. Carlson, M. D. Olmstead, and M. Beck, "Quantum mysteries tested: an experiment implementing Hardy's test of local realism," Am. J. Phys. **74**, 180 (2006).

[L5.6] D. Dehlinger and M. W. Mitchell, "Entangled photon apparatus for the undergraduate laboratory," Am. J. Phys. **70**, 898 (2002); D. Dehlinger and M. W. Mitchell, "Entangled photons, nonlocality, and Bell inequalities in the undergraduate laboratory," Am. J. Phys. **70**, 903 (2002).

This page intentionally left blank

Index