Quantum Mechanics: An Introduction
Quantum Mechanics: An Introduction
# Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>1 • Introduction</th>
<th>2 • Two State Systems: The Ammonia Molecule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>WHAT YOU WILL LEARN IN THIS BOOK</td>
<td>2.1 INTRODUCTION AND MOTIVATION</td>
</tr>
<tr>
<td>1.2</td>
<td>WHAT YOU SHOULD KNOW ABOUT LINEAR ALGEBRA</td>
<td>2.2 THE AMMONIA MOLECULE AS A TWO STATE SYSTEM</td>
</tr>
<tr>
<td>1.3</td>
<td>THE ESSENTIALS OF PROBABILITY THEORY</td>
<td>2.3 PHYSICAL QUANTITIES AS MATRICES</td>
</tr>
<tr>
<td>1.4</td>
<td>PHILOSOPHICAL INTRODUCTION</td>
<td>2.4 OPERATIONS AS MATRICES</td>
</tr>
<tr>
<td>1.4.1</td>
<td>The Essentials of Quantum Mechanics</td>
<td>2.5 THE EIGENVALUES OF A MATRIX</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Unhappening</td>
<td></td>
</tr>
</tbody>
</table>
## Contents

2.6 **THE BORN RULE AND UNCERTAINTY** 38  
2.6.1 Born’s Rule in Pictures 40  
2.7 **WHICH OPERATORS ARE DIAGONALIZABLE?** 42  
2.7.1 Vectors, Functions, and Diagonal Matrices 44  
2.8 **QUANTUM DYNAMICS OF THE AMMONIA MOLECULE** 45  
2.9 **SUMMARY** 52  
2.10 **SKIP THIS UNLESS YOU ARE INTERESTED IN PHILOSOPHY** 53  
2.11 **FOUR STATE SYSTEMS** 56  
2.12 **FURTHER EXERCISES** 60  

**Chapter 3**  
Quantum Mechanics of Single Particle in One-Dimensional Space I 63  

3.1 **INTRODUCTION** 63  
3.2 **TRANSLATION INVARIANCE AND FREE PARTICLE MOTION** 64  
3.2.1 No Experimental Definition of Infinity 64  
3.2.2 Free Particles on an Infinite Continuous Line 68  
3.2.3 Primer on the Fourier Transform and Fourier Series 77  
3.3 **LAGRANGIAN AND HAMILTONIAN MECHANICS: A REMINDER OR A PRIMER** 79  
3.4 **FURTHER EXERCISES** 83  

**Chapter 4**  
Quantum Mechanics of a Single Particle in One-Dimensional Space II 89  

4.1 **INTRODUCTION** 89  
4.2 **A MORE MATHEMATICAL DESCRIPTION OF ONE-DIMENSIONAL MOTION** 89  
4.3 **CONTINUUM OR DELTA FUNCTION NORMALIZATION** 93  
4.4 **TIME EVOLUTION OF A FREE PARTICLE WAVE FUNCTION** 96  
4.5 **THE DOUBLE SLIT EXPERIMENT** 99  
4.6 **A WORKED EXERCISE** 102  
4.7 **FURTHER EXERCISES** 107
## Contents

**Chapter 5**  
The Harmonic Oscillator  

5.1 INTRODUCTION  
5.2 QUANTIZING THE SIMPLE HARMONIC OSCILLATOR  
5.3 QUANTIZATION OF FIELDS AND WAVE–PARTICLE DUALITY  
5.4 FERMI–DIRAC STATISTICS  
5.5 EXERCISES  

**Chapter 6**  
Review of Linear Algebra and Dirac Notation  

6.1 INTRODUCTION  
6.2 SEPARABLE HILBERT SPACES  
6.3 UNITARY TRANSFORMATIONS, UNITARY MATRICES, AND CHANGES OF BASIS  
6.4 NORMAL OPERATORS ARE DIAGONALIZABLE OPERATORS  
6.5 CONTINUUM EIGENVALUES AND UNBOUNDED OPERATORS  
6.6 SUMMARY  
6.7 DIRECT SUMS AND TENSOR PRODUCTS  
6.8 THE GENERALIZED UNCERTAINTY RELATION  
6.9 SYMMETRIES AND CONSERVATION LAWS  
   6.9.1 Dynamical Symmetry Groups  
   6.9.2 Projective Representations of Symmetry Groups  
   6.9.3 Examples  
6.10 EXERCISES  

**Chapter 7**  
Rotation Invariance and the Hydrogen Atom  

7.1 INTRODUCTION  
7.2 UNITS AND SYMMETRIES  
7.3 IRREDUCIBLE REPRESENTATIONS OF THE COMMUTATION RELATIONS  
7.4 ADDITION OF ANGULAR MOMENTA  
7.5 THE HAMILTONIAN OF SPHERICALLY SYMMETRIC POTENTIALS
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.6</td>
<td>THE PERIODIC TABLE</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>7.7</td>
<td>THE SPIN OF THE ELECTRON</td>
<td>186</td>
<td></td>
</tr>
<tr>
<td>7.8</td>
<td>SPIN PRECESSION AND SPIN RESONANCE</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>7.9</td>
<td>STERN–GERLACH EXPERIMENTS</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>7.10</td>
<td>EXERCISES</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>Chapter 8</td>
<td>Scattering Electrons on a Nucleus</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>INTRODUCTION</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>POSITIVE ENERGY EIGENFUNCTIONS AND SCATTERING AMPLITUDES</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>8.3</td>
<td>ANALYSIS OF THE COULOMB SCATTERING AMPLITUDES</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>8.4</td>
<td>EXERCISES</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Charged Particles in a Magnetic Field</td>
<td>207</td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>INTRODUCTION</td>
<td>207</td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>THE LORENTZ FORCE AND LANDAU’S SOLUTION</td>
<td>207</td>
<td></td>
</tr>
<tr>
<td>9.3</td>
<td>EXERCISES</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>Chapter 10</td>
<td>The Meaning of Quantum Measurement and the Emergence of a Classical Reality</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>10.1</td>
<td>INTRODUCTION</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>COUNTING STATES AND COLLECTIVE COORDINATES</td>
<td>216</td>
<td></td>
</tr>
<tr>
<td>10.3</td>
<td>COLLECTIVE COORDINATES</td>
<td>217</td>
<td></td>
</tr>
<tr>
<td>10.4</td>
<td>INTERFERENCE OF CLASSICAL HISTORIES</td>
<td>220</td>
<td></td>
</tr>
<tr>
<td>10.5</td>
<td>SCHRÖDINGER’S CAT AND SCHRÖDINGER’S BOMB</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>10.6</td>
<td>THE EINSTEIN–PODOLSKY–ROSEN (EPR) PARADOX, BELL’S THEOREM, AND OTHER ODDITIES</td>
<td>228</td>
<td></td>
</tr>
<tr>
<td>10.7</td>
<td>THE RELATION BETWEEN THEORY AND EXPERIMENT IN QM</td>
<td>234</td>
<td></td>
</tr>
<tr>
<td>10.8</td>
<td>EXERCISES</td>
<td>236</td>
<td></td>
</tr>
</tbody>
</table>
### Chapter 11: Sketch of Atomic, Molecular, and Condensed Matter Physics

11.1 INTRODUCTION
11.2 THE BORN–OPPENHEIMER APPROXIMATION
11.3 COLLECTIVE COORDINATES, ROTATIONAL LEVELS, AND PHONONS
   11.3.1 Water
   11.3.2 Phonons
11.4 THE HYDROGEN MOLECULAR ION
11.5 DENSITY FUNCTIONAL THEORY
11.6 ELEMENTS OF THE THEORY OF CRYSTALLINE SOLIDS
11.7 BAND STRUCTURE
   11.7.1 A Simple Model
11.8 THE FERMI LIQUID THEORY OF CONDUCTORS
11.9 EXERCISES

### Chapter 12: Quantum Statistical Mechanics

12.1 INTRODUCTION
12.2 QUANTUM FIELD THEORY OF FERMIONS
12.3 STATISTICAL MECHANICS OF BOSONS AND FERMIONS
12.4 THE PARTITION FUNCTION
12.5 THE LOW TEMPERATURE LIMIT
12.6 STATISTICAL MECHANICS OF A FERMI LIQUID
12.7 PLANCK’S RADIATION LAW
12.8 EXERCISES

### Chapter 13: Perturbation Theory: Time Independent

13.1 INTRODUCTION
13.2 BRILLOUIN–WIGNER PERTURBATION THEORY FOR NONDEGENERATE LEVELS
   13.2.1 Relation to Rayleigh–Schrödinger Perturbation Theory
<table>
<thead>
<tr>
<th>Chapter</th>
<th>13.3 DEGENERATE PERTURBATION THEORY</th>
<th>285</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13.4 THE FEYNMAN–HELLMANN THEOREM</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td>13.5 EXAMPLES</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td>13.5.1 The Stark Effect</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td>13.5.2 The Zeeman Effect</td>
<td>289</td>
</tr>
<tr>
<td></td>
<td>13.5.3 Fine Structure and the Weak-Field Zeeman Effect</td>
<td>290</td>
</tr>
<tr>
<td></td>
<td>13.5.4 Coulomb Expectation Values of Powers of $R$</td>
<td>292</td>
</tr>
<tr>
<td></td>
<td>13.5.5 A Three-Dimensional Example</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td>13.5.6 Degenerate Perturbation Theory in a Macroscopic System</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>13.6 EXERCISES</td>
<td>298</td>
</tr>
</tbody>
</table>

**Chapter 14** Perturbation Theory: Time Dependent

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1</td>
<td>INTRODUCTION</td>
</tr>
<tr>
<td>14.2</td>
<td>DYSON'S FORMULA</td>
</tr>
<tr>
<td>14.3</td>
<td>THE DIRAC PICTURE</td>
</tr>
<tr>
<td>14.4</td>
<td>TRANSITION AMPLITUDES</td>
</tr>
<tr>
<td>14.5</td>
<td>ELECTROMAGNETIC RADIATION FROM ATOMS</td>
</tr>
<tr>
<td>14.6</td>
<td>INCOHERENT PERTURBATIONS AND RADIATIVE DECAY</td>
</tr>
<tr>
<td>14.7</td>
<td>SELECTION RULES</td>
</tr>
<tr>
<td>14.8</td>
<td>EXERCISES</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Chapter 15** The Adiabatic Approximation, Aharonov-Bohm, and Berry Phases

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.1</td>
<td>INTRODUCTION</td>
</tr>
<tr>
<td>15.2</td>
<td>ADIABATIC ENERGY EIGENSTATES</td>
</tr>
<tr>
<td>15.3</td>
<td>THE BERRY PHASE</td>
</tr>
<tr>
<td>15.4</td>
<td>THE AHARONOV–BOHM EFFECT</td>
</tr>
<tr>
<td>15.5</td>
<td>ANYONS, FERMIONS, AND THE SPIN STATISTICS THEOREM</td>
</tr>
<tr>
<td>15.6</td>
<td>EXERCISES</td>
</tr>
</tbody>
</table>
CHAPTER 16 • Scattering Theory

16.1 INTRODUCTION
16.2 GENERAL FORMALISM
16.3 POTENTIAL SCATTERING
16.4 THE BORN APPROXIMATION
16.5 PHASE SHIFT ANALYSIS
   16.5.1 The Effective Range Approximation
16.6 RESONANCES
16.7 A PARTIALLY WORKED EXERCISE: THE δ SHELL POTENTIAL
16.8 EXERCISES

CHAPTER 17 • The JWKB Approximation

17.1 INTRODUCTION
17.2 THE JWKB EXPANSION
17.3 THE JWKB APPROXIMATION IN ONE DIMENSION
17.4 COMMENTS ON THE SOLUTIONS
17.5 THE JWKB APPROXIMATION FOR THE PROPAGATOR
17.6 THE JWKB APPROXIMATION FOR ENERGY LEVELS
17.7 THE JWKB APPROXIMATION TO THE WAVE FUNCTIONS OF EIGENSTATES
   17.7.1 Examples
17.8 THE DECAY OF METASTABLE STATES
17.9 MORE EXAMPLES
17.10 THE JWKB APPROXIMATION FOR PHASE SHIFTS
17.11 EXERCISES

CHAPTER 18 • The Variational Principle

18.1 INTRODUCTION
18.2 GENERAL PROPERTIES
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>EXAMPLES</td>
<td>362</td>
</tr>
<tr>
<td>18.4</td>
<td>THE HARTREE AND HARTREE–FOCK APPROXIMATIONS</td>
<td>365</td>
</tr>
<tr>
<td>18.5</td>
<td>THE LANCZOS METHOD</td>
<td>370</td>
</tr>
<tr>
<td>18.6</td>
<td>EXERCISES</td>
<td>372</td>
</tr>
<tr>
<td>19</td>
<td>The Feynman Path Integral</td>
<td>375</td>
</tr>
<tr>
<td>19.1</td>
<td>INTRODUCTION</td>
<td>375</td>
</tr>
<tr>
<td>19.2</td>
<td>TWO DERIVATIONS OF THE PATH INTEGRAL FORMULA</td>
<td>375</td>
</tr>
<tr>
<td>19.3</td>
<td>THE PATH INTEGRAL FOR A HARMONIC OSCILLATOR</td>
<td>380</td>
</tr>
<tr>
<td>19.4</td>
<td>MORE GENERAL POTENTIALS</td>
<td>385</td>
</tr>
<tr>
<td>19.5</td>
<td>PATH INTEGRALS AT FINITE TEMPERATURE</td>
<td>386</td>
</tr>
<tr>
<td>19.6</td>
<td>PATH INTEGRALS AND THE JWKB APPROXIMATION</td>
<td>388</td>
</tr>
<tr>
<td>19.7</td>
<td>PATH INTEGRALS FOR SPIN AND OTHER DISCRETE VARIABLES</td>
<td>389</td>
</tr>
<tr>
<td>19.8</td>
<td>FERMIONS AND GRASSMANN INTEGRATION</td>
<td>391</td>
</tr>
<tr>
<td>19.9</td>
<td>FURTHER EXERCISES</td>
<td>396</td>
</tr>
<tr>
<td>20</td>
<td>Quantum Computation?</td>
<td>399</td>
</tr>
<tr>
<td>20.1</td>
<td>INTRODUCTION</td>
<td>399</td>
</tr>
<tr>
<td>20.2</td>
<td>QUANTUM INFORMATION</td>
<td>400</td>
</tr>
<tr>
<td>20.3</td>
<td>PAGE’S THEOREM REDUX, MONOGAMY, AND CLONING</td>
<td>402</td>
</tr>
<tr>
<td>20.4</td>
<td>QUANTUM KEY DISTRIBUTION</td>
<td>404</td>
</tr>
<tr>
<td>20.5</td>
<td>QUANTUM TELEPORTATION</td>
<td>406</td>
</tr>
<tr>
<td>20.6</td>
<td>GATES FOR QUANTUM COMPUTERS</td>
<td>408</td>
</tr>
<tr>
<td>20.7</td>
<td>COMPUTATIONAL COMPLEXITY</td>
<td>410</td>
</tr>
<tr>
<td>20.8</td>
<td>CAN WE BUILD A QUANTUM COMPUTER?</td>
<td>413</td>
</tr>
<tr>
<td>20.9</td>
<td>EXERCISES</td>
<td>414</td>
</tr>
<tr>
<td>21</td>
<td>L’Envoi: Relativistic Quantum Field Theory</td>
<td>417</td>
</tr>
<tr>
<td>Appendix</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>A</td>
<td>Interpretations of Quantum Mechanics</td>
<td>423</td>
</tr>
<tr>
<td>B</td>
<td>The Dirac Delta Function</td>
<td>447</td>
</tr>
<tr>
<td>C</td>
<td>Noether's Theorem</td>
<td>449</td>
</tr>
<tr>
<td>D</td>
<td>Group Theory</td>
<td>453</td>
</tr>
<tr>
<td>E</td>
<td>Laguerre Polynomials</td>
<td>461</td>
</tr>
<tr>
<td>F</td>
<td>Summary of Dirac Notation and Linear Algebra</td>
<td>463</td>
</tr>
<tr>
<td>G</td>
<td>Answers to Selected Problems</td>
<td>469</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>523</td>
</tr>
<tr>
<td></td>
<td>Index</td>
<td>535</td>
</tr>
</tbody>
</table>
Quantum mechanics is at the basis of most physical science and modern electronic technology, and has increasing relevance to the biological sciences. It’s also the most confusing subject in the world, because it seems to deny the very foundations of logic. Logic is a precise distillation of our intuitive grasp of how things work. It seems to have nothing to do with particular physical situations where we’ve become used to the fact that our intuition is only an approximation. For example, Galileo appreciated that Aristotle’s intuitive notion that rest is the natural state of bodies was wrong, and invented the relativity of frames moving at uniform velocity. Einstein realized that Galileo’s laws relating the kinematics of two frames moving at uniform velocity were only approximately correct, valid when the velocity was much smaller than that of light. All of this is a bit confusing when you first encounter it, but it’s not actually that mind boggling, and once you understand how the correct formulas reduce to the non-relativistic ones (which are straightforward and make intuitive sense) when the velocity is small compared to that of light, it actually can be pretty easy to come to terms with relativity, and even develop an intuition for it. The goal of this text is to help the reader develop a similar understanding of Quantum Mechanics (QM).

While there are many available textbooks on quantum mechanics, they almost uniformly present certain aspects of the subject in a manner that reflects the confusions encountered by inventors of quantum mechanics. These confusions include the meaning of wave–particle duality and the correct interpretation of measurements. *Quantum Mechanics: An Introduction* presents the subject from a modern perspective. It includes an elementary discussion of field quantization, the only proper way to understand wave particle duality, at a very early stage. The essentials of field quantization are not difficult, because fields are just collections of simple harmonic oscillators (the standard example used in elementary texts). The interpretation of particles as excitations of quantized fields is the only way to understand the identity of particles and the peculiar statistical properties of multi-particle states observed in the real world. All extant textbooks either give incorrect explanations of the origin of Bose and Fermi statistics, or introduce these laws as an additional postulate. On the other hand, all working quantum theorists know that the statistics of identical particles is a consequence of quantum field theory.
Field theory is introduced in Chapter 5, and used to simplify the discussion of quantum statistical mechanics in Chapter 12. It’s invoked in Chapter 11 as well, in our brief discussion of density functional theory.

The book also explains the interpretation of measurements in terms of decoherence, including the correct explanation of how the classical world we experience emerges from the underlying quantum formalism (e.g., order of magnitude estimates of deviations from classical behavior) in a new way, to provide a more accurate and rounded picture for the reader. While detailed derivations of the principles of decoherence are difficult, the description of the results of those derivations is straightforward to understand at an elementary level.

The third major innovation in this book is the decision to include a brief discussion in Chapter 11 of the principal approximation methods used in many body physics. Students going on to careers in areas that use quantum mechanics will learn about these in more advanced courses, but all students need a glimpse of the way that quantum mechanics explains the world around us.

Throughout, I’ve attempted to emphasize the key principle that quantum mechanics is a probability theory, in which not all quantities that appear in the equations of motion of the fundamental variables can take definite values at the same time. As a consequence, histories cannot be predicted definitely, even if we have a precise account of a complete set of initial data. Moreover, the quantum formalism is mathematically inevitable, since given any list of data, we can introduce matrices, which change one data point into another. Once we do this, quantum probabilities are defined, even in systems where we’re able to ignore them because the equations of time evolution only require the values of quantities that are simultaneously definite.

Those of a mathematically rigorous turn of mind will find my discussion of continuous spectra and unbounded operators lacking in precision. Von Neumann’s famous book cleared up most of the issues, and there are many fine books on the mathematics of quantum mechanics to which one can turn for the details. Physics students, for the most part, are impatient about such things, and it would distract from their absorption of the already difficult conceptual and computational issues of quantum mechanics.

**ORGANIZATION OF THE BOOK**

I’ve chosen to begin with a careful explanation of the differences between classical and quantum probability. This takes place in terms of the simplest quantum system: a two state system or q-bit, where the algebra involved is elementary. I also describe why the two lowest energy states of the ammonia molecule form a good testing ground for comparing the two theories, and how quantum mechanics wins that test in a decisive manner.

The beginning of the text also introduces a key theme: the relation between symmetries and conservation laws (energy conservation is a consequence of time translation symmetry).
In classical mechanics, this is a sophisticated theorem, proven by Emmy Noether, but in quantum mechanics it follows from the very definition of symmetry. Finally, we’ll understand at this very early stage, the relation between discrete energy levels and the spectrum of light emitted by matter.

My treatment of the free particle is based on symmetry principles: invariance under spatial translations and Galilean boosts. The harmonic oscillator is treated by algebraic methods, which enable us to obtain both the energy eigen-values and the eigen-functions with a minimum of computation. We do this through the introduction of coherent states, a simple topic that is usually reserved for more advanced courses. Apart from computational simplicity, the introduction of coherent states enables us to expose the real connection between particles and physical waves, and to dispel the false notion that the Schrödinger wave function is a physical wave, rather than a device for computing probabilities. Finally, this discussion enables us to introduce photon creation and annihilation operators at an early stage. This is conceptually useful in discussions of transitions among energy levels.

An important choice that must be made in any quantum mechanics text is where and when to go over the mathematics of Hilbert space. I do this by introducing it informally in Chapters 2, 3, and 4, providing a formal introduction in Chapter 6, and summarizing the important rules in a brief appendix. Similarly, group theory, which is not discussed in detail, is split between Chapter 6 and an appendix.

Chapter 7 covers the hydrogen atom. We first solve for the spherical harmonics using the algebraic techniques of angular momentum theory. Then we solve the radial problem two ways, first using the traditional power series solution, and secondly (in a problem set) with an algebraic method, which explains the accidental degeneracy of the hydrogen spectrum.

Chapters 8 and 16 are devoted to scattering theory, the first for the exactly soluble Coulomb potential, while the second is a more general discussion. The book introduces scattering theory in the spherically symmetric context, rather than using artificial one-dimensional examples. One-dimensional scattering is treated through an extensive worked problem set in Chapter 4, on the square well and barrier.

Chapter 9 is about Landau levels. This subject is often omitted from textbooks at this level, but it’s the basis for an enormous amount of modern activity, so it’s important to include.

Chapter 10 finally deals with the thorny problem of the proper interpretation of quantum mechanics, and with measurement theory. The point of view emphasized here is that, while QM always gives us a mathematical definition of probabilities for histories of any given complete commuting set of quantities, these probabilities do not satisfy the “sum over histories rule” for total probability, which leads to Bayes’ notion of conditional probability. The interpretation of quantum predictions in terms of actual experiments depends on the existence of systems with large numbers of variables for which that history sum rule is satisfied with
accuracy exponential in the number of atoms in the subsystem on which those variables operate. This is the phenomenon of decoherence, and we briefly review the order of magnitude estimates necessary to demonstrate its plausibility.

Chapters 13, 14, 17, and 18 sketch the main approximation methods that have been used to solve problems in QM. These chapters follow fairly standard lines. Chapter 15 is on the adiabatic approximation, Berry phases, and the Aharonov–Bohm effect. It also discusses anyon statistics in two spatial dimensions and the idea of changing statistics by flux attachment. We learn that fermions in any dimension can be thought of as bosons coupled to a $\mathbb{Z}_2$ gauge field. In a one-semester course, I usually include the discussion of the Aharonov–Bohm effect along with Landau levels.

Chapter 19 discusses Feynman’s path integral formulation of quantum mechanics in somewhat more detail than is found in most textbooks. Chapter 20 is a brief introduction to quantum information and quantum computer science. While far from complete, its aim is to enable the reader to get a head start on a fascinating, rapidly developing field. Lastly, the first appendix discusses a variety of attempts to interpret quantum mechanics in a “realist” fashion, while the others are devoted to technical and mathematical details.

FOR INSTRUCTORS

This textbook is intended for an advanced (junior or senior level) undergraduate quantum mechanics course, or a first year graduate course, for physics and math majors, depending on the level of preparation of the students. The whole book is intended for a full year course. A single semester course may be constructed using Chapters 1–7, 9 and 12, in addition to an optional lecture briefly presenting the main idea in Chapter 10. It may also be used in a physical chemistry or materials science course, as long as the students have had linear algebra, and will be useful to computer scientists who are interested in studying the subject from a physics standpoint. Readers should be comfortable with basic notions of linear algebra, and are encouraged to review the matrix representation of a linear operator, with particular emphasis on the fact that it depends on a choice of basis. The book uses operator algebra as its primary computational tool, rather than differential equations, because these methods involve much less mathematical manipulation, and are of much greater general utility. The differential equation form of the Schrödinger equation is actually only really useful for artificial problems involving a single particle.

INSTRUCTOR RESOURCES

Solutions to problems are available to course instructors upon request. Please visit the book’s page at http://www.crcpress.com/9781482255065.
Thomas Banks is a Distinguished Professor of Physics at Rutgers University and Emeritus Professor at University of California at Santa Cruz. He was born and brought up in New York City and got his undergraduate education at Reed College in Portland, Oregon, majoring in Physics and Mathematics, and his Ph.D. in physics at M.I.T. in 1973. He then joined the faculty at Tel Aviv University, rising to the rank of full Professor before he left in 1985. In 1986, he became a Professor at U.C., Santa Cruz, and, in 1989, moved to become a founding member of the New High Energy Theory Center at Rutgers University. From 2000 to 2015 Banks split his time between Santa Cruz and Rutgers, and returned to Rutgers full time in 2015. In 2014, Banks married Anne Elizabeth Barnes, a judge on the Georgia Court of Appeals. He has two children from a previous marriage. Banks has held visiting professorships at Stanford, the Stanford Linear Accelerator Center, the Weizmann Institute of Science, and the Institute for Advanced Study in Princeton, New Jersey. He has been a Guggenheim Fellow, a Fellow of the American Physical Society, and in 2011, was elected to the American Academy of Arts and Sciences. His research has mostly been in the fields of particle physics and cosmology.
1.1 WHAT YOU WILL LEARN IN THIS BOOK

The formalism of quantum mechanics (QM) was developed some 90 years ago. Since then, tens of textbooks on the subject have appeared. The reader deserves to know why she/he should choose to learn QM from this book, in preference to all the others. There is no better explanation than a list of the things you should be able to learn from reading it. Comparing it to one of the older texts, you will find some differences in emphasis and some differences in the actual explanations of the physics. The latter were inserted to correct what this author believes are errors, either conceptual or pedagogical, in traditional presentations of the subject. The following list includes both topics where our presentation differs from the traditional one and topics that parallel tradition.

- Any mathematical description of a physical system consists of a list of the possible states the system is in. Physical quantities characterizing the system can be thought of as functions on the space of states. A completely equivalent mathematical description is to view the set of states as the basis of a vector space and the physical quantities as matrices $A$, diagonal in that basis. This description is convenient because operations on the system then become a certain kind of nondiagonal matrix, $U$. If one does not know what state the system is in, one can introduce a probability distribution $\rho$ whose diagonal matrix elements $p_i$ give you the probability of being in the $i$-th state. The expected or average value of the quantity given by the matrix $A$ is just given by the formula

$$\langle\langle A\rangle\rangle = \text{tr} (\rho A).$$

(1.1)

The trace of a matrix means the sum of its diagonal matrix elements. The mathematical result that leads to QM is that we can extend this formula to

$$\langle\langle M\rangle\rangle = \text{tr} (\rho M).$$

(1.2)
for any matrix $M$ which is diagonal in a basis related to the original basis by a transformation that preserves the lengths of (complex) vectors. Such operators\(^1\) are given the name normal. The resulting formula gives a probability for the quantity $M$ to take on one of its eigenvalues. What is remarkable is that if we make this interpretation of the formula, then the quantity $M$ has uncertain values even when we know exactly which of the original states the system is in. Conversely, if we accept the eigenvectors of $M$, for which $M$ takes on a definite value, as allowed physical states of the system, then all of our original physical quantities are uncertain in that state. The essence of QM is that we accept all normal operators as possible physical quantities characterizing the system, so that the theory has an intrinsic uncertainty built into it, not related to our lack of knowledge or failure to measure details. This new kind of probability theory violates some of our intuitions about what a probability theory should do. We will explore this idea in more detail in the second part of this introduction and in Chapter 10.

Another, completely equivalent way of describing this new probability theory is illustrated in Figure 1.1: We view the answer to a Yes/No question as the vertical or horizontal position of a switch, along a pair of axes with positive orientation. If we draw another unit vector in the plane, it defines a pair of positive quantities, $p_{1,2}$, that sum to one: the squares of the dot products of that vector with the original axes. In QM, we consider each of these unit vectors as defining a possible state of the system, with $p_i$ interpreted as the probability that the system will be found to be in each of the Yes/No states. $p_1$ is also the probability that, assuming the answer to our question is No, the system is in the state defined by the unit vector in the picture, and similarly for $p_2$ if the answer is Yes. Operators diagonal in different bases are just physical quantities that have definite values in the states described by those bases. Since the process

\[ a^2 + b^2 = 1 \]

\(^1\) The same linear operation or operator has different matrices if we use different orthonormal bases of the vector space to describe vectors as $n$-tuples of complex numbers.
of diagonalizing matrices involves the solution of algebraic equations, we are forced to consider complex vector spaces, instead of the real space shown in the figure, in order to find the most general allowed state of our system.

- Equations of motion in physics relate the time rate of change of physical quantities to the present values of other physical quantities. Classical mechanics is the special case of QM in which all the physical quantities appearing in the equations of motion have definite values in the same state. Our current state of knowledge suggests that the equations that describe our world at the microscopic level do not have this property.

- QM’s unification of physical quantities characterizing the state of a system and operations on that system leads to a transparent proof of Noether’s theorem, which relates symmetry operations on a system to conserved physical quantities. In particular, time translation invariance leads to a conservation law called energy conservation. The corresponding matrix/operator is called the Hamiltonian $H$ and leads to Heisenberg’s form of the equations of motion of QM, for any physical quantity (normal operator) $A$, whose definition has no explicit time dependence.

$$i\hbar \partial_t A = HA - AH \equiv [H, A]. \quad (1.3)$$

- The eigenvalues of the Hamiltonian operator are the allowed energies of the system in states that have definite energy. In states that correspond to classical motions which extend only into finite regions of space, these eigen-energies are discrete, leading to the famous quantization laws that give the subject its name. For unbounded motions, the energy eigenvalues are continuous.

- A particular example, with only bounded motions, is the harmonic oscillator, a good first approximation to almost any system perturbed a little bit from its lowest energy state. Harmonic oscillators are particularly important because fields obeying linear field equations like Maxwell’s equations are just collections of oscillators. We will see that oscillators have two natural kinds of excitations of their lowest energy state: small quantized oscillations with definite energy and large coherent excitations with indefinite energy. The coherent excitations are parameterized by classical solutions of the system, and if the parameters defining the coherent states are large, then the corresponding quantum states have small uncertainties in the classical values of the coordinates, at all times. When promoted to the field theory context, these two kinds of excitations correspond, respectively, to particles, which are automatically identical and obey what is called Bose-Einstein statistics for multiparticle states and classical fields. This duality between two kinds of states of a quantized field theory is the proper interpretation of the phrase wave-particle duality. Historically, and in many current textbooks, that
phrase is applied to the description of particle states by Schrödinger “wave functions.” You will learn why that use of the phrase is misleading by the fifth chapter of this book.

• Many of the particles in the real world, although identical, do not obey Bose–Einstein statistics, but rather another rule, called Fermi-Dirac (FD) statistics, which was motivated by Pauli’s Exclusion Principle, one of the keys to understanding the Periodic Table of chemical elements. The explanation of FD statistics is quite a bit more complicated, and we will have to learn about the Aharonov–Bohm effect in Chapter 15 before we understand it completely. Leave this as a mysterious teaser for now.

• The application of these ideas to motion of a single particle in a spherically symmetric potential leads to the theory of angular momentum. By Noether’s theorem, this is equivalent to studying how quantum states transform under rotations of coordinates. Since such motions are compact, they lead to quantization rules: in this case, the correct form of Bohr’s rules for quantizing angular momentum. We will then move on to the study of radial motion, which will lead us to Bohr’s formula for the energy levels of hydrogen.

• To study most interesting physical systems, we have to understand the dynamics of multiple particles, and we must abandon exact solutions of the equations. Before moving on to that, we pause to explain the relation between the mathematics of QM and our everyday experience of an apparently classical world. The key to understanding that is the notion of a collective coordinate of a macroscopic system. A macroscopic system is one containing of order $10^{20}$ atoms or more, which is in an energy range where a huge number of different, closely spaced, multiatomic energy levels are excited. A collective coordinate of such a system is an average over all of the atoms, like their center of mass. We will argue that the QM uncertainties in collective coordinates are of order $10^{-10}$ or smaller, and that the violations of the “intuitive” rules of probability theory by the quantum predictions for these quantities are smaller than $e^{-10^{20}}$. This means that the latter violations are unobservable, even in principle. The world we are used to corresponds to observations only of such collective coordinates. A proper understanding of this fact removes much of the mystery of QM, associated with phrases like “Schrödinger’s Cat,” “Collapse of the Wave Function,” “Spooky Action at a Distance,” etc.

Quantitatively detailed treatments of multiparticle QM require the use of large computers and a variety of approximations. The two most important approximations are the Born-Oppenheimer approximation, which exploits the fact that nuclei move much more slowly than electrons, and some form of Hartree’s self-consistent field approximation. The modern form of the latter is called Density Functional Theory, and we will be able to give a quick sketch of it, but in very little detail. It will be enough to give us a rough explanation of the periodic table and the gross properties of solids.
• The quantum treatment of the statistical mechanics of multiparticle systems, which leads to the resolution of Gibbs paradox, the Planck black body spectrum, and the phenomena of Fermi surfaces and Bose–Einstein condensation, will be dealt with in Chapter 12.

• Much of the rest of the book is devoted to discussions of various analytic approximation schemes for quantum problems. Some of this material is presented from a fresh perspective, but for the most part, it follows roughly traditional lines.

• Chapter 15 on Berry phases and the Aharonov–Bohm effect is somewhat novel, in that it presents an explanation of Fermi statistics in terms of the A–B effect and a simple presentation of the properties of anyons, particles that can exist only in two space dimensions, and which obey statistical rules different from either bosons or fermions.

• Chapter 19 on Feynman’s Path Integral formulation of QM presents a topic often omitted from textbooks at the undergraduate level. It also covers Schwinger’s alternative functional differential equation derivation of the path integral as well as path integrals for fermionic variables and other systems with only a finite number of quantum states.

• Chapter 20 is a quick introduction to Quantum Computing. Its sole purpose is to enable you to pick up one of the good texts on the subject and get into it rapidly. For the most part, we have stuck to the notation and nomenclature of this book, rather than introducing a whole chapter written in the foreign language of quantum computer scientists.

1.2 WHAT YOU SHOULD KNOW ABOUT LINEAR ALGEBRA

The basic premise of the approach to QM used in this book is that the mathematics of linear algebra defines a new kind of intrinsic probability theory, in which not all quantities can take definite values at the same time. If the equations of motion relate quantities that cannot be definite at the same time, then the concept of probabilities for histories, which is central to the way that classical physicists and philosophers think about probability, can only be an approximate one, valid for certain systems containing many fundamental degrees of freedom, and only for certain average or collective properties of those systems. The key mathematical notion that defines QM probability theory is the notion of change of basis in a vector space over the complex numbers. It cannot be stressed often enough that the mathematical surprise that leads to QM is that the generalization of Pythagoras’ theorem to $N$-dimensional complex vector space can be interpreted as saying that every unit vector defines a probability distribution over the set of all other unit vectors, the absolute square of the projection of one vector on another. This leads to the idea that any diagonalizable linear operator on the vector space is just as good a candidate for a quantity that can be measured,
as any other. All of the basic properties of quantum systems, and their violation of classical logic, flow from this observation. It is therefore incumbent on any would-be cognoscenta of QM to have a thorough knowledge of linear algebra.

You should know the definition of a complex vector space and understand that the representation of a vector $|v\rangle^2$ in that space as a column of complex numbers:

$$|v\rangle = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix},$$

depends on a choice of orthonormal basis vectors $|e_n\rangle$. The word orthonormal means that the scalar products (Dirac notation again) of these vectors with each other satisfy

$$\langle e_i|e_j \rangle = \delta_{ij}.\)$$

The components $v_n$ are the coefficients in the expansion

$$|v\rangle = \sum_{n=1}^{N} v_n|e_n\rangle.$$ The notation $\langle v|$ for a given (column vector) $|v\rangle$ means the corresponding transposed row vector, but with its elements subjected to complex conjugation

$$\langle v| = (v_1^* \ldots v_N^*).$$

This is the representation for $\langle v|$ in the transposed basis $\langle e_i|$, for row vectors, corresponding to the basis $|e_i\rangle$ for column vectors. The reader would do well to convince him/herself at this point that the expression $\langle v|w \rangle$ is equal to the complex number

$$\langle v|w \rangle = \sum_{i=1}^{N} v_i^* w_i.$$ A key mathematical fact that leads to QM is that if both vectors have unit length, the absolute square of this number obeys all the mathematical properties that one would need to call it “the probability that, if one were in the state of a system represented by the unit vector $|v\rangle$, then a measurement designed to detect whether one was in the state represented by $|w\rangle$ would give a positive answer.” This violates only one rule of classical logic: *The Law of the Excluded Middle*. That law takes as the definition a state that one cannot be in

\[\text{We are here using Dirac’s notation for vectors, which will be explained in Chapter 2, and more extensively in Chapter 6.}\]
two states simultaneously. Ultimately, like any other law in a scientific theory, the Law of the Excluded Middle must be tested by experiment, and it fails decisively for experiments performed on microscopic systems. Mathematically, this law is equivalent to claiming that every state encountered in the world is an element of some particular orthonormal basis.

Orthonormal bases are not unique. In two- or three-dimensional space, we are familiar with the fact that any choice of orthogonal axes is related to any other by a rotation. The corresponding ambiguity in N-dimensional complex space is an N-dimensional unitary transformation. The most important thing to remember is that the column vector notation for \(|v\rangle\) depends on the choice of basis, but the vector itself is independent of that choice. This is like saying that the position of your house on the earth is independent of the longitude latitude coordinates we generally designate it by. One can define another set of longitude latitude coordinates by choosing the “North Pole” to be any other point on earth and drawing the corresponding arcs. The ones we use are convenient because of their relation to the rotation axis and magnetic field of the earth, but the South Pole would be just as good. An observer looking at our solar system from the outside might have chosen the plane of the ecliptic to define Earth’s equator, or the plane of galactic rotation, etc. These changes of the definition of the coordinate system change your house’s designation by a pair of numbers, but they do not change where your house is.

Linear operations on a vector space are mappings that take vectors into vectors and satisfy

\[
A(a|v\rangle + b|w\rangle) = aA(|v\rangle) + bA(|w\rangle).
\]

From now on, we will drop the round brackets around the argument of a linear operator and write \(A(|v\rangle) \equiv A|v\rangle\). Once we have chosen a basis, \(A\) is represented by a matrix, an array of rows and columns of complex numbers. We will see that we can compute those numbers as

\[
A_{ij} = \langle e_i | A | e_j \rangle.
\]

In words, the number in the \(i\)-th row and \(j\)-th column of the matrix is given by the scalar product of the \(i\)-th orthonormal basis vector, with the action of \(A\) on the \(j\)-th orthonormal basis vector. It is the component of \(A|e_j\rangle\) along the \(|e_i\rangle\) axis. The matrix looks different for different choices of bases. In particular, for certain operators, called normal, there is a basis where the matrix is diagonal. This is the basis of eigenvectors, \(|a_i\rangle\) of \(A\) satisfying

\[
A|a_i\rangle = a_i|a_i\rangle.
\]

The process of diagonalizing \(A\), given its matrix in some random orthonormal basis, is simply the process of finding the orthonormal basis of eigenvectors.

To reiterate, the fundamental principle of QM is that Pythagoras’ theorem in \(N\)-dimensional complex vector space can be thought of as a probability theory for unit vectors.
in that space. Each unit vector $|e\rangle$ is a state of a physical system, and it defines a probability
distribution that tells you that if you are in the state $|e\rangle$, the probability to be in any other
state $|f\rangle$ is the absolute square of the scalar product $\langle f|e \rangle$ (called a probability amplitude).
The classical approach to the same system would only allow states that are elements of a
particular orthonormal basis. In that case, the same rule tells us that the probability to be in
some allowed state, given that one is definitely in another, is exactly zero. This is sometimes
called The Law of the Excluded Middle, and it is not true if one allows any unit vector to be an
allowed state. At this point, students should go back to our discussion of defining expectation
values as traces of operators, and try to understand how that discussion of probability relates to the current one. The connection involves the projection operator $P(e)$ on a state $|e\rangle$,
which is defined to equal 1 when applied to $|e\rangle$ and to give zero when acting on the subspace of vectors orthogonal to $|e\rangle$. It is better if you work out that connection for yourself, to learn it more thoroughly.

The intuitive reason that we think that knowing the state of a system should determine
that there is no chance of being in any other state, is that we think of determining the state by “doing all possible measurements of the properties of the system.” This is simply incorrect
in QM. The quantum rule is that one determines a state by doing all possible compatible
measurements. That is, the mathematics of the theory, and the way it assigns values to quantities, is such that not all possible quantities that could be measured can be known with precision at the same time. One can try to make this intuitively plausible by thought experiments first described by Heisenberg. One determines the position of a particle by “looking where it is.” Maxwell’s theory of light tells us that in order to do that with precision $\Delta x$ you have to scatter light off it with wavelength less than or equal to $\Delta x$. Maxwell tells
us that light carries momentum, but in classical electrodynamics one can transfer arbitrarily small momentum with any wavelength. However, the quantum theory of light (see Chapter 5) says that the minimum momentum carried by photons of wavelength $\Delta x$ is $\frac{\hbar}{\Delta x}$, so position and momentum cannot be known simultaneously with arbitrary precision.

If there was anything you did not understand about the mathematics in the previous few paragraphs, you should probably learn it well before starting to try to learn QM from this book.

To summarize: in QM, every vector $|s\rangle$ of length one in the vector space representing the system, is considered a possible state of the system. If that vector is an eigenstate of the operator $A$, with eigenvalue $a$, then the theory predicts that a measurement of the quantity represented by $A$ will find the value $a$ with probability one. If $|s\rangle$ is not an eigenstate of $A$, then the probability of finding the eigenvalue $a_i$ is $|\langle s|a_i \rangle|^2$.\footnote{This assumes all the eigenvalues are different. If some of them are the same, the formula is summed over all eigenvectors with the same eigenvalue.} In Chapter 10, and elsewhere
throughout the book, we will discuss at some length the meaning of the words probability and measurement that we used in the preceding paragraphs.

1.3 THE ESSENTIALS OF PROBABILITY THEORY

Before moving on, let us review the essentials of the classical theory of probability. We will do this for a system that has only a finite number of states, labeled by an integer $1 \leq i \leq N$. For a simple coin flipping experiment, we have $N = 2$. The mathematical definition of infinity and of continuous variables always involves a limit from finite systems, so that measurements of any finite precision will never be able to distinguish a system described by an infinite number of states from such a finite system.

A mathematical probability distribution for a finite system is simply a function $p(i)$ defined on the finite set of integers, such that $p(i) \geq 0$ and $\sum_i p(i) = 1$. There are, roughly speaking, two approaches to the physical interpretation of $p(i)$, which we will call the Bayesian and frequentist interpretations of probability. The Bayesian approach is tied to psychology. $p(i)$ represents the expectation that if one tries to determine the state of the system, one will find the state $i$. In the frequentist approach, one defines the probability in terms of repeated trials. One flips a coin $K$ times, determines the fraction of times it comes up heads $f_K(1)$ and defines $f_K(2) = 1 - f_K(1)$, and then takes the limit $K \rightarrow \infty$ of $f_K(i)$ to be $p(i)$. One can think of the Bayesian interpretation as the theoretical model of the probability distribution and the frequentist definition as the experimental method for testing the theory.

The problem is that we can never really take $K$ to infinity. So no actual experiment can carry out the rigorous frequentist definition of probability. If one has a theory of $p(i)$, like "the coin is not weighted, so $p(1) = p(2)$," and one finds 20 million heads in a row, one cannot say that the theory is wrong, because the theory predicts that there is a probability $2^{-20,000,000}$ that the first 20 million tosses will come up heads. All you can say is that "the probability that the unweighted theory is wrong is very close to one." From a strictly logical point of view, this means that the frequentist definition of probability is circular for any finite number of experiments. On the other hand, there is a clear sense in which, for $K$ this large, one is close to the required limit, and one should simply say that the "equally weighted" theory is wrong.

Indeed, in most real systems, one tries to determine the state one is in by measuring variables that nominally take on all possible real values. This means that the possible values one can find by doing the measurement are distributed more densely than the precision of the measuring apparatus. In such systems, there is an unavoidable measurement error in determining what the state of the system is. Thus, experimental results are quoted with "error bars." This means that the results of any experiment are themselves given by a probability distribution. Experimental physicists work hard to eliminate or estimate "systematic errors,"
which are caused by defects in the apparatus or biased changes in the environment which skew the results in a particular direction. They then assume that the environmental factors over which they do not have systematic control are given by a Gaussian probability distribution $P(x) \propto e^{-a(x-x_0)^2}$, where $a$ is estimated from the results of multiple trials of the experiment. In the exercises below, you will see one example of how Gaussian probabilities arise in a very general way when one is dealing with a large number of random events.

If one has only a probabilistic theory of what the results should be, this probability distribution has to be combined with the experimental probability distribution of what the results actually were. One displays the comparison of theory to experiment as a graph with various bands surrounding a line of theoretical predictions for the functional relationship between two measurable quantities. The bands represent confidence intervals, which take into account both the theoretical and experimental uncertainties in the problem.

As an example, the Large Hadron Collider at the CERN laboratory in Geneva, Switzerland, announced in 2012 that “the Higgs boson had been discovered at 5 standard deviation level with a mass (times the square of the velocity of light) of about $125 \times 10^9$ eV,” the level particle physicists consider a significant discovery of a new particle. Five standard deviations correspond to probability of about 1 in 3.5 million. This is the probability that if the Higgs particle does not exist at a mass close to that, the data that CERN scientists collected in Geneva, Switzerland, would be at least as far from the predictions of a theory where the Higgs particle had a significantly different mass as this data is, because of a random statistical fluctuation. The theoretical predictions here depend on QM, which is intrinsically probabilistic and so that random statistical fluctuation could be either a reflection of that theoretical probability or of some uncontrolled random event in the experimental apparatus.

Readers should note the extreme care and almost legalistic precision with which one must phrase scientific conclusions, if one wants to be completely accurate. Usually, scientists use shorthand phrases like “5 sigma” to refer to such a discovery. The assumption is that anyone professional will understand the small print and ultimately probabilistic guesses that go into any statement about a discovery.

In classical probability theory, the distribution $p(i)$ represents uncertainty about some of the factors that determine the state of the system at a given time. One is given a box into which 6 red, 5 blue, and 17 black balls have been inserted, and blindly reaches in and pulls out a ball. It could be any color, but the probabilities of blue, red, and black are $\frac{5}{28}$, $\frac{6}{28}$, and $\frac{17}{28}$, respectively. Here, the uncertainty is encoded in the word blindly. If the person reaching in were able to feel colors with their hand, then this would not be a problem in probability theory.

A more amusing example is recounted in the book *The Eudaemonic Pie* [4]. Some of the inventors of chaos theory had discovered experimentally that roulette wheels obeyed a form of “low-dimensional chaos.” That is simply the statement that the apparently random fall of the ball into different numbered slots was actually fit to high accuracy, by an equation with
a small number of parameters, which one could determine by observing a reasonably small
number of spins of the wheel. The odds in the casinos in Las Vegas and Reno were set by
assuming that the probability distribution for the fall of the ball was completely random. The
principle of odds making is to set them so that the customer wins sometime, but assuming
the distribution is random, the net flow of money goes to the casino. One does not want to
skew them so badly that no one will play the game. The chaos theorists were able to beat the
odds because they could come up with a better ansatz for the distribution, which overcame
the slight edge the casino owners had built in to the odds.

So, in classical probability theory, the initial distributions themselves must be determined
experimentally, or guessed on the basis of some theory of completely random events. There
is a variety of such general theoretical models, appropriate to different assumptions about
the randomness.

Given a probability distribution for the initial state of some system, one makes predictions
in physics by writing down equations of motion. These take the form

\[ i(t + 1) = g(i(t)). \]  (1.4)

Here, the 1 refers to some unit of time. Our insistence that there are only a finite number of
states means that we can only contemplate discrete time evolution, but the time unit could be
as small as we like. One usually studies evolution equations that are invertible. This means
that one can follow the evolution backwards or forwards, starting from any time \( t \). Note
that this is a much weaker requirement than time inversion symmetry, which says that the
evolution in the backward direction is the same operation as that in the forward direction.
Invertibility is the same as saying that \( g(i) = S(i) \), where \( S \) is some permutation which
exchanges each state label \( i \) with exactly one other label.

Classical physicists assumed that in such a situation, the only detectable properties that
the system has were simply functions \( f(i) \) which take values on each state. The evolution
law for such a function is just

\[ f(i, t + 1) = f(i(t + 1)) = f(g(i)). \]  (1.5)

The probability distribution \( p(i) \) looks just like another function, satisfying some constraint
equations, but its evolution law is different.

\[ p(i, t - 1) \equiv p(g(i)). \]  (1.6)

That is, the time evolution for the probability distribution goes backwards, relative to the
evolution of the detectable quantities. To see why, note that the average or expectation value
of any function at time \( t \) is calculated in terms of the value of the probability distribution at
the initial time, \( t = 0 \) via the equation

\[ \langle f(i, t + 1) \rangle = \sum_i p(i) f(g(i)) = \sum_j p(g^{-1}(j)) f(j) = \sum_j p(j, t + 1) f(j), \]  (1.7)
where we have used the fact that the evolution is invertible to redefine the summation variable by $j = g(i)$. This says that we can view the time evolution of the system either in terms of the time dependence of physical quantities with a fixed probability for initial conditions, or as a time-dependent probability distribution to be in any given state with the value of a given quantity $f(j)$ completely fixed at the initial time. In QM, as we will see, these two different ways of thinking about time evolution are called the Schrödinger and Heisenberg pictures.

The second point of view allows us to think about a more general situation in which some of the parameters that determine the function $g(j)$ are themselves uncertain. Most generally, we can make these transformations depend on time, which we denote by $g_t(j)$ and let the time-dependent transformation be random, with some prescribed probability distribution. One can show that if the probability distributions for those variables at different times are independent of each other, then one gets a similar evolution equation for the time dependence of the probability distribution. Similar means that the equation still relates the probability distribution at time $t+1$ to that at time $t$, and that it is linear as a function of $p(j, t)$. The linearity in $p$ is the crucial property that allows us to define probabilities for histories and formulate Bayes’ rule for conditional probabilities.

A history $i(t)$ is simply some particular sequence of states. The probability of such a history, stretching from $t = 0$ to $t = k$ is simply

$$P[i(t)](k) = \prod_{t=0}^{k} p(i(t), t). \quad (1.8)$$

The square brackets in $P[i(t)]$ indicate that this probability depends on all the intermediate points. We will now show that one can write $p(i, t)$ as a sum over all possible histories, weighted by the probabilities for individual histories.

The space of all complex valued functions $f(i)$ is a vector space. A basis in this space is the set of functions

$$e_j(i) = \delta_{ij}.$$  

We can write the probability distribution as

$$p(i, t) = \sum_{j=1}^{N} p_j(t)e_j(i). \quad (1.9)$$

The mapping $i \rightarrow g(i)$ acts on the basis by

$$e_j(g(i)) = \sum_{l} S_j^l e_l(i), \quad (1.10)$$

Footnote: For the present section, we need to invoke only real-valued functions, but in QM, we will need the more general space.
where $S^l_j$ is a permutation matrix. Its nonzero elements are a 1 in the $j$-th row and $l$-th column, where $j = g(l)$. There is exactly one nonzero element in each row, and they are all in different columns.

Now let us rewrite our time evolution equation

$$p(i, t + 1) = p(g^{-1}(i), t) = \sum_{j,l} p_j(t)(S^{-1})^l_j e_l(i). \quad (1.11)$$

In terms of the coefficients in the expansion of $p(i, t)$ in terms of basis functions

$$p_j(t + 1) = \sum_l (S)^l_j p_l(t). \quad (1.12)$$

We can now iterate this equation to get the full solution in terms of initial conditions

$$p_j(t) = \sum_{l_1(1) \ldots l(t)} S^l_j S^l_{l(1)} \ldots S^l_{l(2)} p_{l(1)}(0). \quad (1.13)$$

You will verify in Exercise 1.8 that the multiple sum over indices in this formula is precisely a sum over probabilities of histories, assuming that the histories are related to the initial condition by the equation of motion $i(t+1) = i(g(t))$. In Exercise 1.9, you will show that if the matrices $S$ at different times are not necessarily the same, but are picked independently from a random ensemble of permutations whose probability distribution is chosen independently at each time, then the same sum over histories formulation is correct as long as we introduce the same randomness into the evolution law of the histories.

The sum over histories solution to the evolution law for probability leads directly to Bayes’ law of conditional probability. Suppose we consider some intermediate time $0 < t_i < t$. Then we can divide all histories into those whose state at $t_i$ lies in some particular subset $\Sigma$ of the set of all states, and those which do not. We can then say that if we make an observation at $t_i$, which verifies that the state at that time lies in $\Sigma$, then we can throw out all the histories that do not satisfy that condition, and define a conditional probability distribution over the subset $\Sigma$. That distribution is the sum over restricted histories, divided by a factor that accounts for the fact that the total probability of the restricted histories is less than one in the original distribution. Bayes’ conditional probability rule is the instruction to construct such a distribution, based on observation.

Readers should be alert to the fact that the discussion above contains the seeds of QM. It shows that we can reformulate all of conventional classical physics in terms of linear transformations, the matrices $S$, on a vector space. In Exercise 1.10, you will show that as a consequence of the special properties of permutation matrices, all functions of the probability distribution satisfy the same equations we have derived above. QM replaces the permutation matrices by more general unitary matrices acting on a complex vector space. Functions of
the original distribution no longer satisfy the sum over histories rule and the functions that do satisfy the rule are complex functions called probability amplitudes and cannot be thought of directly as probability distributions. There is a unique sensible definition of a probability distribution that one can construct from these complex functions. This definition is called Born’s Probability Rule, and it says that a probability is the absolute square of a probability amplitude. Since the absolute square of a sum of complex numbers is not, in general, the sum of their absolute squares, probabilities in QM will not satisfy the sum over histories rule, which allows us to define conditional probabilities. The fact that Born’s rule has the form of an absolute square is just a consequence of the fact that Pythagoras’ theorem in complex inner product spaces defines, for every unit vector, $|e\rangle$, an infinite number of functions, one for each orthonormal basis in the space. These functions are just the absolute squares of the projections $\langle e_i|e\rangle$ of the vector $|e\rangle$ on the basis vectors $|e_i\rangle$. They are all nonnegative, and for each basis they sum up to one. That is, they have the mathematical properties that we would assign to the phrase “$|\langle e_i|e\rangle|^2$ is the probability to find that the system is in the state represented by $|e_i\rangle$, assuming that we have determined it to be in the state represented by $|e\rangle$”. QM follows from assuming that this sentence applies to systems in the real world. It violates the assumption of traditional logic, that we define different states of the system by saying that we have determined that if we are in one state, then we cannot be in another. This assumption is based on an incorrect extrapolation of macroscopic experience, namely that we can always measure all of the properties that might determine the state of a system, simultaneously. The task of “understanding QM” really amounts to demonstrating that certain systems obeying the rules of QM can behave like the idealized systems of classical logic, to a sufficient degree of accuracy to account for our missing the correct rules. We will see that the key to demonstrating this is that typical macroscopic objects are composed of $>10^{20}$ atoms. Such a system has $c10^{20}$ quantum states with $c \geq 2$, and under normal conditions (temperatures far removed from absolute zero), the system explores a double exponentially large number of those states. Collective variables, averages over all the atoms like the center of mass position, then obey the sum over histories rule of classical probability theory, with double exponential accuracy.

1.4 PHILOSOPHICAL INTRODUCTION

The rest of this introduction is important, but it is not important to read it before you start the meat of the book. If the philosophizing makes you impatient, skip to the next chapter and come back to this at your leisure. Its main message can be summarized in a couple of sentences. QM is an intrinsically probabilistic theory. The randomness of the world at the microscopic level cannot be attributed to our inability to measure everything with sufficient accuracy, but stems from the mathematical definition of the theory. The theory of QM identifies certain properties of macroscopic objects, made of large numbers
of constituents, which obey the laws of classical probability theory (where all probability is attributed to ignorance/measurement error) with incredible accuracy. That accounts for the apparent classical nature of the macroworld we live in, which shapes all of our intuitions. Now, go on or skip, at your discretion (but please first read the two Feynman chapters, which are mentioned at the end of this chapter).

1.4.1 The Essentials of Quantum Mechanics

QM is the most confusing subject in the world, because it seems to deny the very foundations of logic. Logic is a precise distillation of our intuitive grasp of how things work. It seems to have nothing to do with particular physical situations where we have become used to the fact that our intuition is only an approximation. For example, Galileo appreciated that Aristotle’s intuitive notion that rest is the natural state of bodies was wrong, and invented the relativity of frames moving at uniform velocity. Einstein realized that Galileo’s laws relating the kinematics of two frames moving at uniform velocity were only approximately correct, valid when the velocity was much smaller than that of light. All of this is a bit confusing when you first encounter it, but it is not mind boggling, and once you understand how the correct formulae reduce to the nonrelativistic ones, which make “intuitive sense,” when the velocity is small compared to that of light, it is pretty easy to come to terms with relativity, and even develop an intuition for it.

But how could a similar situation hold for LOGIC? How could logic be “a little bit wrong”? The key to answering these questions involves the notions of probability and uncertainty. We are used to assuming that at any one time, measurable quantities have definite values, and that given values of enough quantities, one can predict what the values of anything else will be in the future. The essence of QM is that this is not true. The fact that a quantity, which has been measured, had a definite value at the time it was measured, is tautological. However, our experience tells us that in order to predict the future value of something, we need to know not only its current value, but also something else, its “rate of change.” For example, Newton’s equations predict the future motion of a particle, given the present values of both its position and velocity. It turns out that in QM you cannot determine the precise value of both of these quantities at the same time. Measuring one with precision introduces an UNAVOIDABLE uncertainty into the other. This uncertainty has nothing to do with the clumsiness of our measuring apparatus. It is built into the fundamental mathematical structure of QM and the definition of what velocity is in that structure.

At the time QM was invented, probability was a concept familiar to classical physicists, gamblers, and insurance brokers. The world seems to be full of events that seem unpredictable, and we are naturally interested in assessing the risk that some disaster or piece of good luck will occur. Classical physicists always assumed that the reason for this unpredictability was our inability in practice to measure all of the variables necessary in order to make predictions.
This was particularly plausible given the picture of ordinary objects as large collections of microscopic constituents, whose miniscule size and vast numbers made the measurement of their properties problematic. So, by the 19th century, there was a well-developed theory of probability, based on the assumption that everything was precisely measurable and predictable, *in principle*, but rarely in practice. One of the problems of this theory is that the probability distributions cannot usually be known in advance. One must make guesses about what they are, and refine those guesses based on data collected about the system under observation. However, it turns out that fairly general mathematical assumptions about probability distributions enable one to make successful statistical predictions about uncertain events, and to estimate the likelihood of those predictions being right. In the limit of a large number of independent trials, most distributions are well approximated by a normal or Gaussian distribution. A great popular account of this theory can be found in [1] and a classic treatise on the mathematics is the two volume work of Feller [2].

One of the fundamental laws of this classical probability theory is Bayes’ law of conditional probability. Without going into the details, Bayes’ law says, essentially, that once one has measured a definite value of some quantity, one can define new probability distributions in which that quantity has no uncertainty and proceed to calculate the probabilities of other things. Philosophically, this fits perfectly with the assumption that everything could have been known in principle, and that it is only our laziness and incompetence, which prevent us from making completely accurate predictions. A classic example of Bayesian conditional probabilities is the probability that a given hurricane will hit a particular city. Early on in the evolution of hurricane Katrina in 2005, there was some probability, according to the equations used by the National Weather Service of the United States, that it would hit Galveston, Texas instead of New Orleans, Louisiana. These are two cities about 400 miles apart on the Gulf Coast of the United States. Once the hurricane hit New Orleans, one could define a new probability distribution in which the probability of hitting Galveston was zero, and that of hitting New Orleans was one, and make future predictions for the evolution of the storm given that new piece of data.

There is a mathematical fact, associated with the use of Bayes’ law, which will be useful in understanding QM. If we write the equations of probability theory as differential equations, then they have the kind of locality we are used to from the equations for fields in classical physics. The time derivative of the probability distribution at a point depends only on the behavior of the distribution at very nearby points, so that the influence of a change at a point takes time to propagate to distant points. However, when we use Bayes’ law, an event at one point (observation of a hurricane hitting New Orleans) **immediately** changes the value of the probability distribution at a distant point (Galveston), changing the value there to zero in a way that is not prescribed by the equation. There is of course nothing funny going on here. The probability distribution is not a physical field, it is an expression of the uncertainty in the position of the storm. Changing it to zero at Galveston is just adding another piece of data,
to our algorithm for predicting what will happen in the future in this particular sequence of events. The theory does not predict what will happen in that particular sequence, but only the frequency of occurrence of different outcomes in the limit of an infinite sequence of reruns of the same initial data. The philosophy behind the use of Bayes’ rule in this instance is that in principle we could have predicted that the hurricane would hit New Orleans. The reason that we did not is that we did not know all the initial conditions and so our probabilistic equations included initial conditions that would have led the hurricane to hit Galveston.

Mathematically, the reason that we are able to use Bayes’ law on solutions of the National Weather Service equations is that these are linear equations for the probability that the hurricane will be in a certain place at a certain time. If we think about all possible histories of the system, the linearity of the equations leads to a rule

- The probability $P(x, t)$ for the hurricane to be at a particular place at a particular time is equal to the sum of the probabilities of all histories $x(s)$ (for $0 \leq s \leq t$) that have initial conditions whose probability is nonzero in the original distribution $P(x, 0)$, and that have $x(t) = x$.

Probabilities in QM do not obey the laws of classical probability theory, and in particular, they do not in general obey Bayes’ rule or the history sum rule above. Given any system, there is a complete set of compatible detectable quantities, which could, in principle, be measured with absolute precision at the same time. There are actually a continuous infinity of different compatible sets, but when one set is in a state where its members have definite values, the others all have uncertain probability distributions. There are two striking things about these probability distributions, when compared to those of classical probability theory. The first is that they are all quite definite mathematical functions, completely determined by the values of the measured detectables. In classical probability theory, the distributions must be discovered from experiment and only take on a priori functional forms in certain limits. More striking and much more peculiar is the fact that QM probability distributions do not, in general, satisfy the Bayesian conditional probability rule, which allows us to replace uncertainty by certainty when a measurement has been made.

We will see that the last feature has profound philosophical consequences, which are best summarized in a phrase invented by Rutgers professor, Scott Thomas, “Objective Reality is an Emergent Phenomenon.” Emergent concepts are basic objects in an approximate theory of nature, which have no counterpart in a more exact theory which underlies it. The notion of water as a continuous fluid is a prime example. Real water is made up of discrete molecules. The nature of the emergence of reality in QM is still a matter of some debate, and there are many researchers who hope to find at least an interpretation of QM which allows for

---

5 The traditional word to use here is observable, rather than detectable. Detectable is preferable because it does not hint that the “observation” must be performed by a conscious observer.
some kind of underlying reality. We will defer most of our discussion of these abstruse issues to Appendix A. However, it is important to stress that all popular discussions of nonlocal action of things on each other in QM has to do with such interpretations, rather than with the use of QM as a probability theory to make predictions about the results of experiments.

Let us be a little more precise here about the meaning of the phrase Objective Reality in Thomas’ aphorism. In classical probability theory, there are probabilities for histories of any system. If $z_i$ represent all possible detectable quantities characterizing the system, then a history is a time-dependent set $z_i(t)$. Even if we divide the system’s characteristics up into visible quantities $v_a(t)$ and hidden ones $h_A(t)$, there will still be probabilities for histories of the $v_a(t)$. Those probabilities will satisfy the “obvious” rule that if one makes an observation at time $t_0$, one can divide up the histories according to whether they agree with that observation or not and base future predictions on the result of that observation, throwing out those histories which did not agree with that observation. This statement is true in QM as well, as long as we actually make the observation. The italicized phrase is the source of much of the confusion about the meaning of QM, because it seems to imply some connection between human intervention and the basic laws of physics. We will learn later that a more proper form of the phrase is as long as the “observation” is a quantum entanglement between the microscopic property being observed and the collective coordinate of some macroscopic object. The point of this long winded phrase is that the microscopic rules of QM probability theory do not have the property that probabilities of histories can be manipulated in this way. The probability of finding the values $v_a(0)$ and $v_a(T)$ is not the sum of the probabilities of all histories $v_a(t)$ with those initial and final values. However, there are special systems, called macroscopic objects, and certain variables, like the center of mass position of such an object, for which the violations of this probability sum rule are incredibly small ($< 10^{-10^{20}}$). “Realist” interpreters of QM believe that a notion of probabilities for histories, obeying the usual rules, is essential to a notion of Objective Reality. This is clearly untrue if we take the commonplace meaning of objective reality, which is merely a record of the things that actually happen in the universe, with no predictive power. It is only if we want to make a predictive theory of what will happen given the maximum information about what has happened, that we might want to use the notion of probabilities for histories obeying the classical sum rule. Experiment shows that those predictions are wrong. QM is a mathematically beautiful (and somewhat inevitable) alternative theory of prediction, which does agree with experiment.

To those who agree with Prof. Thomas’ epigram, objective reality in the sense described above, is not an exact property of any quantum variable. However, objects that we call macroscopic are made of huge numbers of atoms $N > 10^{20}$. Such macroscopic objects have collective coordinates, like the center of mass, which are defined in terms of weighted averages over all the atoms in the object. One can show that the uncertainties in the values of the collective coordinates are of order $N^{-1/2}$ or smaller. Furthermore, the violations of the rules
of classical probability theory for the distributions of these collective variables, are, under normal circumstances\(^6\), of order \(e^{-N}\). So, collective variables behave a lot like classical objects were supposed to behave in classical physics. Their values are not very uncertain, and the uncertainty that there is can, with extraordinary accuracy, be mistaken for uncertainty due to measurement error. According to this view, our brains and our bodies live in this fictitious macroworld of certainty and it was not until we became sophisticated enough to probe the atomic constituents of the matter around us, that we were forced to recognize the correct, quantum mechanical, rules, which govern the world.

1.4.2 Unhappening

One of the most disturbing features of the emergent nature of the concept of “happening” in QM is what can only be called “unhappening.” In classical probability theory, when we use Bayes’ rule to throw away part of the probability distribution, and renormalize the part we keep, we are doing the correct thing physically. That is, if we are only using probability because of our ignorance, then every new piece of data about the world, which tells us “what really happened,” should be used to reduce the uncertainty in our probability distribution. In QM, this is not true. In QM, when we use Bayes’ rule, upon a single observation of a particular value for some collective coordinate, to throw out the part of the probability distribution that predicted another value, this only gives the correct answer for the probabilities of later observations, as long as the macroscopic object (or some other macroscopic object whose collective coordinates were determined by it) continues to exist. If the object disintegrates into elementary particles, we must go back to the initial probability distribution, before its truncation by the use of Bayes’ rule, to get correct predictions for future observations. We will give a particularly poignant example of this in Chapter 10.

It is important to note that classical probabilities are just a special case of the probabilities defined by QM, and classical mechanics is a special case of QM, defined by the requirement that all quantities appearing in the equations of motion can be definite at the same time. This means that in classical mechanics, we can define probabilities for histories of the system, which allows us to say that some things definitely occurred in the past, by using Bayes’ law to consider only probabilities conditioned on the results of past experiments. No real system exactly obeys these classical rules, but the collective coordinates of macroscopic systems, under “normal” conditions, obey them with such fantastic accuracy that experiments to detect the deviations would take far longer than the current age of the universe (see Chapter 10) and require impossible amounts of isolation of the system and accuracy of the measuring device. As noted in the previous paragraph, these statements only remain valid as long as the macrosystems in question do not disintegrate into elementary particles.

\(^6\) We will be more precise about what we mean by normal circumstances in Chapter 10.
1.4.3 Quantum Mechanics of a Single Bit

We will begin our discussion of QM with the simplest possible system, one which has only two states, corresponding classically to a single Yes/No question. As we will see, the molecule of ammonia, NH$_3$ can be approximated by such a system, in a certain energy regime. We will see that all the machinery of QM, the mathematics of linear algebra, can be introduced in a purely classical discussion of this system, as if we were computer scientists, discussing a single bit. Quantities which a classical physicist/logician would think of as measurable can be modeled as diagonal matrices, while classical operations which change the state of the system are off diagonal matrices. A general classical probability distribution is a diagonal matrix, and the usual formula for the expectation value of a quantity is written as a trace of the product of the matrix for the quantity with that for the distribution. This trace formula immediately generalizes to all matrices, even if they are not diagonal. For those matrices which are diagonal in some orthonormal basis, the trace can be interpreted as a probability distribution for that matrix to take on one of its eigenvalues.

This discussion will show that QM is, in a certain sense, inevitable. That is, for any system, even one that we think of classically, we can introduce quantum variables, which have uncertainty even when we have completely fixed the values of the classical variables. The choice of which variables are definite corresponds, for a two state system, to a choice of basis in a two-dimensional vector space. From the point of view of linear algebra, this choice seems arbitrary. The difference between classical and quantum mechanics is in the nature of their equations of motion. Classical mechanics has initial conditions which can all be definite at the same time, while in more general quantum mechanical systems, only some of the initial variables are compatible with each other. The collective coordinates of macroscopic quantum systems do not obey classical mechanics, but they do, up to fantastically small corrections of size $< e^{-10^{20}}$, obey a classical stochastic theory, in which the uncertainties are very small (of order $10^{-10}$ or smaller). These mathematical facts about the quantum theory are sufficient to explain why our intuitions about the logic of the world are incorrect.

The present author is among those who believe that we will never find a consistent interpretation of the facts of the quantum world, which admits the concept of an underlying reality with exact probabilities for histories. This is not a settled question, and we will try to avoid injecting our prejudices into most of this text.

1.4.4 What is Probability?

QM adds confusion to what one may worry is already a complicated issue in classical probability theory, namely how we should think about probabilities. The inventors of the theory, particularly those whose primary interest was in gambling or other financial transactions, clearly thought of it as a sophisticated way of guessing an unpredictable future. This interpretation
is clearly tied up with human psychology. A discussion of probability from this point of view, which attempted to make very precise rules about how to guess and how to use additional data to assess and improve the quality of one’s guesses, was given by Bayes [3] in the 18th century. In modern times, this view of probability is given the label Bayesian interpretation. If you are a gambler or a financier, this is certainly the way you think of probability. Experimental physicists and theorists who follow their work closely also use Bayesian reasoning quite frequently. Looking at experimental lectures you will often see plots, which include lines indicating the predictions of a theory, and colored stripes following the lines, which indicate things like “the 95% confidence interval.” Translated into full English sentences, this means the region of the graph where, with 95% probability the data actually lie, given all the possible random and systematic errors. In graphs referring to the behavior of microscopic systems, these errors include the fact that the quantum theory does not make definite predictions for the number of events of a certain type, but only predicts (see below) the ratio of the number of events to the number of runs of the experiment in the limit that the number of runs goes to infinity. There may also be different bands telling you that the theoretical calculation has “uncertainty,” but this is a completely different sort of error and stems from the fact that we can usually solve the equations of QM only approximately. The use of the word confidence interval in this context is the reflection of the Bayesian outlook on the meaning of probability.

As probability theory became more and more important in science, scientists searched for a more “objective” way of thinking about it, which removed the human psyche and words like confidence from the unbiased description of nature by the combination of mathematics and observation. This led to what is called the frequentist interpretation of probability. According to this paradigm, you test a prediction which gives probabilistic answers by repeating your experiment \( N \) times, and recording the fraction of times the experiment produces each possible result. As \( N \to \infty \), these fractions converge to the predicted probabilities if the theory in question is correct. This is indeed an objective definition of probabilities, but it is problematic, because it is impossible to take \( N \) to infinity, even if the universe lasts for an infinite amount of time. To illustrate the problem, flip a coin 2000 times and observe that it always comes up heads. The probability for that, assuming the coin is unbiased is \( 2^{-2000} \sim 10^{-500} \), a pretty small number, but this does not prove that someone has weighted the coin. If you think it does, would you bet your life on it? Would you bet the lives of all your loved ones? Would you bet the lives of the entire human race? Obviously, these questions all have subjective answers, which depend on who you are and what your mood is. This is the reason that experimental physicists, who test probabilistic predictions (or even definite predictions that are tested with imprecise machinery) by applying the frequentist rule with a finite number of trials, cite their results in terms of a Bayesian confidence interval. We can try very hard to be completely objective about the data, but no finite amount of effort can completely eliminate the need for “leaps of faith.”
These interpretational problems have nothing to do with QM. They would be there for a completely deterministic theory about which we were ignorant of some of the initial data, and since some of the initial data have to do with the performance of the measuring apparatus itself, or external influences interfering with the machinery (cosmic rays, sound waves, the electromagnetic field generated by a radio 4 km away, etc.), we are always ignorant of some of the data. We continue to do experimental and theoretical physics despite these obstacles, because we believe that we can control these sources of error well enough that we are happy with the small size of the required leap of faith.

The interpretation of QM as a new kind of probability theory is certainly correct, and is the only interpretation that has been tested by experiment. If it is the final word on how to interpret the mathematics, then we will just have to live with the intrinsically indefinite nature of probabilistic predictions. We will explore alternative explanations in Appendix A.

1.5 A NOTE ON MATHEMATICS

This subsection is addressed to two different groups of people. First to teachers of a more conventional course in QM: When the Schrödinger equation for a nonrelativistic QM is written in position representation, it is a partial differential equation. If we are discussing the QM of a single particle, the Schrödinger equation is similar to Maxwell’s equations: it is a differential equation in time and space. It is tempting, and some would argue pedagogically preferable, to utilize the students’ familiarity with electrodynamics as a crutch. Indeed, some students have definitely expressed a preference for this approach. In the long run, it is a mistake. The methods of partial differential equations (PDEs) are practically useless for understanding complicated QM problems involving many particles or an indefinite number of particles, and these are the vast majority of systems of interest in particle physics, nuclear physics, and condensed matter physics. The solution of problems via operator algebra, which is the approach taken in most of this book, is simpler (but more abstract and less familiar) and introduces methods, which are more useful in real applications of QM.

Perhaps equally important is the fact that concentration on the single particle Schrödinger equation and its mathematical analogy to Maxwell’s wave equation, misleads students completely about the nature of what is called wave–particle duality. Let me say it clearly once and for all. The Schrödinger wave function is NOT a classical wave, but instead defines a probability distribution. For $N$ particle systems, the Schrödinger equation is a differential equation in $6N + 1$ variables, and is not a wave in space. Wave particle duality has to do with the fact that multiparticle systems can be described by quantum fields: operators which obey

---

7 More precisely, the density matrix, which contains the entire physical content of the wave function, defines a probability distribution for all normal operators in the Hilbert space of wave functions.
Introduction

wave equations in space and time. These quantum fields have states that behave approximately like classical fields, and other states that behave approximately like particles.

Secondly, for those with a mathematically rigorous turn of mind: I have known a lot of brilliant mathematicians who had a hard time reading QM texts because of the nonrigorous treatment of operators and Hilbert spaces. This book will be no exception. Physicists find that excessive attention to mathematical rigor slows us down, and is difficult for most of our students, who are more interested in the use of physics to understand the world. In my opinion, the correct approach is that of von Neumann: if you are bothered by a statement in a physics book, work out the correct explanation yourself. Much of the necessary rigor is supplied in von Neumann’s famous book [5] of which a beautiful new translation has appeared recently [6].

Finally, it should be emphasized that linear algebra is a prerequisite for this course. We will review it, using the Dirac notation beloved by most physicists and rather less popular among mathematicians, in Chapter 6, but you will be expected to know enough to follow the first five chapters. You can of course skip to Chapter 6 to brush up on things, or in extremis, to learn linear algebra from scratch, but that is not the best way to profit from this book. Another part of mathematical physics that would be helpful to know is the subject called analytical mechanics, which is to say, everything about Lagrangians, Hamiltonians, Poisson brackets, and the Hamilton-Jacobi equation. We will not use much of this in the book, but there is a short summary of it in Chapter 4, and the nomenclature will be helpful in our discussion of particles in magnetic fields and in the chapters on Path Integrals and the JWKB approximation.

1.6 FEYNMAN’S LECTURES

Before you go on any further, you should read the first two chapters of Volume 3 of the Feynman lectures on physics. You can find them here: www.feynmanlectures.caltech.edu. They illustrate the puzzling nature of QM in a beautifully simple way.

1.7 ACKNOWLEDGMENTS

In writing this book, I have benefited from a lot of help from colleagues and students. I would like to thank Eliezer Rabinovici, Patrick Hayden, Daniel Harlow, Subir Sachdev, Sriram Shastry, Michael Dine, and Jacob Barandes for important comments and advice on portions of the manuscript. I would especially like to thank Satish Ramakrishna for providing very careful solutions to many of the problems. Pouya Asadi and Conan Huang are responsible for the figures. Ricardo Mulero provided me with the beautiful cover art. Finally, I would like to thank my beautiful wife Anne for the patience she has shown as I worked on this book and all the love and support she gives me every single day.
1.8 Exercises

1.1 The Let us Make a Deal Problem: On this famous quiz show a contestant is shown three doors and told that behind one of them there is a fabulous prize. After the contestant has chosen a door, one of the other doors is opened, and shown to have no prize (or a booby prize) behind it. The contestant is then given the option to change the door he/she has chosen. What is the best strategy for getting the prize?

1.2 If a couple has three children, what is the probability, given no further information, that two of them are girls? Suppose you know that one of the children is a girl named Florida. What is the probability that two of the children are girls? Does this depend on the choice of name? FYI, Florida is a girl’s name that used to be popular, but has gone out of fashion.

1.3 Drop a needle on a plane surface. Take the origin to be one end of the needle and draw radial lines at angles \( \frac{2\pi k}{N} \), starting in a random direction. What is the probability that the direction of the needle is between the \( k \) and \( k+p \)th radial line? How does the answer change if one takes the origin to be at another point?

1.4 Here is a problem that is easy to state, but hard to analyze. Given a fairly weighted coin, what is the probability that in 100 flips there will be a run of at least \( K \) heads in a row? What if we ask for exactly \( K \) heads in a row? The way to set this up is to define a quantity \( P(i) \) which is 1 if the \( i \)th throw is a head and zero if it is a tail. Each run is characterized by a set of values for these 100 variables, and there are \( 2^{100} \) possible values. The question we are trying to answer is, out of all of these possible “states of the system,” how many of them have \( P(i)P(i+1)\ldots P(i+K−1) = 1 \) for some value of \( i \). The trick is that we have to worry about double counting. E.M. Purcell, the Nobel Prize winning Harvard physicist and author of a classic text on electrodynamics, solved the much harder problem of determining the probability given, “dumb luck” of a player having an \( n \) game hitting streak in baseball. At the time he did the calculation in the 1980s, the only streak in the history of baseball, which was not, “what we could expect from random probabilities for hits and outs” was Joe DiMaggio’s streak in 1941.\(^8\)

1.5 The probability of getting \( K \) heads in \( N \) throws of a fair coin is given by

\[
\frac{N!}{2^N K! (N-K)!}
\]

\(^8\) This calculation is reported in a lovely essay by S.J. Gould \[7\]. You will also be touched by Gould’s essay \[7\] on estimating his own chance of beating the odds after his cancer diagnosis. Note that the Wikipedia article on Dimaggio cites some work disagreeing with Purcell’s calculations.
This is maximized at $K = N/2$ if $N$ is even. Show that as $N \to \infty$ the probability of getting $\frac{N}{2}(1 - x)$ heads, with $x$ kept constant as $N$ goes to infinity, is a Gaussian distribution of the form

$$P(x) = Ae^{-bx^2}.$$ 

Calculate the value of $b$. What happens if $N$ is odd?

1.6 Suppose the probability of finding some quantity $Q$ to have the value $x$ is

$$P(x) = e^{-f(x)},$$

where $f(x) > 0$. Given $N$ independent copies of the system, show that as $N \to \infty$, the fluctuation of $Q$ away from its expectation value goes to zero. For the system with multiple copies, $Q_{AV}$ is defined to be the average of the $Q$ of the individual copies. That is

$$Q_{AV} = \sum \frac{x_i}{N}.$$ 

Show that fluctuations of $Q$ away from that expectation value have, generically, a Gaussian distribution with a width that scales to zero like $N^{-1/2}$. What characterizes the nongeneric exceptions? This result is called the Central Limit Theorem.

1.7 Political polls often quote a “margin of error.” This is defined in the following way. Suppose, in a very large population $N$, a fraction $p$ of the population prefers candidate Thomas Jefferson and the rest prefer his opponent. Imagine you take a random sample of $n \ll N$ voters and find a fraction $p_n$ prefer Jefferson. There are $\frac{N!}{n!(N-n)!}$ different random samples, which give a distribution of $p_n$. A sample of $n$ people voting for Jefferson’s opponent would give $p_n = 0$, whereas a sample containing only Jefferson voters would give $p_n = 1$. Show that for $N \gg n \gg 1$, the distribution of $p_n$ is Gaussian and find its center and width. The margin of error usually quoted is one half of the 95% confidence interval for the true percentage $p$. The margins of error for different confidence intervals are simply related by the value of the inverse error function at those confidence levels.

1.8 Use the fact that the only nonzero matrix element of a permutation matrix is a 1 in each row, with a different column for each row, to show that the formula for the time evolved probability distribution can be thought of as a sum over histories. The contribution from a fixed value of each of the summed matrix indices $l(k)$ is the contribution of a single history.

1.9 Show that the sum over histories formulation still works, if at each time the permutation matrix $S$ is replaced by a time-dependent matrix $S(t)$, chosen from some probability distribution of with no correlation between different times. The dynamics of the histories
is simply replaced by \( i(t) = g(i(t - 1), t) \), where the permutation \( g \) is chosen from the same random distribution.

1.10 Use the fact that the only nonzero matrix element of a permutation matrix is a 1 in each row, with a different column for each row, to show that \( p^2(j, t) \), or indeed any function \( f(p(j, t)) \) satisfies the same sum over histories formula as \( p(j, t) \) itself.


