

Biao Wu

# Quantum Mechanics

A Concise Introduction



北京大学出版社  
PEKING UNIVERSITY PRESS



Springer

# Quantum Mechanics

Biao Wu

# Quantum Mechanics

A Concise Introduction



Biao Wu  
School of Physics  
Peking University  
Beijing, China

*Translated by*  
Ying Hu  
Taiyuan, China

ISBN 978-981-19-7625-4      ISBN 978-981-19-7626-1 (eBook)  
<https://doi.org/10.1007/978-981-19-7626-1>

Jointly published with Peking University Press  
The print edition is not for sale in China (Mainland). Customers from China (Mainland) please order the print book from: Peking University Press.

Translation from the Chinese Simplified language edition: “Jian Ming Liang Zi Li Xue” by Biao Wu, © Peking University Press 2020. Published by Peking University Press. All Rights Reserved.  
The translation was done with the help of artificial intelligence (machine translation by the service [DeepL.com](https://www.DeepL.com)). A subsequent human revision was done primarily in terms of content.

© Peking University Press 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publishers, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publishers nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publishers remain neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*To Yingying, Liangliang, Dingding, and  
Dangdang*

# Foreword

Why learn about quantum mechanics? There are many reasons:

- It is how the physical world works, at a fundamental level. If you want to understand biology at a molecular level, or medicine, or chemistry, or astrophysics, or many of the most exciting frontiers of modern engineering—not to mention physics—you will need to get comfortable with the basic concepts of quantum mechanics.
- It is a great human achievement. It is inspiring to see what humanity at its best can accomplish.
- It is mind expanding. In coming to terms with quantum concepts, you learn that our inner world is a richer, livelier, and far stranger place than superficial appearances and everyday experience might suggest. And you must wrestle with paradoxes that are still being debated by physicists and philosophers.
- It is a revelation of the beauty of the world.

In my experience, everyone who learns some real quantum theory has been glad that they put in the effort.

This short book by my friend and colleague Wu Biao can help introduce you to quantum mechanics. It begins with the human history of the subject. It develops the basic concepts in an honest way using a minimum of mathematical baggage, most of which is explained from scratch (assuming only high-school algebra). From there, it introduces contemporary issues and developments that are still very much alive and in flux: philosophical controversies surrounding the interpretation of quantum theory and the new prospects distinctly quantum ideas are bringing to computing and communication. Although I do not agree with all his conclusions, I enjoyed his account of controversies and prospects for the future. He states things clearly, and he gives you the tools to start thinking for yourself!

Best of all, his love for the subject shines through.

Boston, USA  
August 2022

Frank Wilczek

# Translator's Preface

The initial English translation of the original Chinese edition was done with the help of artificial intelligence (machine translation by the service provider [DeepL.com](https://www.DeepL.com)). A subsequent human revision of the content was done by the translator. The present translation follows closely the text of the original Chinese edition. The translated manuscript has been carefully revised by the author. Therefore, any changes from the original version are due to the author. As well, all the viewpoints in the book are his rather than those of the translator.

The translator wishes to express her deepest gratitude for the privilege of working with the author on the rendering of the original contents of the book into a translation. The translator is also in deep gratitude for the love and support of so many from her family during the translation.

Taiyuan, China  
May 2022

Ying Hu

# Preface

Quantum mechanics, which was established at the beginning of the twentieth century, is one of the most profound revolutions in the history of science. It is a radical departure from classical physics, opening the door to the fascinating world of quantum. Quantum theory not only is capable of describing the strange behaviors of microscopic particles, including electrons, quarks, atoms, and molecules, but also powers a technology revolution that has deeply transformed—and is continuing to change—many aspects of human societies. As important as it is, quantum mechanics attracted little attention outside the professional community of physicists, chemists, and some philosophers for a long time. Recently, the rapid development of quantum information technology has sparked increasing interests in general public in quantum physics.

The purpose of this book is to provide a popular yet serious introduction of quantum mechanics to the general public. By “popular”, I mean the book aims to make the fascinating quantum physics accessible to as many readers as possible. By “serious”, I mean to present the most profound results in quantum mechanics through mathematics. To achieve a balance between the two, I assume that the reader is familiar with high-school mathematics. For mathematics beyond high-school level such as matrices and linear spaces, I will give an introduction at the level just enough for quantum physics covered in this book. The mathematics is not difficult. Determined readers can quickly become proficient after some practices. After all, knowledge is of no value unless you put it into practice.

This book starts with a general introduction of quantum physics and quantum technologies without using mathematics, which is followed by a brief history of quantum mechanics. In 1900, Planck unveiled the first mystery of quantum physics. And the basic framework of quantum mechanics was completed in 1926, the year Schrödinger wrote down his immortal equation. During the first quarter of the twentieth century, great pioneers in quantum physics, with their brilliant intelligence, extraordinary imagination, and tireless efforts, had led mankind into the magical world of quantum and a profound scientific and technological revolution. Chapter 2 is a salute to these pioneers and great minds. In addition, it tells a story how physicists, with the clues from hard experimental facts, came to establish this strange theory of quantum.



The bulk of this book describes the amazing quantum world: quantum states that live in Hilbert spaces, indistinguishable particles, linear superposition of states, Heisenberg's uncertainty relation, quantum entanglement, Bell's inequality, quantum energy levels, Schrödinger's cat, and many-worlds theory. For comparison, a brief introduction of classical mechanics is provided in Chap. 3. The book concludes with some elementary introductions to quantum computing and quantum communication.

This book will also be useful for students in physics major. These students usually get lost in solving the Schrödinger equation: they know a lot of mathematical formulas but do not have a good understanding of essence of quantum mechanics. In addition, traditional textbooks on quantum mechanics do not cover quantum entanglement and quantum information, and this book fills this gap.

Chapters 1–3 of this book are optional. The readers who are unfamiliar with complex numbers and linear algebra should read Chap. 4 carefully. The readers who are familiar with this part of mathematics are advised to go through Chap. 4 quickly, mainly to familiarize themselves with the Dirac notation. Chapters 5–7 are mandatory, as they form the basis for subsequent chapters. There are no deliberately arranged exercises in this book, but you will frequently see sentences that begin with “the interested reader”, reminding the readers to repeat similar derivations or to perform simple proofs.

All hand-drawn figures in this book were the work of Ms. Zhaocheng Sun; some illustrations were made by Mr. Zishuo Han.<sup>1</sup>

Beijing, China

Biao Wu

---

<sup>1</sup> They include Figs. 2.1, 2.3, 3.1, 3.2, 6.2, 6.4, 6.3, and 8.1.

# Acknowledgements

I started writing this book on January 13, 2018, and completed the first draft at the end of August 2018. During this period, since Chaps. 1 and 2 involve few mathematics, they were distributed among friends, who gave me a lot of good advice and pointed out many mistakes. These include Shu Chen, Zhengwei Zhou, Ying Jiang, Qiuyi Guo, Minghui Zuo, Xuebin Chen, He Liu, Hongbo He, Wei Chen, and Ailing Yang. When the manuscript (the original Chinese version) was about to be delivered to the publishing house around the beginning of 2020, Hongqiang Liu, Xiaobing Luo, Ying Jiang, and Qiuyi Guo carefully read the whole manuscript and made many constructive suggestions to improve this book.

The present Chap. 1 was not in the original plan. One day, Ms. Lan Mei, a friend of mine who has no science and engineering background, asked me, “what is quantum?”, expecting me to give a short answer. The first chapter is my answer to this question. A significant revision was made to the first paragraph of this chapter during the English translation thanks to Prof. Xiaofeng Jin’s comments to its Chinese version. I also thank him for his comments and suggestions on other parts of the book.

My professor, Prof. Shouyong Pei of Beijing Normal University, carefully read my unpolished first draft and gave me a lot of valuable advice and great encouragement.

This book was originally lecture notes for my course of quantum mechanics at Peking University. The students are mainly from non-physics majors, including some from liberal arts majors. The questions and suggestions they had in the teaching process are very helpful to the writing and revision of this book. Many students helped me correct small mistakes in the lecture notes. They include Hao Li, Yi Zhang, Zhongqi Guo, Weikang Li, Chenyang Dong, Weiqi Huang, Zhonglin Xie, Xinchun Liu, Jiacheng Liu, Shumei Tan, Jingwen Dong, Yaxuan Liu, and Yu Xiong, Yifei Wang. Please forgive me for any omission. I would also like to thank Runheng Li and Jingwen Dong for their help in the class.

Translating this book into English was first suggested by Ran Cheng, who also helped correcting mistakes in its English translation. Zhenhua Qiao helped to set up the connection with Springer. I sincerely thank Ying Hu for translating the book into English. Although the book was first machine translated into English by the service provider DeepL.com, Ying Hu significantly revised it with great efforts. Lin Dong

read through the whole manuscript, pointing out minor mistakes and making helpful revision suggestions.

Yunkai Zhang helped design the book cover and Xuan Mi helped the final proof reading.

My sincere thanks to Prof. Frank Wilczek, who wrote a beautiful foreword for this book.

Finally, I would like to express my deep appreciations to all my teachers, colleagues, friends, and students who have offered their kind help.

# Contents

<b>1</b>	<b>What Is Quantum?</b> .....	1
<b>2</b>	<b>A Brief History of Quantum Mechanics</b> .....	9
2.1	The Birth of Quantum .....	9
2.2	The Difficult Start .....	12
2.3	Hydrogen Atom .....	14
2.4	The Crisis .....	18
2.5	Identical Particles .....	20
2.6	It's Matrix .....	24
2.7	Particles Are Waves and Waves Are Particles .....	26
2.8	Retrospect .....	28
<b>3</b>	<b>Classical Mechanics and the Old Quantum Theory</b> .....	31
3.1	Free Fall .....	31
3.2	Phase Space .....	33
3.3	Calculus for Velocity and Acceleration .....	36
3.4	Harmonic Oscillator .....	37
3.5	The Old Quantum Theory .....	38
<b>4</b>	<b>Complex Number and Linear Algebra</b> .....	41
4.1	Complex Number .....	41
4.2	Linear Algebra .....	45
4.2.1	Linear Space .....	45
4.2.2	Hilbert Space .....	48
4.2.3	Matrix .....	52
4.2.4	Eigenstates and Eigenvalues .....	60
4.2.5	Direct Product .....	61
<b>5</b>	<b>Into the Quantum World</b> .....	63
5.1	The Stern-Gerlach Experiment .....	63
5.2	Spin .....	65
5.3	Quantum State and Its Statistical Interpretation .....	67
5.4	Observables and Operators .....	72

- 5.5 Spin Along an Arbitrary Direction ..... 73
- 5.6 Theoretical Framework of Quantum Mechanics ..... 77
- 6 Quantum Dynamics** ..... 81
  - 6.1 Schrödinger Equation ..... 81
  - 6.2 Wave Function ..... 82
  - 6.3 Hamiltonian Operator and Unitary Evolution ..... 84
  - 6.4 Quantum Energy Levels and Eigenstates ..... 87
  - 6.5 Superposition Principle of Quantum States and No-cloning Theorem ..... 93
  - 6.6 Double-Slit Interference ..... 96
- 7 Quantum Entanglement and Bell’s Inequality** ..... 101
  - 7.1 System of Two Spins ..... 101
  - 7.2 Quantum Entanglement ..... 105
    - 7.2.1 Bell’s Inequality ..... 106
    - 7.2.2 Loss of Individuality ..... 114
- 8 Quantum Measurement** ..... 117
  - 8.1 The Uncertainty Relation ..... 118
  - 8.2 Collapse of a Wave Function ..... 124
  - 8.3 Many-Worlds Theory and Schrödinger’s Cat ..... 128
  - 8.4 Flaws in Feynman’s Argument ..... 138
- 9 Quantum Computation** ..... 143
  - 9.1 Classical Computer ..... 144
    - 9.1.1 Basic Framework ..... 144
    - 9.1.2 The Unbearable Quantum ..... 147
  - 9.2 Quantum Computer ..... 148
  - 9.3 Reversible Classical Computers ..... 154
  - 9.4 How Powerful Is a Quantum Computer ..... 158
  - 9.5 Technical Difficulties of Building a Quantum Computer ..... 161
- 10 Quantum Communication** ..... 165
  - 10.1 Polarization of Photons ..... 165
  - 10.2 Quantum Teleportation ..... 168
  - 10.3 Classic Encryption ..... 172
  - 10.4 Quantum Key Distribution ..... 174
  - 10.5 Future Quantum Technologies ..... 177
- Further Reading** ..... 181
- Index** ..... 185

# Chapter 1

## What Is Quantum?



Advances in the field of quantum information have brought unprecedented attention to quantum physics. The physics term, quantum, has become a buzzword around the world. Major international companies are gearing up to build quantum computers which are far superior to classical computers. People are curious about what quantum is, hoping that experts can provide a short answer. Some technical terms can be explained precisely in one or two sentences. For example, a lightyear is the distance travelled by light in one year. The speed of light is 299,792,458 m per second and there are about  $3.15 \times 10^7$  s in a year, therefore, a lightyear is about  $9.46 \times 10^{12}$  km. Unfortunately, quantum is not such a technical term and there is no short answer to the question, “What is quantum?”.

Nevertheless, I will try to give a short answer. In physics, quantum, originating from the Latin word “quantus” for “how much”, is the minimum amount of excitation in a field. For instance, the smallest excitation of an electromagnetic field is a photon, i.e. the quantum of an electromagnetic field is a photon. All elementary particles, including electrons, quarks, neutrinos, and gluons, are quanta (minimum excitations) of some fields. Protons are not quanta because protons are composite particles made of quarks. Similarly, a hydrogen atom, which is made of an electron and a proton, is not a quantum. In 1900, Planck first discovered this fundamental aspect of nature when he found that the energy of light must be divided evenly into a number of the smallest elements.

This is the shortest answer I can give. It is already much longer than the answer for lightyear. If you are not familiar with physics, you are probably still confused, raising even more questions: what is a field? What is an excitation? Why does it have anything to do with computer? To answer these new questions, I will have to introduce new terminologies, and you will raise more new questions. In particular, the word “quantum” in quantum computer and many other phrases, such as quantum chemistry and quantum entanglement, has no direct relation to excitations of fields, and it just means something related to quantum mechanics. So, the question,

“what is quantum?”, is essentially the question, “what is quantum mechanics?”. It can be answered only with a book or a course.

As quantum mechanics has been a standard course for physics major in colleges around the world for many decades, there are already many excellent textbooks on this subject. However, almost all of them are written with difficult mathematics, such as calculus, differential equations, and linear algebra. This is necessary for anyone who wants to apply quantum mechanics and use it to solve problems, but it poses an intimidating barrier for many people who are just curious and want to appreciate the strangeness or magics of quantum mechanics in a meaningful way. This book is for these curious people. Anyone with high-school level mathematics should be able to follow my book into the beautiful world of quantum. You will have to learn some new mathematics, such as matrices, and do some exercises. Galileo said that the universe, which is governed largely by quantum mechanics at the fundamental level as we understand it today, is written in the language of mathematics. To appreciate its inner working, you have to use some mathematics. On the journey to the quantum world, it is very likely that the difficulties you may encounter will start to compete with your curiosity. I hope that your curiosity will win over and you will finish reading the book.

In this chapter, I take a compromised strategy, introducing quantum mechanics without mathematics. I hope that the readers will come away with a general picture about quantum mechanics, and strengthen their wills to go further down the road to the quantum world. As quantum becomes a buzzword around the world, many merchants try to sell their products, which have nothing to do with genuine quantum technologies, with quantum on their labels. I hope that, after reading this chapter, you will have a clear understanding as to

What is **not** quantum?

so that you can see through these fake “quantum” products in everyday life.

Planck’s discovery in 1900 was groundbreaking. Known as the father of quantum mechanics, he inadvertently opened the door to quantum physics. Many great physicists, inspired by hard experimental facts, followed the path that Planck had blazed and established a new theory, quantum mechanics, by the end of 1926. They include Einstein, Bohr, de Broglie, Dirac, Fermi, Heisenberg, and Schrödinger.

It is now customary to use “quantum” to name or define something related to quantum mechanics. For instance, to be distinguished from the information entropy, the entropy in quantum mechanics is called quantum entropy. Quantum chemistry is a branch of chemistry where chemists rely on quantum mechanics to understand the spectra of atoms and molecules, or the molecular bonds, etc. Quantum dots are devices a few nanometers in size fabricated in the laboratory; as electrons in quantum dots are confined in a small space, they exhibit discrete energy levels and exotic properties due to quantum mechanics.

Quantum mechanics represents a revolutionary departure from almost all the ideas in classical physics represented by Newtonian mechanics. This revolution, in my opinion, is even more shocking and profound than the theory of relativity, as is reflected in the following six aspects.

- 1. Quantum**—A quantum is the smallest unit of excitation in a fundamental field. The term ‘excitation’ means to perturb the field by injecting some energy. It is like stirring a big bowl of water. From our experience, if we are careful enough, in principle, we can continuously increase the stirring strength from zero to infinity; accordingly, water in the bowl develops from just a few ripples into a splash. In other words, in classical physics, the excitation can be arbitrarily small. But this is forbidden by quantum mechanics, which says that an excitation in a fundamental field, such as electromagnetic field, cannot be smaller than the smallest unit—quantum. This was one of the first discovered features that distinguish quantum physics from classical physics, and it is how quantum mechanics got its name.
- 2. Heisenberg’s uncertainty relation**—When you sit on the couch watching the World Cup, your position is known and you have a zero velocity. Your navigator on your car can tell you where you should turn right and remind you when you exceed the speed limit. These everyday experiences tell us that an object can have definite position and velocity simultaneously. It agrees completely with Newtonian mechanics. In fact, Newtonian mechanics demands that a particle has perfectly determined position and velocity at the same time, otherwise we would not be able to determine the particle’s motion. However, things are completely different in quantum mechanics. According to Heisenberg’s relation of uncertainty, a particle cannot have precisely determined position and velocity simultaneously: if the position is determined, its velocity is uncertain; and vice versa. Yet in our everyday world, we do not feel Heisenberg’s uncertainty relation. This is because our measurement of position and velocity are still far from the atomic accuracy. One of the physical consequences of Heisenberg’s relation of uncertainty is that an atom cannot be “frozen” even at absolute zero (the lowest temperature allowed by nature). If it were completely “frozen”, both its position and velocity would be perfectly determined (i.e., zero), violating Heisenberg’s uncertainty relation. Therefore, an atom cannot be at rest, and it vibrates even at the temperature of absolute zero. This is called the zero-point vibration. A helium atom has particularly prominent zero-point vibration due to its small mass. As a result, helium remains in the liquid state at the temperature of absolute zero, where all other substances become solid. This property makes liquid helium a necessary working substance in all low-temperature cryogenic refrigerators.
- 3. Superposition principle**—In a classical world, an object has a definite position at any moment, which is consistent with our general observations. You can’t be at home the same time you work in office. No one believes you were sleeping at home if the surveillance video shows that you were at the crime scene at that time.  
But in the world governed by quantum mechanics, an object can simultaneously appear in two different places or have different velocities. For instance, an electron in a hydrogen atom can simultaneously be at the left and right sides of a proton at the same time, or simultaneously revolves clockwise and counterclockwise around a proton. Schrödinger’s cat is a dramatic example of this amazing and exotic quantum phenomenon: a cat can be both alive and dead at the same time. Similarly, a pot of water can be simultaneously cold and hot, and the sun can simultaneously



rise in the east and set in the west. Of course, you've never encountered these situations in daily life. But according to quantum mechanics, all these phenomena, in principle, can happen. Physicists are still debating as to why these miraculous quantum phenomena occur only on the microscopic scale but not in everyday world. Many drastically different theories are proposed with no consensus to be reached in any foreseeable future. We will discuss and compare two of them, the Copenhagen theory of collapse of wave function and many-worlds theory, in Chap. 8.

4. **Quantum randomness**—Suppose a particle is in a superposition of two positions, i.e., it is both at points A and B. Let us measure its position to determine where it really is. Quantum mechanics tells us that the outcome of the measurement is random: it could be A or B. More importantly, this randomness is fundamentally different from the randomness that we usually experience.

The randomness that we normally encounter in our daily life originates from our ignorance or incomplete knowledge. Consider a box that contains balls of two colors, red and white. If the box is transparent, you can get exactly a red ball if you want. But if the box is opaque, you can only hope that Lady Luck is on your side to get a red ball. In quantum mechanics, the randomness of a measurement outcome is intrinsic, resulting from the aforementioned superposition principle. If the balls in a box are quantum and in a superposition state of red and white. Then there is no guarantee that you get a red ball every time even if the box is transparent.

5. **Identical particles**—In the macroscopic world, being identical is approximate. When we say two objects are identical, it is in the sense that the two objects are the same for the properties that we are interested in, but they can be distinguished if we observe them closely enough. For example, suppose we have two coins of ten cents. When we only use them for purchasing, the two coins are identical to us, although one looks new and the other looks old. Even if both coins are brand new, we can distinguish them with a good enough microscope. As another example, a pair of twins may look identical for most of the people, but their parents can always tell them apart because they have observed more carefully.

As a fundamental contrast to the classical world, being identical is absolute in quantum mechanics. Two electrons are completely the same, and cannot be distinguished in any way. To emphasize this absolute sameness, we call electrons identical particles. All the microscopic particles are identical particles. Photons are absolutely the same, protons are absolutely identical to each other, etc. Such quantum indistinguishability manifests in their statistical properties. For example, two ordinary coins can have four possible states: both are heads, both are tails, coin 1 is head and coin 2 is tail, coin 1 is tail and coin 2 is head. But for two identical quantum coins, there is absolutely no way that you can tell which is coin 1 and which is coin 2. Therefore, there are three states at most: both are heads, both are tails, one is head and one is tail. For ordinary coins, each of the four possibilities occurs with a probability of  $1/4$ , so the probability to find a coin being head and a coin being tail is  $1/2$ . But for identical quantum coins, the probability to find one being head and one being tail is  $1/3$  or  $1$  (because there

are situations where two identical coins are forbidden to be heads or tails at the same time; see discussion in Chap. 2).

6. **Quantum entanglement**—When you depart on a trip in a hurry and find a right-handed glove in your bag when arriving at the destination, then no matter how far away you are from home, you know instantly that the glove left at home is left-handed. On February 9th, 1796, the Qianlong Emperor of Qing dynasty announced his abdication to Jiaqing. From this moment, all edicts issued by the emperor Jiaqing was effective immediately in the whole country, even though someone in Xinjiang, a place far from the capital, knew it a few days later. This instantaneous correlation in information is common in our daily life. For convenience, we shall call it correlation at distance or, more generally, nonlocal correlation as it occurs over a distance without taking time. A prerequisite for such nonlocal effect is that we know something *a priori*: a pair of gloves is made up of a left-handed glove and a right-handed glove; the laws issued by the Qing emperor are effective unconditionally and immediately in the country.

Correlation at distance can also occur in the quantum world, which is known as quantum entanglement. Suppose two quantum particles, A and B, are in a quantum state where one has a velocity  $v$  and the other has a velocity  $-v$ . If we know from the measurement that particle A has the velocity  $-v$ , we immediately know that particle B has the velocity  $v$ , no matter how far away particle B is. Such nonlocal correlation in quantum entanglement is so similar to the classical correlation at distance that in a long time physicists thought that they were the same thing. Only when Bell proved a famous inequality in 1964, physicists began to realize that they are different. Bell found that classical correlations always satisfy this inequality, but quantum entanglement can violate it. Physicists have carefully measured quantum entanglement in experiments and confirmed Bell's prediction. In Chap. 7, we will prove Bell's inequality and discuss its implications. Apart from nonlocal correlation, quantum entanglement has another remarkable feature: entangled particles lose their individual status. Suppose that there are two people, Alice and Bob. Alice is sitting and Bob is standing. If they were entangled quantum mechanically, Alice would be either sitting or standing, and similarly Bob would be either sitting or standing. That is, they would have lost their own individual status, not certain whether they were sitting or standing. Fortunately, quantum entanglement disappears completely in our daily world, otherwise our lives would be very interesting. Physicists still have not yet fully understood why quantum entanglement disappears in the macroscopic world.

The seemingly bizarre quantum mechanics turns out to be one of the most successful theories in physics, and is now one of the two theoretical pillars of modern physics along with relativity. Not only can it precisely describe the behavior of microscopic particles such as quarks, neutrinos, and atoms, it also explains why metals conduct electricity and why magnets are magnetic. Moreover, quantum mechanics has nurtured a revolution in technology.

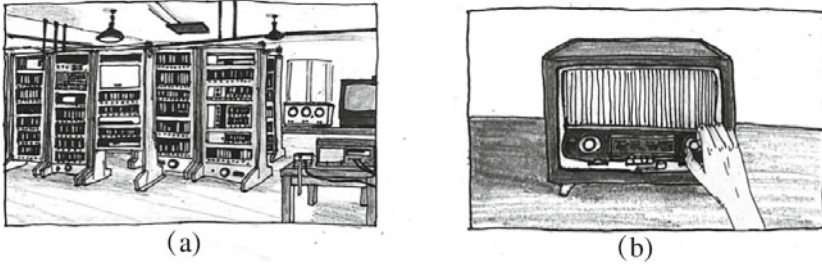
To glimpse how quantum mechanics has revolutionized our technology, consider a mobile phone chip. In a modern cell phone, billions of transistors are squeezed into a

chip the size of a fingernail, performing a billion operations per second! But it would not have been possible without quantum mechanics. It has long been known that metals can conduct electricity, whereas various gemstones, such as diamond, can not. But physicists were unable to explain these phenomena with classical physics. Using quantum mechanics, physicists not only were able to explain why some materials can conduct electricity and some can not, but also discovered semiconductors—a new class of materials with electrical conductivity between conductor and insulator. Using physical means, one can easily control the transport property of a semiconductor, switching it quickly between conductive and non-conducting. Building on the unique property of semiconductors, physicists invented transistor in 1947. In the decades to follow, engineers have continued to refine and develop the transistor technology, making transistors smaller and smaller. Today transistors on a modern computer chip are only around a dozen nanometers (about one hundred millionth of a meter) in scale. Many other technologies that we use in our daily life, such as fiber communication and magnetic resonance imaging (MRI), are also technologies brought by quantum mechanics.

Quantum communication and quantum computer that have sparked intensive interests in general public are new generations of quantum technology. In contrast to the classical communication in our daily life, quantum communication exploits the quantum effects to encrypt communications. The basic principles of quantum communication have been established in 1990s, and its technical implementation has relatively matured. However, there is still a lot of room for improvement at the technical level, and its potential applications demand further exploration.

A quantum computer functions like an ordinary computer, but operates on a very different principle—quantum mechanics. To emphasize the difference, we shall call the conventional computers that we use in daily life as classical computers. Scientists have discovered that quantum computers could be more powerful than classical computers. But so far scientists have only been able to demonstrate the supremacy of quantum computers in a few problems, such as the integer factorization and the random search, and it is not entirely clear as to why quantum computers are more powerful than classical computers. More importantly, to build a useful quantum computer is technically very challenging. Governments and large technology firms have invested heavily in quantum computing research for decades. But a quantum computer that can outperform classical computer in a useful and practical task is yet to emerge. In my opinion, a general-purpose quantum computer that can surpass classical computers are at least 50 years away. We will introduce and discuss quantum computation and quantum communication in detail in Chaps. 9 and 10.

There is an important and interesting difference between quantum technologies represented by mobile phone chips and quantum technologies represented by quantum computers. To facilitate the discussion, I shall refer to the former as implicit quantum technology and the latter explicit quantum technology. In both technologies, quantum mechanics plays a crucial role. However, an explicit quantum technology exploits unique quantum effects, such as the state superposition, quantum randomness, and quantum entanglement, to do things that ordinary classical technology can not even in principle. For example, the quantum key distribution in quantum



**Fig. 1.1** a Vacuum tube computer; b vacuum tube radio

communication and certain gate operations in quantum computers are infeasible, in principle, with any classical techniques. By contrast, what an implicit quantum technology can do, in principle, can also be done by classical technology. In other words, implicit quantum technology does not change the working principles that underlie classical technologies, it just makes the devices often dramatically much smaller, faster, and more sophisticated. Any logic gate operation in a modern transistor-based computer, which is an implicit quantum technology, can be also implemented in a vacuum-tube computer, which was already built before the transistors were invented. But a vacuum-tube computer was very large, housed in several rooms, and it was very slow (see Fig. 1.1a). Old radios, built with vacuum-tubes, are also very large in size (see Fig. 1.1b) compared to transistor radio. The hard drive we use nowadays is also an implicit quantum technology. This magnetic storage technology is deeply rooted in quantum mechanics: physicists have discovered that electrons have spins, which is a quantum property of matter. With spins, electrons behave like tiny compasses. With this discovery, physicists are able to explain why magnets are magnetic and further develop magnetic storage technology. Now, an ordinary hard disk can store thousands of books. Of course, we can also use the classical technologies, papers or films, to store the books. But these technologies are clumsy and slow compared to magnetic storage.

In implicit quantum technology, quantum mechanics is like a back stage boy: it plays a crucial role while never appears in the functionalities of the technology. In contrast, explicit quantum technology, quantum mechanics is really a star on the stage. An interesting fact is that the word ‘quantum’ is not used in naming or marketing implicit quantum technology. Transistor is called semiconductor technology, not quantum semiconductor technology; laser has never been called quantum light. Related products have never been marketed as quantum technology. When IBM sold its first transistor computers on market in 1955, it could have called them quantum computers. Fortunately, it did not.

Quantum communication can never replace classical communication because it is very fragile. Due to technical difficulties and complexities, quantum computers that outperform classical computers are still far away. As a result, quantum information technology has not yielded any product that can be used in our daily life. However, it has made “quantum” a buzzword in popular media, which has been taken advantage

by some merchants. They are marketing their commercial products with “quantum” explicit in the labels. The truth is, as far as I know, almost all existing commercial products with “quantum” explicit in the labels have nothing to do with quantum technology. Quantum dot display may be the only exception, which is interestingly an implicit quantum technology. All products of quantum technologies in our everyday life today are implicit quantum technology, they do not have ‘quantum’ in their names and brands.

Quantum mechanics has been very successful. Yet the world that it describes is very weird, and radically differs from our daily experience as demonstrated by the six unique attributes of quantum mechanics that we have summarized above. This kind of “quantum weirdness” has sparked profound discussion among philosophers. Unfortunately, it has also stirred wild speculations: some people relate quantum weirdness superficially to some yet-to-be-fully-understood phenomena, such as consciousness, and some even relate it to the religion, proposing things like “quantum buddism”. All of these are far-fetched. Quantum mechanics is a science, which has been rigorously tested by various experiments, and will continue to be driven and tested by experiments. As time passes by, the hullabaloo surrounding quantum or quantum mechanics will disappear, and only the true science of quantum will survive.

# Chapter 2

## A Brief History of Quantum Mechanics



The development of quantum mechanics is an exhilarating history packed with legendary heroes and stories. Each figure deserves to be commemorated and each breakthrough deserves to be celebrated. Let us remember these heroes:

**Planck, Einstein, Bohr, Bose, de Broglie, Pauli, Dirac, Fermi, Heisenberg, Schrödinger, Born ...**

Their achievements transcend the borders and belong to the whole world, and deserve to be documented and celebrated in books, music, movies, etc. Due to the space limitation, I can only give a brief introduction to these heroes and their contributions to the establishment of quantum theory.

### 2.1 The Birth of Quantum

Planck (Max Planck, 1858–1947) was by all accounts a traditional intellectual. He was born in 1858 in an intellectual family. His great-grandfather and grandfather were both theology professors; his father was a law professor. He grew up with a very good education. He composed songs, played piano, organ and cello, but he eventually chose to study physics. Planck had a very successful scientific career. He received his doctorate degree at age 21. At age 27, he was appointed as an associated professor of the University of Kiel. At age 31, Planck was named the successor to the position of Kirchhoff (Gustav Robert Kirchhoff, 1824–1887) in Humboldt University of Berlin, and became a full professor three years later. He was a man of integrity and honesty without any eccentricities or anecdotes. If he had not discovered “quantum”, he would prob-



MAX PLANCK.  
Planck (1858-1947)

ably have been buried in the dustbin of history like other typical intellectuals and professors of prestigious universities.

In 1894, Planck decided to study the problem of black-body radiation, which eventually leads to a revolution in physics. A black body is an object that absorbs all incident light. For example, the dark window in a distant building is approximately a black body as the light that gets in the window is deflected by the furnitures, irregular walls and other objects in the room and is very unlikely to get out from the same window. A black body emits light as it has temperature. Earliest studies of black-body radiation were made by Planck's predecessor Kirchhoff, who showed that black-body radiation is a universal phenomenon that does not depend on the material that the black body is made of. Wien (Wilhelm Wien, 1864–1928) later discovered a universal relation between the intensity and the frequency of the radiation. In the following five years, Planck published a series of articles on the black-body radiation, which did not contain substantial breakthrough, but merely new approaches for reproducing the known results, such as Wien's law.

At that time, experimental physicists at the Physikalisch-Technische Reichsanstalt in Berlin were performing measurements on the black-body radiation spectrum. These experimental studies were funded by industries. At that time, it was believed that the study of black-body radiation could help improve the lighting and heating technology. The physicists at the Reichsanstalt first measured the radiation at high frequencies, confirming Wien's law. After improving their measurement techniques, the experiments reached lower frequencies. By 1899, experimentalists have already noticed small deviations between Wien's law and the experimental results. By the fall of 1900, more significant deviations were observed at low frequencies, which could not be explained by experimental noise. Planck was among the first to know these experimental results. Facing the compelling experimental data, Planck revisited his derivation. He found out that by slightly revising the expression of entropy in the derivations of Wien's law, he could obtain a new formula for the black-body radiation

$$u(\nu) = \frac{8\pi b\nu^3}{c^3} \frac{1}{e^{a\nu/T} - 1}, \quad (2.1)$$

where  $\nu$  is the radiation frequency, and  $a$  and  $b$  are two constants. Planck found this formula is in perfect agreement with the experimental data. He announced his result at a meeting of the Berlin Academy of Sciences on October 19, 1900. But Planck was not satisfied with this result, because he did not understand why the entropy formula must be changed, and he tried to grasp the physics behind it. After more than a month, Planck found the answer. He assumed that the energy of an electric dipole oscillator in the radiation field is a multiple of an elementary energy unit  $h\nu$ , where  $\nu$  is the oscillation frequency and  $h$  is a constant. Using this assumption and Boltzmann's law of entropy, Planck re-derived the law of black-body radiation that he obtained one month earlier. His new result was

$$u(\nu) = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/k_B T} - 1}. \quad (2.2)$$

Compared to Eqs. (2.1), (2.2) only replaced the constants  $a$  and  $b$  by  $h$  and  $k_B$ . This replacement, while mathematically trivial, is revolutionary in physics.<sup>1</sup> The  $h$  is now known as Planck's constant, and  $k_B$  is Boltzmann's constant. By comparing with the experimental data, Planck found  $h = 6.55 \times 10^{-27}$  erg·sec, and  $k_B = 1.346 \times 10^{-16}$  erg/K.<sup>2</sup> According to the latest standards from the International System of Unites (SI), these two constants are defined as  $h = 6.62607015 \times 10^{-27}$  erg·sec and  $k_B = 1.380649 \times 10^{-16}$  erg/K.

Planck presented this result in a meeting of the Berlin Academy of Sciences on December 14th, 1900, marking the birth of the quantum theory.

Let us recap how Planck introduced "quantum" in his paper. Planck wrote in German [Annalen der Physik, vol. 4, p. 553 (1901)],

Es kommt nun darauf an, die Wahrscheinlichkeit  $W$  dafür zu finden, dass die  $N$  Resonatoren insgesamt die Schwingungsenergie  $U_N$  besitzen. Hierzu ist es notwendig,  $U_N$  nicht als eine stetige, unbeschränkt teilbare, sondern als eine discrete, aus einer ganzen Zahl von endlichen gleichen Teilen zusammengesetzte Grösse aufzufassen. Nennen wir einen solchen Teil ein Energieelement  $\epsilon$ , so ist mithin zu setzen

$$U_N = P\epsilon$$

wobei  $P$  eine ganze, im allgemeinen grosse Zahl bedeutet, während wir den Wert von  $\epsilon$  noch dahingestellt sein lassen.

In English, this paragraph says

It is now necessary to find the probability  $W$  that the total energy of  $N$  dipole oscillators is  $U_N$ . Here, the  $U_N$  must be interpreted as a discrete variable that is an integer number of some common energy unit, rather than a continuous and infinitely divisible one. We shall call this common energy unit  $\epsilon$ , so that we have

$$U_N = P\epsilon$$

Here  $P$  is an integer that is generally very large, whereas the value of  $\epsilon$  remains to be determined.

Through generations of efforts, quantum theory has developed into an immensely rich body of science that has been tested rigorously by experiments. It includes the theories that we commonly call quantum mechanics, as well as quantum field theories that describe fundamental interactions. At the same time, technologies based on the quantum physics, such as semiconductors and lasers, have changed and continue to change our daily lives. And it all started, remarkably, with this short paragraph. This is a monumental demonstration of the power and magic of a beautiful idea.

---

<sup>1</sup> Planck's derivation was still wrong. It was not until 1924 when Indian physicist Bose found the correct derivation for the black-body radiation for the first time.

<sup>2</sup> Erg is an old energy unit, 1erg =  $10^{-7}$  J.



## 2.2 The Difficult Start

Planck's black-body radiation law was a huge success, confirmed by more and more experiments. But Planck's quantum,  $h\nu$ , did not attract as much attention. At that time, physicists, including Planck himself, were not aware that the door of quantum mechanics had been opened. Nor did they anticipate that a gathering storm of quantum mechanics would sweep through physics in the years to come, revolutionizing human understanding of nature. Over the next few years, instead of developing the concept of quantum, Planck tried to find an explanation with classical physics, but failed. Since then, Planck had made no substantial contribution to the development of quantum theory. But such great results cannot simply be ignored. Lorentz (Hendrik Antoon Lorentz, 1853–1928), who began to study this problem in 1903, concluded that Planck's quantum and classical theories could not be reconciled. Due to the importance of Lorentz in physics at the time, Planck's quantum theory began to attract more attention from physicists, but was still largely ignored.

These developments caught the attention of a young clerk at the Swiss patent office in Bern, whose name was Einstein (Albert Einstein, 1879–1955). This man was gifted with a remarkable ability to uncover the new physics behind the familiar formula. Let us recall the Planck's black-body radiation law (2.2). For large frequencies,  $h\nu \gg k_B T$ , we have  $e^{h\nu/k_B T} \gg 1$ , so the unity in the denominator can be neglected, giving

$$u(\nu) = \frac{8\pi h\nu^3}{c^3} e^{-h\nu/k_B T}. \quad (2.3)$$

This is exactly Wien's law. Note that this formula contains Planck's constant  $h$ . Modern physicists know that if a formula contains Planck's constant  $h$ , it describes a quantum phenomenon or process. Wien first derived this formula in 1896, and Planck re-derived it a few years later. But neither Wien nor Planck saw the quantum physics hidden behind this famous law.

In 1905, Einstein discovered the quanta hidden behind this well-known formula. By making an analogy with the entropy of a classical gas, Einstein found that the black-body radiation could be regarded as a special kind of gas consisting of "photons", each with an energy  $h\nu$ . In the paper published in 1905 [Ann. Phys., 1905, 17: 132], Einstein used the term energy quanta or light quanta instead of "photon". But, obviously, he was clear that light has the properties of a particle. Einstein's understanding of light was a remarkable advance with respect to Planck's. In this paper, Einstein stated explicitly that the behavior of a particle and wave is fundamentally different, and that although light is widely regarded as a wave, it behaves more like a particle in many phenomena, such as the black-body radiation, fluorescence, and photocathode radiation. He stated that the



Einstein (1879-1955)

purpose of his paper is to clarify this understanding and to establish the underlying principles. In the second half of his paper, Einstein explained the photoelectric effect in terms of light quanta: When these “photons” collide with electrons in a metal, they are either absorbed completely or not absorbed at all. In 1905, Einstein also introduced special relativity. But in his private correspondence with friends, he considered his quantum theory of light, instead of special relativity, as “revolutionary”, because almost everyone in the physics community at that time viewed light as electromagnetic waves that obey Maxwell’s equations, not as particles.

Planck, Lorentz and Einstein had very different attitudes towards “quantum”. Planck was somewhat reluctant, believing that “quantum” was just a trick he had to borrow temporarily in the derivation and that it would automatically disappear in some improved derivation. Lorentz was also skeptical about “quantum” at first, but after some investigations, Lorentz was convinced that “quantum” could not be reconciled with classical physics. However, he did not further develop or advertise the idea of “quantum”. Genius Einstein, on the other hand, immediately recognized that “quantum” was a revolutionary idea. Not only did he further develop the concept, he also immediately applied it to explain the photoelectric effect, for which he was awarded the 1921 Nobel Prize in physics.

Planck incidentally nudged to open the door of quantum mechanics, and then returned to classical physics. Lorentz realized that there was a very different world behind the door, but he had no intention or power, to step inside. Einstein, on the other hand, completely pushed the door open and bravely stepped inside. In 1905, Einstein also proposed the special theory of relativity that made him world famous. However, in the next five years, Einstein devoted more time to develop the quantum theory instead of the relativity.

By then, Einstein was just a young clerk at the Swiss patent office in Bern. His quantum theory of light and the explanation of the photoelectric effect did not cause an immediate impact, and were hardly discussed in the physics community. But young Einstein continued to forge ahead in the quantum world. In 1907, Einstein made a major progress when he applied Planck’s quantum theory to an entirely different subject, the specific heat of solids. Einstein believed that the energy of atomic vibrations in solids was also quantized, and that it should also obey Planck’s law of black-body radiation. Physicists had been able to lower the temperature to  $-250^{\circ}\text{C}$  in laboratories. They found in experiments that the specific heats fell markedly at low temperatures. The classical theory could not explain this phenomenon at all. By employing Planck’s law, Einstein found that the specific heat of solids does indeed go down with temperature, and his own derivation agreed very well with the published experimental results. Still, Einstein’s new result was not immediately celebrated, and majority of physicists were not interested in quantum theory. But Einstein’s work caught the attention of a chemist, Nernst (Walther Nernst, 1864–1941), who immediately recognized the significance of Einstein’s quantum theory. Nernst not only began to develop and apply Einstein’s quantum theory himself, he also encouraged his colleagues and assistants to do so. This was already in the year of 1910.

In 1911, the first Solvay Conference, advocated by Nernst, was held in Brussels. The subject was “Radiation and the Quanta”. Lorenz was chairman of the conference. Einstein was invited to give a talk on “The Problems of Specific Heat”. This Solvay Conference marks a turning point in the history of quantum theory, after which began the full development of quantum mechanics.

## 2.3 Hydrogen Atom

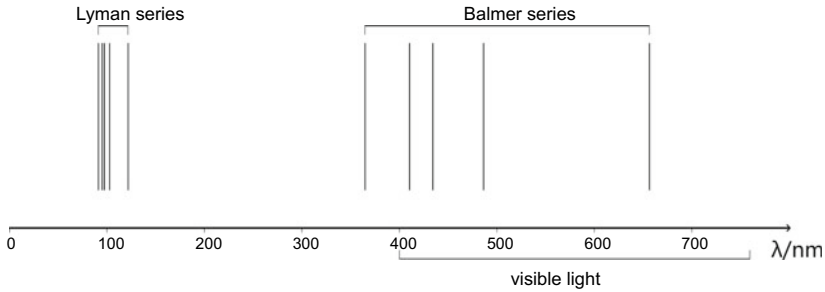
After the first Solvay conference, quantum theory became the forefront of physics. The number of papers on the subject grew explosively. In 1913, Bohr proposed the quantum theory of hydrogen atom, which was another major milestone in the development of quantum physics. To understand Bohr’s work, we need to first review a bit of history.

By the end of the 19th century, classical physics has been so accomplished that there were prevailing optimism among physicists that what remains was just to decorate this well-built edifice of physics. In a famous speech in April 1900, Sir Kelvin (William Thomson, 1st Baron Kelvin, 1824–1907) declared that only two dark clouds still obscured the sky of physics—the ether problem and the specific heat problem.<sup>3</sup> However, not everyone was optimistic, because classical physics did not answer a very fundamental question—what is the world made of? Through the studies of thermodynamics and statistical mechanics, many physicists accepted the idea that matter is made up of atoms and molecules. But there was no direct experimental evidence as to their existence, and not everyone agreed to this viewpoint. For example, Mach (Ernst Mach, 1838–1916) famously declared, “I don’t believe that atoms exist!” Even one accepted the atomic hypothesis, it remained unclear what an atom is: is it made of smaller particles or is it a vortex of the ether?

Before and around the rise of quantum theory, there were continuous developments in experimental techniques. Experimental physicists had increased the spectral resolution, achieved lower temperatures, and realized better vacuums. These developments had drastically improved the experimental precision, expanding the scopes of observation and allowing for more accurate results. As already mentioned, with the capability to reach lower temperatures, physicists discovered changes in the specific heat of solids or gases. Furthermore, by replacing Newton’s prism with the grating, physicists were able to analyze in detail the spectra of many atomic or molecular gases, finding them were discrete (see Fig. 2.1). Based on these experiments, Balmer (Johann Balmer, 1825–1898) discovered an empirical relation among some of the spectral lines of hydrogen atom in 1885. In 1888, Rydberg (Johannes Rydberg, 1854–

---

<sup>3</sup> There is a widely circulated claim that the two dark clouds refer to the ether problem and the black-body radiation problem. This is false! Sir Kelvin’s lecture was finally collated and published [The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(7): 1–40]. In the paper, entitled *Nineteenth Century Clouds Over the Dynamical Theory of Heat and Light*, Sir Kelvin did not mention the black-body radiation at all.



**Fig. 2.1** Spectral lines of hydrogen. For clarity, only the Balmer series and Lyman series are shown here

1919) summarized these findings into a more general empirical relation that is now known as the Rydberg formula,

$$\frac{1}{\lambda} = R_H \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right), \quad (2.4)$$

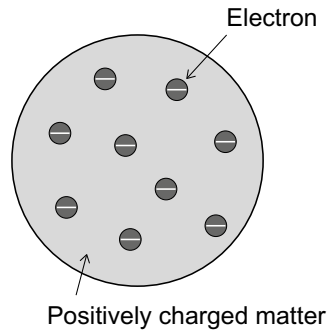
where  $\lambda$  is the wavelength,  $R_H$  is a constant, and  $n_1$  and  $n_2$  are positive integers. These regular spectral lines were mysterious to physicists at the time. As we will see, this empirical relation actually reflects the motion of electrons inside the atom, and it is as important to quantum mechanics as the Kepler' three laws of planetary motion to the law of gravity.

Thomson (J. J. Thomson, 1856–1940) was a teenage genius who was admitted to university at the young age of 14. In 1884, the 28-year-old Thomson was appointed Cavendish Professor of Physics at the University of Cambridge, where he switched from the study of mathematics and theoretical physics to experiments. In 1897, Thomson carefully studied cathode rays. By creating a vacuum cathode tube, he was able to measure the properties of cathode ray particles with high precision. Thomson found that the mass and the charge of the emitted particles were the same no matter what atoms they came from, and that the mass of the particle was less than one thousandth of the mass of a hydrogen atom. The particle discovered by Thomson is the electron. Based on this discovery, Thomson started to theorize about a model of the atom using his deep knowledge. He hypothesized the atom was a ball of positively charged matter in which point-like electrons were rather uniformly scattered (see Fig. 2.2). While Thomson's model of atom only qualitatively agreed with the experimental results, it was the most accepted model until 1910.



Thomson (1856-1940)

**Fig. 2.2** Thomson's model of atom. Positive charges are uniformly distributed in a ball, and electrons with negative charges are scattered rather evenly in this uniform positively charged medium



Thomson was not only successful in his own research, winning the Nobel Prize, he also trained eight Nobel Prize winners, including Rutherford and Bohr to be introduced below. His son also won a Nobel Prize.

Rutherford (Ernest Rutherford, 1871–1937) was born in New Zealand. At age 24, he traveled to England to study at University of Cambridge, where he was a graduate student of J. J. Thomson. At age 27, Rutherford became a professor at McGill University, Canada, thanks to Thomson's recommendation. In 1907, Rutherford returned to Britain to take the position of professor at the University of Manchester. At McGill, Rutherford systematically explored the radioactivity and discovered phenomena such as alpha-ray and beta-ray, and studied radioactive element's half-life. Because of these work, he was awarded the Nobel Prize in chemistry in 1908.



Rutherford (1871-1937)

After he became the Nobel laureate, Rutherford continued to work relentlessly. In 1909, already the chair of physics at the University of Manchester, Rutherford performed his most famous experiment. Along with his assistants Geiger (Johannes "Hans" Geiger 1882 -1945) and Marsden (Ernest Marsden, 1889–1970), he bombarded a thin gold foil using alpha particles. Surprisingly, they found that alpha particles could be deflected with very large angles. According to Thomson's model, the positive charge is uniformly distributed inside the atom. Because the mass of the electron is much smaller than the positively charged alpha particle, alpha particles were expected to pass right through the atom with very small deflection angles. Based on the experimental findings, Rutherford boldly abandoned the Thomson's model and created his own. Rutherford postulated that the atom had a very small nucleus at the center, containing much of the atom's mass. But Rutherford did not specify how the electrons were distributed in the atom. The Rutherford model did not immediately attract attention. In 1912, a young Danish man named Bohr came to his laboratory, who put the understanding of atom in a totally new perspective.

Bohr (Niels Bohr, 1885–1962) was born in Denmark. His father was a professor of physiology at the University of Copenhagen, and his mother was also well-educated. He received a good education at an early age and was a passionate footballer. Bohr had a young brother, Harald Bohr (1887–1951). Although Harald was two years younger, he seemed to be better than his brother in everything. Harald was a better footballer than his brother and played for the Danish national team at the 1908 Olympics. He received his master's degree one year earlier than his brother and his doctorate one year earlier as well. But in the end, his brother Niels Bohr became the more famous Bohr.

Bohr received his doctorate in April 1911. His thesis was on the electron theory of metals. In the thesis, Bohr came to a very important conclusion: the electron theory of metals at that time could not explain the magnetic properties of irons. In modern terms, the classical theory can not explain the magnetic properties of materials, which is today known as the Bohr-Van Leeuwen theorem. This result left a very deep impression on the young Bohr that the classical theory was flawed. In September of the same year, Bohr, supported by a fellowship from the Carlsberg Foundation (yes, that beer company), did some research on cathode rays in Thomson's Cavendish Laboratory.

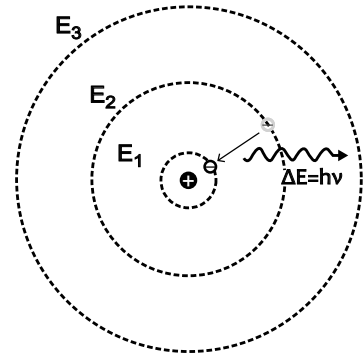
Having failed to impress Thomson, Bohr received an invitation from Rutherford to conduct research at the University of Manchester in early 1912. Bohr was immediately attracted to the Rutherford's model of atom. As Bohr studied electrons during his PhD, he began to think how the electrons are distributed in order to stabilize the atom. By the summer of 1912, Bohr had formulated his model, and described his ideas to Rutherford in text. Bohr believed that in order to stabilize the atom, it was necessary to introduce the concept of quanta. In 1913, Bohr published a series of three papers, announcing his model of atom to the world.

Bohr's model hinges on two key postulates: (1) the electron can only be in certain quantized orbits, which have discrete energy levels  $E_1, E_2, E_3, \dots$ ; (2) an electron can jump to a higher-energy orbit by absorbing a photon, or it can drop to a lower-energy orbit by emitting a photon. The energy of the absorbed or emitted photon equals to the energy difference between the levels:  $h\nu = |E_i - E_j|$  (see Fig. 2.3). Again Planck's constant  $h$  appears. Einstein had described the atomic vibrations in solids in terms of quanta; now Bohr had a quantum description for the inner structure of an atom. It was a milestone development. Upon a closer inspection, Bohr's model seemed rather weird and artificial. Why must the energy levels of electrons be discrete? But physics is not mathematics. Physicists are more concerned about whether the theory agrees with the experiment. Bohr's theory not only explained the known spectral lines of the hydrogen atom, reproducing the Rydberg Formula (2.4), it also predicted new spectral lines in the ultraviolet regime, which was verified by



Bohr (1885-1962)

**Fig. 2.3** Bohr's model of hydrogen atom: the electron can only be in certain discrete or quantized orbits with discrete energy levels  $E_1, E_2, E_3, \dots$ . When an electron jumps from energy level  $E_2$  to  $E_1$ , it emits a photon with frequency  $\nu = (E_2 - E_1)/h$



experiments one year later. Bohr's model also explained the Pickering series of the ionized helium, which had puzzled physicists for a long time.

Unlike Einstein's quantum theory of light, Bohr's work was quickly acknowledged, attracting more physicists to the quantum theory. Among them, a prominent figure is Sommerfeld (Arnold Sommerfeld, 1868–1951), who quickly extended Bohr's concept of discrete energy levels to more physical systems, providing a more general "quantization" rule. Using this generalized theory, Sommerfeld found electrons should have three quantum numbers instead of just one as in Bohr's model. Sommerfeld's theory could explain more atomic phenomena, such as the Stark effect and the Zeeman splitting.

When the younger generation of physicists bravely explored in the quantum world, making one success after another, some older physicists either felt lost or stood on the fence. Lorentz was never actively engaged in the development of quantum theory, even though he already recognized the inadequacy of classical theory in 1903. Planck opened the door to quantum mechanics, but he tried to derive the black-body radiation law from the classical theory until 1914.

## 2.4 The Crisis

Bohr and Sommerfeld's quantum theory was far from perfect despite its great success. Bohr himself was well aware of it. Bohr's model worked very well for explaining the frequencies of the spectral lines of hydrogen atom, but it could predict neither the intensity of the spectral lines nor the polarization of the emitted photons. To refine his theory, Bohr used his physical intuition to formulate a correspondence principle, which assumes that the electronic transition probability between energy levels obey Maxwell's classical equations. Combined with Einstein's theory of spontaneous and stimulated emissions, Bohr obtained a selection rule for the transition between energy levels. Using Bohr's correspondence principle, the Dutch physicist Kramers (Hendrik Anthony "Hans" Kramers, 1894–1952) successfully explained the intensities of all

the spectral lines and the polarization of emitted photons in a hydrogen atom, which agree well with the experimental results.

But these efforts were still inadequate. The Bohr-Sommerfeld theory could not explain a lot of experimental observations. In particular, it could not describe atoms or molecules with two or more electrons. For example, it could not give the correct spectral lines of helium atom and could not describe the covalent bonds between molecules. Moreover, the framework of the theory appeared rather handwaving. By 1924, atomic physicists felt that the Bohr-Sommerfeld model needed some major revision. In a paper published in 1924, Born (Max Born, 1882–1970) began to call forth a new kind of “quantum mechanics”. Two years later, a new quantum theory was indeed constructed, solving all the difficulties that stymied the Bohr-Sommerfeld theory.

The years from 1900 to 1924 saw the early development of quantum physics, with limited progress achieved. At that time, almost all discussions were centered on the “quantumness” of energy: the radiation energy was discrete; electrons could only be in discrete energy levels. Einstein’s theory of light quanta is an exception, which provided the starting point for what is now known as the wave-particle duality. But at that time, no one took a further step to develop Einstein’s idea. In retrospect, the quantum theory during this period was actually quite ugly, full of flaws and inconsistencies: the derivation of Planck’s black-body radiation law was wrong; Einstein’s theory of the specific heat of solids was obtained by stretchy analogy; Bohr obtained the energy levels of hydrogen atoms in a heuristic way. These shortcomings made many older physicists uncomfortable that they chose to stand by. The younger generation, albeit aware of these flaws, looked at the positive side, in particular, the ability of quantum theory to explain the experiments that could not reconcile with the classical theory.

From 1924 to 1926, with one brilliant breakthrough after another, quantum physics took a great leap and blossomed into a beautiful theory that has stood till today. In the three years, a group of bright, hard-working, and brave young physicists, with diverse personalities and little direct collaboration, developed the theoretical foundations for quantum mechanics. These young physicists came from all over the world, communicating only through personal letters and academic journals. The basic concepts and theoretical framework described in today’s books on quantum mechanics can all be found in papers published before the end of 1926. *The Principles of Quantum Mechanics*, a monograph on quantum mechanics by Dirac (Paul Adrien Maurice Dirac, 1902–1984) that was first published in 1930, remains to this day a must for students in physics major. It is no exaggeration to say that these three years were not only one of the most glorious chapters in the history of science but also in human history. Unfortunately, little is known about this glorious history to the general public.



## 2.5 Identical Particles

Our everyday experience tells us that two objects, no matter how similar they are to each other, can always be distinguished if we observe them carefully enough. For example, two coins of dime can be told apart in many situations with naked eyes. If not, we can resort to microscopes with large amplification. When we say two objects are identical, what we really mean is that the difference between these two objects is not important and can be ignored for the matter we are concerned with. When we use a coin to buy something, we don't care if it is slightly defected; we simply treat it as the same as any other coin, because it buys an object of equal value. We are more careful when we want to bet on winning or losing by tossing a coin. If there are two coins, one with a dent and the other intact, we use the intact coin. If both coins are intact, we regard them as the same and choose one at random, even though we know that the two coins look different under a microscope. To sum up, two objects are only approximately identical, and we can always distinguish them if we are careful enough. We ignore these small differences only because they are not important for our purpose.

But physicists have found that two photons are completely identical: there is no way to distinguish two photons. We can only say that one photon has a frequency  $\nu_1$  and another photon has a frequency  $\nu_2$ ; we cannot say that photon 1 has frequency  $\nu_1$  and photon 2 has frequency  $\nu_2$ . Similarly, two electrons are identical, two water molecules are identical, two fullerene ( $C_{60}$ ) molecules are identical, and so on. That is, this identity at the microscopic level is perfect and absolute.

This is one of the most fundamental differences between quantum mechanics and classical mechanics. In quantum mechanics, the indistinguishability of particles are absolute, not approximate. The first person to discover the quantum indistinguishability of microscopic particles was Indian physicist Bose (Satyendra Nath Bose, 1894–1974). In early 20th century, science in India lagged far behind Europe. Still, new scientific results, including the emerging quantum physics, found their way into India through journals and books, and they stimulated intensive interests among the young generations in India.

Bose was born in Calcutta, India. His father worked in the East Indian Railway Company, but later started his own company. His mother was from a family of lawyers and was well educated. His schooling began at the age of five. Bose excelled in school. In 1909, Bose entered Presidency College, where he received a bachelor degree of science in 1913 and his master's degree in 1915. As India was not developed in science and education, Bose was not able to pursue further studies. After working as a private tutor for a year, he took the opportunity to join the Science College in Calcutta University as one of the first lecturers in physics. He and his colleagues borrowed physics



SATYENDRA NATH BOSE  
Bose (1894-1974)

books and journals from a friend who studied in Germany, and taught themselves before teaching students. In 1921, Bose joined University of Dhaka with a high salary, where he set up a new physics department. Here Bose wrote his most famous paper.

In this paper, Bose used a novel method to derive Planck's law of black-body radiation. As mentioned earlier, Planck was not satisfied with his own derivation and tried various approaches to improve it. In retrospect, Planck's efforts were doomed to fail because he always attempted to go back to classical physics. Bose introduced a new concept that represented a radical departure from the classical ideas, i.e., photons are identical particles. Based on this new concept, and light quanta, Bose gave the first correct derivation of the black-body radiation law in human history.

Bose's breakthrough was revolutionary. Up until that time, no one had realized that quantum and classical physics could be so fundamentally different: in the quantum world, being identical is absolute; in the classical world, being identical is only an approximation.

But Bose had difficulty publishing his article. He submitted his article to a British journal but it was rejected. On June 4, 1924, he sent the article directly to Einstein, hoping he could help publish it in a German journal. Einstein immediately recognized the importance of the paper. On July 2, 1924, in a postcard to Bose, Einstein wrote that he had translated the article into German and had it published in a German journal. Moreover, Einstein immediately extended the idea of identical particles from photons to particles with mass. Einstein published three consecutive papers on this subject, in which Einstein predicted the famous Bose-Einstein condensation. Seventy years later, in 1995 in a laboratory at JILA, a joint institute of University of Colorado, Boulder, and NIST, physicists confirmed Einstein's prediction in ultracold atomic gases.

So how did Bose make this breakthrough? It was, in my opinion, by accident. Let's take a look at Bose's letter to Einstein on June 4, 1924,

Respected Sir:

I have ventured to send you the accompanying article for your perusal and opinion. I am anxious to know what you think of it. You will see that I have tried to deduce the coefficient  $8\pi\nu^2/c^3$  in Planck's law independent of classical electrodynamics, only assuming that the ultimate elementary region in the phase-space has the content  $h^3$ . I do not know sufficient German to translate the paper. If you think the paper worth publication I shall be grateful if you arrange for its publication in *Zeitschrift für Physik*<sup>4</sup>.

Though a complete stranger to you, I do not feel any hesitation in making such a request. Because we are all your pupils though profiting only by your teachings through your writings...

Yours sincerely

N. Bose

Bose did not mention that photons are indistinguishable in his letter, nor did he explicitly mention it in his paper. Here is one plausible explanation. In his derivation, Bose needed to put photons into what he called "ultimate elementary regions"

---

<sup>4</sup> A German physics journal.

and calculated all possible combinations. In this process, he treated photons as indistinguishable particles without actually realizing it. Had he treated the photons as distinguishable particles, he would have obtained a different number of combinations and could not reproduce Planck's law. But Einstein realized it at once and quickly extended the idea.

I find it hard to resist the temptation to speculate that Bose might not have made such a "brilliant" mistake if he were able to continue his studies (in India or in Europe) to improve his academic training.

At the same time, three young geniuses independently worked on the problem of identical particles. They were Pauli (Wolfgang Pauli, 1900–1958), Fermi (Enrico Fermi, 1901–1954), and Dirac. Pauli was born in Austria to a chemist; his mother is a writer's daughter. His godfather is the famous physicist Mach. Pauli showed his gift at a very young age. He published his first paper at age 18 on the theory of general relativity, just two months after graduating from the high school. He worked under Sommerfeld and received his doctorate in 1921. Pauli was a perfectionist, not only trying to be perfect himself, but also criticizing with no mercy the "imperfect" work of others. Perhaps he was so obsessed with perfection that he seldom published papers, and many of his contributions can only be found in his personal letters to colleagues.



Pauli (1900-1958)

Fermi was born in Rome to a government employee, and his mother was an elementary school teacher. As a young boy, Fermi was interested in playing with electrical and mechanical toys, and read any books on physics and mathematics that he could get his hands on. After graduating from high school, Fermi took the university entrance exam, which included an essay on the theme "Specific Characteristics of Sounds". Fermi chose to use Fourier analysis to solve the differential equation for a vibrating rod. The chief examiner was so impressed and gave him the highest score. Although Italy was Galileo's homeland, Italian physics at that time was far behind Germany, England and France. In the university, Fermi remained largely self-taught. The university professors found that they had nothing to teach Fermi. Instead, they often asked his help for solving problems and even assigned him to organize seminars on quantum physics. Fermi received his doctorate in 1922. He was one of the few physicists who was proficient in both theory and experiment.



Fermi (1901-1954)

Dirac was born in 1902 in Bristol, England. His father was a teacher and his mother worked at the library. Dirac was educated in the technical school where his father taught French. Apart from the usual curriculum, Dirac took technical subjects like drafting and metal work. Dirac came first in almost every class. He then studied electrical engineering in the University of Bristol. In the university, in addition to the mandatory courses, Dirac taught himself a lot of physics and mathematics, including the theory of relativity. When he completed his degree in 1921, he had an opportunity to study at Cambridge University, but the scholarship fell short of the amount of money



Dirac (1902-1984)

required to live and study at Cambridge. With bachelor's degree in engineering, he was unable to find work as an engineer. So he returned to the University of Bristol to study mathematics. In 1923, Dirac graduated again and was offered a higher scholarship, which was enough for him to pursue his interests in science at Cambridge. Dirac was a loner, with taciturn nature, and was awkward in socializing with people. There were many anecdotes about Dirac. Once in a seminar, someone asked Dirac, "I don't understand the equation in the upper right corner of the blackboard." Dirac did not respond for a considerable time. The host tried to break the ice and prodded Dirac politely. Dirac replied: "That wasn't a question, it was just a comment." By modern medical standards, Dirac was probably with autism, but that miraculously didn't affect his scientific research at all.

After receiving his doctorate, Pauli moved to the University of Göttingen as the assistant to Born. In 1922, Bohr visited Göttingen and gave a series of comprehensive lectures on how he used quantum theory to explain the periodic table of elements. Despite some progress, Bohr was unable to solve one of the greatest difficulties: why don't electrons aggregate at the lowest energy levels? This question had been on Pauli's mind ever since. After more than three years of thinking and research, and inspired by the results of others, Pauli solved this problem in 1925. To explain the periodic table, Pauli made two assumptions: (1) apart from the spatial degrees of freedom, electrons have a strange new degree of freedom; (2) no two electrons can exist in the same quantum state. The first hypothesis was quickly confirmed: the new degree of freedom is spin. The second hypothesis is now known as the Pauli exclusion principle.

Fermi had been thinking about whether electrons can be distinguished since 1924. As mentioned earlier, the quantum theory of Bohr and Sommerfeld was unable to explain the spectral lines of helium atom. Fermi conjectured that the main reason was that the two electrons in the helium atom were identical and could be distinguished, but he didn't know how to pursue this idea in a quantitative manner. After reading Pauli's paper, Fermi immediately realized what he had to do. In 1926, he published two papers with almost identical results. The first paper was short and written in Italian; the second paper was written in German, containing more detailed derivations and results. In his papers, Fermi described a new quantum gas of identical particles,

where each quantum state can only be occupied by at most one particle. What is the difference between the Fermi gas and the identical particles discussed by Bose and Einstein? For the identical particles discussed by Bose and Einstein, many particles can occupy the same quantum state.

A few months later, Dirac revisited this problem with a new approach, providing a systematic description of identical particles. Dirac proved that there are only two kinds of microscopic particles: bosons and fermions. Photons and hydrogen atoms are bosons, while electrons and protons are fermions. Bosons obey the Bose-Einstein statistics, where multiple identical particles can occupy the same quantum state. Fermions obey the Fermi-Dirac statistics, where multiple identical particles cannot occupy the same quantum state.

## 2.6 It's Matrix

While Bose, Einstein, Fermi, and Dirac were developing the idea of identical particles, other people including Heisenberg (Werner Heisenberg, 1901–1976) and Born were making breakthroughs in a different direction, formulating “quantum mechanics” that Born had dreamed.

Heisenberg was born in Germany in 1901. His father was a secondary school teacher who later became a professor at the University of Munich. His mother was the daughter of a headmaster. In his late teenage years, Heisenberg excelled in his academic studies, studied classical music, and was an accomplished pianist. In 1920, he studied at the University of Munich where his father was a professor. At first, he wanted to study mathematics with an old professor Ferdinand von Lindemann (1852–1939), but was rejected. After discussing with his father, he chose to study physics under Sommerfeld, and became Pauli's classmate. Like Thomson, Sommerfeld has trained many Nobel Prize winners, most notably Heisenberg and Pauli. Sommerfeld let Heisenberg attend his advanced seminars with other senior students. Heisenberg did not disappoint his teacher. One year later, he proposed a new model of the atom that explained the outstanding problem of the anomalous Zeeman effect. While this model still had many flaws from the modern point of view, Heisenberg showed his unique character in this work: he was willing to abandon the old theory to explain the experiment. A credo of the quantum theory at that time is: the quantum number must be an integer. Heisenberg's model introduced half-integers. This not only shocked his teacher Sommerfeld, but also Pauli, who objected vehemently: If  $1/2$  can be a quantum number, so can  $1/4$ ,  $1/8$ ,  $1/16$ , . . . , and there would be no discrete energy levels. Heisenberg's reply to this criticism was “Success sanctifies means.”



Heisenberg (1901-1976)

This work, while controversial, earned Heisenberg many opportunities. He was invited by Born to Göttingen for one year, where he met Bohr and had in-depth discussion. Heisenberg impressed both Born and Bohr. Born wanted Heisenberg to work in Göttingen after his doctorate, and Bohr invited him to visit Copenhagen at an appropriate time. From 1922 to 1925, Heisenberg commuted among the three centers of quantum theory: Copenhagen, Göttingen, and Munich, learning from the masters of quantum theory. He developed a deep understanding about the difficulties of the old Bohr-Sommerfeld quantum theory and began to think about how to break through. Despite his rapid growth, there were shortfalls in Heisenberg's structure of knowledge, some of which appeared to be quite alarming. For example, Heisenberg was unable to answer a few simple questions from Professor Wien when defending his doctoral thesis: What determines the resolution of a microscope? How does a battery work? But Heisenberg, with his obvious shortcomings, was to accomplish a complete breakthrough in quantum theory.

By June 1925, Heisenberg had formulated an early version of the new quantum theory. He realized that he had to abandon old concepts like electron orbits, and focus only on quantities that can be observed experimentally. In classical physics, we can observe the planet orbits around the sun or record the path of a flying spacecraft, but electron orbits are not observable. At that time, only the intensity of the electronic transition between energy levels could be observed. Heisenberg began to construct a theory on the transition intensity. Soon, he ran into trouble with the calculations, because the multiplication of observables was odd. Around the time Heisenberg was suffering from hay fever allergy and he decided to leave Göttingen to recuperate on an island called Helgoland, where there was few plants. After staying there for ten days, Heisenberg not only recovered from his illness but also completed his calculations.

In September 1925, Heisenberg published a seminal paper in *Zeitschrift für Physik*, entitled "Quantum theoretical re-interpretation of kinematic and mechanical relations". Heisenberg wrote that this paper aims to "to establish the foundation of quantum mechanics, which only involve relations between the observables". Heisenberg found that these observables could be denoted by variables with two indices and the multiplication of these observables is non-commuting. That is, if  $A$  and  $B$  are two observables, then  $AB \neq BA$ . Heisenberg himself was not sure what these variables were and did not have much confidence in his new theory. These variables with two indices made the calculations very complicated, and Heisenberg did not know how to obtain the energy levels of hydrogen atoms with his new theory. So in this paper he turned to a simple example, a harmonic oscillator.

Born immediately recognized the importance of Heisenberg's work. He realized that the odd observables proposed by Heisenberg are in fact matrices. Along with his student Jordan (Pascual Jordan, 1902–1980), Born quickly proved the commutation relation between the two observables—momentum and position. In November 1925, Born, Heisenberg and Jordan published a joint paper that clearly established the basic framework of matrix mechanics. By early 1926, Pauli and Dirac independently derived the energy levels of the hydrogen atom using matrix mechanics.

At age 20, Heisenberg bravely introduced half-integer quantum numbers. At age 24, Heisenberg made a breakthrough and created the matrix mechanics.

## 2.7 Particles Are Waves and Waves Are Particles

While Heisenberg commuted among the golden triangle of quantum physics at that time—Göttingen, Copenhagen and Munich—in search of a new quantum theory, a completely different line of thought was pursued outside the golden triangle. These efforts eventually led to the emergence of an alternative version of quantum mechanics—the Schrödinger equation for wave function.

De Broglie (Louis Victor Pierre Raymond de Broglie, 1892–1987) belonged to the famous aristocratic family of Broglie in France. When his brother, 6th Duke de Broglie, died in 1960, he became the 7th Duke de Broglie. De Broglie's early interest was in literature and history, and received his first degree in history at age 18. Afterwards he turned his attention toward science, and received a degree in science at age 21. With the outbreak of the First World War, de Broglie served the army, developing radio communications in the Eiffel Tower. This experience with waves had a long-lasting influence on de Broglie. When the war ended in 1918, he began to study physics, participating the research at his brother's laboratory. But de Broglie personally was more interested in theoretical physics, especially, the newly emerged quantum physics.



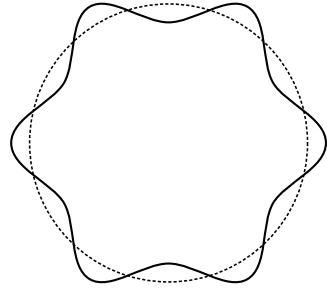
de Broglie (1892-1987)

In 1923, de Broglie made progress on quantum physics, and wrote several papers in succession. But his theories attracted little attention. In early 1924, de Broglie assembled these results in his doctoral thesis, and sent it to the famous French physicist Langevin (Paul Langevin, 1872–1946) for review. When Langevin read his paper, he found that de Broglie's ideas were quite new. To avoid jumping to conclusions, he asked for another copy of the thesis from de Broglie and sent it to Einstein. Einstein immediately recognized the importance of de Broglie's work, and he wrote to Langevin that de Broglie had "lifted a corner of the great veil". In his paper on the Bose statistics in 1925, Einstein brought de Broglie's theory to the attention of the world. So what kind of novel theory did de Broglie proposed in his doctoral thesis?

Let us recapitulate Einstein's seminal paper on light quanta in 1905, where Einstein proposed that light is a particle and used it to explain the photoelectric effect. By 1916, experimental physicists have unambiguously verified Einstein's formula for the photoelectric effect. Yet still, the majority of physicists rejected Einstein's idea that light is a particle. The reason was simple: a large number of experiments and Maxwell's equations tell us that light is a wave. How can something be both a wave and a particle? Almost all physicists at the time thought this was impossible. De Broglie seemed unaffected by this traditional view and took a more positive attitude. He conjectured that, if light, which everyone thought was a wave, could be a particle, then a particle could also be a wave. For example, an electron can be a wave. In his doctoral thesis, de Broglie developed an extensive mathematical formulation around



**Fig. 2.4** According to de Broglie, electrons are waves. The electron in a hydrogen atom forms a standing wave for each quantized orbit (dashed circle) in Bohr's model



this idea. First, he argued that if the momentum of a particle is  $p$ , then its wavelength is  $\lambda = h/p$ . Second, he argued that since electrons are waves, electrons can form standing waves around protons (see Fig. 2.4). Following this line of thought, de Broglie magically re-derived the orbits and energy levels in Bohr's model of hydrogen atom. Finally, de Broglie predicted that electrons could also interfere just like other waves. This prediction of de Broglie was later confirmed by experiments, for which he was awarded the Nobel Prize in 1929.

As a former student in liberal arts, de Broglie was unknown in physics community at that time. After proposing the wave-particle duality, de Broglie became the only French physicist who made foundational contribution to quantum mechanics.

It was time for Schrödinger to take the stage to finish the last, yet extremely important, chapter of quantum mechanics. Schrödinger (Erwin Rudolf Josef Alexander Schrödinger, 1887–1961) was born in Vienna in August 1887. His father was a botanist, and his mother was daughter of a professor. Schrödinger's early academic career was quite similar to Planck's. Although he was successful and became a full professor at the University of Zurich, he did not have particularly impressive achievements. Different from Planck, though, Schrödinger had many lovers in his life, and lived openly with his wife and lovers.



Schrödinger (1887-1961)

Schrödinger was greatly inspired by de Broglie's wave-particle duality. He first learned about de Broglie's ideas through Einstein's 1925 paper on Bose statistics. He then studied de Broglie's doctorate thesis, which was published later in a journal. If electrons are waves, there should be a corresponding wave equation. With this in mind, Schrödinger left Zurich for Arosa in 1925 before the Christmas. Schrödinger returned to Zurich in January, with his famed equation and many calculations. With the assistance of Weyl (Hermann Weyl, 1885–1995), Schrödinger fixed the last few mathematical problems. On January 27, 1926, he submitted his paper to *Annalen der Physik*. In this paper, he presented the wave equation, which now bears his name, and showed that it gave the correct energy levels for a hydrogen atom.



What really happened in Arosa is a mystery in history. Schrödinger kept a diary, in which he recorded not only his research but also his dates with lovers. But Schrödinger's diary in 1925 disappeared. In his memoirs, Schrödinger did not reveal how he discovered this famous equation in Arosa in 1925. What we are now certain are: (1) Schrödinger was accompanied by an anonymous lover; (2) two notebooks full of equations were left behind; (3) Schrödinger did not go skiing during this holiday as he usually did because he was completely absorbed in his work. Whatever happened, the Schrödinger equation was born, and it was shown later to be equivalent to Heisenberg's matrix equation. Schrödinger published three more papers to develop and refine his wave dynamics. In the fourth paper (June 1926), he introduced complex numbers in his formulation and wrote down the time-dependent Schrödinger's equation.

Dirac put the finishing touches on the entire framework of the new quantum theory. After reading Schrödinger's paper, Dirac quickly recognized the equivalence of Schrödinger's wave dynamics and Heisenberg's matrix mechanics. In September 1926, Dirac published a paper entitled "On the Theory of Quantum Mechanics". In the paper Dirac established the equivalence of the two theories, and made it clear that there are only two kinds of particles in the quantum world—bosons and fermions—through the exchange symmetry of the multi-particle wave function.

## 2.8 Retrospect

The history of quantum mechanics is exhilarating and instructive. Here I will focus on our heroes, summarizing how they played their roles in different ways in this history.

Planck was a typical university professor, who had solid knowledge and tried to get the bottom of a problem as much as he could. But he was by nature conservative, preferring to improve an existing theory rather than break it. He made the discovery of "quantum" only when faced with the cold fact that the experimental data no longer fit the old formula. After the discovery, instead of developing the new idea of "quantum", Planck always tried to eliminate it by refining a classical theory. In other words, he had tried to close the door to the quantum world that he had opened.

Einstein was a rare genius. He not only independently developed the theory of relativity, but also made important contributions to the development of quantum mechanics. Before Bohr proposed the quantum theory of atom, Einstein was almost alone in the development of quantum theory. He developed Planck's concept of "quantum" to reveal the particle nature of light. Einstein never looked back. He moved further down the road, applying Planck's law of black-body radiation to the problem of specific heat of solids. His theory of spontaneous radiation refined the old Bohr-Sommerfeld quantum theory. In the second stage of the development of quantum theory, Einstein played an active role by helping Bose and de Broglie, who both were unknown at the time. Einstein immediately recognized the importance of their work and actively introduced them to the physics community. Einstein had

a second thought about quantum mechanics only in his later years. Interestingly, even his loud questioning of quantum mechanics contributed to the development of quantum mechanics, leading to a deep insight of quantum entanglement—a concept that had been neglected in the early development of quantum mechanics.

Bohr was a soldier in his early career. His atomic model marked a turning point in the development of quantum theory, and moved it to the forefront of physics research. In the development of new quantum theory, Bohr was more like a mentor. Thanks to Bohr, Sommerfeld, and Born, Copenhagen, Munich and Göttingen became the most important centers in the development of quantum theory. A group of exceptional young physicists studied quantum physics there, with big names like Heisenberg and Pauli. It was in these these centers that Heisenberg learned the limitations of the old quantum theory and developed the matrix mechanics.

It is always interesting to compare Einstein and Bohr. They both made revolutionary contributions to quantum theory in its early years and played decisive roles in its development. Later, both of them actively helped young physicists, albeit in a very different way. Einstein was not a good mentor. He was unwilling to be around students, and seldom collaborated with them in research. However, he helped young talents with his keen insight—discovering the importance of their work. Bohr, on the other hand, liked to talk to his students and discuss problems with them. Like Sommerfeld and Born, he first discovered the young talents and then guided them in their scientific explorations.

During the later development of quantum mechanics, a group of young and talented physicists made rapid breakthroughs, establishing a completely new quantum theory in just three years. Each of them made great contribution in his unique way.

De Broglie was an aristocrat. He switched his study from literature to science, motivated purely by his interests in physics. With such a background, he was not as deeply trapped in the traditional views as his contemporaries. He proposed the revolutionary concept of particle-wave duality and paved a way for the wave mechanics.

In contrast to de Broglie, Dirac came from an ordinary middle-class family. He was awkward at social communication, and almost never collaborated or discussed physics with others. But with his absolute genius, he stood out in the world of physics. Not only did he make the important contribution described above, he also proposed a new wave equation, known as Dirac's equation, which combined the special relativity and quantum mechanics. With this equation, Dirac predicted the existence of antiparticles, which have the same masses and spins but opposite charges to their counterparts. For example, a positron is the antiparticle of an electron and it has the same mass and spin as an electron but carries a positive charge.

Fermi was also a rare genius, almost self-taught. Unlike Dirac, Fermi was a good communicator and focused more on physics intuition than elegant mathematics. Later in his life, Fermi made many outstanding contributions to nuclear physics, such as building the first nuclear reactor and co-leading the Manhattan Project.

Pauli and Heisenberg had similar career path. Both of them studied under Sommerfeld, and then became assistants to Born and Bohr. Pauli was one year older, and was more knowledgeable in physics as evident in his studying general relativity at age 18. Heisenberg, on the other hand, had obvious deficiencies in physics, for exam-

ple, he could not explain the working principle of microscope when defending his doctorate thesis. But eventually, Heisenberg made greater contribution to the development of quantum theory. The reason is that Heisenberg was more brave and more willing to abandon the old theory. In this sense, Heisenberg's incomplete knowledge may have actually helped him: the more unknown, the fewer constraints.

The success of Bose was unique. He had a passion for physics, but did not have the opportunity go to Europe where there was the best physics education available at the time. It was probably for this reason that he made the "mistake", and proposed new quantum statistics by accident.

Schrödinger was the oldest among the pioneers of the new quantum theory. In 1926, when all the aforementioned young men were in their twenties, he was already 39 years old. Like Planck, he had few influential achievements before his scientific breakthrough. But unlike Planck, he actively participated the development of quantum mechanics. He was one of the first to notice quantum entanglement. Schrödinger's book "*What Is Life?*" has a profound influence in the biological community. Watson (James Watson, 1928–), the biologist who discovered the double helix structure of DNA, was initially interested in ornithology but switched to genetics after reading Schrödinger's book.

All this clearly tells us one thing: there are no assembly lines for scientific breakthroughs. Each track leading to a breakthrough is unique and different.

# Chapter 3

## Classical Mechanics and the Old Quantum Theory



We encounter all kinds of motion in our daily lives: a speeding car, a walking pedestrian, a rolling football, a flying bird. Our experience tells us that in order to accurately describe the motion of an object, both its position and velocity must be known. Knowing only position is inadequate. Suppose there is a bullet in front of you. If it has a zero velocity, you have nothing to worry about; but if it is traveling at a high speed, you'd better already put on a bullet-proof vest. Knowing only velocity is inadequate as well. For a bullet flying at a high speed, if it is 100 km away, you do not feel any threat; but if it is right in front of you, it is a threat to your life. In the world described by classical mechanics, an object can have well-defined position and velocity at the same time. When both are known, the state of the object at any instant of time is precisely determined. In quantum mechanics, as we will see, a particle can never have definite position and velocity (more accurately, momentum) simultaneously.

In this chapter I will briefly review the well-known system of free fall, using it as an example to introduce some important concepts in classical mechanics, such as phase space and Hamiltonian. Based on this, I will summarize the main features of classical mechanics. These features are so obvious that they have been taken for granted and seldom mentioned. In subsequent chapters, we will see that almost all of these properties are radically changed in quantum mechanics. At the end of this chapter, I introduce the old quantum theory of Bohr and Sommerfeld and apply it to harmonic oscillators.

### 3.1 Free Fall

Free fall is a simple and famous problem. Story has it that the famous Italian physicist Galileo (Galileo Galilei, 1564–1642) performed the experiment of free fall himself on the Leaning Tower of Pisa. The truth is that Galileo never did this experiment. But he did think seriously about the problem of free fall and did a thought experiment.

Imagine two balls, one heavier than the other, which are connected to each other by a short string. If the heavy ball falls faster than the light ball, the string will soon pull taut and the heavy ball is dragged by the slower light ball. Therefore, as a whole, they will fall slower than the heavy ball. But on the other hand, the system considered as a whole is heavier than the heavy ball alone, and therefore should fall faster. The two conclusions contradict each other. So the wise Galileo proved that the heavy and light balls must fall as fast as each other without actually climbing the Leaning Tower of Pisa.

Galileo's thought experiment is very clever, but its application is very limited, valid only when the force applied is proportional to the mass. The interested reader may extend Galileo's thought experiment to other systems, such as the spring oscillator, and you will quickly find that Galileo's method would give wrong results. The universal law of classical motion was discovered by Newton. Let us stand on the shoulders of Newton (Isaac Newton, 1643–1727)<sup>1</sup> and revisit the problem of free fall, using the classical mechanics created by the giant of science. According to Newton's second law, the force is equal to mass multiplied by acceleration,

$$F = ma. \quad (3.1)$$

Here  $m$  is the mass of an object,  $F$  is the force applied on the object, and  $a$  is the acceleration. This equation clearly shows that motion is in general not independent of mass. For a free-falling heavy object (say an iron ball), the air resistance can be ignored and the object only feels gravity  $mg$ , where  $g$  is the acceleration of gravity. Since  $F = ma$ , we have  $a = g$ . Therefore, all free-falling bodies accelerate with a rate  $g$ , regardless of their masses. This agrees with Galileo's thought experiment. Here we see that the motion of an object is independent of its mass only when the force acting on it is proportional to mass. In general, motion and mass are closely related. Galileo appeared smarter on the problem of free fall, drawing the correct conclusion without calculations. But in general, Newton's method is more powerful. His formula  $F = ma$  can be applied to arbitrary classical systems, not just free fall. Even for free fall, Newton's theory can predict precisely the instantaneous position and velocity of the falling body whereas Galileo could not. Let's look at it in detail below.

Assume that the initial velocity of a free-falling body is zero and its initial height is  $x_0$ . Since its acceleration is a constant  $g$ , what is its velocity at time  $t$ ? The acceleration is the rate of change of the velocity of an object with respect to time. A constant acceleration means the velocity changes uniformly in time. This gives  $v = -gt$  at time  $t$ , where the negative sign indicates the direction of velocity is downward. We can continue to calculate the position of the object at this time. Because the velocity changes at a constant rate  $g$ , the average velocity over this period of time is  $\bar{v} = (0 - gt)/2 = -gt/2$ . During this period, the distance that the object falls is  $s = |\bar{v}|t = gt^2/2$ . Because the initial height of the object is  $x_0$ , the height of the

---

<sup>1</sup> Newton's birthday was December 25, 1642 according to the old Julian calendar. For the modern calendar, Newton was born on January 4, 1643.

object at time  $t$  is  $x = x_0 - gt^2/2$ . Thus we obtain the position  $x$  and the velocity  $v$  of the free-falling object at any instant of time  $t$ ,

$$x = x_0 - gt^2/2, \quad v = -gt, \quad (3.2)$$

which completely determine the state of the object. Such detailed results cannot be obtained from Galileo's thought experiments.

The above discussion reveals an obvious fact: a falling body has well-defined position and velocity at every instant of time. This seems obvious. In fact, we, the people who live in the macroscopic world, cannot imagine that an object does not simultaneously have definite position and velocity. Things are radically different in quantum mechanics, where a particle is impossible to have definite position and velocity at the same time.

## 3.2 Phase Space

While the free fall has nothing to do with the mass of the falling object, we shall put the mass back in the equation, as it allows for new insight and understanding. With mass  $m$ , we can calculate the kinetic energy of an object at time  $t$  as

$$K = mv^2/2 = mg^2t^2/2 = mg(x_0 - x), \quad (3.3)$$

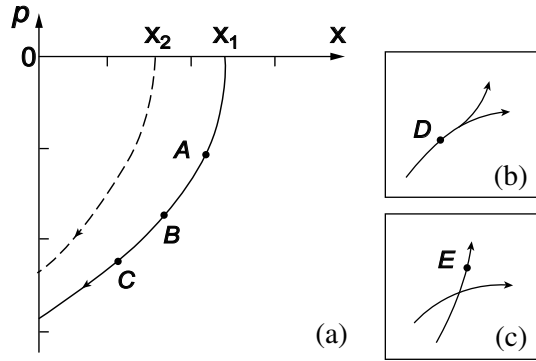
where we used Eq. (3.2). Moving the second term on the right to the left, we obtain

$$K + mgx = mgx_0. \quad (3.4)$$

The right-hand side of this equation is a constant. Therefore, although both the kinetic energy and  $mgx$  vary with time, their sum is a constant of motion. This sum is now known as the energy  $E = K + mgx$ , and  $mgx$  is called the potential energy, which is generally denoted as  $V(x)$ , i.e.,  $V(x) = mgx$ . When an object falls, its kinetic energy increases at the price of decreasing potential energy, but their sum stays constant. This is the conservation of energy. This result can be extended to arbitrary systems without friction, where the total energy is composed of kinetic energy and potential energy. During the motion, the kinetic energy and potential energy are transformed into each other, but the total energy is conserved over time. Initially, the total energy is  $E = mgx_0$ , containing only potential energy. Since the initial position  $x_0$  can be continuously changed, it follows from  $E = mgx_0$  that the energy of the system can also be continuously changed. In classical mechanics, that energy can vary continuously is an obvious fact. But in quantum mechanics, energy can be discrete.

With mass, we can further define the momentum  $p = mv$ . Usually, the momentum and velocity of an object are equivalent concepts. But for modern physicists, the concept of momentum is more fundamental. In particular, the momentum and veloc-

**Fig. 3.1** Phase space. **a** Two phase-space trajectories of a free-falling body. The two trajectories have different initial heights  $x_1, x_2$ . **b, c** Two impossible phase-space trajectories



ity are very different for particles without the rest mass. For example, all the photons have the same speed, i.e. the speed of light. But photons of different frequencies can have different momenta.

In terms of momentum, we can rewrite the kinetic energy as  $K = p^2/(2m)$ . Modern physicists like to write the expression of the energy  $E$  in terms of position and momentum as follows usually denoted by  $H$ , i.e.,

$$H = \frac{p^2}{2m} + V(x). \quad (3.5)$$

Here  $H$  is called Hamiltonian. There are many reasons for introducing Hamiltonian, one of which is that Hamiltonian plays a key role in quantum mechanics, as we will see in Chap. 6.

Having defined the momentum, we can introduce a powerful tool for studying classical mechanics, phase space. In a phase space as shown in Fig. 3.1a, the vertical axis denotes the momentum and the horizontal axis denotes the position. Every point in the phase space, with definite position and momentum, represents a state of the object. For the free fall, we have  $p^2/(2m) + mgx = E$ , where  $E$  is the conserved energy of the falling body. For each position  $x$ , we have a well-defined momentum

$$p = -\sqrt{2mE - 2m^2gx}, \quad (3.6)$$

where the negative sign means that the momentum  $p$  of the object always points downwards. Equation (3.6) represents a trajectory in a phase space. We have plotted two such trajectories (see Fig. 3.1a), corresponding to different initial heights  $x_1, x_2$ , respectively. For each point in a phase space, we can obtain an energy from the Hamiltonian in Eq. (3.5). Due to the conservation of energy, all the points on the phase-space trajectory have the same energy. As such, these trajectories are also called the isoenergetic lines. In Fig. 3.1a, all points on the solid curve have the same energy, and all points on the dashed curve have the same energy. But the points on the solid and dashed curves have different energies, which are determined by the initial

conditions according to Eq. (3.6). Since  $E$  can vary continuously, the trajectory can change continuously in a phase space.

Although Fig. 3.1a is a simplest example of a phase space, it displays many common features of the classical motion. Let us look at the solid curve in Fig. 3.1a, where we choose three arbitrary points  $A$ ,  $B$ ,  $C$ . For the middle point  $B$ , it evolves from point  $A$ , and it will evolve into point  $C$ . In other words, both its past and future are determined; we can know what happened in the past, as well as accurately predict the future. If the phase-space trajectory of an object bifurcates or intersects another trajectory as in Fig. 3.1b, c, the corresponding motion becomes uncertain: the point  $D$  can have two possible future states; and there are two possibilities for the past of the point  $E$ . In classical mechanics, we can prove rigorously in mathematics that the phase-space trajectories of motion cannot exhibit the bifurcation or intersection seen in Fig. 3.1b, c, meaning the classical motion has a determined past and future.

Based on the above discussion, we now summarize the properties of the classical motion.

- In classical mechanics, particles have deterministic trajectories, represented by a smooth curve in phase space. At every instant of time, a particle has well-defined position and momentum. Knowing the present position and momentum, one can predict the position and momentum in the future, or infer the position and momentum in the past. This manifests in the fact that the curve in phase space representing the trajectory of a moving particle do not bifurcate or intersect.
- The energy of a particle is continuous.
- If a system has two particles, we need to know the momentum and position of each particle to determine the state of the system. Namely, we need to specify four variables: the momentum and position of particle 1,  $p_1, x_1$ , and the momentum and position of particle 2,  $p_2, x_2$ . Thus the phase space of the system is four-dimensional. By analogy, if a system has  $n$  particles, each moving in  $d$  dimensions, a state of the system is represented by a point in a  $2dn$ -dimensional phase space. Therefore, the dimension of phase space grows linearly with the number of particles.

In addition, our daily experience tells us that both the position and velocity are observable quantities, and that their simultaneous measurement does not affect each other. The human or animal eyes can determine the instantaneous position and velocity of an object fairly accurately; a monitor on a highway can record exactly how much your car has exceeded the speed limit and where your car is; a ground control center can know precisely the position and velocity of every satellite at any time. These experiences tell us

- Variables that describe a classical mechanical motion, i.e., position and momentum, can be directly observed in experiments.
- The measurement outcomes of the position and momentum are certain.
- Measurements of position and momentum can be made simultaneously, without affecting each other in principle.



Like the first four axioms of Euclidean geometry, these properties of classical mechanics have long been taken as a matter of self-evident. In classes of classical mechanics, neither lecturers nor textbooks would emphasize these properties. Before the emergence of quantum mechanics, physicists did not pay special attention to these properties, either. In quantum mechanics, all these properties vanish: the variables describing the state of a system cannot be measured directly in experiment; a particle cannot have precise position and momentum at the same time; the outcome of a measurement is no longer determined; energy can be discrete; and so on. The reader is advised to come back to these properties after reading Chaps. 5, 6, 7 and 8 to appreciate how quantum mechanics is radically different.

### 3.3 Calculus for Velocity and Acceleration

In classical mechanics, velocity and acceleration are closely related to differential calculus. In fact, Newton invented differential calculus when thinking about velocity and acceleration. We now take the example of free fall to demonstrate how to calculate the velocity and acceleration with calculus. Suppose the position of an object at time  $t$  is  $x = x_0 - gt^2/2$ , and the position at time  $t + \delta t$  is  $x' = x_0 - g(t + \delta t)^2/2$ . The displacement over the time interval between  $t$  and  $t + \delta t$  is

$$\delta x = x' - x = -gt\delta t - g\delta t^2/2, \quad (3.7)$$

and the average velocity is

$$\tilde{v} = \frac{\delta x}{\delta t} = -gt - g\delta t/2. \quad (3.8)$$

Imagine a limiting process where  $\delta t$  gets closer and closer to zero. We will find  $\tilde{v}$  gets closer and closer to  $v = -gt$ , i.e., the velocity at time  $t$ . This limit is known as differential calculus in mathematics. Using the notations from calculus, we can write

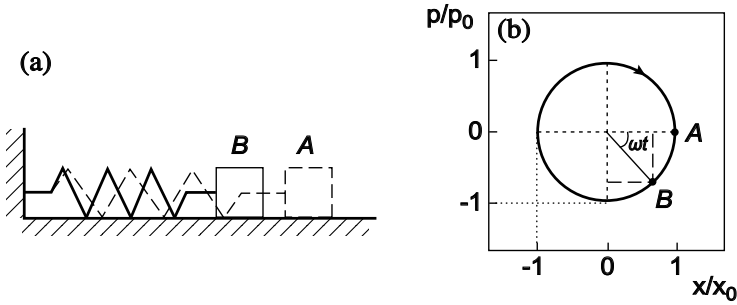
$$v = \frac{dx}{dt} \approx \frac{\delta x}{\delta t}. \quad (3.9)$$

Thus a velocity is the derivative of position with respect to time. Similarly, the acceleration is the derivative of velocity with respect to time

$$a = \frac{dv}{dt}. \quad (3.10)$$

With calculus, we can rewrite Newton's second law of motion as

$$F = ma = m \frac{dv}{dt} = \frac{dp}{dt}. \quad (3.11)$$



**Fig. 3.2** a A schematic of a harmonic oscillator (ignoring friction); b its phase space with one trajectory. The initial state is represented by point A. The state at time  $t$  is represented by point B.  $\omega$  is the vibration frequency

It states that the force results in the change of momentum with time.

Below are four basic formulas for calculating the derivatives of a trigonometric function

$$\frac{d \sin(t)}{dt} = \cos(t) , \quad \frac{d \cos(t)}{dt} = -\sin(t) , \quad (3.12)$$

$$\frac{d \sin(\omega t)}{dt} = \omega \cos(\omega t) , \quad \frac{d \cos(\omega t)}{dt} = -\omega \sin(\omega t) . \quad (3.13)$$

If you have already learned calculus, you should know these formulas. If you have not, just take them as facts. We will use these formula in the next section.

### 3.4 Harmonic Oscillator

Let's consider another example of classical mechanics. This time we start directly with phase space. In the phase space shown in Fig. 3.2, let us draw a circle around the origin. When an object has a positive momentum (or velocity), its position  $x$  increases with time. Thus if this circle represents a trajectory of an object, this trajectory should rotate clockwise along the circle. Suppose this rotation has an angular velocity of  $\omega$ . Then an initial point A will evolve to point B after some time  $t$ , where the position and momentum, respectively, are

$$x = x_0 \cos(\omega t), \quad p = -p_0 \sin(\omega t). \quad (3.14)$$

Does this circle in phase space represent a physical path of motion? Let's take a close look.

As said earlier, the velocity is the derivative of position with respect to time. Using the differentiation formula in Eq. (3.13), we obtain

$$v = \frac{dx}{dt} = -x_0\omega \sin(\omega t). \quad (3.15)$$

If the object has mass  $m$ , then its momentum is

$$p = mv = -mx_0\omega \sin(\omega t). \quad (3.16)$$

For  $p_0 = m\omega x_0$ , this result is consistent with previous result  $p = -p_0 \sin(\omega t)$ , and Eq. (3.14) describes a physical motion. Using Newton's second law, we can further analyze the force that acts on the object. According to the differential equation (3.11), we obtain

$$F = \frac{dp}{dt} = -p_0\omega \cos(\omega t) = -p_0\omega x/x_0 = -m\omega^2 x. \quad (3.17)$$

This means that the magnitude of the force on a particle is proportional to its displacement from the equilibrium point, while the negative sign indicates the direction of the force points towards the equilibrium point. This is exactly the dynamics of a spring oscillator (also called harmonic oscillator) (see Fig. 3.2a).

We can also write down the Hamiltonian for the harmonic oscillator. Using the trigonometric relation  $\sin^2 \theta + \cos^2 \theta = 1$ , we have

$$\frac{p^2}{p_0^2} + \frac{x^2}{x_0^2} = 1. \quad (3.18)$$

Substituting  $p_0 = mx_0\omega$  into above equation, we obtain

$$\frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2 = \frac{1}{2}m\omega^2 x_0^2. \quad (3.19)$$

Both sides of Eq. (3.19) are energies. Since  $E = m\omega^2 x_0^2/2$  on the right hand side is a constant, Eq. (3.19) shows the energy of a harmonic oscillator is conserved. The term on the left hand side of Eq. (3.19) is the Hamiltonian of the harmonic oscillator

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2. \quad (3.20)$$

By comparing with Eq. (3.5), we find  $V(x) = \frac{1}{2}m\omega^2 x^2$ , which is the potential energy of the harmonic oscillator.

### 3.5 The Old Quantum Theory

In Eq. (3.19) for the harmonic oscillator, the  $x_0$  on the right hand side is the maximum displacement of an oscillator from its equilibrium point. A different value of  $x_0$  in Eq. (3.19) yields a different phase-space trajectory. In classical mechanics,  $x_0$  is allowed

to vary continuously from zero to the infinity, so that the corresponding trajectories can fill the entire phase space. But according to the old quantum theory developed by Bohr and Sommerfeld, only the orbits<sup>2</sup> obeying the quantization rule are allowed. The Bohr-Sommerfeld quantization rule is:

**A quantized orbit encloses an area  $S$  in phase space which is an integer multiple of Planck's constant  $h$ .**

For readers familiar with calculus, this rule can be written mathematically as

$$S = \oint p dx = nh, \quad n = 1, 2, 3, \dots \quad (3.21)$$

Obviously, these quantized orbits are discrete (see Fig. 3.3a).

Now we apply this quantization rule to a harmonic oscillator. For a given  $x_0$ , the corresponding trajectory encloses an area

$$S = \pi p_0 x_0 = \pi m x_0^2 \omega = 2\pi E / \omega, \quad (3.22)$$

where  $E = m x_0^2 \omega^2 / 2$  is the energy of the oscillator. If  $x_0$  is associated with the  $n$ th quantized energy  $E_n$ , we have

$$2\pi E_n / \omega = nh. \quad (3.23)$$

As a result, we can obtain ( $h = 2\pi \hbar$ )

$$E_n = n\hbar\omega. \quad (3.24)$$

This is the quantized energy of a harmonic oscillator. For each discrete energy  $E_n$ , there is a phase-space trajectory, as schematically illustrated in Fig. 3.3a. These trajectories are the quantized orbits. According to the old quantum theory, other trajectories are not allowed. Applying the Bohr-Sommerfeld quantization rule to the hydrogen atom, we can obtain its energy levels and quantized orbits. As it requires more advanced mathematics, we will not discuss it here.

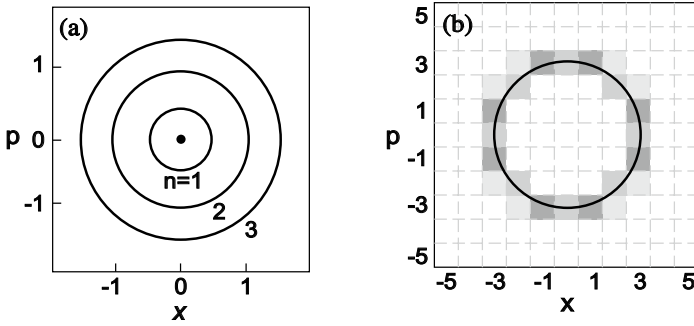
In modern quantum theory, the energy levels can be obtained by solving the Schrödinger equation, where every energy level corresponds to an eigenfunction. Solving the Schrödinger equation is beyond the scope of this book, so we only provide some results here so that the reader can have a glimpse of the new theory. For example, by solving the Schrödinger equation of a harmonic oscillator, we obtain

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \quad n = 0, 1, 2, \dots \quad (3.25)$$

which differs from Eq. (3.24) by  $\hbar\omega/2$ . The interpretation of this difference is also beyond the scope of this book. Figure 3.3b illustrates the eigenfunction of the 30th

---

<sup>2</sup> In quantum mechanics, it is customary to call the trajectories as orbitals.



**Fig. 3.3** **a** Schematic of quantized orbits of a harmonic oscillator in phase space. **b** Quantum phase space and the eigen wave function of a harmonic oscillator. In classical phase space, every point represents a state of an object; in quantum phase space, every small square represents a quantum state. The area of each small square lattice is Planck's constant  $h$ . They are usually called Planck cell. The darker the cell, the larger the value of the wave function on that square. The black circle denotes an orbit obeying the Bohr-Sommerfeld quantization rule. The results in **b** are from [Fang Y, Wu F, and Wu B. *J. Stat. Mech.* (2018) 023113]. Note that the trajectory is rendered circular by properly choosing the units of  $x$ ,  $p$

energy level of a harmonic oscillator in phase space. According to modern quantum theory, a particle cannot have simultaneously determined momentum and position. Thus a quantum state cannot be represented as a phase-space point, but rather as a small square with a size of Planck's constant  $h$ . As a result, the phase space in Fig. 3.3b is divided into a number of small squares, which are called Planck cells. On each Planck cell, the wave function has a value; the larger this value is, the darker the cell appears. The black circles in Fig. 3.3b denote the quantized orbital from the old quantum theory, and we see that the wave function is concentrated around this orbit. Chapter 6 will present a more detailed discussion on the Schrödinger equation and the eigenfunction.

# Chapter 4

## Complex Number and Linear Algebra



The full description of quantum mechanics needs a lot of advanced mathematics, which includes calculus and partial differential equations. However, the essence of quantum mechanics can be described with just complex number and some basics of linear algebra on top of high-school mathematics. In this chapter I shall briefly introduce complex number and linear algebra. The latter is mostly about Hilbert spaces and matrices.

### 4.1 Complex Number

Number began as a very practical matter. Even without searching historic records, it is not difficult to imagine how integers, fractional numbers, and negative numbers originated. In the early days, men had to keep track of how many preys they hunted, and woman needed to count how many fruits they picked. Thus appeared the concept of integer. To share things, people naturally began to use fractional numbers. When commercial and tax activities emerged in a human society, negative numbers were adopted to record debts and taxes. In ancient China, a red counting rod<sup>1</sup> represents a positive number and a black counting rod represents a negative number.

The discovery of irrational number represents an important advance in human understanding of numbers, as well as a great triumph of human's ability of abstract reasoning. In our daily lives, we only encounter integers and fractions (positive or negative). Mathematically, they are known as rational numbers. In a practical measurement, no matter how accurate it is, the outcome can only be rational numbers; in calculations, while engineers may use irrational numbers such as  $\pi$ , the final results that they send to the manufacturers can only be rational numbers; any computer can only handle rational numbers due to the limited number of bits.

---

<sup>1</sup> A counting rod is called Suanchou in Chinese.

The discovery of irrational numbers is usually attributed to an ancient Greek philosopher Hippasus (Hippasus, about the 5th century BC), who got interested in the side length of the isosceles right triangle. He discovered that the ratio of the hypotenuse to a leg could not be a rational number. Hippasus made this discovery by abstract logic reasoning, without measuring the length of the hypotenuse. In fact, if he had made the measurement, he could not have discovered the irrational number. Hippasus was a Pythagorean. At that time, Pythagoreans preached that all numbers could be expressed as integers or the ratios of integers. Hippasus's discovery shocked them. A popular story tells that Hippasus was drowned at sea as a punishment. An alternative story is that Hippasus was expelled from the Pythagorean community. For Hippasus, it is a tragic story; for mankind, it is a giant progress: one can discover something, which is as real as the earth under our feet, by abstract reasoning.

Imaginary number was also discovered from logical reasoning. The ancient Greek mathematician and engineer Hero of Alexandria (Hero of Alexandria, about 10–70 A.D) discovered the imaginary number<sup>2</sup> when he thought about solving equations like  $x^2 + 1 = 0$ . Hero did not suffer from this discovery probably because the society had become more tolerant. But for a long time imaginary numbers had been regarded, even by mathematicians, as fictitious or useless. This is reflected in its name 'imaginary', which was meant to be derogatory. This attitude continued until the great mathematician Euler (Leonhard Euler, 1707–1783) studied complex numbers in depth. Euler used  $i$  to stand for  $\sqrt{-1}$ , i.e.,  $i^2 = -1$ . This notation is still in use to this day. For arbitrary real numbers  $x$  and  $y$ ,  $yi$  is called imaginary number and  $x + yi$  is called complex number. In subsequent chapters, we will see that quantum mechanics describes this colorful world with complex numbers or groups of complex numbers. The great Hero and Euler, although exceptionally brilliant, must have never thought that the world is ultimately described by complex numbers.

For a complex number  $z = x + yi$ ,  $x$  is its real part and  $y$  is its imaginary part. It can be represented as a vector  $(x, y)$  in the complex plane (see Fig. 4.1), where the horizontal axis is in general for the real part and the vertical axis is for the imaginary part. Like other vectors, a complex number  $z$  has a length and a direction (i.e., angle). As illustrated in Fig. 4.1, the length of complex number  $z$  is  $r = \sqrt{x^2 + y^2}$ , and its angle  $\theta$  satisfies  $\tan \theta = y/x$ . Here  $r$  is called the modulus of complex number  $z$ , denoted by  $r = |z|$ ;  $\theta$  is the argument of  $z$ , denoted by  $\theta = \arg(z)$ .

Basic arithmetic operations on imaginary and complex numbers are as follows. Note that  $x_{1,2}$  and  $y_{1,2}$  in the following equations are real numbers.

*Addition of imaginary numbers:*

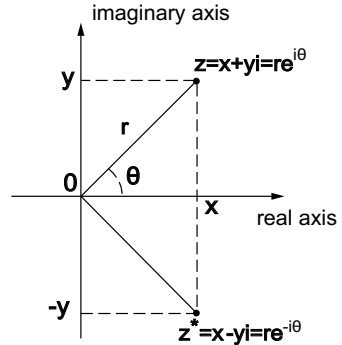
$$x_1i + x_2i = (x_1 + x_2)i. \quad (4.1)$$

Example:  $-5i + 2i = -3i$ .

---

<sup>2</sup> See two books: Hargittai, *Fivefold symmetry* (World Scientific, 2nd ed., 1992), p. 153; Roy, *Complex numbers: lattice simulation and zeta function applications* (Horwood, 2007), p. 1.

**Fig. 4.1** Complex number  $z$ .  $x$  is the real part,  $y$  is the imaginary part,  $r$  is the modulus, and  $\theta$  is the argument.  $z^*$  is the complex conjugate of  $z$



*Multiplication of imaginary numbers:*

$$(y_1i) \times (y_2i) = (y_1 \times y_2) \times (i \times i) = -y_1y_2. \tag{4.2}$$

Example:  $-3i \times 2i = 6$ .

*Division of imaginary numbers:*

$$(y_1i) \div (y_2i) = y_1 \div y_2 = y_1/y_2. \tag{4.3}$$

Example:  $3i \div 5i = 3/5$ .

*Addition of complex numbers:*

$$z_1 + z_2 = (x_1 + y_1i) + (x_2 + y_2i) = (x_1 + x_2) + (y_1 + y_2)i. \tag{4.4}$$

Example:  $(4.2 + 5i) + (2.1 - 2.3i) = 6.3 + 2.7i$ .

*Multiplication of complex numbers:*

$$\begin{aligned} z_1 \times z_2 &= (x_1 + y_1i) \times (x_2 + y_2i) \\ &= x_1x_2 + x_1y_2i + (y_1i)x_2 + (y_1i) \times (y_2i) \\ &= x_1x_2 - y_1y_2 + (x_1y_2 + x_2y_1)i. \end{aligned} \tag{4.5}$$

Example:  $(3 + 4i) \times (3 + 4i) = 9 + 12i + 12i - 16 = -7 + 24i$ .

Division of complex numbers is a bit more complicated. The division of two complex numbers  $z_1 \div z_2$  is equivalent to that  $z_1$  multiplies the inverse of  $z_2$ , i.e.,  $z_1 \times \frac{1}{z_2}$ . Because the inverse of  $z_2$  is given by

$$\frac{1}{z_2} = \frac{1}{x_2 + iy_2} = \frac{x_2 - iy_2}{(x_2 + iy_2)(x_2 - iy_2)} = \frac{x_2 - iy_2}{x_2^2 + y_2^2}, \tag{4.6}$$



we have

$$z_1 \div z_2 = \frac{x_1 + iy_1}{x_2 + iy_2} = \frac{x_1x_2 + y_1y_2 + i(x_2y_1 - x_1y_2)}{x_2^2 + y_2^2}. \quad (4.7)$$

Note that arithmetic operations of real and imaginary numbers can be considered as special cases of the arithmetic of complex numbers.

There is an operation on complex number that is absent for real numbers, complex conjugate. The complex conjugate of a complex number  $z = x + yi$  is  $z^* = (x + yi)^* = x - yi$ . As demonstrated in Fig. 4.1, the complex number  $z$  and its complex conjugate  $z^*$  is symmetric about the real axis. Clearly,  $z^*z = |z|^2$ .

Now we introduce a special but frequently used complex number

$$e^{i\theta} = \cos \theta + i \sin \theta. \quad (4.8)$$

Here  $\theta$  is a real number. If you know calculus, you can prove the validity of Eq. (4.8); if not, just accept it as a fact. For any complex number, we always have

$$z = x + yi = \sqrt{x^2 + y^2} \left( \frac{x}{\sqrt{x^2 + y^2}} + \frac{y}{\sqrt{x^2 + y^2}}i \right), \quad (4.9)$$

Let  $r = \sqrt{x^2 + y^2}$  and  $\cos \theta = x/\sqrt{x^2 + y^2}$ , we obtain

$$z = re^{i\theta}. \quad (4.10)$$

where  $r$  is the modulus of the complex number  $z$ , and  $\theta$  is the argument of  $z$ . Physicists prefer to call  $\theta$  the phase of  $z$ . Equation (4.10) provides another representation of a complex number. It allows to easily calculate the inverse of  $z$  as  $z^{-1} = 1/z = e^{-i\theta}/r$ . In this form, the complex conjugate of  $z$  can be written as  $z^* = re^{-i\theta}$ .

Later we will see many applications of complex numbers in quantum mechanics. Below are two simple applications of complex numbers in mathematics.

- By using  $e^{i(\theta_1+\theta_2)} = e^{i\theta_1} e^{i\theta_2}$  and Eq. (4.8), we can derive the familiar trigonometric relations

$$\sin(\theta_1 + \theta_2) = \sin(\theta_1) \cos(\theta_2) + \cos(\theta_1) \sin(\theta_2), \quad (4.11)$$

$$\cos(\theta_1 + \theta_2) = \cos(\theta_1) \cos(\theta_2) - \sin(\theta_1) \sin(\theta_2). \quad (4.12)$$

- We know that a quadratic equation  $ax^2 + bx + c = 0$  admit the following roots

$$x_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (4.13)$$

Your high school math teacher may have told you that this equation has no solutions if  $b^2 < 4ac$ . However, with complex numbers, a quadratic equation has two solutions even when  $b^2 < 4ac$ . In this situation, the equation admits two complex solutions. It was in solving this kind of equations that Hero discovered complex numbers.

## 4.2 Linear Algebra

Linear algebra is a systematic generalization of mathematical concepts such as coordinates, vectors, and vector transformations. We shall first introduce linear spaces and then discuss transformation between vectors in linear spaces, i.e., matrices. According to quantum mechanics, all matters in the universe live in Hilbert spaces, one type of linear spaces.

### 4.2.1 Linear Space

Let us start by recapitulating vectors in a two-dimensional plane. When the axes are chosen, each point in the plane is represented by two real coordinates  $x$  and  $y$ . A vector pointing from the origin to this point can be expressed as

$$\mathbf{r} = (x, y). \quad (4.14)$$

Vectors have some well known and obvious properties.

- A vector multiplied by a constant results in another vector

$$\mathbf{r}' = a\mathbf{r} = a(x, y) = (ax, ay). \quad (4.15)$$

The constant  $a$  is usually called scalar. If  $a = -1$ , the vector  $\mathbf{r}'$  is antiparallel with the vector  $\mathbf{r}$ ; if  $0 < a < 1$ ,  $\mathbf{r}'$  is parallel with  $\mathbf{r}$ , but is with a shorter length; if  $a > 1$ ,  $\mathbf{r}'$  is parallel with  $\mathbf{r}$ , but with a longer length.

- The addition of two vectors yields another vector

$$\mathbf{r}_1 + \mathbf{r}_2 = (x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2). \quad (4.16)$$

- The dot product of two vectors is defined as

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = (x_1, y_1) \cdot (x_2, y_2) = x_1x_2 + y_1y_2. \quad (4.17)$$

The dot product is also called scalar product. Consider a vector  $\mathbf{r} = (x, y)$ . The dot product with itself is

$$\mathbf{r} \cdot \mathbf{r} = x^2 + y^2, \quad (4.18)$$

which is exactly the square of the length of vector  $\mathbf{r}$ . So we can use the dot product to calculate the length of a vector. For two vectors of unit length,  $\mathbf{r}_1$  and  $\mathbf{r}_2$ , we have

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = \cos \theta. \quad (4.19)$$

Here  $\theta$  is the angle between vector  $\mathbf{r}_1$  and vector  $\mathbf{r}_2$ . If  $\theta = \pi/2$ , then  $\mathbf{r}_1 \cdot \mathbf{r}_2 = 0$ , and we say that these two vectors are perpendicular to each other.

Based on the above properties, mathematicians have generalized two-dimensional vectors to the concept of linear space (also called vector space). When we introduced vectors in the above, we first defined a two-dimensional space and established a coordinate system, allowing us to define a vector in terms of the coordinates of a point. Mathematicians reverse the line of thought. Rather than defining a space before defining a point, they regard a space as a collection of points and defined by the relation between these points.

We introduce linear spaces with a simple example, the two-dimensional linear space. Suppose that there are a set of points, each with two components. They are written in columns as

$$\begin{pmatrix} x \\ y \end{pmatrix}. \quad (4.20)$$

The multiplication between a point and a constant  $a$  is defined as

$$a \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax \\ ay \end{pmatrix}. \quad (4.21)$$

The addition of two points is defined as

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}. \quad (4.22)$$

All the points with two components, which satisfy the above two relations, make up a two-dimensional linear space, where each point is called a vector<sup>3</sup>. Compared to the two-dimensional vectors we reviewed at the beginning, the representation of vectors is changed, they are now expressed as columns as in Eq. (4.20). Such an expression facilitates the introduction of matrix, as we will describe later.

In order to describe the length of a vector and the angle between two vectors, something similar to the dot product needs to be defined. For this purpose, the concepts of column vector and row vector are introduced. The vector in Eq. (4.20) is called column vector. The corresponding row vector is defined as

---

<sup>3</sup> In the rigorous definition of a linear space, mathematicians also require that the multiplication and addition satisfy certain relations, such as the addition of vector 1 and vector 2 is equivalent to the addition of vector 2 and vector 1. In this book, mathematical rigor is not emphasized; the interested reader is referred to a textbook on linear algebra.

$$(x \ y) . \quad (4.23)$$

The operation that transforms a column vector to a row vector is called transpose. Multiplication between a row vector and a column vector is defined as

$$(x_1 \ y_1) \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = x_1x_2 + y_1y_2. \quad (4.24)$$

That is, we multiply every entry of a row vector with the corresponding entry of a column vector and then add them together. To obtain the dot product of two vectors,

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \quad (4.25)$$

we multiply the transpose of one of the vectors with another vector according to Eq. (4.24). The dot product is also called inner product. Using the inner product, we can calculate the square of length  $r$  of a vector as

$$r^2 = (x \ y) \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2. \quad (4.26)$$

Such a re-formulation of the two-dimensional vectors makes it convenient for further generalization, and the multiplication between a row vector and a column vector paves the way for introducing matrix. First of all, it allows extension to other dimensions. By considering a vector with  $n$  components, and defining similarly the addition and multiplication operations, we can construct a  $n$ -dimensional linear space. Consider the vector with four components as an example. Addition of two such vectors can be defined as

$$\begin{pmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \end{pmatrix} + \begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 \\ b_1 + b_2 \\ c_1 + c_2 \\ d_1 + d_2 \end{pmatrix}, \quad (4.27)$$

and the inner product of these two vectors is defined as

$$(a_1 \ b_1 \ c_1 \ d_1) \begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} = a_1a_2 + b_1b_2 + c_1c_2 + d_1d_2. \quad (4.28)$$

Another possible generalization is by using complex numbers. So far, both scalars and components of a vector are real numbers. By allowing them to be complex numbers, we obtain an important type of linear space, Hilbert space, which is the mathematical foundation for quantum mechanics.

### 4.2.2 Hilbert Space

For simplicity, we begin with a two-dimensional Hilbert space. The vectors that make up this space have two components, which can be written as

$$|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix}, \quad (4.29)$$

where both  $a$  and  $b$  are complex numbers. In the above, we have used the Dirac notation,  $|\psi\rangle$ , to denote a vector in a Hilbert space. The  $|\ \rangle$  is called a ket. The Dirac notation is adopted here for the sake of clarity and convenience. In a high dimensional Hilbert space, vectors have multiple-components, but it is not necessary to list all of them in most cases. It also prepares readers for quantum mechanics, where the Dirac notation is commonly used.

Multiplying a vector  $|\psi\rangle$  by a constant is given by

$$c|\psi\rangle = c \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} ca \\ cb \end{pmatrix}, \quad (4.30)$$

where  $c$  is a complex number. For two vectors in a two-dimensional Hilbert space

$$|\psi_1\rangle = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \quad \text{and} \quad |\psi_2\rangle = \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}, \quad (4.31)$$

their addition is defined as

$$|\psi_1\rangle + |\psi_2\rangle = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 \\ b_1 + b_2 \end{pmatrix}. \quad (4.32)$$

Because the vector components can be complex numbers, the relation between a column vector and a row vector in a Hilbert space is a bit more complicated. The corresponding row vector of a column vector

$$|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \quad (4.33)$$

is denoted as  $\langle\psi|$ . The form  $\langle\ |$  is called bra<sup>4</sup>. To obtain  $\langle\psi|$ , one first transposes  $|\psi\rangle$  and then take the complex conjugate, i.e.,

$$\langle\psi| = (a^* \ b^*). \quad (4.34)$$

---

<sup>4</sup> The pronunciation of  $\langle |$  and  $| \rangle$  comes from the English word “bracket”. By breaking it up and removing the unimportant c, we are left with “bra” and “ket”.

The inner product of two vectors,  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , is defined as

$$\langle\psi_1|\psi_2\rangle = (a_1^* \ b_1^*) \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} = a_1^*a_2 + b_1^*b_2, \quad (4.35)$$

or, alternatively as

$$\langle\psi_2|\psi_1\rangle = (a_2^* \ b_2^*) \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = a_2^*a_1 + b_2^*b_1. \quad (4.36)$$

It is clear that the order how one computes the inner product of two vectors  $|\psi_1\rangle$  and  $|\psi_2\rangle$  matters. The inner product  $\langle\psi_1|\psi_2\rangle$  is in general not equal to  $\langle\psi_2|\psi_1\rangle$ , but rather the complex conjugate of the other, i.e.,

$$\langle\psi_1|\psi_2\rangle = \langle\psi_2|\psi_1\rangle^*. \quad (4.37)$$

We have  $\langle\psi_1|\psi_2\rangle = \langle\psi_2|\psi_1\rangle$  only if the inner product of two vectors is a real number. The length  $r$  of a vector  $|\psi\rangle$  is defined as the square root of the inner product of  $|\psi\rangle$  with itself, i.e.,

$$r^2 = \langle\psi|\psi\rangle = |a|^2 + |b|^2. \quad (4.38)$$

When  $\langle\psi_1|\psi_2\rangle = 0$ , we say vector  $|\psi_1\rangle$  is perpendicular or orthogonal to vector  $|\psi_2\rangle$ . One more commonly uses “orthogonal” in the context of Hilbert space.

Recall that in our review of two-dimensional vectors, we have mentioned that one needs to establish a coordinate system before defining a vector. For generic linear spaces including Hilbert spaces, we also need to establish a “coordinate system” in order to write down the components of a vector. This “coordinate system” is provided by the basis of a linear space. A two-dimensional Hilbert space has two basis vectors. Previously, in writing the elements of  $|\psi\rangle$ , we have used implicitly the following two basis vectors

$$|e_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |e_2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (4.39)$$

With these two basis vectors, a vector  $|\psi\rangle$  can be written as

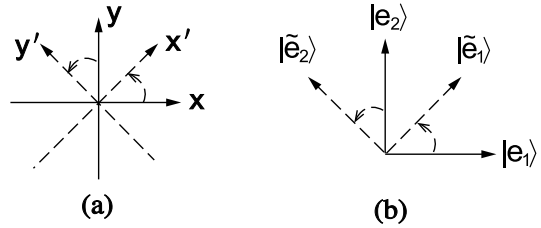
$$|\psi\rangle = a|e_1\rangle + b|e_2\rangle. \quad (4.40)$$

The following relations can be verified by straightforward calculations

$$\langle e_1|e_1\rangle = \langle e_2|e_2\rangle = 1, \quad \langle e_1|e_2\rangle = \langle e_2|e_1\rangle = 0. \quad (4.41)$$

They show that both two basis vectors have a unit length, and they are orthogonal to each other. We call a set of such basis vectors as an orthonormal basis.

**Fig. 4.2** Change of orthonormal basis.  
**a** Rotation of a two-dimensional coordinate system in a real space;  
**b** “rotation” of orthonormal basis in a two-dimensional Hilbert space



The choice of coordinate system is not unique. A coordinate system can be transformed to another by rotations (see Fig. 4.2a). Similarly, we can construct a different orthonormal basis by “rotation”. A “rotation” in a Hilbert space can be mathematically represented by a unitary matrix, as will be introduced shortly, which generically involves complex numbers. For example, consider two vectors,  $|\tilde{e}_1\rangle$  and  $|\tilde{e}_2\rangle$ , in a two-dimensional Hilbert space

$$|\tilde{e}_1\rangle = \frac{1}{\sqrt{2}}(|e_1\rangle + i|e_2\rangle), \quad |\tilde{e}_2\rangle = \frac{1}{\sqrt{2}}(|e_1\rangle - i|e_2\rangle). \quad (4.42)$$

After straightforward calculations, we can verify that

$$\langle \tilde{e}_1 | \tilde{e}_1 \rangle = \langle \tilde{e}_2 | \tilde{e}_2 \rangle = 1, \quad \langle \tilde{e}_1 | \tilde{e}_2 \rangle = \langle \tilde{e}_2 | \tilde{e}_1 \rangle = 0. \quad (4.43)$$

So, they also form a set of orthonormal basis, and can be regarded as being obtained from  $|e_1\rangle$  and  $|e_2\rangle$  by “rotation”. But the above expressions contain complex numbers. Therefore, such rotation is different from, and is conceptually richer than, the familiar rotation in the real space. In this new orthonormal basis, we have

$$|\psi\rangle = a|e_1\rangle + b|e_2\rangle = \frac{a - ib}{\sqrt{2}}|\tilde{e}_1\rangle + \frac{a + ib}{\sqrt{2}}|\tilde{e}_2\rangle. \quad (4.44)$$

Any two vectors of unit length that are orthogonal to each other can be used as an orthonormal basis for a two-dimensional Hilbert space.

All these results can be straightforwardly generalized to  $n$ -dimensional Hilbert spaces. Consider two vectors in an  $n$ -dimensional Hilbert space

$$|\phi\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}, \quad |\psi\rangle = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix}. \quad (4.45)$$

Multiplying the vector  $|\phi\rangle$  by a constant  $c$  is given by

$$c|\phi\rangle = c \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ca_1 \\ ca_2 \\ ca_3 \\ \vdots \\ ca_n \end{pmatrix}. \quad (4.46)$$

The addition of these two vectors is

$$|\phi\rangle + |\psi\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \\ \vdots \\ a_n + b_n \end{pmatrix}. \quad (4.47)$$

The inner product of  $|\phi\rangle$  and  $|\psi\rangle$  is

$$\langle\phi|\psi\rangle = (a_1^* \ a_2^* \ a_3^* \ \cdots \ a_n^*) \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix} = a_1^* b_1 + a_2^* b_2 + a_3^* b_3 + \cdots + a_n^* b_n. \quad (4.48)$$

An  $n$ -dimensional Hilbert space is spanned by  $n$  orthonormal basis vectors. A natural choice is

$$|e_1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |e_2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |e_3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots, \quad |e_n\rangle = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad (4.49)$$

Using this basis, we can write  $|\phi\rangle$  as

$$|\phi\rangle = a_1|e_1\rangle + a_2|e_2\rangle + a_3|e_3\rangle + \cdots + a_n|e_n\rangle = \sum_{j=1}^n a_j|e_j\rangle, \quad (4.50)$$



where we have used the summation symbol  $\sum_{j=1}^n$ , which means adding terms from  $j = 1$  to  $j = n$ . The complex conjugate of a vector  $|\phi\rangle$  can be written as

$$\langle\phi| = \sum_{j=1}^n a_j^* \langle e_j|. \quad (4.51)$$

The inner product of  $|\phi\rangle$  and  $|\psi\rangle$  is given by

$$\langle\phi|\psi\rangle = \left( \sum_{j=1}^n a_j^* \langle e_j| \right) \left( \sum_{k=1}^n b_k |e_k\rangle \right) = \sum_{i=1}^n a_i^* b_i, \quad (4.52)$$

where we have used

$$\langle e_j | e_j \rangle = 1, \quad \langle e_j | e_k \rangle = 0 \quad (j \neq k). \quad (4.53)$$

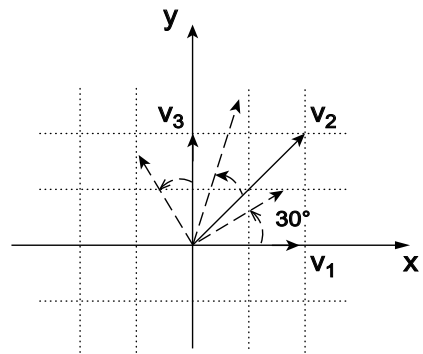
### 4.2.3 Matrix

Before formally introducing matrices, let us revisit a typical example of the transformation of two-dimensional vectors. In Fig. 4.3, there are three vectors,  $\mathbf{v}_1$ ,  $\mathbf{v}_2$ , and  $\mathbf{v}_3$ . Consider first  $\mathbf{v}_1$  and  $\mathbf{v}_3$ , which can be written in the form of column vectors as

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (4.54)$$

Let us rotate these two vectors counterclockwise by  $30^\circ$  as illustrated in Fig. 4.3. By some simple calculations with trigonometric functions, we obtain two new vectors

**Fig. 4.3** Counterclockwise rotation of vectors by  $30^\circ$



$$R_{30} \mathbf{v}_1 = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}, \quad R_{30} \mathbf{v}_3 = \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}. \quad (4.55)$$

Here  $R_{30}$  is an abstract notation, representing a counterclockwise rotation by  $30^\circ$ . Mathematicians have found that  $R_{30}$  can be expressed in the following form

$$R_{30} = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}, \quad (4.56)$$

which is called matrix. The following is the rule for the multiplication of a matrix and a column vector

$$\begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}x}{2} - \frac{y}{2} \\ \frac{x}{2} + \frac{\sqrt{3}y}{2} \end{pmatrix}. \quad (4.57)$$

That is, the first entry of the new vector is the multiplication of the first row of the matrix and the column vector, and the second entry of the new vector is the multiplication of the second row of the matrix and the column vector. In other words, the matrix has been regarded as composed of two row vectors in the multiplication. Applying this rule, one can easily verify that Eq.(4.55) holds when Eq.(4.56) is inserted.

Now let us consider the vector  $\mathbf{v}_2$  in Fig.4.3. We rotate it also counterclockwise by  $30^\circ$ . Using the matrix  $R_{30}$ , we can easily obtain the new vector as

$$R_{30} \mathbf{v}_2 = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} - \frac{1}{2} \\ \frac{1}{2} + \frac{\sqrt{3}}{2} \end{pmatrix}. \quad (4.58)$$

Interested readers can verify this result using other methods such as triangulation.

Matrices can represent not only rotations but also many other transformations. As an example, we consider shear transformation. The following matrix

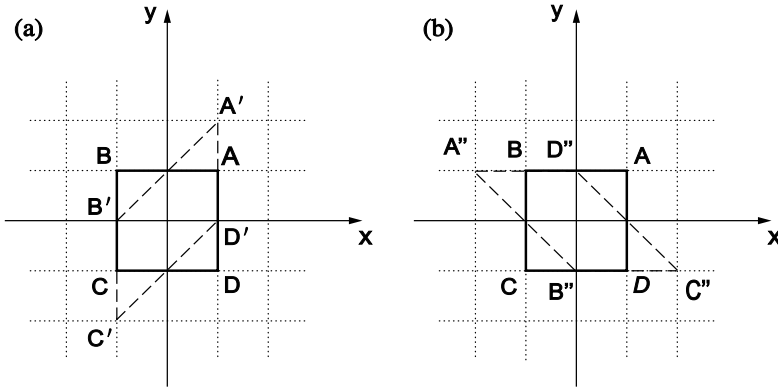
$$Q = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad (4.59)$$

represents a shear transformation along the  $y$  axis. Its action on an arbitrary vector is as follows

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y + x \end{pmatrix}. \quad (4.60)$$

It does not change the  $x$  component, but the  $y$  component is changed into  $y + x$ . For point A in Fig.4.4a, whose  $x$  and  $y$  components are both 1, we have

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad (4.61)$$



**Fig. 4.4** **a** Shear transformation of square ABCD; **b** shear transformation followed by a counter-clockwise  $90^\circ$  rotation

That is, the shear transformation  $Q$  on point A results in  $A'$ . Similarly, the action of  $Q$  on point B results in  $B'$ , the action of  $Q$  on point C gives  $C'$ , and the action of  $Q$  on point D gives  $D'$ . To sum up, the action of  $Q$  on square ABCD in Fig. 4.4a results in parallelogram  $A'B'C'D'$ .

We often need to perform a series of transformations on a vector. As each transformation corresponds to a matrix, this raises an issue of matrix multiplication. As an example, let us consider that a two-dimensional vector is rotated first with  $R_{30}$  and then sheared with  $Q$ . This can be calculated as follows

$$\begin{aligned} QR_{30} \begin{pmatrix} x \\ y \end{pmatrix} &= Q \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{2}x - \frac{y}{2} \\ \frac{x}{2} + \frac{\sqrt{3}}{2}y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2}x - \frac{y}{2} \\ \frac{\sqrt{3}+1}{2}x + \frac{\sqrt{3}-1}{2}y \end{pmatrix}. \end{aligned} \quad (4.62)$$

Introducing a new matrix

$$W = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{\sqrt{3}+1}{2} & \frac{\sqrt{3}-1}{2} \end{pmatrix}. \quad (4.63)$$

Straightforward calculations show that

$$QR_{30} \begin{pmatrix} x \\ y \end{pmatrix} = W \begin{pmatrix} x \\ y \end{pmatrix}. \quad (4.64)$$

The above identity indicates that the action of two consecutive matrices is equivalent to one single matrix. Alternatively, one can regard it as the multiplication of two matrices  $QR_{30}$  leads to another matrix  $W$ , i.e.,  $QR_{30} = W$ .

To illustrate the rules of matrix multiplication, let us consider two generic matrices

$$M_1 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad M_2 = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}. \quad (4.65)$$

The result of performing transformation  $M_2$  followed by transformation  $M_1$  is

$$\begin{aligned} M_1 M_2 \begin{pmatrix} x \\ y \end{pmatrix} &= M_1 \begin{pmatrix} b_{11}x + b_{12}y \\ b_{21}x + b_{22}y \end{pmatrix} \\ &= \begin{pmatrix} (a_{11}b_{11} + a_{12}b_{21})x + (a_{11}b_{12} + a_{12}b_{22})y \\ (a_{21}b_{11} + a_{22}b_{21})x + (a_{21}b_{12} + a_{22}b_{22})y \end{pmatrix}. \end{aligned} \quad (4.66)$$

Direction calculations indicate that the above two-step transformation equals to the following transformation

$$M_3 = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}, \quad (4.67)$$

that is,

$$\begin{aligned} M_1 M_2 &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix} = M_3. \end{aligned} \quad (4.68)$$

Equation (4.68) clearly demonstrates the rules of matrix multiplication: the entry on the first row and first column of matrix  $M_3$  is the product of the first row of  $M_1$  and the first column of  $M_2$ ; the entry on the first row and second column of matrix  $M_3$  is the product of the first row of  $M_1$  and the second column of  $M_2$ ; the entry on the second row and first column of matrix  $M_3$  is the product of the second row of  $M_1$  and the first column of  $M_2$ ; the entry on the second row and second column of matrix  $M_3$  is the product of the second row of  $M_1$  and the second column of  $M_2$ . In the other words, in the multiplication of two matrices, the left one is regarded as made of two row vectors and the right one is viewed as made of two column vectors. Multiplication of these row vectors and column vectors, respectively, results in the new matrix. The interested reader can use these rules to prove  $QR_3 = W$  in the above example.

Matrix multiplication has a very important property of noncommutativity, that is, the order of multiplication is important. Let us calculate  $R_3Q$ ,

$$R_3Q = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}-1}{2} & -\frac{1}{2} \\ \frac{\sqrt{3}+1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}. \quad (4.69)$$

It is clear that  $R_{30}Q \neq QR_{30}$ , i.e., the order of matrix multiplication matters. For two arbitrary matrices,  $M_1$  and  $M_2$ , we have  $M_1M_2 \neq M_2M_1$  in general. Only in some special cases the order can be changed, for instance,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (4.70)$$

In the two-dimensional real spaces, the non-commutativity of matrix multiplication has a clear geometric meaning. To see it, let us introduce the matrix of two dimensional rotation,

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (4.71)$$

The action of  $R_\theta$  on a column vector represents a counterclockwise rotation around the origin by an angle  $\theta$ . Previously we have considered one special case  $\theta = 30^\circ$ . For another special case  $\theta = 90^\circ$ , we have

$$R_{90} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (4.72)$$

Consider again square ABCD in Fig. 4.4a. Let us operate two different sets of transformations on it: (1) rotation  $R_{90}$  followed by shearing  $Q$ ; (2) shearing  $Q$  followed by rotation  $R_{90}$ . In the first case, rotation  $R_{90}$  does not cause any actual change due to the symmetry of a square. Therefore, the overall effect is the same as a single shear transformation  $Q$ , which results in parallelogram  $A'B'C'D'$  in Fig. 4.4a. By contrast, the second set of operations rotates the parallelogram  $A'B'C'D'$  counterclockwise by  $90^\circ$ , yielding a different parallelogram  $A''B''C''D''$  as shown in Fig. 4.4b. Thus, by changing the order of operations, we obtain different results. We can also directly verify that  $R_{90}Q \neq QR_{90}$  using matrix multiplication. This is the geometric meaning of non-commutativity of matrix multiplication. The interested reader can compare another two sets of transformations: (1) rotation  $R_{30}$  followed by shearing  $Q$ ; (2) shearing  $Q$  followed by rotation  $R_{30}$ . These two sets of transformations on square ABCD will also result in two different parallelograms.

A matrix is often called a linear transformation or a linear operator. The mathematics about matrix and linear vector space is called linear algebra. It is called linear for perhaps two reasons: (1) a linear combination of any two vectors,  $|\psi\rangle$  and  $|\phi\rangle$ , which is  $c_1|\psi\rangle + c_2|\phi\rangle$ , is still a vector; (2) a transformation represented by a matrix never turns a straight line into a curve. The interested reader can think about why.

A generic matrix has  $n$  rows and  $n$  columns, and the matrix elements can be complex numbers,

$$M = \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n1} & M_{n2} & \cdots & M_{nn} \end{pmatrix}. \quad (4.73)$$

Here  $M_{ij}$  denotes the matrix element at the  $i$ th row and the  $j$ th column. Matrix elements with identical row and column indices, such as  $M_{11}$  and  $M_{22}$ , are called diagonal elements; Other entries, with different row and column indices, are called off-diagonal matrix elements, such as  $M_{12}$  and  $M_{29}$ .

For an  $n \times n$  matrix  $M$ , multiplying by a constant amounts to multiplying every matrix element by the constant. For two  $n \times n$  matrices,  $M$  and  $P$ , the addition  $G = M + P$  is defined as

$$G_{ij} = M_{ij} + P_{ij}, \quad (4.74)$$

which amounts to the addition of corresponding matrix elements. The product of two matrices,  $D = MP$ , is given by

$$D_{ij} = \sum_{k=1}^n M_{ik} P_{kj}. \quad (4.75)$$

That is, the entry  $D_{ij}$  of matrix  $D$  is the multiplication of the  $i$ th row of matrix  $M$  and the  $j$ th column of matrix  $P$ . As an example, we consider two  $4 \times 4$  matrices

$$\gamma^0 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix}. \quad (4.76)$$

Adding these two matrices yields a new matrix

$$\gamma^0 + \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & i & -i \\ 0 & 0 & -i & i \\ i & i & 0 & 0 \\ i & i & 0 & 0 \end{pmatrix}. \quad (4.77)$$

Their multiplication or product is

$$\gamma^1 \gamma^0 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (4.78)$$

$$\gamma^0 \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \quad (4.79)$$

Obviously,  $\gamma^0 \gamma^1 \neq \gamma^1 \gamma^0$ . Note that since both matrices contain imaginary elements, a geometrical meaning of noncommutativity of matrix multiplication in this example

is no longer clear. As we will see, the property that matrix multiplication is in general non-commutative provides the mathematical foundation for the non-commutability of operators in quantum mechanics, giving rise to quantum effects completely incomprehensible in classical mechanics, such as Heisenberg's uncertainty relation.

In the following we introduce two very important matrix operations.

- **Transpose.** The transpose of matrix  $M$ , denoted by  $M^T$ , is the same set of elements, but with rows and columns interchanged. Mathematically, we have  $M_{ij}^T = M_{ji}$ . The following two matrices are the transpose of each other,

$$\begin{pmatrix} 1 & 2 & 3 \\ i & 2i & 3i \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & i & 4 \\ 2 & 2i & 5 \\ 3 & 3i & 6 \end{pmatrix}^T. \quad (4.80)$$

- **Hermitian conjugate.** The Hermitian conjugate of matrix  $M$ , denoted by  $M^\dagger$ , is the complex conjugate of the transpose  $M^T$ . Mathematically, we have  $M_{ij}^\dagger = M_{ji}^*$ . The following two matrices are the Hermitian conjugate of each other.

$$\begin{pmatrix} 1 & 2 & 3 \\ i & 2i & 3i \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & -i & 4 \\ 2 & -2i & 5 \\ 3 & -3i & 6 \end{pmatrix}^\dagger. \quad (4.81)$$

Using the above two operations, we can define several important types of matrices.

- **Diagonal matrix.** All off-diagonal entries of a diagonal matrix are zero. For instance,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 8 \end{pmatrix}. \quad (4.82)$$

If matrices  $M_1$  and  $M_2$  are diagonal matrices, they commute, i.e.,  $M_1 M_2 = M_2 M_1$ . Interested readers are encouraged to prove it.

- **Identity matrix.** A diagonal matrix whose diagonal entries are all equal to 1 is an identity matrix, and is denoted by  $I$ .
- **Inverse matrix.** For a matrix  $M$ , its inverse is denoted by  $M^{-1}$ . Inverse matrix satisfies  $MM^{-1} = M^{-1}M = I$ . Not every matrix has an inverse.
- **Symmetric matrix.** A symmetric matrix  $M$  is a matrix that satisfies the property  $M = M^T$ . The following is an example of symmetric matrix:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 3 & 4 & 6 \end{pmatrix}. \quad (4.83)$$

- **Hermitian matrix.** A Hermitian matrix  $M$  is a matrix that satisfies the property  $M = M^\dagger$ . Following matrices are Hermitian matrices.

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (4.84)$$

These three matrices are called Pauli matrices. It is obvious that a symmetric real matrix (i.e., all entries are real numbers) is a Hermitian matrix.

- *Unitary matrix.* A unitary matrix  $M$  satisfies  $M^\dagger M = MM^\dagger = I$ . The following matrix is a unitary matrix.

$$U = \begin{pmatrix} \cos \theta & i \sin \theta \\ i \sin \theta & \cos \theta \end{pmatrix}. \quad (4.85)$$

One can show straightforwardly that it satisfies  $U^\dagger U = UU^\dagger = I$ .

As we will see, matrices are widely used in quantum mechanics. They have also many applications in mathematics. Here we consider an example, using matrix to solve two linear equations with two variables. For the following linear equations,

$$2x + 3y = 2, \quad (4.86)$$

$$x - 3y = 1, \quad (4.87)$$

we usually add these two equations to eliminate variable  $y$ , obtaining  $x = 1$ . We then insert it back to the equation to yield  $y = 0$ . Now let us solve it using matrix. We rewrite the equations into a matrix form as follows

$$\begin{pmatrix} 2 & 3 \\ 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \quad (4.88)$$

For the matrix

$$M = \begin{pmatrix} 2 & 3 \\ 1 & -3 \end{pmatrix}, \quad (4.89)$$

its inverse is

$$M^{-1} = \begin{pmatrix} 1/3 & 1/3 \\ 1/9 & -2/9 \end{pmatrix}. \quad (4.90)$$

Multiplying both sides of Eq. (4.88) by  $M^{-1}$ , we obtain

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 \\ 1/9 & -2/9 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (4.91)$$

which is exactly the solution of the equation. While for this simple problem we do not really need matrix, it illustrates a general approach for solving linear equations with matrix.



### 4.2.4 Eigenstates and Eigenvalues

Given a matrix  $M$ , there exist some special vectors  $|\psi\rangle$  which satisfy

$$M|\psi\rangle = v|\psi\rangle. \quad (4.92)$$

Namely, applying a matrix  $M$  to these vectors is equal to multiplying these vectors by a scalar. In mathematics,  $|\psi\rangle$  is known as the eigenvector of matrix  $M$ , and  $v$  is the eigenvalue associated with  $|\psi\rangle$ . For instance, the matrix  $Q$  has an eigenvector

$$|\alpha\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (4.93)$$

and the corresponding eigenvalue is 1. This can be directly verified as follows

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (4.94)$$

Evidently, vector  $c|\alpha\rangle$  ( $c$  is an arbitrary complex number) is also an eigenvector of  $Q$ . In mathematics,  $c|\alpha\rangle$  is not considered as a new eigenvector. As such, the shear matrix  $Q$  has only one eigenvector. The proof is left for the interested reader.

In quantum mechanics, one is primarily concerned with the eigenvectors and eigenvalues of a Hermitian matrix. For an arbitrary  $n \times n$  Hermitian matrix  $O$ , its eigenvectors and eigenvalues have the following general properties.

- $O$  have  $n$  eigenvectors  $|\phi_1\rangle, |\phi_2\rangle, \dots, |\phi_n\rangle$ , which correspond to eigenvalues  $v_1, v_2, \dots, v_n$ , i.e.,  $O|\phi_j\rangle = v_j|\phi_j\rangle$ , ( $j = 1, 2, \dots, n$ ).
- The eigenvalues  $v_1, v_2, \dots, v_n$  are real.
- Eigenvectors  $|\phi_1\rangle, |\phi_2\rangle, \dots, |\phi_n\rangle$  are orthogonal to each other, i.e.,  $\langle\phi_i|\phi_j\rangle = 0$ , ( $i \neq j$ ).
- Since  $c|\phi_j\rangle$  is also an eigenvector, we can use this property to achieve  $\langle\phi_j|\phi_j\rangle = 1$  by choosing appropriate  $c$ . This is the normalization of an eigenvector.

If two eigenvectors  $|\phi_i\rangle$  and  $|\phi_j\rangle$  correspond to the same eigenvalue,  $v_i = v_j$  ( $i \neq j$ ), we say eigenvectors  $|\phi_i\rangle$  and  $|\phi_j\rangle$  are degenerate. Proofs of the first three properties are beyond the scope of this book. Interested readers can find it in any standard linear algebra book or quantum mechanics textbook. Because an eigenvector in quantum mechanics represents to a quantum state, physicists prefer to refer to eigenvectors as eigenstates.

As an example, we consider a  $4 \times 4$  Hermitian matrix

$$\Omega_x = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (4.95)$$

It has four eigenstates

$$|\phi_1\rangle = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad |\phi_2\rangle = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \quad |\phi_3\rangle = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \quad |\phi_4\rangle = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \quad (4.96)$$

which correspond to eigenvalues 1,1,-1,-1. One can verify these four eigenstates are orthonormal by direct calculation. Moreover, eigenstates  $|\phi_1\rangle$  and  $|\phi_2\rangle$  are degenerate with eigenvalue 1, and eigenstates  $|\phi_3\rangle$  and  $|\phi_4\rangle$  are degenerate with eigenvalue -1.

### 4.2.5 Direct Product

In quantum mechanics, we often encounter composite systems, which are composed of multiple subsystems. To construct the Hilbert space of a composite system, we need a mathematical tool called direct product.

Two Hilbert spaces,  $V_1$  and  $V_2$ , can make up a new Hilbert space  $V = V_1 \otimes V_2$ . Here  $\otimes$  is called direct product. If the dimension of Hilbert space  $V_1$  is  $n_1$ , the dimension of Hilbert space  $V_2$  is  $n_2$ , then the direct product  $V$  of the two spaces has a dimension  $n = n_1 n_2$ . Suppose  $|\phi\rangle$  is a vector in Hilbert space  $V_1$  and  $|\varphi\rangle$  is a vector in Hilbert space  $V_2$ , then their direct product is a vector in Hilbert space  $V$ ,  $|\psi\rangle = |\phi\rangle \otimes |\varphi\rangle$ . A vector in the space  $V$  can always be expressed as the direct product of vectors in  $V_1$  and  $V_2$ , or a linear superposition of several direct product, i.e.,

$$|\Psi\rangle = c_1 |\phi_1\rangle \otimes |\varphi_1\rangle + c_2 |\phi_2\rangle \otimes |\varphi_2\rangle + \cdots + c_j |\phi_j\rangle \otimes |\varphi_j\rangle + \cdots \quad (4.97)$$

Direct product has the following properties:

*Rules of direct product* They are similar to the ordinary multiplication (or product)

$$\begin{aligned} & (|\phi_1\rangle + |\phi_2\rangle) \otimes (|\varphi_1\rangle + |\varphi_2\rangle) \\ &= |\phi_1\rangle \otimes |\varphi_1\rangle + |\phi_1\rangle \otimes |\varphi_2\rangle + |\phi_2\rangle \otimes |\varphi_1\rangle + |\phi_2\rangle \otimes |\varphi_2\rangle. \end{aligned} \quad (4.98)$$

A special example is  $(|\phi_1\rangle + |\phi_2\rangle) \otimes |\varphi\rangle = |\phi_1\rangle \otimes |\varphi\rangle + |\phi_2\rangle \otimes |\varphi\rangle$ .

*Non-commutativity*  $|\phi\rangle \otimes |\varphi\rangle \neq |\varphi\rangle \otimes |\phi\rangle$ .

*Multiplication by scalar*  $c|\phi\rangle \otimes |\varphi\rangle = |\phi\rangle \otimes c|\varphi\rangle$ .

There are some other operation rules about direct product, especially about the inner product. We will introduce them in Chapter 7 in the context of the double spin system.

# Chapter 5

## Into the Quantum World



Now let us enter the world of quantum mechanics. In this world, almost all the concepts that have been taken for granted in classical mechanics are discarded, and all the intuitions that are gained from our daily life are challenged. Quantum mechanics is a wholly new world. There, particles no longer have definite trajectories and are described by magical wave functions; future can only be predicted with probabilities; structureless particles can exhibit spin; the sun can in principle simultaneously rise in the east and set in the west; two particles can display mysterious correlations; energy can be discrete. There you describe nature with complex numbers and matrices. I will demonstrate all these amazing quantum phenomena to you mostly through the Stern-Gerlach experiment and spin. Quantum interference will be illustrated with the famed double-slit experiment. In this chapter, I will introduce the basic framework of quantum mechanics.

### 5.1 The Stern-Gerlach Experiment

In 1922, two German physicists, Stern (Otto Stern, 1888–1969) and Gerlach (Walther Gerlach, 1889–1979), carried out an experiment that strongly influenced later developments in modern physics.<sup>1</sup> Figure 5.1 is a sketch of this experiment. Stern and Gerlach generated a beam of neutral silver atoms by evaporating silver in a hot furnace, which was then sent through a spatially varying magnetic field, before they struck a detection screen. They found that the beam of silver atoms was deflected

---

<sup>1</sup> In 1922 quantum mechanics was still in its early stage of development (see Chap. 2), and the existence of the electron spin was not even known. People had no idea how to explain the experimental results for a while. But we shall skip the messy history and explain this famous experiment using contemporary quantum theory.

by the non-uniform magnetic field, and split into two parts, resulting in two separate spots on the detection screen, rather than a continuous stripe predicted by classical physics.

Stern and Gerlach knew that each silver atom carries a magnetic moment and can be regarded as a small magnet (see Fig. 5.1a). When this small magnet is in a magnetic field, the field will exert a force on the north pole and an opposing force on the south pole.<sup>2</sup> If the magnetic field is spatially homogenous, the forces exerted on opposite ends of the magnet cancel each other out, and the silver atom does not feel any net force. However, if the magnetic field is non-uniform as in the experiment, the forces on the two ends will be different, so that there is a net force which deflects the atom's trajectory. This net force is determined by the angle between the orientations of the magnetic moment and the magnetic field, the strength of which increases when the angle decreases. If two silver atoms have opposite magnetic moments (e.g. the first two silver atoms in Fig. 5.1b), they will feel opposite forces and they are deflected in opposite directions. Since the silver atoms are emitted from a high-temperature furnace, their magnetic moments are randomly orientated with equal probability in every direction. It means that there is a continuous distribution of the net force acting on these silver atoms. Therefore, according to classical physics, the resulting distribution on the detector screen should be a continuous stripe. Instead, two separate spots were observed by Stern and Gerlach in their experiment.

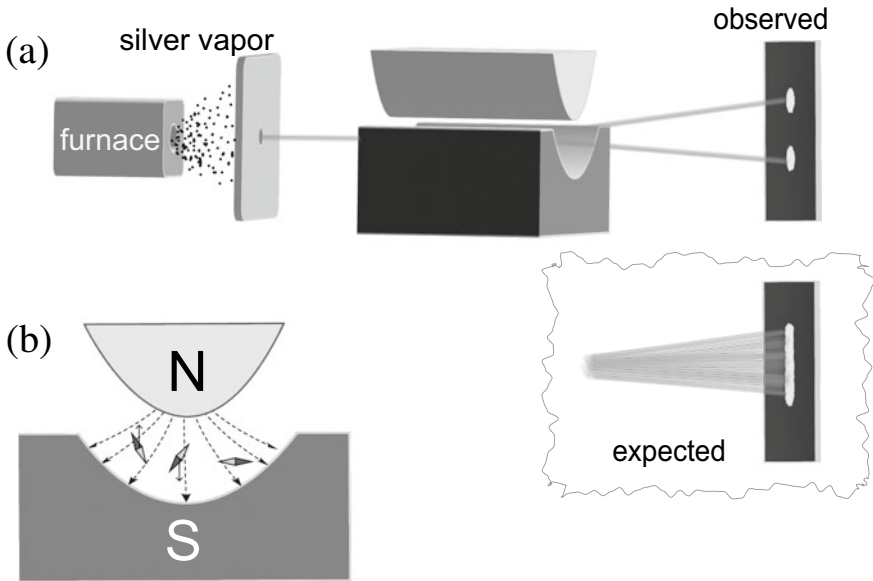
Let us simplify this experiment to see what really happens to the silver atom. In 1922, the silver atoms in the Stern-Gerlach experiment were produced from a high-temperature furnace. The flux of the atomic beam was so strong that there were a large number of silver atoms traveling together through the non-uniform magnetic field before they struck the detection screen. Today the technology has been significantly improved so that it is possible to have only one silver atom passing through the magnetic field each time. Suppose the magnetic moment of the silver atom still orients randomly. What will be the result? A silver atom would be deflected either up or down, each with a probability of 1/2. This is like tossing a coin, the probability to find heads or tails is 1/2.

However, the analogy between a silver atom and a coin quickly breaks down upon further analysis. If someone rolls a dice in front of you (note that it has six sides!) and the dice shows only 1 or 6 each time, you immediately begin to suspect that the dice has been loaded. As emphasized earlier, the silver atom comes from a hot furnace and has a randomly orientated magnetic moment with equal probability in any direction. The Stern-Gerlach experiment is like rolling a dice with an infinite number of sides (which is essentially a ball!). So how can it be possible that there are only two outcomes as indicated by two separate spots on the detection screen (see Fig. 5.1)? The silver atom must have been loaded!

It is loaded by quantum mechanics. The silver atom has a degree of freedom that cannot be described by classical mechanics—spin. As a result, it has a magnetic

---

<sup>2</sup> This explanation is given by drawing an analogy with an electric dipole in a non-uniform electric field. Here scientific rigor, which requires advanced physics and mathematics that is beyond the scope of this book, is sacrificed for simplicity.



**Fig. 5.1** **a** The Stern-Gerlach experiment. Two separated spots are observed on the detection screen instead of an expected continuous stripe. **b** A silver atom has an unpaired electron. Because of the spin of this unpaired electron, a silver atom has a small magnetic moment and behaves as a small magnet, feeling a force in a non-uniform magnetic field. Three typical examples are given: the first one feels an upward force, the second feels a downward force, and the third feels no force

moment. Although spin (or magnetic moment) can point in any direction in space, quantum mechanics demands that there are only two measurement outcomes, and therefore only two spots on the detection screen.

## 5.2 Spin

Almost every microscopic particle has a special type of angular momentum, spin. When an object rotates around another object (such as the revolution of the Earth around the Sun), or around its own axis (such as a gyroscope), it has an angular momentum. The angular momentum associated with the spatial rotation of an object is called orbital angular momentum. For distinction, spin is referred to as an intrinsic form of angular momentum. In classical physics, an object has only orbital angular momentum; in quantum physics, a particle or object can have both orbital angular momentum and spin. We must use quantum mechanics to describe spin. When a charged particle rotates, it carries an orbital angular momentum and thus a magnetic moment, so that the particle behaves like a small magnetic compass. For most of the particles with spin, such as electron and neutron, they have magnetic moments.

Because of this property, we can regard spin as a very small compass. Although this analogy is not rigorous, it allows us to gain some intuitive understanding.

There are many types of spin. For simplicity, we will only consider the simplest spin, spin  $1/2$ . Unless explicitly stated otherwise, a spin means a spin  $1/2$  in the following discussion. An electron has spin  $1/2$ . A silver atom has 47 electrons, among which 46 are paired without manifesting the effect of spin. This leaves one unpaired electron in the energy level  $5s$ . It is exactly the spin of this unpaired electron that gives rise to the magic phenomena observed in the Stern–Gerlach experiment.

As said, spin is a special form of angular momentum. Physicists find that the orbital angular momentum is always an integer multiple of Planck’s constant,  $m\hbar$ , where  $m$  is an integer (positive or negative). The angular momentum corresponding to spin can be  $\pm\hbar/2$ ,  $\pm\hbar$ ,  $\pm3\hbar/2$ , etc. Moreover, the spin angular momentum of a particle is found to have a maximum value. For example, the maximum spin angular momentum of an electron is  $\hbar/2$ . Spin  $1/2$  means that the maximum spin angular momentum is  $\hbar/2$ . Proton and neutron are also spin- $1/2$  particles. Photons are spin 1 particles, which means that the maximum spin angular momentum of a photon is  $\hbar$ . In quantum mechanics, both orbital angular momentum and spin angular momentum take on discrete values. For example, the angular momentum of spin  $1/2$  can only be  $-\hbar/2$ ,  $\hbar/2$ , and the angular momentum of spin 1 can only be  $-\hbar$ ,  $0$ ,  $\hbar$ .

Spin is intimately connected to fermions and bosons introduced in Chap. 2. Fermions have half-integer spins, such as  $1/2$ ,  $3/2$ ,  $5/2$ , etc.; bosons have integer spins, such as  $0$ ,  $1$ ,  $2$ , etc.

There is no spin  $1/4$ . Dirac discovered that the electron spin naturally arises from the combination of the special relativity and quantum mechanics and it is  $1/2$ . If experimental physicists discovered spin- $1/4$  particles in nature some day, then theoretical physicists would have to revise either relativity or quantum mechanics, or both.

Similar to the rest mass and electric charge, spin is an intrinsic property of microscopic particles. But spin is conceptually richer because it is also a degree of freedom. If a particle can move in a real space, physicists say that this particle has spatial degrees of freedom. Physicists have discovered that particles can also move in an abstract space associated with spin. Each “point” in this space represents a quantum state of spin (briefly, spin state). Therefore, spin is also a degree of freedom. Such spin degrees of freedom do not exist at all in classical mechanics and can only be described by quantum mechanics. For spin  $1/2$ , this abstract space is a two-dimensional Hilbert space, where a vector represents a spin state. We shall begin with two special spin states, spin up  $|u\rangle$  and spin down  $|d\rangle$ ,

$$|u\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |d\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (5.1)$$

A straightforward calculation verifies that  $\langle u|d\rangle = 0$  and  $\langle u|u\rangle = \langle d|d\rangle = 1$ , i.e.,  $|u\rangle$  and  $|d\rangle$  are orthonormal. As discussed in Chap. 4, these two states  $|u\rangle$  and  $|d\rangle$  form an orthonormal basis of the two-dimensional Hilbert space. So, an arbitrary vector  $|\psi\rangle$  in this space can be expressed as a linear superposition of  $|u\rangle$  and  $|d\rangle$ ,

$$|\psi\rangle = c_1 |u\rangle + c_2 |d\rangle = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \quad (5.2)$$

We emphasize that  $c_1$  and  $c_2$  here are complex numbers.

Quantum mechanics states that, if we measure the above state, the probability to find spin up (i.e.,  $|u\rangle$ ) is  $|c_1|^2$  and the probability to find spin down (i.e.,  $|d\rangle$ ) is  $|c_2|^2$ . Because the total probability should be 1, we require  $|c_1|^2 + |c_2|^2 = 1$ . This condition is called the normalization condition. In quantum mechanics, a quantum state is required to satisfy the normalization condition.

The brief theory of spin described above is already sufficient for us to explain the Stern-Gerlach experiment. Suppose when a silver atom comes out of a high-temperature furnace, it is in the following quantum state

$$|\psi_{1/6}\rangle = \sqrt{\frac{1}{6}} |u\rangle + \sqrt{\frac{5}{6}} |d\rangle. \quad (5.3)$$

When this silver atom is measured, according to the above theory, the probability to finding spin up is  $1/6$  and the probability for spin down is  $5/6$ . That is, if there are 60 such silver atoms passing through the non-uniform magnetic field in the Stern-Gerlach experiment, roughly 10 will be deflected up and 50 will be deflected down. Overall, we will observe two spots of the same size on the detection screen, the upper one being smaller than the one below. In the real experiments, the orientation of the magnetic moment of a silver atom is random, corresponding to a random spin state, i.e.  $c_1$  and  $c_2$  in Eq. (5.2) are arbitrary. Thus we should see two spots of the same size on the detection screen. This is exactly what is observed in the Stern-Gerlach experiment.

Since  $c_1$  and  $c_2$  in Eq. (5.2) can be used to store information, spin  $1/2$  is often used as a quantum bit (briefly, qubit) in the field of quantum information, which be introduced in Chaps. 9 and 10.

### 5.3 Quantum State and Its Statistical Interpretation

Consider a classical particle; for simplicity, we focus on the one dimensional case. As discussed in Chap. 3, the state of this particle is represented by a point  $(x, p)$  in its phase space, where  $x$  and  $p$  are both real with  $x$  being its spatial position and  $p$  its momentum. If we measure this particle, we can obtain definite values for its position  $x$  and momentum  $p$ . There is no indeterminacy, and therefore, no probability in the measurement outcome.<sup>3</sup>

---

<sup>3</sup> In a realistic experiment, there is always some noise, which causes some uncertainty for the measurement outcome. However, this kind of uncertainty can be reduced in principle as small as one wishes.

The situation changes drastically in quantum mechanics. The state of a quantum particle, i.e. its quantum state, is a vector in a Hilbert space, which is mathematically described by a set of complex coordinates. For spin-1/2, this Hilbert space is two dimensional, and thus its quantum state is specified with two coordinates,  $c_1$  and  $c_2$  as in Eq. (5.2). But such coordinates are very different from the real coordinates  $(x, p)$  in a phase space. Mathematically,  $c_1$  and  $c_2$  are complex numbers; physically, neither  $c_1$  nor  $c_2$  is directly measurable, and they only give the probability to obtain a certain measurement outcome. For a single spin, the outcome is probabilistic: it could be spin up (with probability  $|c_1|^2$ ) or spin down (with probability  $|c_2|^2$ ).

We encounter probability very often in our daily lives. For example, if you throw a die, there are 6 possible outcomes, each with a probability 1/6. If we only throw once, we do not know the exact result, for there are 6 possible results. Similar thing happens when we make a measurement on a quantum state. However, this similarity is only on surface. When we look deep and more carefully, we find that the probability in quantum mechanics is fundamentally different from the probability we encounter in our daily life. For clarity, we call the former quantum probability and the latter classical probability. The classical probability results from our lack of knowledge or accidents, whereas the quantum probability is intrinsic to the system.

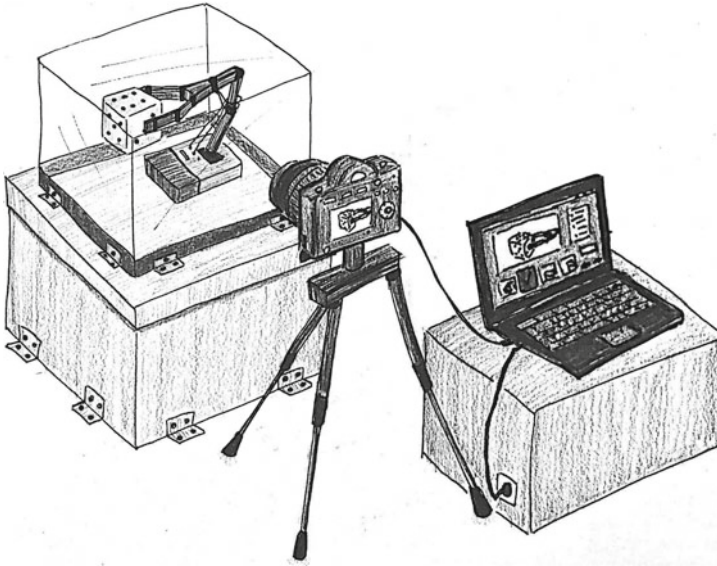
Let us use dice to illustrate the origin of classical probability. When a dice is rolled, some random events may occur. For instance, someone may have accidentally moved the table, or someone's jewelry may have suddenly dropped and touched the dice, etc. These external accidental elements create uncertainties. But we can eliminate these uncertainties by preventing accidents. As shown in Fig. 5.2, we use a robotic hand to roll the dice in a transparent box, which is firmly mounted on a heavy table. Furthermore, we pump the air out of the box and make it ultra-high vacuum. In this way, we are able to prevent the aforementioned accidents to happen. But even so, we still cannot tell which side of the die will face up, for the die may bounce multiple times on the table (and possibly against the walls of the box), rendering its movement so complicated that it is beyond our comprehension. As a result, we cannot predict the exact outcome. All we know is that every side has a probability to face up. In this case, the probability arises because we do not have sufficient knowledge. If one could find a way to know all the details of bouncing, the probability would disappear.

There is a clever physicist named Xiaoliang. He carefully studies the materials making up the die, the table, and the box walls, such as the elastic properties of the materials. He also comprehensively investigates the movement of a die after it bounces off from the table or a box wall at a given angle and velocity. Xiaoliang also sets up a high-speed camera outside the box to record the state of the dice, which includes its position, velocity, and rotation, at the moment of leaving the robotic hand. Note that the camera is used only to record the initial state of the dice and it is no longer used afterwards. With all the preparation and the relevant knowledge, he is able to write a computer code to predict exactly which side would face up when it comes to rest<sup>4</sup>. For Xiaoliang, dice rolling is no longer a random event whose

---

<sup>4</sup> It takes about 2–3 seconds for the dice to stop. With modern computer, it takes only a fraction of a second to finish the computing and predict the outcome.





**Fig. 5.2** A setup for accurately predicting the outcome of a rolling dice. The drawing is not to scale

outcome occurs with a probability; for him, the outcome can be precisely predicted as soon as the dice is being thrown by the robotic hand.

Now let us take a closer look at quantum probability. Consider a spin in the following state,

$$|\psi'_{1/6}\rangle = \sqrt{\frac{1}{6}} |u\rangle + i\sqrt{\frac{5}{6}} |d\rangle. \quad (5.4)$$

We use the Stern-Gerlach set-up in Fig. 5.1 to measure this spin state by replacing the furnace with a particle source so that the emitted silver atoms are always in the above spin state. According to quantum mechanics, there is a probability of  $1/6$  to find spin up and a probability of  $5/6$  to find spin down. Assume that there are 6000 such silver atoms emitted from the source and passed through the non-uniform magnetic field. We will observe two spots on the detection screen, with about 1000 atoms in the upper spot and 5000 atoms in the lower one. This observation can be reproduced exactly using the dice. To compare this quantum spin system to a classical system, we manufacture a special kind of die with one of six faces marked with “up” and five other faces marked with “down”. When 6000 such dices are tossed, there will be about 1000 dices showing “up” and 5000 dices showing “down”. From this comparison, there seems no distinction between the Stern-Gerlach experiment and rolling dice. So, is the probability in quantum mechanics (quantum probability) same as the probability of dice rolling (classical probability)? The answer is no. Here is our analysis.

Let us examine the various elements in the Stern-Gerlach experiment that may affect the outcome of measurement. First, the particle source may be imperfect, so that the spin state of emitted silver atoms may differ slightly from Eq. (5.4); second, the silver atoms may be affected by some accidental events during their flight, such as collisions with air molecules; finally, there may be small vibrations in the magnets generating the magnetic field. All these random elements, often referred to as the measurement noise, can affect the experimental observations: imperfections in the particle source can affect the exact number of silver atoms accumulated in the spots on the detection screen; collisions with air molecules and the vibrations of the magnets can affect the shape and size of the spot, etc. However, these effects are not substantial, and we can always minimize them by improving the experimental setup, such as by improving the particle source, performing the experiment in a vacuum environment, and fixing the magnet on a very heavy table. For dice rolling, Xiaoliang can predict with certainty which side of the dice will face up after eliminating all the measurement noises. For a spin, however, even after eliminating these noises, we still cannot predict the exact outcome of the spin measurement. Instead, we can only predict an outcome with probability.

So far we have carefully analyzed the physical process of dice rolling. This example demonstrates that the classical probability arises because of an incomplete knowledge of the relevant physical processes and factors, such as the initial state of the die, the elastic properties of materials constituting the die and the table. Could the probabilistic observation in the Stern-Gerlach experiment also arise from our ignorance of some processes? Modern physics tells us, other than those described above, we have not ignored any relevant physical processes and properties. The quantum probability is fundamental and intrinsic. According to quantum mechanics, a spin state, specified by a vector in a Hilbert space, can only indicate the probability of a certain outcomes in a given measurement.

Many famous physicists are very displeased with the probabilistic nature of quantum states, even till this day. Their sentiment is concisely expressed by Einstein's famous quote "God doesn't play dice!". These physicists suggest that there may exist some variables that are not observable with current technology, hence giving rise to the probability in quantum mechanics. As these variables are not included in quantum theory, they are referred to as hidden variables. They believe that there exists a more fundamental theory that incorporate these hidden variables such that one can eliminate probability and accurately predict the outcome of a measurement. Such a theory is called hidden variable theory.

The above discussion may seem a bit abstract. Let us again use dice to explain what is the hidden variable theory. We have in fact discussed two different theories for predicting the outcome of a rolling dice. The first one is the probability theory, which we all are familiar with. This theory predicts which side will appear with what probability. This theory is clearly affected by the shape and mass distribution of the die. If the die is not a cube, such as one face is significantly smaller than the others, the probability of a given side will be changed. If the material making up the dice is not uniform or the die has been deliberately loaded, the center of the mass will not be exactly at the cube center and the probability of an outcome will be affected.

This theory also loosely depends on the material made of the dice. If the material is too soft or even sticky, the outcome is also affected. In short, we can summarize the theory as

$$P_{\text{dice}}(\text{shape, mass distribution, dice material}), \quad (5.5)$$

which shows explicitly that the probability theory of a dice depends on three variables. Other than these three, the whole dice rolling process is affected by many other other factors, such as the material of the table, the material of the box walls, and the initial state of the thrown dice. However, these factors do not enter the probability theory of dice and they are hidden variables for this particular theory.

The second theory for the rolling dice is from Xiaoliang, that clever physicist. All the factors are taken into account in his theory, which can be formally expressed as

$$T_{\text{dice}}(\text{shape, mass distribution, dice material}; \\ \text{table material, box wall material, initial state}). \quad (5.6)$$

Based on this theory, Xiaoliang is able to write a computer code to predict exactly which side of the dice will face up. As a counterpart to the probability theory  $P_{\text{dice}}$ ,  $T_{\text{dice}}$  is regarded as the theory of hidden variables. With these extra variables in  $T_{\text{dice}}$ , the probability is eliminated.

For many physicists such as Einstein, quantum theory is like the probability theory  $P_{\text{dice}}$  of a dice and is an incomplete theory. Since for a dice there is a better and complete theory  $T_{\text{dice}}$  that can predict the outcome with certainty, they believe that there is also a complete theory that is more fundamental than quantum theory and can predict the outcome of an experiment such as the Stern-Gerlach experiment with certainty. For these physicists, there is no fundamental difference between classical probability and quantum probability.

The debate on hidden variables has remained philosophical for a long time, and either side could not convince the other. In 1964, Bell showed that an experimental measurement involving just one particle or spin cannot distinguish the quantum probability from the classical probability. To distinguish them, at least two particles or spins need be involved. Bell proved a famous inequality, which can be violated by the correlation between the probabilistic outcomes of spins, but not by the correlation between the probabilistic outcomes of dices. This provides a way to test the argument experimentally. To date, all Bell tests have found that the hypothesis of local hidden variables is not valid. That is, a quantum state is completely described by a vector in Hilbert space, which only provides probabilistic prediction for the outcome of a measurement. In Chap. 7, we will prove the Bell's inequality and further clarify the probabilistic nature of quantum states.

As already stated, a quantum state is represented by a vector of Hilbert space. But quantum states and vectors of Hilbert spaces do not have a one-to-one correspondence: vectors  $|\psi\rangle$  and  $|\tilde{\psi}\rangle = c|\psi\rangle$  are different mathematically but correspond to the same quantum state. The normalization condition demands  $\langle\psi|\psi\rangle = \langle\tilde{\psi}|\tilde{\psi}\rangle = 1$ ,

so  $|c|^2 = 1$ . As to why they are the same state, it will be explained at the end of this chapter. Careful readers may have noticed the difference between the two spin states in Eqs. (5.3) and (5.4): the coefficient of  $|d\rangle$  in the former is a real number while the counterpart coefficient in the latter is imaginary. Although they give the same probabilistic prediction for an outcome, with a probability of  $1/6$  for spin up and a probability of  $5/6$  for spin down, they are different spin states. For  $|\psi_{1/6}\rangle$  and  $|\psi'_{1/6}\rangle$  to represent the same quantum state, we must have

$$\sqrt{\frac{1}{6}}|u\rangle + \sqrt{\frac{5}{6}}|d\rangle = c\sqrt{\frac{1}{6}}|u\rangle + ci\sqrt{\frac{5}{6}}|d\rangle. \quad (5.7)$$

The identity of the coefficients before  $|u\rangle$  requires  $c = 1$ ; and the identity of the coefficients before  $|d\rangle$  requires  $c = -i$ . These two conditions cannot be simultaneously fulfilled, and therefore,  $|\psi_{1/6}\rangle$  and  $|\psi'_{1/6}\rangle$  represent two different quantum states. For the Stern-Gerlach experiment in Fig. 5.1, these two spin states yield the same result. But if we change the orientation of the magnetic field in the experiment, the two spin states will give different predictions. Further discussions will be presented in Sect. 5.5.

## 5.4 Observables and Operators

We have introduced that a spin has two possible states, spin up  $|u\rangle$  and spin down  $|d\rangle$ . Accordingly, we observe two spots in the Stern-Gerlach experiment. In the discussion, I have deliberately been vague for the sake of simplicity. In particular, I have not explained why  $|u\rangle$  and  $|d\rangle$  describe the spin up and spin down states, respectively.

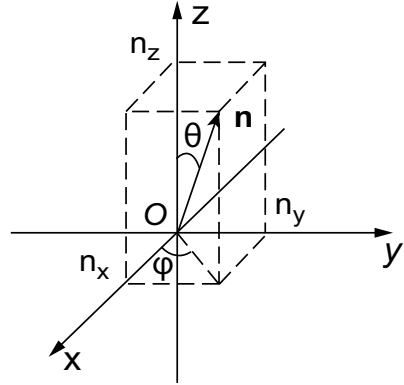
The state of a quantum system is described by a vector  $|\psi\rangle$  in an abstract Hilbert space, and  $|\psi\rangle$  cannot be observed directly in experiments. In order to relate the abstract  $|\psi\rangle$  to physical observations in the real world, quantum mechanics introduces the concept of observables and operators. For the Stern-Gerlach experiment in Fig. 5.1, the observable is the component of spin along the  $z$ -direction, and the corresponding operator is the Pauli matrix  $\hat{\sigma}_z$ . How are these operators related to experimental observations? This connection is established through the eigenstates and eigenvalues of matrices (mathematical forms of operators).

The Pauli matrix  $\hat{\sigma}_z$  is a  $2 \times 2$  Hermitian matrix, with two eigenvectors and two real eigenvalues. Through direct calculations one can easily verify

$$\hat{\sigma}_z |u\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |u\rangle, \quad (5.8)$$

$$\hat{\sigma}_z |d\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} = -|d\rangle. \quad (5.9)$$

**Fig. 5.3** A unit vector in the three-dimensional real space



This shows that both  $|u\rangle$  and  $|d\rangle$  are the eigenstates of  $\hat{\sigma}_z$  with corresponding eigenvalues being 1 and  $-1$ , respectively. For an observable, the outcome of its measurement is an eigenvalue of the corresponding operator. Thus the outcome associated with the measurement of  $\hat{\sigma}_z$  can only be  $\pm 1$ , corresponding to the upper and lower spots in the Stern-Gerlach (SG) experiment. If all the silver atoms in the SG experiment are in the state  $|u\rangle$  whose eigenvalue is 1, they will fly upward, forming a spot in the upper part of the screen. If all the silver atoms are in the state  $|d\rangle$  corresponding to the eigenvalue  $-1$ , they will fly downward, forming a spot in the lower part of the screen. If the spin is in the superposition state  $|\psi_{1/6}\rangle = \sqrt{1/6}|u\rangle + \sqrt{5/6}|d\rangle$ , then the silver atoms fly upwards and downward simultaneously, with a probability of  $1/6$  hitting the upper part of the screen and a probability of  $5/6$  hitting the lower part.

From the above brief introduction, it is clear that quantum mechanics is radically different from classical mechanics. In classical mechanics, the state of a particle is described by the position  $x$  and the momentum  $p$ , the observables are  $x$  and  $p$ , and the observed values are also  $x$  and  $p$ . In quantum mechanics, however, the quantum state, the observable, and the observed values are different concepts: the quantum state is a vector in Hilbert space; the observable is an operator (mathematically, a matrix); and the observed value is an eigenvalue of the operator.

### 5.5 Spin Along an Arbitrary Direction

In the previous section we have discussed about the spin component along the  $z$  axis. If there is only one magnetic field, we can always choose a coordinate system such that the magnetic field is along the  $z$  axis. But we will encounter more complicated problems in the future, such as the double-spin Stern-Gerlach experiment in Sect. 7.1, where there are two magnetic fields with different directions. In this case, we must consider how to describe a spin component along an arbitrary direction.

We use the unit vector  $\mathbf{n} = \{n_x, n_y, n_z\}$  in the real space to represent an arbitrary direction. The vector  $\mathbf{n}$  has a unit length,  $n_x^2 + n_y^2 + n_z^2 = 1$ . Using  $\mathbf{n}$  and the Pauli matrices, we construct the following operator

$$\mathbf{n} \cdot \hat{\sigma} \equiv n_x \hat{\sigma}_x + n_y \hat{\sigma}_y + n_z \hat{\sigma}_z = \begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}. \quad (5.10)$$

This matrix is evidently a Hermitian matrix. Consider a special case,  $\mathbf{n}$  along the  $z$  direction, i.e.,  $\mathbf{n} = \{0, 0, 1\}$ . We have  $\mathbf{n} \cdot \hat{\sigma} = \hat{\sigma}_z$ . As discussed earlier, the eigenstates of  $\hat{\sigma}_z$  are  $|u\rangle$  and  $|d\rangle$ , corresponding to eigenvalues 1 and  $-1$ , respectively.

Consider another special case,  $\mathbf{n}$  along the  $x$  direction, i.e.,  $\mathbf{n} = \{1, 0, 0\}$ . We have  $\mathbf{n} \cdot \hat{\sigma} = \hat{\sigma}_x$ . The observable represented by the operator  $\hat{\sigma}_x$  is the spin component along the  $x$  axis. We define two spin states

$$|f\rangle = \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad (5.11)$$

and

$$|b\rangle = \frac{1}{\sqrt{2}}(|u\rangle - |d\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \quad (5.12)$$

After straightforward calculations, we find that

$$\hat{\sigma}_x |f\rangle = |f\rangle, \quad \hat{\sigma}_x |b\rangle = -|b\rangle. \quad (5.13)$$

This shows that  $|f\rangle$  and  $|b\rangle$  are the eigenstates of the operator  $\hat{\sigma}_x$ , corresponding to the eigenvalues  $\pm 1$  and describing the forward and backward components of the spin, respectively. In the Stern-Gerlach experiment, this means that the magnetic field is oriented along the  $x$  axis, and we will observe two spots on the front and the back. For  $\hat{\sigma}_y$ , we also have two eigenstates, which are

$$|r\rangle = \frac{1}{\sqrt{2}}(|u\rangle + i|d\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad (5.14)$$

and

$$|l\rangle = \frac{1}{\sqrt{2}}(|u\rangle - i|d\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}. \quad (5.15)$$

Similarly, we can prove

$$\hat{\sigma}_y |r\rangle = |r\rangle, \quad \hat{\sigma}_y |l\rangle = -|l\rangle. \quad (5.16)$$

So the corresponding eigenvalues are  $\pm 1$ , respectively. For this case, the magnetic field in the Stern-Gerlach experiment is oriented along the  $y$  axis; one will observe two spots on the right and left parts of the screen.

Consider a general case, where the magnetic field is along an arbitrary direction. We use the polar angle to rewrite the direction as (see Fig. 5.3)

$$\mathbf{n} = \{\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta\}, \quad (5.17)$$

and we have

$$\mathbf{n} \cdot \hat{\sigma} = \begin{pmatrix} \cos \theta & \sin \theta e^{-i\varphi} \\ \sin \theta e^{i\varphi} & -\cos \theta \end{pmatrix}. \quad (5.18)$$

It can be shown that the eigenstates of the operator  $\mathbf{n} \cdot \hat{\sigma}$  are

$$|n_+\rangle = \begin{pmatrix} \cos \frac{\theta}{2} \\ e^{i\varphi} \sin \frac{\theta}{2} \end{pmatrix}, \quad |n_-\rangle = \begin{pmatrix} \sin \frac{\theta}{2} \\ -e^{i\varphi} \cos \frac{\theta}{2} \end{pmatrix}, \quad (5.19)$$

which correspond to eigenvalues  $\pm 1$ , respectively. That is,

$$\mathbf{n} \cdot \hat{\sigma} |n_+\rangle = |n_+\rangle, \quad \mathbf{n} \cdot \hat{\sigma} |n_-\rangle = -|n_-\rangle. \quad (5.20)$$

These results show that only two spots will be observed on the screen of the Stern-Gerlach experiment regardless of the orientation of the magnetic field. From the physics point of view, this can be easily understood: after all, there is nothing special about the  $z$ -axis.

It is important to note that sometimes only one spot is observed in the Stern-Gerlach experiment. For example, when the spin is in the  $|u\rangle$  state, only one spot will be observed experimentally if the magnetic field is oriented along the  $z$ -axis. In fact, for any spin state described by Eq. (5.2), we can always find a direction  $\mathbf{n}$  such that

$$\mathbf{n} \cdot \hat{\sigma} |\psi\rangle = |\psi\rangle. \quad (5.21)$$

By comparing Eq. (5.2) with Eq. (5.19), we can determine the relation between  $\mathbf{n}$  and  $c_1, c_2$  as  $c_1 = \cos \frac{\theta}{2}$ ,  $c_2 = \sin \frac{\theta}{2} e^{i\varphi}$ . The physical implication of this result is as follows. In the Stern-Gerlach experiment, if we replace the high-temperature furnace with a more sophisticated device that produces silver atoms always in the same spin state, we can then orient the magnetic field to a certain direction so that there is only one spot on the detection screen. In literature, people often say that a spin is along a certain direction  $\mathbf{n}$ . What this means is that the spin is in a quantum state which is the eigenstate of operator  $\mathbf{n} \cdot \hat{\sigma}$  with eigenvalue 1.

At the end of Sect. 5.3, we stated that the spin states (5.3) and (5.4) are mathematically different. Now let us discuss their physical differences. In the Stern-Gerlach experiment, we reorient the magnetic field to be along the  $x$ -axis, so that the observable is the spin components along the  $x$ -axis. According to Chap. 4, any two orthonormal vectors can be used as the basis vectors of a two-dimensional Hilbert space. Choosing the two eigenstates of  $\hat{\sigma}_x$  as the basis, we expand the spin state (5.3) as

$$|\psi_{1/6}\rangle = c_1 |f\rangle + c_2 |b\rangle. \quad (5.22)$$

To obtain  $c_1$ , we multiply the two sides by  $\langle f|$  from the left. Using  $\langle f|f\rangle = 1$  and  $\langle f|b\rangle = 0$ , we obtain

$$c_1 = \langle f|\psi_{1/6}\rangle = \frac{\sqrt{5} + 1}{2\sqrt{3}}. \quad (5.23)$$

Similarly, we have

$$c_2 = \langle b|\psi_{1/6}\rangle = \frac{1 - \sqrt{5}}{2\sqrt{3}}. \quad (5.24)$$

So the probabilities to observe the forward and backward components of the spin (i.e., the two components of the spin along the  $x$  direction), respectively, are

$$|c_1|^2 = \frac{3 + \sqrt{5}}{6} \approx 0.873, \quad |c_2|^2 = \frac{3 - \sqrt{5}}{6} \approx 0.127. \quad (5.25)$$

Similarly, we can expand the spin states (5.4) as

$$|\psi'_{1/6}\rangle = c'_1 |f\rangle + c'_2 |b\rangle, \quad (5.26)$$

where

$$c'_1 = \langle f|\psi'_{1/6}\rangle = \frac{1 + i\sqrt{5}}{2\sqrt{3}}, \quad c'_2 = \langle b|\psi'_{1/6}\rangle = \frac{1 - i\sqrt{5}}{2\sqrt{3}}. \quad (5.27)$$

Thus the probabilities to observe the forward and backward components, respectively, are

$$|c'_1|^2 = |c'_2|^2 = \frac{1}{2}. \quad (5.28)$$

These calculations show that  $|\psi_{1/6}\rangle$  and  $|\psi'_{1/6}\rangle$  are different states: for  $|\psi'_{1/6}\rangle$ , the probability of the forward component equals the probability of the backward component, whereas for  $|\psi_{1/6}\rangle$ , the two probabilities differ significantly.

Can we measure  $\hat{\sigma}_x$  and  $\hat{\sigma}_z$  at the same time? Would we see four spots? The answer is no. Such a measurement requires simultaneous presence of a magnetic field along the  $x$  axis and a magnetic field along the  $z$  axis. Instead of two magnetic fields, it turns out you will obtain a single magnetic field along the  $\mathbf{n} = \{n_x, 0, n_z\}$  direction. Therefore, you will not observe four spots on the screen, but rather two spots along the direction  $\mathbf{n} = \{n_x, 0, n_z\}$ .

In addition, theoretically, we find that the operator  $\hat{\sigma}_x$  does not commute with the operator  $\hat{\sigma}_z$

$$[\hat{\sigma}_x, \hat{\sigma}_z] \equiv \hat{\sigma}_x \hat{\sigma}_z - \hat{\sigma}_z \hat{\sigma}_x = -2i\hat{\sigma}_y, \quad (5.29)$$

where  $[\hat{\sigma}_1, \hat{\sigma}_2] \equiv \hat{\sigma}_1 \hat{\sigma}_2 - \hat{\sigma}_2 \hat{\sigma}_1$  is called the commutator of operators  $\hat{\sigma}_1$  and  $\hat{\sigma}_2$ . Similarly, we have

$$[\hat{\sigma}_x, \hat{\sigma}_y] = 2i\hat{\sigma}_z, \quad [\hat{\sigma}_y, \hat{\sigma}_z] = 2i\hat{\sigma}_x. \quad (5.30)$$



So the three operators  $\hat{\sigma}_x$ ,  $\hat{\sigma}_y$ , and  $\hat{\sigma}_z$  do not commute with each other. According to quantum mechanics, this means that the three observables that they represent cannot be precisely determined simultaneously. This is the famous Heisenberg's uncertainty relation. Does this have anything to do with the fact that the spin components along the  $x$ -axis and the  $z$ -axis cannot be measured simultaneously in the Stern-Gerlach experiment? I do not think so. We will discuss this in detail in Chap. 8.

Before concluding our discussion, it is necessary to emphasize once more that, although the spin state represented by Eq. (5.2) can only give a probabilistic prediction for the outcome of a measurement, it does not mean that Eq. (5.2) is incomplete. In quantum mechanics, Eq. (5.2) gives a complete description of the spin state, and we cannot give a more accurate and better description.

In the theory of probability, if the probability to obtain the value  $w_j$  out of  $n$  possibilities is  $p_j$ , then the average of all the possible values is  $\bar{w} = \sum_{j=1}^n w_j p_j$ . For a quantum state  $|\psi\rangle$ , we can also define an average value for an observable, which is called the expectation value. For a spin state  $|\psi\rangle$ , the expectation value of the spin operator  $\mathbf{n} \cdot \hat{\boldsymbol{\sigma}}$  is defined as

$$\langle \psi | \mathbf{n} \cdot \hat{\boldsymbol{\sigma}} | \psi \rangle. \quad (5.31)$$

We can directly verify the following two expectation values

$$\langle u | \hat{\sigma}_z | u \rangle = 1, \quad \langle u | \hat{\sigma}_x | u \rangle = 0. \quad (5.32)$$

These results are consistent with our usual understanding of average value. Because  $|u\rangle$  is an eigenstate of  $\hat{\sigma}_z$ , every measurement on it will give the same result, so  $\langle u | \hat{\sigma}_z | u \rangle = 1$ . Because  $|u\rangle = (|f\rangle + |b\rangle)/\sqrt{2}$ , the measurement of  $\hat{\sigma}_x$  will yield 1 with a probability of 50%, and  $-1$  with a probability of 50%. As a result, the expectation value is zero.

## 5.6 Theoretical Framework of Quantum Mechanics

We have introduced some basics of quantum mechanics with the example of spin. We now provide the general theoretical framework for quantum mechanics.

**Quantum state** A state of a quantum system is represented by a vector in Hilbert space. For the spin discussed earlier, the corresponding Hilbert space is two-dimensional. In general, the dimension of a Hilbert space is  $n$ , which can be infinite. For an  $n$ -dimensional Hilbert space, we can always find  $n$  orthonormal basis vectors  $|e_n\rangle$ . Any quantum state can then be expressed as a linear superposition of these basis vectors as

$$|\psi\rangle = \sum_{j=1}^n c_j |e_j\rangle. \quad (5.33)$$

The expansion coefficient  $c_j$  expresses the projection of the state  $|\psi\rangle$  on the basis vector  $|e_j\rangle$ , and it is given by the inner product of  $|\psi\rangle$  and  $|e_j\rangle$ , i.e.,  $c_j = \langle e_j|\psi\rangle$ . The expansion coefficients must satisfy the normalization condition

$$\sum_{j=1}^n |c_j|^2 = 1. \quad (5.34)$$

**Observable** An observable in a quantum system is represented by an operator, which is expressed mathematically as a matrix. The Pauli matrices discussed earlier represent the observables, spin components along different axes. For an observable  $\hat{O}$ , its eigenstate  $|\phi_j\rangle$  satisfies  $\hat{O}|\phi_j\rangle = v_j|\phi_j\rangle$ , with the eigenvalue  $v_j$  corresponding to a possible outcome in an experiment. For a quantum state  $|\psi\rangle$ , its expectation value is  $\langle\psi|\hat{O}|\psi\rangle$ .

**Probabilistic interpretation** Any quantum state can be expressed as a linear superposition of the eigenstates  $|\phi_j\rangle$  of an observable  $\hat{O}$ <sup>5</sup>

$$|\psi\rangle = \sum_{j=1}^n a_j |\phi_j\rangle, \quad (5.35)$$

where  $|a_j|^2 = |\langle\phi_j|\psi\rangle|^2$  is the probability of finding the system in an eigenstate  $|\phi_j\rangle$ . If we measure  $\hat{O}$  in this quantum state, the probability of the outcome being  $v_j$  is  $|a_j|^2$ .

This is the basic framework of quantum mechanics, but it is not complete because there is no dynamics, i.e., the time evolution of a quantum state. We will introduce quantum dynamics in Chap. 6.

According to the above framework, the overall phase of a quantum state has no physical meaning. In other words, the vectors  $|\psi\rangle$  and  $|\psi'\rangle = e^{i\theta}|\psi\rangle$  ( $\theta$  is a constant real number) represent the same quantum state. The expectation values of  $|\psi\rangle$  and  $|\psi'\rangle$  are the same,

$$\langle\psi'|\hat{O}|\psi'\rangle = \langle\psi|e^{-i\theta}\hat{O}e^{i\theta}|\psi\rangle = \langle\psi|e^{-i\theta}e^{i\theta}\hat{O}|\psi\rangle = \langle\psi|\hat{O}|\psi\rangle. \quad (5.36)$$

The probability of finding the system in the eigenstate  $|\phi_j\rangle$  is also the same

$$|\langle\phi_j|\psi'\rangle|^2 = |\langle\phi_j|e^{i\theta}|\psi\rangle|^2 = |\langle\phi_j|\psi\rangle|^2 |e^{i\theta}|^2 = |a_j|^2. \quad (5.37)$$

Thus  $|\psi\rangle$  and  $|\psi'\rangle = e^{i\theta}|\psi\rangle$  have no physical difference and represent the same quantum state.

---

<sup>5</sup> This book will not discuss the cases where the number of eigenstates is smaller than the dimension of the Hilbert space.

**Table 5.1** Comparison between classical mechanics and quantum mechanics

	Classical mechanics	Quantum mechanics
State of a system	Momentum $p$ , position $x$	A vector $ \psi\rangle$ in a Hilbert space
Observable	Momentum $p$ , position $x$	Operator (matrix), for instance, momentum operator $\hat{p}$ , position operator $\hat{x}$ , spin operator $\mathbf{n} \cdot \hat{\sigma}$
Observed values	Momentum $p$ , position $x$	An eigenvalue of the matrix operator
Measurement outcome	Certain	Probabilistic

The basic framework of quantum theory is very different from that of classical mechanics. The following table is a direct comparison between classical mechanics and quantum mechanics.

As shown in Table 5.1, in classical mechanics, the state of a system, the observable, and the observed value are the same thing. In contrast, they are independent concepts in quantum mechanics. In textbooks or classes on classical mechanics, no one would emphasize this kind of trinity status of momentum and position, which is noticed only after the emergence of quantum mechanics. This makes quantum mechanics strange and difficult to understand.

We will introduce the position operator  $\hat{x}$  and momentum operator  $\hat{p}$  in Chap. 6.

# Chapter 6

## Quantum Dynamics



The centerpiece of classical mechanics is to describe the motion of particles or systems, i.e., how their states vary with time. The time evolution of the state of a system is called dynamics. In classical mechanics, the dynamics is governed by the Newton's second law. In this chapter we describe how quantum states change with time, i.e. quantum dynamics. The equation that describes the time evolution of a quantum state is called the Schrödinger equation. To avoid complicated mathematics, I will focus on the properties of the Schrödinger equation and some of its solutions, and will not discuss how to solve the Schrödinger equation. I will also discuss the principle of superposition of states, the quantum no-cloning theorem, and quantum interference with the double-slit experiment.

### 6.1 Schrödinger Equation

Schrödinger was the first physicist who correctly describes how a quantum state evolves with time. Before Schrödinger, Heisenberg had proposed a quantum dynamical equation, which is known as Heisenberg's equation of motion. But Heisenberg's equation describes how an observable or operator evolves in time, not a quantum state. Dirac later showed that the Schrödinger equation is mathematically equivalent to Heisenberg's equation. In practice, the Schrödinger equation is more convenient to use in most cases, so we will focus on the Schrödinger equation.

The original equation written down by Schrödinger was three dimensional; for the sake of simplicity, we here consider its one-dimensional form

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi(x, t) + V(x)\psi(x, t). \quad (6.1)$$

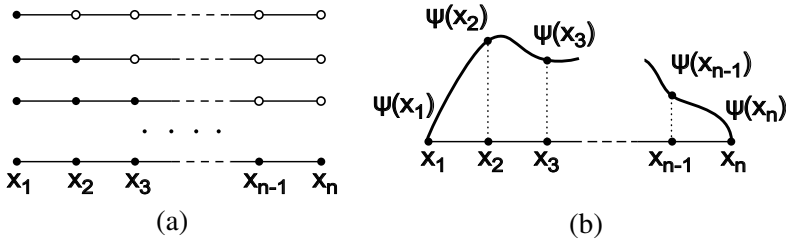
This type of equation is known as partial differential equation in mathematics, and how to find its solutions is beyond the scope of this book. We shall only briefly introduce some of its properties. The  $\psi(x, t)$  in Eq. (6.1) is called wave function, describing the quantum state of a particle at time  $t$ . The square modulus of the wave function,  $|\psi(x, t)|^2$ , predicts the probability of finding the particle at position  $x$  at time  $t$ . The left hand side of Eq. (6.1) describes the rate of change of a wave function  $\psi(x, t)$  over time; the right hand side describes the variation of  $\psi(x, t)$  in space. The above equation establishes a relation between the two. The parameter  $m$  is the mass of the particle, and  $V(x)$  is the potential which the particle experiences. In Chap. 3, we have introduced two systems, a free-falling particle and a harmonic oscillator. In these two cases,  $V(x)$  is given by  $-mgx$  and  $\frac{1}{2}m\omega^2x^2$ , respectively. Interestingly, while the mass and the potential are typical concepts associated with a particle, the wave function  $\psi(x, t)$  describes the behavior of a wave. The Schrödinger equation magically unifies these two distinct aspects, providing a concrete mathematical formulation for the wave-particle duality.

Notice that both sides of the Schrödinger equation contain Planck's constant  $\hbar$ . Thus most results obtained from this equation involves  $\hbar$ . However, there are cases when  $\hbar$  does not appear in the solutions. Consider a harmonic oscillator as an example. We have discussed its classic motion in Chap. 3, where both its position and momentum vary periodically in time, with the period given by  $T = 2\pi/\omega$ . In quantum mechanics, the state of a harmonic oscillator is described by a wave function  $\phi(x, t)$  governed by the Schrödinger equation. By solving the Schrödinger equation of a harmonic oscillator, we find that its wave function also changes periodically in time, with a period  $T$ , i.e.,  $\phi(x, t) = \phi(x, t + T)$ . Interestingly, this property has nothing to do with Planck's constant  $\hbar$ : the classic oscillation period equals the quantum oscillation period. Interested readers can go to Dirac's *Principles of Quantum Mechanics* to learn how to solve the Schrödinger equation of a harmonic oscillator.

This particular example illustrates an important general relation: although the Schrödinger equation and Newton's equations appear to be distinct, the motion they describe are somewhat related. Physicists find that a classical motion can be seen as an approximation of a quantum motion; the higher the energy of the system has, the better the approximation is. In the limit of high energy, the behavior of a particle described by quantum dynamics becomes classical. The harmonic oscillator is a special example, where the quantum mechanical behavior is similar to the classical physics even at low energy. We will revisit the quantum-classical correspondence when we introduce the Hamiltonian operator.

## 6.2 Wave Function

One of the fundamental aspects of quantum mechanics introduced in Chap. 5 is: a quantum state of particles or systems is represented by a vector in a Hilbert space. In the Schrödinger equation, the quantum state of a particle is described by the wave



**Fig. 6.1** **a** One-dimensional lattice. From top to bottom, the number of lattice sites that the particle can occupy increases from 1 to  $n$ . **b** The wave function  $\psi(x)$  in a one-dimensional continuous space

function  $\psi(x)$ . While it appears different at first glance, the wave function  $\psi(x)$  is in fact a vector in a Hilbert space with infinite dimensions. To see why this is the case, let us consider a simple system—a particle moving on a one-dimensional lattice (see Fig. 6.1a). In the simplest case, the particle can only stay at the lattice site  $x_1$ . We say the particle at this site  $x_1$  is in a quantum state  $|x_1\rangle$ , which is a vector in a one-dimensional Hilbert space. This case is quite boring and physically trivial, as the particle can only stay at one lattice site.

A slightly more complicated case is that the particle can be at two lattice sites  $x_1$  and  $x_2$ . Namely, the particle can be in two possible quantum states:  $|x_1\rangle$  or  $|x_2\rangle$ . These two quantum states  $|x_1\rangle$  and  $|x_2\rangle$  satisfy the orthonormal condition,  $\langle x_1 | x_1 \rangle = \langle x_2 | x_2 \rangle = 1$  and  $\langle x_1 | x_2 \rangle = 0$ , spanning a two-dimensional Hilbert space. Any quantum state in this Hilbert space can be written as  $\psi(x_1) |x_1\rangle + \psi(x_2) |x_2\rangle$ , meaning that the probability of finding the particle at  $x_1$  is  $|\psi(x_1)|^2$  and the probability of finding the particle at  $x_2$  is  $|\psi(x_2)|^2$ . Now the physics gets more interesting: not only can the particle be at two different lattice sites simultaneously, it can also hop between them according to the Schrödinger equation.

Following this line of logic, if the particle can be in  $n$  lattice sites, its possible quantum states are

$$|x_1\rangle, |x_2\rangle, |x_3\rangle, \dots, |x_{n-1}\rangle, |x_n\rangle. \quad (6.2)$$

The quantum states  $|x_j\rangle$  also satisfy the orthonormal conditions <sup>1</sup>

$$\langle x_j | x_j \rangle = 1, \quad \langle x_i | x_j \rangle = 0 \quad (i \neq j). \quad (6.3)$$

These states span a  $n$ -dimensional Hilbert space. An arbitrary vector in this Hilbert space can be written as

$$|\psi\rangle = \sum_{j=1}^n \psi(x_j) |x_j\rangle, \quad (6.4)$$

<sup>1</sup> The quantum state  $|x_j\rangle$  can be roughly understood as a special kind of function  $\delta(x - x_j)$ , which approaches infinity at  $x_j$  and vanishes at other points. Using the  $\delta$  function, one can write the orthonormal condition as  $\langle x_i | x_j \rangle = \delta(x_i - x_j)$  in the continuum limit. A mathematically rigorous discussion of the  $\delta$  function can be found in Dirac's book *Principles of Quantum Mechanics*.

where the coefficients satisfy the normalization condition  $\sum_{j=1}^n |\psi(x_j)|^2 = 1$ . The quantum state described by the vector  $|\psi\rangle$  predicts the probability of finding the particle at  $x_j$  is  $|\psi(x_j)|^2$ .

Now we imagine a limiting process. We maintain the distance between the left- and right-most lattice sites, and increase the number of lattice sites to infinity, obtaining a continuous line segment. Accordingly, the coefficient  $\psi(x_j)$  becomes a continuous function  $\psi(x)$  defined on this line segment as seen Fig. 6.1b. Hence we see that the wave function  $\psi(x)$  is indeed a vector in an infinite-dimensional Hilbert space. This result can be further generalized to an infinitely long line and to higher dimensional spaces. Multiplying  $\langle x_i |$  from the left on both sides of Eq. (6.4), we obtain  $\psi(x_i) = \langle x_i | \psi \rangle$ . In the continuous limit, it becomes a function

$$\psi(x) = \langle x | \psi \rangle. \quad (6.5)$$

This formula shows how to represent the wave function  $\psi(x)$  with the Dirac notation.

### 6.3 Hamiltonian Operator and Unitary Evolution

The Schrödinger equation can be written in a more compact form as

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle, \quad (6.6)$$

where  $\hat{H}$  is called Hamiltonian operator. The Schrödinger equation in Eq.(6.1) involves only the spatial degrees of freedom of particles; to describe how the spin state evolves in time, one needs to use the above Schrödinger equation. The Hamiltonian operator corresponding to Eq. (6.1) is

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x). \quad (6.7)$$

If we define the momentum operator as

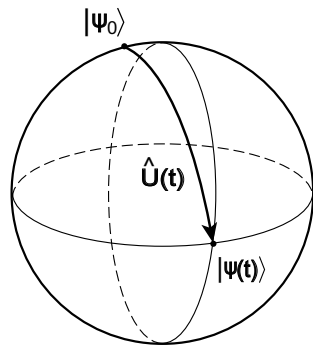
$$\hat{p} = -i\hbar \frac{\partial}{\partial x}, \quad (6.8)$$

we have

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(x). \quad (6.9)$$

This is similar to the classical Hamiltonian  $H = p^2/2m + V(x)$  introduced in Chap. 3, except that the classical momentum  $p$  is replaced by an operator  $\hat{p}$ . We have mentioned that the quantum dynamics of a system approaches classical

**Fig. 6.2** Time evolution in a Hilbert space. A quantum state  $|\psi_0\rangle$  is a vector in a Hilbert space, which transforms into another quantum state  $|\psi(t)\rangle$  under a unitary evolution  $\hat{U}(t)$ . The trajectory of evolution is always on the sphere, indicating that the length of the vector is conserved during the unitary evolution



physics when the energy of the system becomes very high. This generic quantum-classical correspondence is rooted in this close relation between the quantum Hamiltonian and the classical Hamiltonian. For spin, the Hamiltonian operator  $\hat{H}$  has a different form. For example, when spin is in a magnetic field along the  $z$ -direction, its Hamiltonian is given by  $\hat{H} = \mu_b B \hat{\sigma}_z$ , where  $\mu_b$  is the magnetic moment carried by the spin and  $B$  is the strength of the magnetic field. The Hamiltonian of spin has no correspondence in classical mechanics.

The Hamiltonian or Hamiltonian operator of a system plays a central role in modern physics. It was introduced by mathematician Hamilton (Sir William Rowan Hamilton 1805–1865) in 1833. Hamilton found that, starting from the Hamiltonian, he could rigorously reformulate Newtonian mechanics. In other words, Hamilton developed a new formulation of Newtonian mechanics. Although quantum mechanics has abandoned many concepts of classical mechanics (i.e., Newtonian mechanics), the Hamiltonian is retained in the form of the Hamiltonian operator. The Hamiltonian operator is the center piece of the Schrödinger equation, forming the basis for understanding the physical properties of quantum systems.

Consider a quantum system initially in the quantum state  $|\psi_0\rangle$ . According to the Schrödinger equation (6.6), it evolves in a Hilbert space with time into the state  $|\psi(t)\rangle$  at time  $t$  (see Fig. 6.2). We can describe this evolution in terms of an operator or matrix as

$$|\psi(t)\rangle = \hat{U}(t) |\psi_0\rangle. \quad (6.10)$$

Using the Schrödinger equation (6.6), we can prove rigorously that  $\hat{U}(t)$  is a unitary operator or unitary matrix,<sup>2</sup> i.e., it satisfies

$$\hat{U}^\dagger(t) \hat{U}(t) = I. \quad (6.11)$$

The unitarity of the evolution operator  $\hat{U}(t)$  has profound physical implications. Suppose initially there are two quantum states  $|\psi_0\rangle$  and  $|\phi_0\rangle$ , which evolve into  $|\psi(t)\rangle$

<sup>2</sup> Readers interested in the detailed proof can refer to Dirac's book *Principles of Quantum Mechanics* or other textbooks of quantum mechanics.



and  $|\phi(t)\rangle$  after some time  $t$ , respectively,

$$|\psi(t)\rangle = \hat{U}(t) |\psi_0\rangle, \quad |\phi(t)\rangle = \hat{U}(t) |\phi_0\rangle. \quad (6.12)$$

Using the unitarity of  $\hat{U}(t)$ , we have

$$\langle \psi(t) | \phi(t) \rangle = \langle \psi_0 | \hat{U}^\dagger(t) \hat{U}(t) | \phi_0 \rangle = \langle \psi_0 | \phi_0 \rangle, \quad (6.13)$$

which shows that the inner product of two quantum states does not change with time. If  $|\psi_0\rangle$  and  $|\phi_0\rangle$  are orthogonal, i.e.  $\langle \psi_0 | \phi_0 \rangle = 0$ , then  $\langle \psi(t) | \phi(t) \rangle = 0$ . This means that if two quantum states are orthogonal at the initial time, they will remain orthogonal at every instant of time. If  $|\psi_0\rangle = |\phi_0\rangle$ , then we have

$$\langle \psi(t) | \psi(t) \rangle = \langle \psi_0 | \hat{U}^\dagger(t) \hat{U}(t) | \psi_0 \rangle = \langle \psi_0 | \psi_0 \rangle. \quad (6.14)$$

As mentioned before, the inner product between a vector and itself gives the length of the vector. The above equation shows that the length of a vector describing a quantum state is conserved during the time evolution (see Fig. 6.2). Physically, this means that the total probability is conserved with time. For the spin state that we discussed earlier

$$|\psi(t)\rangle = c_1(t) |u\rangle + c_2(t) |d\rangle, \quad (6.15)$$

this statement implies that  $|c_1(t)|^2 + |c_2(t)|^2$  is invariant with time. If  $|c_1(0)|^2 + |c_2(0)|^2 = 1$  initially, then  $|c_1(t)|^2 + |c_2(t)|^2 = 1$  at every instant of time.

Unitary evolution is an important feature of quantum information technology. In both quantum computation and quantum communication, the operation protocol is as follows: (1) set up a certain number of quantum bits (also called qubit);<sup>3</sup> (2) initialize the system of qubits in a quantum state; (3) perform a series of operations on the qubits, generating evolution or propagation of the quantum state; (4) reach the target quantum state. In quantum computation and quantum communication, the operations for evolution must be unitary, otherwise there would be no quantum computation and quantum communication. But there is an important difference between the unitary evolution in quantum information technology and the unitary time evolution we described previously: the former is usually generated by manipulating the qubit with some external devices,<sup>4</sup> whereas the latter evolves according to the Schrödinger equation. The manipulations using external devices will inevitably generate “extra” perturbations to the qubits, entangling them with the environment, so that the qubits no longer have a well defined quantum state. This is known as decoherence. A challenge facing quantum information technology is to achieve efficient manipulations on the qubits while minimizing decoherence. It is instructive to make an analogy. In a summer night, we may want to open the windows to let in the cool breeze, on the other hand we want to close them to prevent mosquitoes. A solution to this dilemma

<sup>3</sup> If you don't understand what a qubit is, just regard it as a spin.

<sup>4</sup> Quantum adiabatic computers are exceptions.

is to use window screens. However, it is challenging to build a “window screen” in quantum technology, particularly in quantum computation, where the challenge is outstanding. We will have more detailed discussion on this in Chaps. 9 and 10, where quantum computation and quantum information are introduced, respectively.

The basic principles of quantum mechanics introduced in Chap. 5, along with the Schrödinger equation, form the complete theoretical framework for quantum mechanics. If we make an analogy between quantum mechanics and Go, then we have now finished introducing the basic rules of Go. The more you play and think about the game, the better you will play and the more you will enjoy. Similarly, the more you practice and think about quantum mechanics, the deeper you understand quantum mechanics and the nature.

Before concluding this section, let us introduce the Heisenberg equation

$$i\hbar \frac{d}{dt} \hat{O}(t) = [\hat{O}(t), \hat{H}], \quad (6.16)$$

where  $\hat{O}$  is an operator associated with a given observable. As the operator is mathematically represented by a matrix, the Heisenberg equation is also called matrix equation. For momentum operator  $\hat{p}$ , we have

$$i\hbar \frac{d}{dt} \hat{p}(t) = [\hat{p}(t), \hat{H}]. \quad (6.17)$$

This equation describes how the momentum of a particle evolves with time, which can be seen as the quantum counterpart of Newton’s second law. Mathematically, physicists can prove that there exist a very interesting and profound quantum-classical correspondence between them. Detailed discussion of the Heisenberg equation and its equivalence to the Schrödinger equation are beyond the scope of this book, and the interested reader is referred to Dirac’s *Principles of Quantum Mechanics*.

## 6.4 Quantum Energy Levels and Eigenstates

In quantum mechanics, an observable or a physical quantity is described by an operator, and the eigenvalues are the possible measured values of the observable. The energy of a system is represented by the Hamiltonian operator, which has its own eigenstates and eigenvalues

$$\hat{H} |\psi_n\rangle = E_n |\psi_n\rangle. \quad (6.18)$$

Here  $|\psi_n\rangle$  is the energy eigenstate, and  $E_n$  is the corresponding eigenenergy or energy level. The physical properties of a quantum system are entirely contained in the Hamiltonian, so finding the eigenstates and eigenvalues of a Hamiltonian is the key to understanding a quantum system. We illustrate the energy eigenstates and

energy levels of a quantum system with two simple examples. To avoid complicated mathematics, we directly write down the solutions without explaining how they are obtained.

In the first example, we consider a spin-1/2 particle in a magnetic field  $\mathbf{B}$ . Its Hamiltonian operator reads

$$\hat{H}_s = \mu_b \mathbf{B} \cdot \hat{\boldsymbol{\sigma}}, \quad (6.19)$$

where  $\mu_b$  is the magnetic dipole moment. Different particles possess different magnetic moments associated with their spins. For example, the proton's magnetic moment is less than one thousandth of the electron's magnetic moment although a proton and an electron are both spin-1/2. The Hamiltonian in Eq. (6.19) is equivalent to the spin operator  $\mathbf{n} \cdot \hat{\boldsymbol{\sigma}}$  introduced in Chap. 5. If we also use angles  $\theta$  and  $\beta$  to specify the direction of  $\mathbf{B}$ , i.e.,  $\mathbf{B} = B(\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$ , the two eigenstates of  $\hat{H}_s$  are exactly given by Eq. (5.19).

$$|E_+\rangle = \begin{pmatrix} \cos \frac{\theta}{2} \\ e^{i\varphi} \sin \frac{\theta}{2} \end{pmatrix}, |E_-\rangle = \begin{pmatrix} \sin \frac{\theta}{2} \\ -e^{i\varphi} \cos \frac{\theta}{2} \end{pmatrix}, \quad (6.20)$$

which correspond to the energy levels  $E_{\pm} = \pm \mu_b B$ , respectively. Imagine a spin-1/2 particle interacts with a beam of light (or electromagnetic wave) of frequency  $\nu$ . When the photon energy equals the difference between the energy levels, i.e.,  $h\nu = \Delta E = E_+ - E_- = 2\mu_b B$ , the spin flips its direction by absorbing a photon. This phenomenon, known as spin resonance absorption, is at the core of magnetic resonance imaging (MRI) technology.

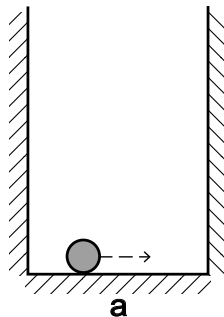
It is convenient to use the energy eigenstates to solve the time evolution of a quantum state. If the initial spin state is  $|\phi_0\rangle = c_1 |u\rangle + c_2 |d\rangle$ , we can find the spin state at time  $t$  as follows. First, we use the energy eigenstates  $|E_{\pm}\rangle$  to expand  $|\phi_0\rangle$  as

$$|\phi_0\rangle = \left( c_1 \cos \frac{\theta}{2} + c_2 e^{-i\varphi} \sin \frac{\theta}{2} \right) |E_+\rangle + \left( c_1 \sin \frac{\theta}{2} - c_2 e^{-i\varphi} \cos \frac{\theta}{2} \right) |E_-\rangle. \quad (6.21)$$

We then insert a prefactor  $e^{-iE_{\pm}t/\hbar}$  before the energy eigenstate  $|E_{\pm}\rangle$  and obtain the spin state at time  $t$

$$\begin{aligned} |\phi(t)\rangle &= \left( c_1 \cos \frac{\theta}{2} + c_2 e^{-i\varphi} \sin \frac{\theta}{2} \right) e^{-i\frac{E_+t}{\hbar}} |E_+\rangle \\ &\quad + \left( c_1 \sin \frac{\theta}{2} - c_2 e^{-i\varphi} \cos \frac{\theta}{2} \right) e^{-i\frac{E_-t}{\hbar}} |E_-\rangle. \end{aligned} \quad (6.22)$$

**Fig. 6.3** A ball in a box with two impenetrable walls



It is beyond the scope of this book to explain why this method is correct. Readers familiar with calculus can simply insert this solution into the Schrödinger equation (6.6) to verify its correctness. Since this is the simplest system in quantum mechanics, let us continue to play with it. We re-express  $|\phi(t)\rangle$  in terms of  $|u\rangle$  and  $|d\rangle$  as

$$|\phi(t)\rangle = (c_1 \cos \omega t - i c_1 \cos \theta \sin \omega t - i c_2 e^{-i\varphi} \sin \theta \sin \omega t) |u\rangle + (c_2 \cos \omega t + i c_2 \cos \theta \sin \omega t - i c_1 e^{i\varphi} \sin \theta \sin \omega t) |d\rangle, \quad (6.23)$$

where frequency  $\omega = \mu_b B / \hbar$ . When  $t = \pi/\omega$ , we have  $|\phi(t)\rangle = -|\phi_0\rangle$ , i.e., the spin returns to its initial state. So the spin state oscillates at frequency  $\omega$ , known as the spin precession frequency. Alternatively, we can express the dynamics in terms of the unitary evolution operator,  $|\phi(t)\rangle = \hat{U}_s(t) |\phi_0\rangle$ , where

$$\hat{U}_s(t) = \begin{pmatrix} \cos \omega t - i \cos \theta \sin \omega t & -i e^{-i\varphi} \sin \theta \sin \omega t \\ -i e^{i\varphi} \sin \theta \sin \omega t & \cos \omega t + i \cos \theta \sin \omega t \end{pmatrix}. \quad (6.24)$$

Interested readers can verify that it is a unitary matrix (Fig. 6.3).

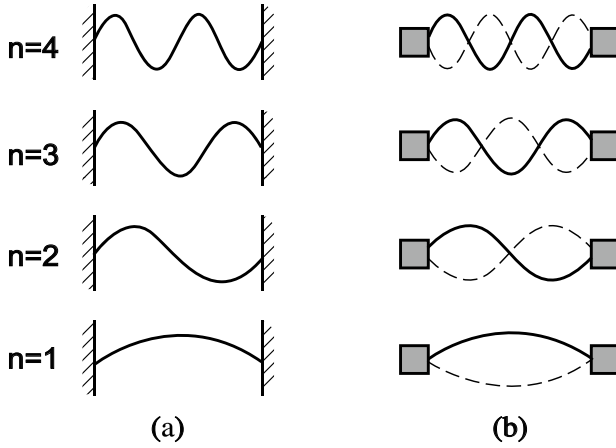
The second example is a ball with mass  $m$  moving in a one-dimensional box. The box has a length  $a$  and two impenetrable walls. The collision between the ball and the walls is perfectly elastic (i.e., the collision conserves the kinetic energy of the ball). Moreover, we shall ignore the friction for simplicity. As the net force on the ball is zero, which is equivalent to  $V(x) = 0$  in Eq. (6.7), the Hamiltonian operator for this ball is

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}. \quad (6.25)$$

The eigen-equation (6.18) for this problem is as follows

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi_n(x) = E_n \psi_n(x). \quad (6.26)$$

Finding its solutions requires knowledge of calculus and we shall directly give the answers. Let us set up a coordinate system with its origin at the left wall of the



**Fig. 6.4** **a** The first four eigenstates of a ball in an one-dimensional box. **b** Vibrational modes of strings

box. As the ball moves in space, its energy eigenstate  $\psi_n$  is a wave function and is therefore also called eigenfunction. The eigenfunction for this ball can be written as

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right), \quad (6.27)$$

with the corresponding eigenenergy given by

$$E_n = \frac{n^2\pi^2\hbar^2}{2ma^2}. \quad (6.28)$$

Here  $n$  can only take integer numbers, i.e.,  $n = 1, 2, 3, \dots$ . Interested readers can verify this solution by inserting the above results into Eq. (6.26) and using Eqs. (3.12, 3.13) in Chap. 3. Interested readers can also try to calculate the energy levels of this ball using the Bohr-Sommerfeld quantization rule from Chap. 3 and compare them with the results here.

Now let us examine closely the physical implications of these results. As  $n$  is a positive integer, the eigenenergies  $E_n$  are discrete, with  $E_1$  being the lowest energy. As  $E_1 > 0$ , this indicates that the ball is not stationary even in its lowest energy level. This is in radical contrast to the classical case: according to classical mechanics, the ball can completely come to rest, so the lowest energy of a classical ball is zero. This non-zero lowest energy is the zero-point energy that we have mentioned in Chap. 1; it is a purely quantum effect, a consequence of Heisenberg's uncertainty relation.

For eigenfunctions, we notice that  $\psi_n(0) = \psi_n(a) = 0$ , which reflects that the walls are impenetrable and the ball never escapes the box. We plot the first four eigenfunctions  $\psi_n(x)$  ( $n = 1, 2, 3, 4$ ) in Fig. 6.4a. Readers familiar with acoustics know that sound waves in various instruments form standing waves of different

modes. Mathematically, the standing wave of sound is the same as the eigenfunction of the Schrödinger equation. For comparison, we show the four vibrational modes of a string in Fig. 6.4b, which appear identical to the first four eigenfunctions shown in Fig. 6.4a. Because of this similarity, you can regard each quantum system as a musical instrument, and the entire universe as a remarkable symphony played by these “instruments”. This magnificent symphony began with a Big Bang.<sup>5</sup> It has been performed for nearly 14 billion years, and will continue to be performed forever. Physicists only understand a small part of it so far. I hope that some of the readers will join our efforts to understand and appreciate this symphony better.

In the papers published in 1926, Schrödinger not only wrote down his equation but also solved it for the hydrogen atom and found its eigenfunctions and eigenenergies. Schrödinger reproduced the energy levels obtained by Bohr in 1913 using the old quantum theory, thus explaining the spectrum of the hydrogen atom. However, the eigenfunctions obtained by Schrödinger are very different from Bohr’s quantized orbits. Finding the eigenfunctions of hydrogen is beyond the scope of this book, and we give the results directly. Figure 2.3 shows the first three energy eigenstates of an electron in the hydrogen atom. They all look very differently from Bohr’s quantum orbits as well as de Broglie’s standing waves as shown in Fig. 2.4. In particular, the  $2p$  wave function has a rather weird shape. The  $1s$  eigenfunction has the lowest energy and is often referred to as the ground state wave function. Chemists like to call these wave functions electronic orbitals, which are fundamental to understand the structure of molecules and chemical reactions.

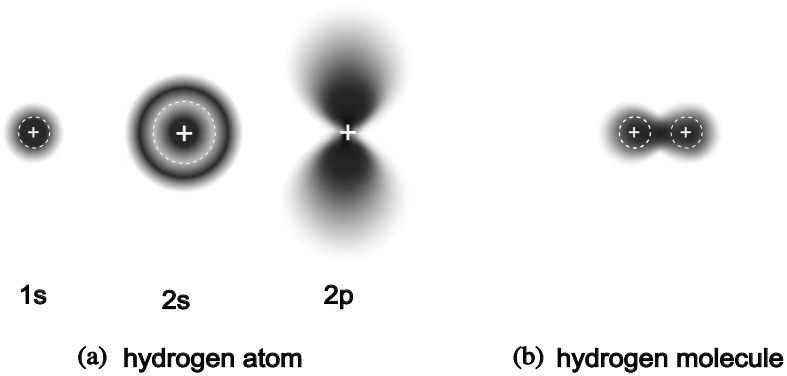
A wave function is a vector in an infinite-dimensional Hilbert space that is abstract and can not be perceived directly by either by human or machine, but its presence can be felt all the time in our daily lives. A piece of wood, a grain of sand, and a glass of water all have a certain volume. The volume of an object originates from the wave function. Let us see how we feel the presence of a wave function indirectly through volume.

The wave functions in the left panel of Fig. 6.4 are ordinary sine functions. Mathematically, they are quite trivial. But when you connect them with the object that they describe, you will see how remarkable their physical meaning is. These functions describe a single particle—a structureless ball of no size, which nonetheless spreads over the whole space in the box. If we regard the ball as a soccer ball, then these functions actually say that the soccer ball can be in the left and right sides of the field simultaneously. How can this be possible? For a soccer ball, of course, this is impossible. But for particles in the microscopic world, they can indeed be at different places simultaneously.<sup>6</sup> In comparison, functions in Fig. 6.4b are not odd: they spread in space because they describe the vibration of a string that is made of billions of atoms and has a length. The wave functions in Fig. 6.4a are not special at

---

<sup>5</sup> This is now the most popular theory about the origin of our universe. Its strongest experimental evidence is the cosmic microwave background radiation permeating the whole universe, which in effect a black-body radiation around 2.7 K.

<sup>6</sup> In Chap. 8, we will explain why a macroscopic object and a microscopic particle have this distinction.



**Fig. 6.5** **a** The three energy eigenfunctions of the hydrogen atom; **b** the ground state wave function of the hydrogen molecule. The “+” denotes a positively charged proton. The darker the color, the larger the amplitude of the wave function. It is interesting to compare them with Bohr’s orbitals in Fig. 2.3 and de Broglie’s electronic standing waves in Fig. 2.4

all, they reflect a general feature in quantum mechanics: a single particle can appear at different places at the same time. For example, the wave functions of electrons in Fig. 6.5 also spread in space, indicating that although the hydrogen atom contains only one electron, the electron can be anywhere around the proton simultaneously.

More importantly, these abstract wave functions, although spreading out in the entire space, is not like a soft cloud. Rather, it has rigidity. For example, consider the  $1s$  wave function of an electron in the hydrogen atom shown in Fig. 6.5. It represents the ground state of the hydrogen atom, i.e., the atom has the lowest energy in this state. Any attempt to modify the shape of the wave function will lead to an increase of the energy of the hydrogen atom. This means that you must apply some force on the hydrogen atom, and give it some energy to change the shape of its wave function. So we should regard the electronic ground state in Fig. 6.5 as an elastic ball, instead of a fluffy cloud.

By solving the Schrödinger equation, physicists find that the spatial region occupied by the ground state wave function is roughly a sphere with radius  $0.53 \times 10^{-10}$  m. This is regarded as the radius of a hydrogen atom. The radius of a proton is about  $0.877 \times 10^{-15}$  m,<sup>7</sup> which is about 5 orders of magnitude smaller than the radius of a hydrogen atom. Electron is an elementary particle. According to quantum mechanics, it is a point particle with no spatial size<sup>8</sup>. If we view the proton as a dust particle in the

<sup>7</sup> A proton is made of three quarks, so it has a size.

<sup>8</sup> There are estimations about the so-called classical electron radius: assume that the electron charge is uniformly distributed in a small sphere, so that the electron has an electrostatic energy. Assume further that all the electron mass comes from this electrostatic energy, we can obtain the electron radius using Einstein’s mass-energy equation. This gives a radius of about  $2.828 \times 10^{-15}$  m, which is larger than the radius of the proton. There is no experimental evidence to support this estimation. The more widely accepted viewpoint is that the radius of the electron is zero.

air,<sup>9</sup> then the hydrogen atom is about the size of a basketball. Quantum mechanics offers a brilliant and economic way to make a bouncing basketball: place just one dust particle at the center and fill it with the wave function of an electron.

The above property of a wave function is not limited to the hydrogen atom, and is universal: the spatial extension of the electron wave function defines the radii of all atoms and molecules, thus the volumes of all macroscopic physical objects. The hydrogen molecule in Fig. 6.5b consists of two hydrogen atoms. By solving the Schrödinger equation, physicists find that the two electrons are in the ground state when the two hydrogen nuclei are separated by  $0.74 \times 10^{-10}$  m. A larger or smaller separation will change the wave function of the electrons, increasing the molecular energy. This way, nature creates from two very small protons and two point-like electrons a 100,000 times larger hydrogen molecule with “rigidity”, through the wave function. By the same token, atoms and molecules can form even larger object, such as a piece of wood, a grain of sand, and a glass of water. If we want to change the volumes of these objects, we have to push very hard. This is when we feel the wave function of electrons. In this sense, the wave function is something tangible in our lives although it is a vector in the abstract infinite-dimensional Hilbert space.

## 6.5 Superposition Principle of Quantum States and No-cloning Theorem

Consider two initial states,  $|\phi_1(0)\rangle$  and  $|\phi_2(0)\rangle$ , of a quantum system. If the system starts out in  $|\phi_1(0)\rangle$ , it evolves into the state  $|\phi_1(t)\rangle$  after time  $t$ ; if it starts out in  $|\phi_2(0)\rangle$ , it evolves into  $|\phi_2(t)\rangle$ . By virtue of Eq. (6.10), we have

$$|\phi_1(t)\rangle = \hat{U}(t) |\phi_1(0)\rangle, \quad |\phi_2(t)\rangle = \hat{U}(t) |\phi_2(0)\rangle, \quad (6.29)$$

where  $\hat{U}(t)$  is the unitary evolution operator of the considered quantum system. Consider another initial state, which is a superposition of the first two initial states, i.e.,  $c_1 |\phi_1(0)\rangle + c_2 |\phi_2(0)\rangle$ . According to the following derivations, the system will evolve into the state  $c_1 |\phi_1(t)\rangle + c_2 |\phi_2(t)\rangle$ .

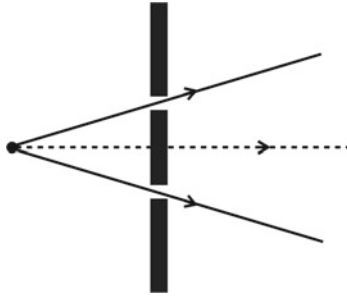
$$\begin{aligned} \hat{U}(t)[c_1 |\phi_1(0)\rangle + c_2 |\phi_2(0)\rangle] &= c_1 \hat{U}(t) |\phi_1(0)\rangle + c_2 \hat{U}(t) |\phi_2(0)\rangle \\ &= c_1 |\phi_1(t)\rangle + c_2 |\phi_2(t)\rangle. \end{aligned} \quad (6.30)$$

This is the superposition principle of quantum states: for a quantum system, the linear superposition of its two quantum evolutions is still a legitimate quantum evolution. This is another fundamental and important feature of quantum mechanics that is distinct from classical mechanics.

---

<sup>9</sup> Technically it is called particulate matter with diameters normally between 2.5 and 10  $\mu\text{m}$ .





**Fig. 6.6** Linear superposition of classical trajectories. There is no force on the particle (solid circle) except when it hits on the wall with two slits. The solid curves represent two possible trajectories of the particle. The dashed curve denotes the equal-weight superposition of these two trajectories. Clearly, the dashed line is not a physically possible trajectory

In classical mechanics, if there are two trajectories,  $\{x_1(t), p_1(t)\}$  and  $\{x_2(t), p_2(t)\}$ , in general, their linear superposition  $\{a_1x_1(t) + a_2x_2(t), a_1p_1(t) + a_2p_2(t)\}$  (where  $a_1$  and  $a_2$  are real numbers) is not a trajectory that obeys Newton's second law. Let us look at Fig. 6.6, where there is an impenetrable wall with two slits. The moving particle is not subject to any external forces (including gravity). Consider two different initial conditions,  $(\mathbf{x}_1, \mathbf{p}_1)$  and  $(\mathbf{x}_2, \mathbf{p}_2)$ , starting from which the particle can pass through either of the two slits. The two solid lines in Fig. 6.6 represent the two possible trajectories. Assume that the speed is the same on the two trajectories, i.e.,  $|\mathbf{p}_1| = |\mathbf{p}_2|$ . We construct an equal-weight superposition of these two trajectories. After the superposition, the initial position is still  $\mathbf{x}_0$ , but the velocity only has the horizontal component. The resulting trajectory is indicated by the dashed line in Fig. 6.6, which shows that the particle will pass through the impenetrable wall. This evidently violates the Newton's second law, which predicts the particle will be reflected. This example shows that the superposition principle in general does not apply in classical mechanics.

The superposition principle of quantum states has profound implications. We will first discuss one of them, the no-cloning theorem, and then discuss the famous interference phenomenon.

**The no-cloning theorem**—Clone is an identical copy of the original. It is common in our daily life: create a copy of one document and you have two identical documents; back up your data to an external hard drive and you have two identical copies of the data. Interestingly, such a common operation is fundamentally forbidden in the quantum world by the superposition principle. Let us prove it by contradiction. Suppose that we have two systems, one in a quantum state  $|\psi\rangle$  and the other in an empty state  $|\emptyset\rangle$ . The total system composed of these two systems is thus in the quantum state  $|\psi\rangle \otimes |\emptyset\rangle$ . Here  $\otimes$  is the direct product introduced at the end of Chap. 4. If quantum cloning is possible, then we can achieve the following transition through a quantum operation

$$|\psi\rangle \otimes |\emptyset\rangle \longrightarrow |\psi\rangle \otimes |\psi\rangle. \quad (6.31)$$

No matter how complex or simple the operation is, it should be a unitary transformation  $\hat{U}$ . Otherwise it is not quantum cloning. So we have

$$\hat{U}(|\psi\rangle \otimes |\emptyset\rangle) = |\psi\rangle \otimes |\psi\rangle. \quad (6.32)$$

Similarly, for a different quantum state  $|\phi\rangle$  we have

$$\hat{U}(|\phi\rangle \otimes |\emptyset\rangle) = |\phi\rangle \otimes |\phi\rangle. \quad (6.33)$$

Without loss of generality, we assume  $\langle\phi|\psi\rangle = 0$ . It is important to note that the cloning operation on two different quantum states  $|\psi\rangle$  and  $|\phi\rangle$  must correspond to the same unitary operator  $\hat{U}$ . This is similar to the familiar photocopying or copying: the same copy machine can make copies of different documents and the same copying code in a computer can copy different sets of data.

We now want to clone a new quantum state  $|\varphi\rangle = (|\phi\rangle + |\psi\rangle)/\sqrt{2}$ . There are two possible ways to do this

1. Using the superposition principle, we add Eqs. (6.32) and (6.33) and then divide it by  $\sqrt{2}$ ,

$$\hat{U}\left[\frac{|\psi\rangle + |\phi\rangle}{\sqrt{2}} \otimes |\emptyset\rangle\right] = \frac{1}{\sqrt{2}}(|\psi\rangle \otimes |\psi\rangle + |\phi\rangle \otimes |\phi\rangle). \quad (6.34)$$

2. According to the definition of  $\hat{U}$ , we should

$$\hat{U}\left[\frac{|\psi\rangle + |\phi\rangle}{\sqrt{2}} \otimes |\emptyset\rangle\right] = \frac{1}{2}(|\psi\rangle + |\phi\rangle) \otimes (|\psi\rangle + |\phi\rangle). \quad (6.35)$$

Evidently, the two approaches, which are both legitimate if quantum cloning is possible, lead to two different results that contradict each other. Therefore, quantum cloning is not allowed. This is the no-cloning theorem. One of its important implications is that a quantum computer can not in general store its current state. While on a classical computer, we often temporarily store its current state to be called for later use or analysis, and this is not allowed on a quantum computer.

Our world is made up of microscopic particles, such as atoms and molecules, that evolve according to quantum mechanics. As cloning is forbidden in quantum mechanics, why we can create replicas or clones in our daily lives? There is a fundamental difference between an ordinary copying and quantum cloning. Usual copying is not a unitary operation. As was said before, a unitary operation is represented by a unitary matrix  $\hat{U}$ , which is reversible. This means that any unitary operation can be reversed and the reversed operation is represented by matrix  $\hat{U}^\dagger$ . If the ordinary copy operation were unitary, it would mean that we could put a piece of copied paper, already with words on it, back into the copy machine and reverse the operation, a white clean paper would come out the machine, and the ink on the paper would return to the cartridge. That is everything would return to its beginning. Obviously, this is

impossible in an ordinary copy machine. As a non-unitary operation, the ordinary copying does not need to satisfy the no-cloning theorem.

## 6.6 Double-Slit Interference

Another important consequence of the superposition principle of quantum states is the well-known double-slit interference. A typical double-slit interference experiment is shown in Fig. 6.7. The setup is similar to Fig. 6.6, except that the classical particles are replaced by electrons. If the distance between the slits and the size of the slits are properly chosen, the electron beams passing through the slits will produce a pattern of bright and dark fringes on the screen. There is a coil of wire on the right side of the double-slit plate. When an electric current runs through the wire, a magnetic field is produced in the coil, which affects the phase difference between the upper and lower beams of electrons, shifting the interference pattern.

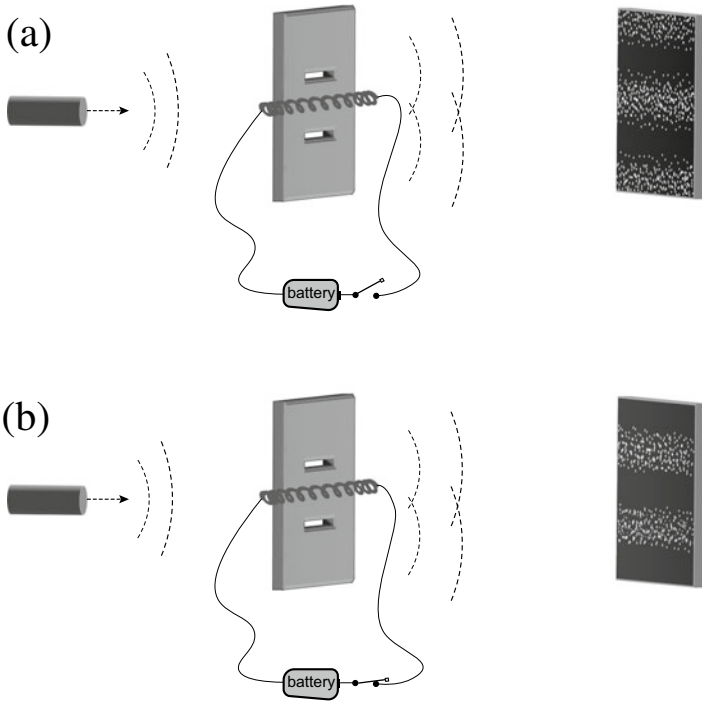
A detailed analysis of the interference pattern in Fig. 6.7, such as the width and intensity of fringes, as well as the positions of the bright and dark fringes, requires rather complicated mathematics that is beyond this book. Here we shall focus on analyzing the intensity in the middle of the screen, which is mathematically simple but nonetheless captures the essential physics. To this end, we further simplify the experiment in Fig. 6.7 by assuming that the electrons incident on the double-slit plate, except at the slits, are absorbed. Moreover, we replace the detection screen by nine detectors  $d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9$  (see Fig. 6.8). Finally, we assume that all electrons in the electron beam are in the quantum state  $|\psi_0\rangle$ .

Let us first consider the case where no electric current is in the wire. After being generated from the source, the electron beam reaches the double-slit plate. There, only the wave functions at the two slits can continue evolving to the right, while all others are absorbed by the plate. We will represent this evolution as

$$|\psi_0\rangle \longrightarrow \frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle), \quad (6.36)$$

where  $|\psi_1\rangle$  is the wave function (or quantum state) of electron at the slit  $s_1$  and  $|\psi_2\rangle$  is the wave function of electron at the slit  $s_2$ . As the electron can be absorbed by the plate, the above evolution is not unitary. Since the two slits are symmetric, the electronic state is a symmetric equal-weight superposition of the two quantum states. After some additional time evolution, the electron reaches the detectors. There, the initial quantum state  $|\psi_1\rangle$  at the slit  $s_1$  evolves into a superposition of quantum states at each detector, i.e.,

$$|\psi_1\rangle \longrightarrow \sum_{j=1}^9 a_j |d_j\rangle, \quad (6.37)$$



**Fig. 6.7** The double-slit experiment. **a** Electrons are fired from the left side and travel to the screen on the right side, forming an interference pattern. **b** The interference fringes are shifted by turning on the electric current in the coil of wire

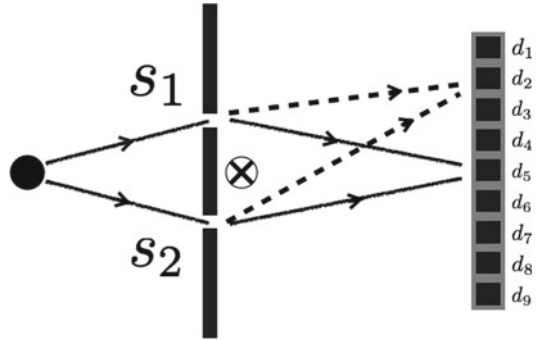
where  $|d_j\rangle$  denotes the quantum state of the electron at the detector  $d_j$ , similar to the quantum state  $|x_j\rangle$  introduced earlier. The above expression indicates if there are a total of  $N/2$  electrons passing through the slit  $s_1$ ,  $N|a_j|^2/2$  of them will reach the detector  $d_j$ . Accordingly, the quantum state  $|\psi_2\rangle$  evolves as

$$|\psi_2\rangle \longrightarrow \sum_{j=1}^9 b_j |d_j\rangle. \tag{6.38}$$

Similarly, if there are a total of  $N/2$  electrons passing through the slit  $s_2$ , the detector  $d_j$  will detect  $N|b_j|^2/2$  electrons. Both evolutions in Eqs. (6.37, 6.38) are unitary evolutions. According to the superposition principle, the overall evolution is the superposition of these two evolutions and is given by

$$\frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle) \longrightarrow \frac{1}{\sqrt{2}} \sum_{j=1}^9 (a_j + b_j) |d_j\rangle. \tag{6.39}$$

**Fig. 6.8** The double-slit experiment. Filled circle: the source of electrons; square blocks: detector; circle with cross: coil of wire



This means that the detector  $d_j$  will detect a total of  $N(|a_j + b_j|^2)/2$  electrons. We consider the middle detector  $d_5$ . Due to symmetry, there should be  $a_5 = b_5$ , so detector  $d_5$  will detect  $2N|a_5|^2$  electrons. If electrons were classical, for  $N|a_5|^2/2$  electrons from slit  $s_1$  and  $N|b_5|^2/2$  electrons from the slit  $s_2$ , the total number of electrons coming to the detector  $d_5$  would be  $N|a_5|^2/2 + N|b_5|^2/2 = N|a_5|^2$ . So we see that the quantum and classical results are very different. This effect is called quantum interference. Let us take a closer look and see where the difference originates. Expanding  $|a_j + b_j|^2$ , we have

$$|a_j + b_j|^2 = (a_j^* + b_j^*)(a_j + b_j) = |a_j|^2 + |b_j|^2 + a_j^*b_j + b_j^*a_j. \quad (6.40)$$

If only the first two terms on the right hand side were considered, we would have the classical results. The last two terms  $a_j^*b_j + b_j^*a_j$  are called the interference term, which is the origin of the quantum interference. The results on other detectors are slightly more complicated to analyze due to the lack of symmetry, and we will not discuss them. The overall result of interference is shown in Fig. 6.7, where the electrons form a pattern with alternating light and dark fringes on the screen.

We now turn on the electric current in the wire, producing a magnetic field perpendicular to the paper but parallel to the double slits. This magnetic field can affect the phases of the upper and lower electron wave functions, but not the magnitude. We choose an appropriate current intensity such that the upper and lower wave functions differ by a negative sign,<sup>10</sup> i.e.,

$$|\psi_0\rangle \longrightarrow \frac{1}{\sqrt{2}}(|\psi_1\rangle - |\psi_2\rangle). \quad (6.41)$$

<sup>10</sup> Calculation and explanation of this phase difference is beyond the scope of this book, and here we give the results directly.

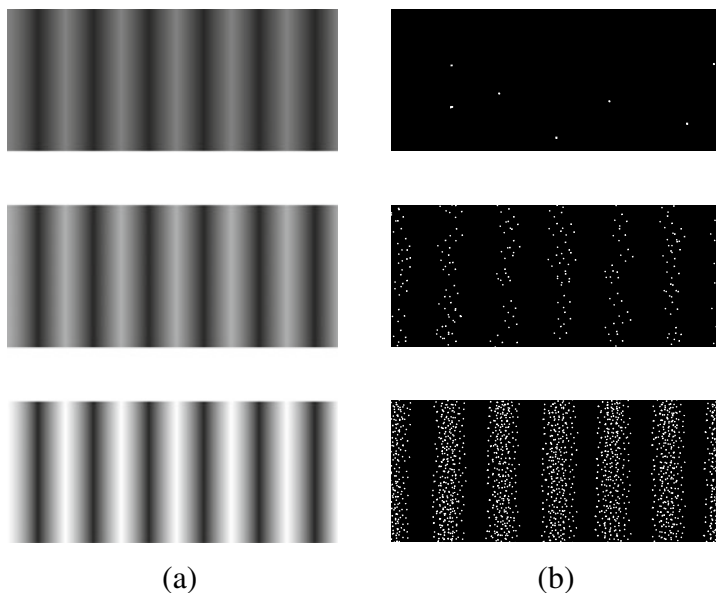
In this case, the superposition of the wave functions after the two slits is given by

$$\frac{1}{\sqrt{2}}(|\psi_1\rangle - |\psi_2\rangle) \longrightarrow \frac{1}{\sqrt{2}} \sum_{j=1}^9 (a_j - b_j) |d_j\rangle \quad (6.42)$$

So  $N(|a_j - b_j|^2)/2$  electrons reach the detector  $d_j$ . As  $a_5 = b_5$ , the central detector  $d_5$  will not detect any electrons. This is the destructive quantum interference, which is utterly inexplicable using classical mechanics. Classically, as there is no magnetic field outside the solenoid, the electrons do not feel any difference with the electric current on or off, and the pattern on the screen would be the same.

Interference is a common phenomenon of waves, which has long been observed with classical waves, such as sound wave and water wave. It is usually considered as the simplest and most direct evidence of wave behavior. Therefore, the interference pattern in Fig. 6.7 provides a direct evidence about the wave nature of electron. But it is important to note that there is a crucial difference between classical waves and quantum waves. In classical physics, a wave is the collective motion of a large number of particles: a sound wave in air is the collective vibration of a large number of atoms and molecules in the air; a water wave is the collective motion of a large number of water molecules. Or, the wave can occur as the oscillation and propagation of fields, such as electromagnetic waves. In quantum mechanics, even a single particle, described by the wave function, can display wave behavior. All the wave functions that we used in discussing the electron interference experiment are single-electron wave functions, including the initial wave function  $|\psi_0\rangle$ , the wave function at the double slits  $|\psi_1\rangle$  and  $|\psi_2\rangle$ , and the wave function at the detector. As such, quantum interference is the interference of a particle with itself, where the fringe contrast reflects the probability of particles. By contrast, classical interference is the interference of different classical waves, where the fringe contrast reflects the oscillation strength. This fundamental difference between quantum and classical interferences can be tested experimentally. Let us see how this is achieved.

The double-slit interference experiment was first performed, using light, by Young (Thomas Young, 1773–1829) in 1801. This experiment was considered at that time as a demonstration that light is wave. Indeed, when the number of photons is large, i.e., the light intensity is strong, the double-slit interference experiment cannot distinguish whether the light is a classical wave or a quantum wave. To distinguish them, one can gradually reduce the light intensity. If the light is a classical wave, the interference fringes will become fainter and fainter with decreasing light intensity, but they are always present, as shown in Fig. 6.9a. If, however, light consists of particles and is a quantum wave described by a wave function, we can reduce the light intensity so that only one photon passes through a double-slit apparatus each time. At first, the photon hits on the screen rather randomly; only after many photons are accumulated can a pattern be discerned as shown in Fig. 6.9b. The interference pattern eventually becomes clear when the number of photon is very large. This experiment, which has been actually performed, demonstrates the particle nature of light.



**Fig. 6.9** Comparison between classical and quantum double-slit interferences. **a** Classical interference: the contrast between bright and dark fringes varies with the wave intensity, and the interference fringes are always present. **b** Quantum interference: when the number of particles is small, interference is not observed. An interference pattern emerges only when more and more particles hit the screen. A clear interference pattern is eventually observed when the number of particles is very large. The white dots in the top right panel are deliberately made larger for clarity

The most mysterious and debated part of the quantum double-slit interference experiment is the following: which slit the electron passes through before it hits the screen? At the two slits, the electron is in a superposition state  $|\psi_1\rangle \pm |\psi_2\rangle$ , indicating that the electron passes through both slits  $s_1$  and  $s_2$  simultaneously. We have already seen this odd behavior of electrons in previous sections. For example, the electron wave function in a hydrogen atom is spread out in space, i.e., a single electron is at many places at the same time. Because of the wave property of a single electron, a hydrogen atom has a radius, a hydrogen molecule has a size, and normal objects have volumes. In our daily life, an object always has a definite position: a flying tennis ball has a precise position at any time; no one can be at home and the office simultaneously; we cannot observe sunrise in the east and sunset in the west simultaneously. However, the double-slit interference experiment tells us that if we were electrons, we would be able to rest at home and work at the office simultaneously, and the “sun” could set and rise in all directions simultaneously. Is there a fundamental difference between these macroscopic objects and electrons in the microscopic world? In Chap. 8, we will revisit the double-slit interference experiment and related issues in the context of quantum measurement.

# Chapter 7

## Quantum Entanglement and Bell's Inequality



Quantum entanglement is a nonlocal correlation between two or more particles in a quantum system. Nonlocality means that the correlation does not depend on the distance between the particles. Nonlocal correlations can also occur in a classical system. But quantum entanglement is different from the classical nonlocal correlation: quantum entanglement violates Bell's inequality whereas classical nonlocal correlation does not. Moreover, quantum entanglement has an equally important but less mentioned feature: the loss of individuality. In classical mechanics, if we want to know the state of a many-particle system, we must know the position and momentum of each particle in the system: we can only describe the state of a many-particle system if we have a complete description of every particle in it. Quantum systems are strikingly different. You can write down a many-body wave function that completely describes a quantum many-body system. But if there is entanglement between particles, the state of at least one of the particles becomes uncertain, losing its own individuality in the entangled quantum state. In short, you must know individuals to know the whole in classical physics; you can know the whole without the knowledge of individuals in quantum mechanics. In this chapter, I will illustrate these two basic features of quantum entanglement with a system of two spins.

### 7.1 System of Two Spins

Entanglement involves at least two particles. Here we consider the simplest example of a quantum many-body system, a system of two spins. We use the operator  $\hat{\sigma} = \{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$  to denote spin 1, and the operator  $\hat{\tau} = \{\hat{\tau}_x, \hat{\tau}_y, \hat{\tau}_z\}$  to denote spin 2. The three components of  $\hat{\tau}$  are also Pauli matrices,



$$\hat{\tau}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\tau}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\tau}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (7.1)$$

We use  $\hat{\tau}$  just to distinguish it from spin 1. The two spins constitute a composite system. According to Chap. 4, if spin 1 is in a quantum state  $|\psi\rangle = a_1 |u\rangle + b_1 |d\rangle$ , and spin 2 is in a quantum state  $|\phi\rangle = a_2 |u\rangle + b_2 |d\rangle$ , then the quantum state of this double-spin system can be written in terms of the direct product  $\otimes$  as

$$\begin{aligned} |\Psi_{12}\rangle &= |\psi\rangle \otimes |\phi\rangle = \underbrace{(a_1 |u\rangle + b_1 |d\rangle)}_1 \otimes \underbrace{(a_2 |u\rangle + b_2 |d\rangle)}_2 \\ &= a_1 a_2 \underbrace{|u\rangle}_1 \otimes \underbrace{|u\rangle}_2 + a_1 b_2 \underbrace{|u\rangle}_1 \otimes \underbrace{|d\rangle}_2 + b_1 a_2 \underbrace{|d\rangle}_1 \otimes \underbrace{|u\rangle}_2 + b_1 b_2 \underbrace{|d\rangle}_1 \otimes \underbrace{|d\rangle}_2. \end{aligned} \quad (7.2)$$

Direct product  $\otimes$  is like multiplication: multiplying two terms with another two terms yields four terms. But there is a key difference: the order matters in direct product. The state before  $\otimes$  is for spin 1 and the state after  $\otimes$  is for spin 2. For example,  $\underbrace{|u\rangle}_1 \otimes \underbrace{|d\rangle}_2$  represents that spin 1 is up and spin 2 is down;  $\underbrace{|d\rangle}_1 \otimes \underbrace{|u\rangle}_2$  represents that spin 1 is down and spin 2 is up. Therefore,  $\underbrace{|u\rangle}_1 \otimes \underbrace{|d\rangle}_2$  and  $\underbrace{|d\rangle}_1 \otimes \underbrace{|u\rangle}_2$  are different double-spin states, and cannot be combined in the above equation.

In the above equation we have deliberately used underlined labels to tell which state is for spin 1 and which state for spin 2. For simplicity, henceforth we will remove the labels, and take it as a rule that the left one is for spin 1 and the right one for spin 2 in the product state. In practical calculations, omitting the direct product symbol  $\otimes$  will not cause confusion in most cases, much the same way we omit the multiplication symbol  $\times$ . With these considerations in mind, we simplify the notation as follows.

$$\underbrace{|u\rangle}_1 \otimes \underbrace{|u\rangle}_2 \equiv |uu\rangle, \quad \underbrace{|u\rangle}_1 \otimes \underbrace{|d\rangle}_2 \equiv |ud\rangle, \quad \underbrace{|d\rangle}_1 \otimes \underbrace{|u\rangle}_2 \equiv |du\rangle, \quad \underbrace{|d\rangle}_1 \otimes \underbrace{|d\rangle}_2 \equiv |dd\rangle. \quad (7.3)$$

The double-spin state  $|\Psi_{12}\rangle$  then becomes

$$|\Psi_{12}\rangle = a_1 a_2 |uu\rangle + a_1 b_2 |ud\rangle + b_1 a_2 |du\rangle + b_1 b_2 |dd\rangle. \quad (7.4)$$

Similarly, the corresponding Hermitian conjugates are simplified as

$$\underbrace{\langle u\rangle}_1 \otimes \underbrace{\langle u\rangle}_2 \equiv \langle uu|, \quad \underbrace{\langle u\rangle}_1 \otimes \underbrace{\langle d\rangle}_2 \equiv \langle ud|, \quad \underbrace{\langle d\rangle}_1 \otimes \underbrace{\langle u\rangle}_2 \equiv \langle du|, \quad \underbrace{\langle d\rangle}_1 \otimes \underbrace{\langle d\rangle}_2 \equiv \langle dd|. \quad (7.5)$$

As well, in this expression, the one on the left describes spin 1 and the one on the right describes spin 2. In many books and papers on quantum mechanics, a comma is added between the two spins, such as  $|u, u\rangle \equiv |uu\rangle$  and  $|\psi, \phi\rangle \equiv |\psi\phi\rangle$ . The choice of convention is a matter of taste. Here we do not use the comma.

For two double-spin states,  $|\psi_1\phi_1\rangle$  and  $|\psi_2\phi_2\rangle$ , we calculate their inner product as follows

$$\langle\psi_1\phi_1|\psi_2\phi_2\rangle = \langle\psi_1|\psi_2\rangle\langle\phi_1|\phi_2\rangle. \quad (7.6)$$

With this rule it is straightforward to find  $\langle uu|uu\rangle = \langle u|u\rangle\langle u|u\rangle = 1$ ,  $\langle dd|ud\rangle = \langle d|u\rangle\langle d|d\rangle = 0$ , etc. These relations indicate that  $|uu\rangle$ ,  $|ud\rangle$ ,  $|du\rangle$ , and  $|dd\rangle$  form a set of orthonormal basis. Any double-spin quantum state  $|\Phi\rangle$  can be expanded in this basis as

$$|\Phi\rangle = c_1|uu\rangle + c_2|ud\rangle + c_3|du\rangle + c_4|dd\rangle, \quad (7.7)$$

where the expansion coefficients satisfy the normalization condition  $|c_1|^2 + |c_2|^2 + |c_3|^2 + |c_4|^2 = 1$ . For a pair of double-spin states

$$|\Phi_1\rangle = a_1|uu\rangle + a_2|ud\rangle + a_3|du\rangle + a_4|dd\rangle \quad (7.8)$$

and

$$|\Phi_2\rangle = b_1|uu\rangle + b_2|ud\rangle + b_3|du\rangle + b_4|dd\rangle, \quad (7.9)$$

we can calculate their inner product  $\langle\Phi_1|\Phi_2\rangle$  as follows

$$\begin{aligned} \langle\Phi_1|\Phi_2\rangle &= \left( a_1^*\langle uu| + a_2^*\langle ud| + a_3^*\langle du| + a_4^*\langle dd| \right) \\ &\quad \left( b_1|uu\rangle + b_2|ud\rangle + b_3|du\rangle + b_4|dd\rangle \right) \\ &= a_1^*b_1 + a_2^*b_2 + a_3^*b_3 + a_4^*b_4. \end{aligned} \quad (7.10)$$

In the above computation, the second line have sixteen terms after expansion and only four of them are not zero. The other inner product  $\langle\Phi_2|\Phi_1\rangle$  can be calculated in a similar way. In addition, one can prove that  $\langle\Phi_1|\Phi_2\rangle = \langle\Phi_2|\Phi_1\rangle^*$ .

The aforementioned  $|\Psi_{12}\rangle$  is a direct product of two single-spin states, and is called a product state. Not all double-spin states are product states. For example,

$$|S_3\rangle = \frac{1}{\sqrt{2}}\left(|ud\rangle + |du\rangle\right) \quad (7.11)$$

is not a product state. The following is a proof by contradiction. Assume that  $|S_3\rangle$  is a product state. We can then choose the coefficients  $a_1, b_1, a_2, b_2$  in  $|\Psi_{12}\rangle$  in such a way that  $|S_3\rangle = |\Psi_{12}\rangle$ . Comparing the coefficients in Eqs. (7.4) and (7.11), we obtain

$$a_1a_2 = b_1b_2 = 0, \quad a_1b_2 = a_2b_1 = 1/\sqrt{2}. \quad (7.12)$$

From the first identity we have  $a_1a_2b_1b_2 = 0$ , while the second identity leads to  $a_1a_2b_1b_2 = 1/2$ . These two results contradict each other, indicating that the previous assumption, i.e.,  $|S_3\rangle$  is a product state, is not valid. Therefore,  $|S_3\rangle$  is not a product

state. A double-spin state like  $|S_3\rangle$  that is not a product state is defined as an entangled state. In the next section, we will discuss entangled state and its implications in detail. Before that, let us introduce the operators for a double-spin system and their actions on the double-spin state.

There are two kinds of operators in a double-spin system: single spin operator, such as  $\hat{\sigma}_x$  and  $\hat{\tau}_y$ ; double-spin operator, such as  $\hat{\sigma}_z \otimes \hat{\tau}_x$  and  $\hat{\sigma}_y \otimes \hat{\tau}_z$ . When the operator of spin 1 acts on the double-spin state  $|\Phi\rangle$ , it only acts on the state of spin 1, for instance,

$$\begin{aligned}\hat{\sigma}_z |\Phi\rangle &= c_1(\hat{\sigma}_z |u\rangle) \otimes |u\rangle + c_2(\hat{\sigma}_z |u\rangle) \otimes |d\rangle + c_3(\hat{\sigma}_z |d\rangle) \otimes |u\rangle + c_4(\hat{\sigma}_z |d\rangle) \otimes |d\rangle \\ &= c_1 |uu\rangle + c_2 |ud\rangle - c_3 |du\rangle - c_4 |dd\rangle.\end{aligned}\quad (7.13)$$

Similarly, an operator of spin 2 only acts on the state of spin 2,

$$\begin{aligned}\hat{\tau}_x |\Phi\rangle &= c_1 |u\rangle \otimes (\hat{\tau}_x |u\rangle) + c_2 |u\rangle \otimes (\hat{\tau}_x |d\rangle) + c_3 |d\rangle \otimes (\hat{\tau}_x |u\rangle) + c_4 |d\rangle \otimes (\hat{\tau}_x |d\rangle) \\ &= c_1 |ud\rangle + c_2 |uu\rangle + c_3 |dd\rangle + c_4 |du\rangle.\end{aligned}\quad (7.14)$$

When a double-spin operator, such as  $\hat{\sigma}_z \otimes \hat{\tau}_x$ , acts on a double-spin state, the operator of spin 1 acts only on the state of spin 1, and the operator of spin 2 acts only on the state of spin 2. Below is an example

$$\begin{aligned}\hat{\sigma}_z \otimes \hat{\tau}_x |\Phi\rangle &= c_1(\hat{\sigma}_z |u\rangle) \otimes (\hat{\tau}_x |u\rangle) + c_2(\hat{\sigma}_z |u\rangle) \otimes (\hat{\tau}_x |d\rangle) + \\ &\quad c_3(\hat{\sigma}_z |d\rangle) \otimes (\hat{\tau}_x |u\rangle) + c_4(\hat{\sigma}_z |d\rangle) \otimes (\hat{\tau}_x |d\rangle) \\ &= c_1 |ud\rangle + c_2 |uu\rangle - c_3 |dd\rangle - c_4 |du\rangle.\end{aligned}\quad (7.15)$$

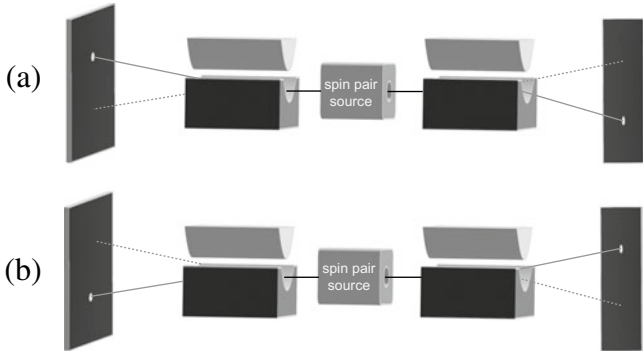
We now use  $\langle\Phi|\hat{\sigma}_z \otimes \hat{\tau}_x|\Phi\rangle$  as an example to illustrate how to calculate the expectation value for a system of two spins. As seen in Eq. (7.15), the action of the operator  $\hat{\sigma}_z \otimes \hat{\tau}_x$  on vector  $|\Phi\rangle$  yields a new vector, whose inner product with vector  $|\Phi\rangle$  gives the expectation value of this operator. For a product state, the calculation of the expectation value can be further simplified. Consider the following two examples. For a double-spin operator, we have

$$\begin{aligned}\langle\Psi_{12}|\hat{\sigma}_z \otimes \hat{\tau}_x|\Psi_{12}\rangle &= \left(\langle\psi| \otimes \langle\phi|\right) \hat{\sigma}_z \otimes \hat{\tau}_x \left(|\psi\rangle \otimes |\phi\rangle\right) \\ &= \langle\psi|\hat{\sigma}_z|\psi\rangle \langle\phi|\hat{\tau}_x|\phi\rangle = (a_1^* a_1 - b_1 b_1^*)(a_2^* b_2 + a_2 b_2^*).\end{aligned}\quad (7.16)$$

For a single-spin operator, we have

$$\langle\Psi_{12}|\hat{\tau}_x|\Psi_{12}\rangle = \langle\psi| \otimes \langle\phi| \hat{\tau}_x |\psi\rangle \otimes |\phi\rangle = \langle\psi|\psi\rangle \langle\phi|\hat{\tau}_x|\phi\rangle = a_2^* b_2 + a_2 b_2^*.\quad (7.17)$$

In fact, for any quantum system, the expectation value of an operator can be obtained by combining the rule of an operator acting on a quantum state with the rule of calculating inner product.



**Fig. 7.1** Double-spin Stern-Gerlach experiment. Different from Fig. 5.1, the evaporation furnace is replaced by a more elaborate device capable of producing pairs of entangled spins. Each pair is in the singlet state, where two spins have opposite momentum, with spin 1 flying to the left and spin 2 to the right. There are only two possible outcomes: **a** an upper spot on the left and a lower spot on the right; **b** a lower spot on the left and an upper spot on the right

## 7.2 Quantum Entanglement

Our discussion of entanglement will be based on the following double-spin state

$$|S\rangle = \frac{1}{\sqrt{2}}(|ud\rangle - |du\rangle), \quad (7.18)$$

which is known as the spin singlet state. Apparently it is an entangled state. Let us use the Stern-Gerlach experiment to reveal the physical properties of this entangled state. We first need to upgrade the experimental apparatus. We replace the high-temperature furnace with a more sophisticated atom source, which is capable of producing a pair of spins in a singlet state. The two spins in the pair carry opposite momenta with spin 1 flying to the left and spin 2 flying to the right. To observe the double-spin state, a non-uniform magnetic field is placed on each side of the source. Figure 7.1 shows a schematic of the new experimental setup.

Here is what we will observe in this double-spin Stern-Gerlach experiment for which we have turned down the source's intensity so that only one pair of spins is emitted from the source each time. The spin singlet state (7.18) has two components: the first component is  $|ud\rangle$ , which indicates that if spin 1 is up, then spin 2 is down; the second component is  $|du\rangle$ , which means that if spin 1 is down then spin 2 is up. This implies a magical phenomenon in the double-spin Stern-Gerlach experiment: (1) if the spin flying to the left is detected at the upper spot, the spin flying to the right will always appear in the lower spot; (2) if the spin flying to the left appears in the lower spot, the spin flying to the right must appear on the upper spot. There will never be situations when both spins appear on the upper or lower spots. This indicates a remarkable correlation between the two spins: if spin 1 is up, we immediately know

spin 2 is down for sure; if spin 1 is down, we instantly know spin 2 is up. This is one of the characteristic feature of quantum entanglement—nonlocal correlation. The non-locality is reflected by Eq. (7.18) having nothing to do with the positions of the spins. Alternatively, the non-locality is intuitively manifested in Fig. 7.1: the experimental results are clearly independent of the distance between the two detection screens.

Interestingly, similar nonlocal correlation also occurs in the classical world. Suppose there are two identical boxes, one containing a red ball and the other containing a white ball. The two boxes are given to a pair of twins, Dingding and Dangdang. Neither Dingding nor Dangdang saw how the ball was put into the box, so they do not know the color of the ball in the box in their hands. Then Dangdang flies to the Mars on a spaceship while Dingding stays on the Earth. If Dingding opens his box, finding a red ball inside, he immediately knows that Dangdang has a white ball; if he finds a white ball inside, he immediately knows that Dangdang has a red ball. The shortest distance between the Mars and the Earth is  $54.6 \times 10^6$  km. It takes about 3 minutes for light to travel to the Mars from the Earth. Dingding, of course, does not need to wait 3 minutes to know what ball is in Dangdang's box. So, the correlation is clearly nonlocal. Such nonlocal correlation is in fact quite common, and appears not different from the nonlocal correlation in quantum entanglement. For a long time, physicists also thought that these two types of nonlocal correlations were the same. In 1964, Bell proved that an inequality which is strictly obeyed by classical nonlocal correlation can be violated by quantum entanglement.

### 7.2.1 Bell's Inequality

We shall see that the proof of Bell's inequality is purely mathematical, involving no physics at all. Yet its magic and thought-provoking aspect comes from the connection to physics: quantum entanglement violates this inequality, but classical nonlocal correlation does not.

We begin by revisiting single-spin and double-spin operators in the context of double-spin Stern-Gerlach experiment. As the name suggests, the single-spin operator is an operator involving only one spin. In the double-spin Stern-Gerlach experiment, this corresponds to only one magnetic field. The double-spin operator, on the other hand, involves two spins, which in the double-spin Stern-Gerlach experiment corresponds to two magnetic fields, one on each side. The case illustrated in Fig. 7.1, where both magnetic fields point in the  $z$  direction, is associated with the double-spin operator,  $\hat{\sigma}_z \otimes \hat{\tau}_z$ . Using the operation rule of the double-spin operator introduced before, we can verify

$$\hat{\sigma}_z \otimes \hat{\tau}_z |S\rangle = -|S\rangle. \quad (7.19)$$

It indicates that the spin singlet state  $|S\rangle$  is an eigenstate of the double-spin operator  $\hat{\sigma}_z \otimes \hat{\tau}_z$  with eigenvalue  $-1$ . We have mentioned earlier that the eigenvalue corresponds to a possible outcome of measurement. For a single spin, the outcome can only be spin up or spin down, i.e.,  $1$  or  $-1$ . For two spins, if the outcome is  $-1$ , it

means that the measurement outcome of spin 1 and the measurement outcome of spin 2 are opposite: if spin 1 is up, then spin 2 is down; if spin 1 is down, then spin 2 is up. This is consistent with our earlier analysis using the components of the singlet state  $|S\rangle$ .

Now consider a new double-spin operator,  $\mathbf{n} \cdot \hat{\sigma} \otimes \mathbf{n} \cdot \hat{\tau}$ . In the double-spin Stern-Gerlach experiment, this corresponds to both magnetic fields oriented along the direction  $\mathbf{n}$ . It can be straightforwardly shown that the spin singlet state  $|S\rangle$  is an eigenstate of  $\mathbf{n} \cdot \hat{\sigma} \otimes \mathbf{n} \cdot \hat{\tau}$ , i.e.,

$$\mathbf{n} \cdot \hat{\sigma} \otimes \mathbf{n} \cdot \hat{\tau} |S\rangle = - |S\rangle. \tag{7.20}$$

The eigenvalue  $-1$  means that, in the experiment shown in Fig. 7.1, regardless of the direction  $\mathbf{n}$  of the magnetic fields, two spins in a pair will appear on the opposite spots on the screen. In other words, if you measure along the direction  $\mathbf{n}$  and find spin 1 is up, then spin 2 is down, or vice versa. Previously we have discussed a special case of  $\mathbf{n}$  along the  $z$  direction. Interested readers can verify the following identity,

$$|S\rangle = -\frac{e^{-i\varphi}}{\sqrt{2}}(|n_+n_- \rangle - |n_-n_+ \rangle), \tag{7.21}$$

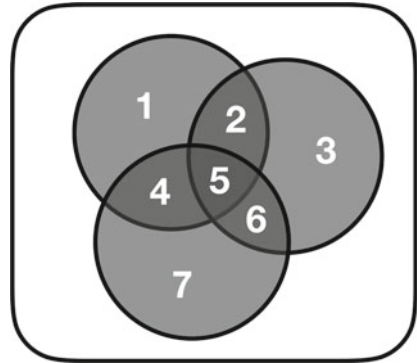
where  $|n_+\rangle$ ,  $|n_-\rangle$  are the two eigenstates of the spin operator along the  $\mathbf{n}$  direction  $\mathbf{n} \cdot \hat{\sigma}$  (see Eq. (5.19)). The two components on the right side of the above equation give the same conclusion: in the double-spin Stern-Gerlach experiment, if spin 1 is found up in a measurement along the direction  $\mathbf{n}$ , then we can say with absolute certainty that spin 2 is observed down, and vice versa.

We have already pointed out that nonlocal correlation exists in classical systems as well. A natural question arises, is the nonlocal correlation in quantum entanglement the same as the classical nonlocal correlation? Since the classical nonlocal correlation can be described in terms of classical probability theory (that is, the theory that describes probabilistic events such as dice, slot machines, etc.), the question can be formulated in an alternative way, can the classical probability theory fully explain the correlations in quantum entanglement? In 1964, Bell (John Stewart Bell, 1928–1990) proved that this was impossible. Bell began by assuming that this was possible, and showed that the nonlocal correlation in quantum entanglement should then satisfy an inequality. He then gave a counterexample, showing the nonlocal correlation of two spins in a singlet state violates this inequality. Bell’s result demonstrates the correlation in quantum entanglement is inexplicable with the classical probability theory. The inequality proved by Bell is expressed in terms of the expectation value of the double-spin operator, and his proof is not easy for people with no sophisticated



Bell (1928–1990)

**Fig. 7.2** A set of elements. The box encloses all the elements; the upper left circle contains the elements with attribute  $A$ ; the upper right circle the elements with attribute  $B$ ; the bottom circle the elements with attribute  $C$



mathematical knowledge. Susskind (Leonard Susskind, 1940–) creatively adapted Bell's proof and made it much easier to understand.<sup>1</sup> We present his proof below.

**Bell's inequality**—Suppose we have a set, whose elements may have three possible attributes  $A$ ,  $B$ , and  $C$ . Define a subset of elements  $S(A, \neg B)$ , which has attribute  $A$  but does not have attribute  $B$ . By the same token, we define  $S(B, \neg C)$  and  $S(A, \neg C)$ . They satisfy the following inequality

$$S(A, \neg B) + S(B, \neg C) \geq S(A, \neg C). \quad (7.22)$$

**Proof** As shown in Fig. 7.2, the three attributes  $A$ ,  $B$ , and  $C$  divide the total set into eight subsets:  $K_1, K_2, K_3, K_4, K_5, K_6, K_7$ , and the subset that excludes  $A$ ,  $B$ , and  $C$  (the white region in Fig. 7.2). Evidently,  $S(A, \neg B) = K_1 + K_4$ ,  $S(B, \neg C) = K_2 + K_3$ ,  $S(A, \neg C) = K_1 + K_2$ . So

$$S(A, \neg B) + S(B, \neg C) = K_1 + K_4 + K_2 + K_3 = S(A, \neg C) + K_3 + K_4 \geq S(A, \neg C). \quad (7.23)$$

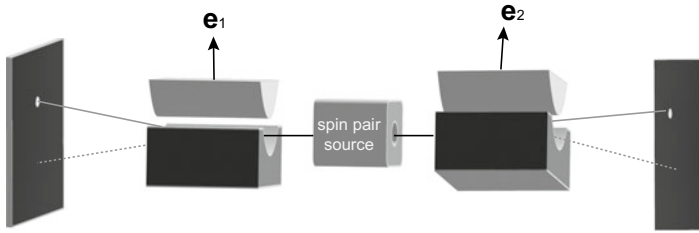
This completes the proof. Dividing both sides of the above inequality by the total number of elements in the whole set, we get

$$p(A, \neg B) + p(B, \neg C) \geq p(A, \neg C), \quad (7.24)$$

where  $p(A, \neg B)$  is the probability to have attribute  $A$  without having attribute  $B$ , and similarly for  $p(B, \neg C)$  and  $p(A, \neg C)$ . In the following, we will show in detail that this seemingly absolute inequality is violated due to the correlation between two spins in the singlet state  $|S\rangle$ .

In the double-spin Stern-Gerlach experiment shown in Fig. 7.1, the magnetic field on each sides is oriented along the same direction. We can, of course, orient them along different directions. For instance, the magnetic field on the left side is along direction  $\mathbf{e}_1$ , and the magnetic field on the right side is along direction  $\mathbf{e}_2$  as

<sup>1</sup> Please see the video recording of Susskind's course of quantum mechanics at Stanford on YouTube.



**Fig. 7.3** The double-spin Stern-Gerlach experiment. The left magnetic field is along the  $\mathbf{e}_1$  direction and the right magnetic field is along the  $\mathbf{e}_2$  direction. Note that the two spots on the screens represent only one of the four possible outcomes

shown in Fig. 7.3. For simplicity, we assume both  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are in the  $xz$  plane, without component in the  $y$  direction. Because spin 1 flies leftwards, its observable operator is  $\mathbf{e}_1 \cdot \hat{\boldsymbol{\sigma}}$ ; similarly, the observable operator for spin 2 is  $\mathbf{e}_2 \cdot \hat{\boldsymbol{\tau}}$ . In this case, the double-spin operator takes the form  $\mathbf{e}_1 \cdot \hat{\boldsymbol{\sigma}} \otimes \mathbf{e}_2 \cdot \hat{\boldsymbol{\tau}}$ . The two eigenstates of the operator  $\mathbf{e}_1 \cdot \hat{\boldsymbol{\sigma}}$  are *cf.* Eq. (5.19)

$$\mathbf{e}_1 \cdot \hat{\boldsymbol{\sigma}} |e_1^+\rangle = |e_1^+\rangle, \quad \mathbf{e}_1 \cdot \hat{\boldsymbol{\sigma}} |e_1^-\rangle = -|e_1^-\rangle, \quad (7.25)$$

and the eigenstates of  $\mathbf{e}_2 \cdot \hat{\boldsymbol{\tau}}$  are

$$\mathbf{e}_2 \cdot \hat{\boldsymbol{\tau}} |e_2^+\rangle = |e_2^+\rangle, \quad \mathbf{e}_2 \cdot \hat{\boldsymbol{\tau}} |e_2^-\rangle = -|e_2^-\rangle. \quad (7.26)$$

Using these eigenstates, we can construct the four eigenstates of the double-spin operator  $\mathbf{e}_1 \cdot \hat{\boldsymbol{\sigma}} \otimes \mathbf{e}_2 \cdot \hat{\boldsymbol{\tau}}$ ,  $|e_1^+ e_2^+\rangle, |e_1^+ e_2^-\rangle, |e_1^- e_2^+\rangle, |e_1^- e_2^-\rangle$ . These four eigenstates constitute an orthonormal basis for the Hilbert space of two spins. We can expand the singlet state on this basis as

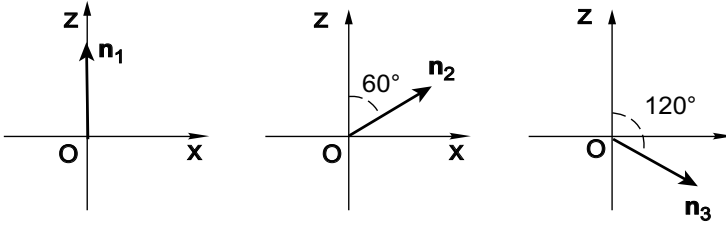
$$|S\rangle = g_1 |e_1^+ e_2^+\rangle + g_2 |e_1^+ e_2^-\rangle + g_3 |e_1^- e_2^+\rangle + g_4 |e_1^- e_2^-\rangle. \quad (7.27)$$

We multiply this equation with  $\langle e_1^+ e_2^+ |$  from the left. By virtual of the orthonormal property of the basis, only the first term on the right hand side of the equation is nonzero. Therefore, we have

$$p(\hat{e}_1, \hat{e}_2) = |g_1|^2 = |\langle e_1^+ e_2^+ | S \rangle|^2. \quad (7.28)$$

This is the probability of finding two spins pointing up along direction  $\mathbf{e}_1$  and direction  $\mathbf{e}_2$ , respectively. This probability apparently depends on the angle between  $\mathbf{e}_1$  and  $\mathbf{e}_2$ . For simplicity, we assume that  $\mathbf{e}_1$  is along the  $z$  axis, and  $\mathbf{e}_2$  is in the  $xz$  plane with an angle  $\theta$  from the  $z$  axis, i.e.,  $\hat{e}_1 = \{0, 0, 1\}$  and  $\hat{e}_2 = \{\sin \theta, 0, \cos \theta\}$ . Thus we have *cf.* Eq. (5.19)





**Fig. 7.4** One example of violating Bell's inequality.  $\mathbf{n}_1$ ,  $\mathbf{n}_2$  and  $\mathbf{n}_3$  are three different directions

$$|e_1^+ e_2^+\rangle = |u\rangle \otimes \left( \cos \frac{\theta}{2} |u\rangle + \sin \frac{\theta}{2} |d\rangle \right) = \cos \frac{\theta}{2} |uu\rangle + \sin \frac{\theta}{2} |ud\rangle, \quad (7.29)$$

and

$$\begin{aligned} p(\hat{e}_1, \hat{e}_2) &= \left| \cos \frac{\theta}{2} \langle uu| + \sin \frac{\theta}{2} \langle ud| \right| |S\rangle \|^2 \\ &= \sin^2 \frac{\theta}{2} |\langle ud|S\rangle|^2 = \frac{1}{2} \sin^2 \frac{\theta}{2}. \end{aligned} \quad (7.30)$$

With this result we are ready to demonstrate that the nonlocal correlation in quantum entanglement can violate Bell's inequality.

We assign attributes  $A$ ,  $B$ , and  $C$  to situations where spin 1 is found up along directions  $\mathbf{n}_1 = \{0, 0, 1\}$ ,  $\mathbf{n}_2 = \{\sqrt{3}/2, 0, 1/2\}$ , and  $\mathbf{n}_3 = \{\sqrt{3}/2, 0, -1/2\}$  (see Fig. 7.4), respectively. Accordingly, attribute  $\neg B$ , the negation of attribute  $B$ , corresponds to finding spin 1 down along direction  $\mathbf{n}_2$ . In the singlet state  $|S\rangle$ , when spin 1 is down along direction  $\mathbf{n}_2$ , we are certain to find spin 2 up in the  $\mathbf{n}_2$  direction. So attribute  $\neg B$  also corresponds to the situation where spin 2 is found up along the direction  $\mathbf{n}_2$ . Similarly, attribute  $\neg C$  can be regarded as finding spin 2 up along the direction  $\mathbf{n}_3$ . Based on these understandings, we obtain from Eq. (7.30)

$$p(A, \neg B) = p(\mathbf{n}_1, \mathbf{n}_2) = \frac{1}{2} \sin^2 \frac{\pi}{6} = \frac{1}{8}, \quad (7.31)$$

$$p(B, \neg C) = p(\mathbf{n}_2, \mathbf{n}_3) = \frac{1}{2} \sin^2 \frac{\pi}{6} = \frac{1}{8}, \quad (7.32)$$

and

$$p(A, \neg C) = p(\mathbf{n}_1, \mathbf{n}_3) = \frac{1}{2} \sin^2 \frac{\pi}{3} = \frac{3}{8}. \quad (7.33)$$

Obviously, we have

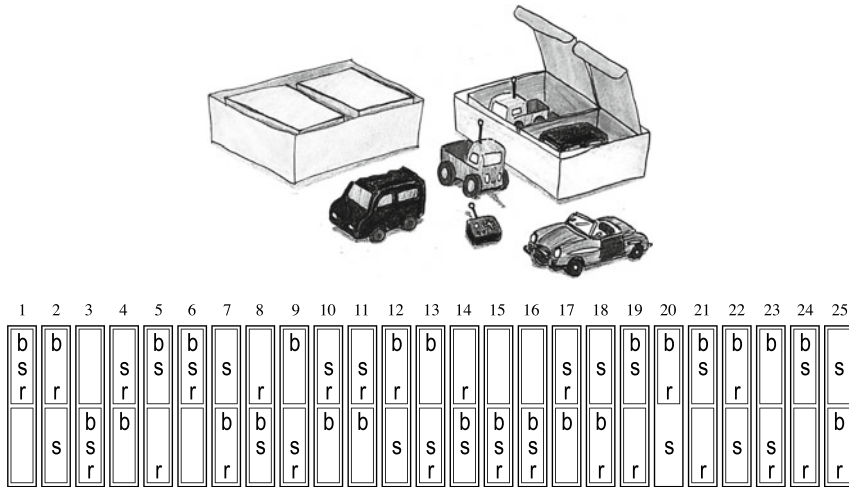
$$p(A, \neg B) + p(B, \neg C) = 1/4 < p(A, \neg C). \quad (7.34)$$

Bell's inequality is violated! We can, of course, define a new set of attributes  $A$ ,  $B$ , and  $C$  (or,  $\mathbf{n}_1, \mathbf{n}_2$ , and  $\mathbf{n}_3$ ), which do not violate Bell's inequality. But what matters is that there does exist an example violating Bell's inequality. The violation originates from the nonlocal correlation between the two spins: if spin 1 is up, spin 2 must be down; and vice versa. In the previous discussions, we have repeatedly used this correlation to convert the single spin functions  $p(A, \neg B)$ ,  $p(B, \neg C)$ , and  $p(A, \neg C)$  to the double-spin correlation functions  $p(\mathbf{n}_1, \mathbf{n}_2)$ ,  $p(\mathbf{n}_2, \mathbf{n}_3)$ , and  $p(\mathbf{n}_1, \mathbf{n}_3)$ , respectively. It is exactly this nonlocal correlation that enables a violation of Bell's inequality.

In stark contrast, the classical nonlocal correlation never violates Bell's inequality. Consider the following example. Dingding and Dangdang are twins. Before Dangdang is leaving to work in another city, Dingding takes out a lot of identical rectangular boxes (see Fig. 7.5). Dingding says, "Here are the 50 toy cars we played with when we were kids. Among them, there are exactly 25 black cars, 25 sports cars, and 25 remote-controlled cars. I have put all these toy cars into 25 identical rectangular boxes". Dingding takes out a rectangular box, inside which there are two identical smaller boxes, saying, "Each of the small boxes has one car in it. When packing, I made sure that no two sports cars were in the same big box, no two black cars were in the same big box, and no two remote-controlled cars were in the same big box (see Fig. 7.5 for illustration). Now you select one small box randomly from every big box, and I will take the other". Dangdang randomly takes one small box from every big box and departs with a total of 25 small boxes.

In another city, Dangdang often looks at the 25 toy cars, which remind him all the happy moments that he had with Dingding. One day, Dangdang suddenly realizes that Dingding is playing a math game with him. From the 25 toy cars that he has, Dangdang deduces this: there are 6 big boxes, from which he took a black car while leaving Dingding with a sports car; there are 7 big boxes, from which he took a sports car while leaving Dingding with a remote controlled car; there are 7 big boxes, from which he took a black car while leaving Dingding with a remote controlled car. These three numbers obviously satisfy an inequality:  $6 + 7 > 7$ . This is illustrated in the lower half of Fig. 7.5 with Dangdang's cars in the top row. Dangdang finds that it is no coincidence that there is such an inequality. Regardless his choice at the time, the numbers of the first two types of big boxes always add up to be larger than the number of the third type of big boxes. This is in fact Bell's inequality. Let us see why.

The toy cars can have three attributes, black, sports, and remote-controlled. We denote them, respectively, as  $A$ ,  $B$ , and  $C$ . We focus on three types of big boxes. From the first type of big box, if Dangdang takes a black car, Dingding is left with a sports car. There are  $D_{A,B}$  such boxes. From the second type, if Dangdang takes a sports car, Dingding is left with a remote-controlled car. There are  $D_{B,C}$  such boxes. From the third type, if Dangdang takes a black car, Dingding is left with a remote-controlled car. There are  $D_{A,C}$  such boxes. Because the two cars in the same big box cannot both be black cars, or sports cars, or remote-controlled cars, and because the number of black cars, or sports cars, or remote-controlled cars is exactly half of the total number of cars, there is a nonlocal correlation between the toy cars that Dingding has and the ones that Dangdang has. For example, when Dangdang got a sports car, it



**Fig. 7.5** Toy cars. They may be black, remote-controlled, or sports cars, or they may have other properties. There are 50 cars: 25 of them are black; 25 of them are remote-controlled; 25 of them are sports. They are placed in 25 pairs of small boxes according to the following rule: for each pair, when the car in one of the boxes is black, the car in the other box is not black; when the car in one of the boxes is sports, the car in the other box is not sports; when the car in one of the boxes is remote-controlled, the car in the other box is not remote-controlled. For instance, for the first big box, the upper small box contains a black, remote-controlled sports car while the car in the lower small box may be a red fire-engine, which does not have any of these three attributes. For the 16th big box, the car in the upper is remote-controlled while the one in the lower is a black sports car. One of such arrangements is shown with b=black, r=remote-controlled, s=sports

is certain that Dingding did not get a sports car from the same big box; the inverse is also true, if Dangdang did not get a sports car, then Dingding must have got a sports car from the same big box. Therefore, the number of Dangdang's cars that are black but not sports,  $S(A, \neg B)$ , is equal to the number of the first type of big boxes,  $D_{A,B}$ . Similarly, we have  $S(B, \neg C) = D_{B,C}$  and  $S(A, \neg C) = D_{A,C}$ . According to Bell's inequality  $S(A, \neg B) + S(B, \neg C) \geq S(A, \neg C)$ , we have  $D_{A,B} + D_{B,C} \geq D_{A,C}$ . For the example in Fig. 7.5,  $D_{A,B} = 6$ ,  $D_{B,C} = 7$ ,  $D_{A,C} = 7$ , which is only a special case of this inequality.

At first glance, this toy car example is quite similar to the spin singlet state. If two toy cars in a big box are regarded as a pair of spins, that two cars in a big box cannot simultaneously be black cars, or sports cars, or remote controlled cars correspond to that two spins in the singlet state must point to opposite directions. However, the nonlocal correlation between toy cars does not lead to the violation of Bell's inequality, but the nonlocal correlation of two spins in a singlet state does. So the classical nonlocal correlation is different from the nonlocal correlation in quantum entanglement. This is revealed by Bell's inequality.

In Chap. 5, we have demonstrated the origin of probability in classical physics with dice rolling. There we showed that the classical probability is not intrinsic but stems from our ignorance. We use probability to predict the outcome when there exist some random elements (e.g., accidental vibration of the table), or when we are

not able to completely describe the various aspects of dice rolling (e.g., elasticity of the dice material, elasticity of the table material, the angle at which the dice collide with the table, etc.). If we, like the serious Xiaoliang, try all out to eliminate every uncertain factors and carefully take into account every aspects of dice rolling, we can eliminate probability and make precise prediction of the outcome (see the discussion in Sect. 5.3).

We have repeatedly emphasized that the quantum probability is different from the classical probability, in that it is fundamental and intrinsic of quantum systems, and does not arise because of some random or unknown factors. However, such emphasis seems somewhat ambiguous and not convincing because it is entirely possible that the probabilistic nature of quantum state is due to some “hidden” factors that are beyond the detection capabilities of current techniques. This is the hidden variable theory advocated by many physicists like Einstein. Consider a single spin, and assume that it is in the aforementioned quantum state

$$|\psi_{1/6}\rangle = \sqrt{\frac{1}{6}} |u\rangle + \sqrt{\frac{5}{6}} |d\rangle. \quad (5.3)$$

To understand the possible outcome of a measurement on this spin state, we can regard it as a special dice, in which five faces are marked with “down” and one face is marked with “up”. In the measurement, since we don’t know the “hidden” factors, we can only make probabilistic predictions about the outcome. Such an understanding may be reasonable when only a single spin is concerned. Bell told us with his inequality that such an understanding can not be right (5.3).

If the hidden variable theory were correct, the probability correlation measured in any two-spin quantum state had to satisfy the Bell’s inequality (7.24). This is because the hidden variable theory implies that the quantum probability is identical to the classical probability and is caused by incomplete knowledge. Therefore, the quantum probability would comply with the classical probability theory. The Bell’s inequality (7.24) strictly holds in classical probability theory. But we have seen earlier that the nonlocal correlation between two spins in the singlet state does not satisfy Bell’s inequality (7.24) in many cases. Such an violation has a profound implication: the probability in quantum mechanics is fundamental and intrinsic and the hidden variable theory does not exist.<sup>2</sup>

Physicists represented by Einstein and Bohr had heated debates about the probabilistic nature of quantum mechanics, and argued whether the hidden variable theory exists. Before the discovery of Bell’s inequality, these debates had remained rather philosophical. Bell turned the debate into a concrete mathematical expression, allowing physicists to test it experimentally. So far, all experiments have shown agreement with quantum mechanics: the Bell’s inequality can be violated. God does play dice,

---

<sup>2</sup> It is generally agreed that Bell’s inequality rules out only the *local* hidden variable theory; *non-local* hidden variable theory may still be right. We do not discuss the subtlety between these two versions of hidden variable theory.

but the dice is no ordinary dice and it is quantum. Bell has lifted the veil of the mystery and demonstrated the “quantumness” of such a dice.

### 7.2.2 Loss of Individuality

There is another distinctive feature of quantum entanglement that is rarely mentioned in popular articles on quantum entanglement. I call this feature the loss of individuality, which we discuss in detail below.

We have repeatedly said before that a quantum state is represented by a vector in a Hilbert space, which completely describes the state of a system. The singlet spin state  $|S\rangle$  in Eq. (7.18) is a vector in a four-dimensional Hilbert space that gives a complete description of the double-spin system. So what is the quantum state of each spin in this singlet state? Suppose each of the two spins has a definite quantum state: the state of spin 1 is described by a vector  $|\phi_1\rangle$  in the two-dimensional Hilbert space, while the state of spin 2 is described by another vector  $|\phi_2\rangle$  in the two-dimensional Hilbert space. The state of the two spins, therefore, can be described by direct product  $|\phi_1\rangle \otimes |\phi_2\rangle$ , which is a product state. Since the singlet state is an entangled state, this is impossible. This analysis shows that, although we know perfectly which quantum state the double-spin system is in, we do not know which quantum state an individual spin is in. In simple words, each spin loses itself in the spin singlet state  $|S\rangle$ . This is an important and generic feature of quantum entangled states, which is fundamentally different from classical mechanics. If a system contains two classical particles and we know the state of the system exactly, it automatically indicates that we know the state of particle  $(x_1, p_1)$  and the state of particle  $(x_2, p_2)$ . In classical mechanics, the state of the whole system is known precisely only if the state of each particle is known precisely. However, in quantum mechanics, the situation is radically different. In an entangled state, we know exactly the state of the total system, but not the state of individual particle. I call this feature of an entangled state loss of individuality.

Let us try to understand the loss of individuality from a more physical perspective. According to the discussion related to Eq. (5.21) in Chap. 5, for any single spin state  $|\psi\rangle$ , we can always find a direction  $\mathbf{n}$  such that  $\mathbf{n} \cdot \hat{\sigma} |\psi\rangle = |\psi\rangle$ , which is equivalent to

$$\langle \psi | \mathbf{n} \cdot \hat{\sigma} | \psi \rangle = 1. \quad (7.35)$$

This means that in the usual Stern-Gerlach experiment, if every silver atom emitted from the source is in the spin state  $|\psi\rangle$ , we can always choose a proper orientation of the magnetic field so that only one spot appears on the screen.

Let us turn to the double-spin Stern-Gerlach experiment (see Fig. 7.1). If the spin pair emitted from the source is in the product state  $|\Psi_{12}\rangle$ , what will be observed? Let us first do the calculation,

$$\langle \Psi_{12} | \mathbf{n} \cdot \hat{\sigma} | \Psi_{12} \rangle = \langle \psi | \mathbf{n} \cdot \hat{\sigma} | \psi \rangle \langle \phi | \phi \rangle = \langle \psi | \mathbf{n} \cdot \hat{\sigma} | \psi \rangle, \quad (7.36)$$

$$\langle \Psi_{12} | \mathbf{n} \cdot \hat{\boldsymbol{\tau}} | \Psi_{12} \rangle = \langle \psi | \psi \rangle \langle \phi | \mathbf{n} \cdot \hat{\boldsymbol{\tau}} | \phi \rangle = \langle \phi | \mathbf{n} \cdot \hat{\boldsymbol{\tau}} | \phi \rangle. \quad (7.37)$$

According to the above discussion, we can always find two directions,  $\mathbf{n}_1$  and  $\mathbf{n}_2$ , such that  $\langle \psi | \mathbf{n}_1 \cdot \hat{\boldsymbol{\sigma}} | \psi \rangle = \langle \phi | \mathbf{n}_2 \cdot \hat{\boldsymbol{\sigma}} | \phi \rangle = 1$ . Physically this means that, when we orient the magnetic field on the left side to be along the direction  $\mathbf{n}_1$ , and the magnetic field on the right side to be along the direction  $\mathbf{n}_2$ , only one spot will appear in each detection screen. As we shall see below, it is very different for an entangled state.

We consider the spin singlet state  $|S\rangle$ . Let us calculate  $\langle S | \mathbf{n} \cdot \hat{\boldsymbol{\sigma}} | S \rangle$ . For  $\langle S | \hat{\sigma}_x | S \rangle$ , we have

$$\begin{aligned} \langle S | \hat{\sigma}_x | S \rangle &= \frac{1}{2} (\langle ud | - \langle du |) \hat{\sigma}_x (|ud\rangle - |du\rangle) \\ &= \frac{1}{2} (\langle ud | - \langle du |) (\hat{\sigma}_x |ud\rangle - \hat{\sigma}_x |du\rangle) \\ &= \frac{1}{2} (\langle ud | - \langle du |) (|dd\rangle - |uu\rangle) \\ &= \frac{1}{2} (\langle ud | dd\rangle - \langle ud | uu\rangle - \langle du | dd\rangle + \langle du | uu\rangle) \\ &= \frac{1}{2} (\langle u | d \rangle \langle d | d \rangle - \langle u | u \rangle \langle d | u \rangle - \langle d | d \rangle \langle u | d \rangle + \langle d | u \rangle \langle u | u \rangle) \\ &= 0. \end{aligned} \quad (7.38)$$

With similar calculations we have

$$\langle S | \hat{\sigma}_y | S \rangle = \langle S | \hat{\sigma}_z | S \rangle = 0. \quad (7.39)$$

Thus

$$\langle S | \mathbf{n} \cdot \hat{\boldsymbol{\sigma}} | S \rangle = n_x \langle S | \hat{\sigma}_x | S \rangle + n_y \langle S | \hat{\sigma}_y | S \rangle + n_z \langle S | \hat{\sigma}_z | S \rangle = 0. \quad (7.40)$$

That the expectation is zero means that spin 1 has equal probability being up or down. Since  $\mathbf{n}$  is arbitrary, one always observes two discrete spots of the same size on the left screen regardless how one orients the left magnets. The result is the same for spin 2, and we have

$$\langle S | \mathbf{n} \cdot \hat{\boldsymbol{\tau}} | S \rangle = 0. \quad (7.41)$$

Similarly, this means that regardless how you orient the magnets on the right, you always observe two discrete spots of the same size on the right screen. This is strikingly different from the product state for which one can always turn the magnets so that there are only one spot on each of the screens. That two spots of the same size on each screen means that the individual spin in the spin singlet state  $|S\rangle$  completely loses its own individuality, with no knowledge which quantum state it is in.

The above analysis is done for a pair of spins, but the result is general. The product state and the entangled state are not only different mathematically, and they are also physically very different with different observable consequences. For a single particle

(or spin) in an entangled state, it loses its individuality and its quantum state becomes uncertain. In classical mechanics or in our daily life, to know the whole we have to know each individual; in quantum mechanics, one often knows the whole without knowing each individual. This difference is fundamental.

As a single particle in an entangled state has no well-defined quantum state, it cannot be described by a vector in the Hilbert space. One is forced to wonder what is its actual state and how to describe it? The answer is that a single particle in an entangled state is in a mixed state, described by a density matrix. For the spin singlet state, spin 1 is in the following mixed state:

$$\hat{\rho}_1 = \frac{1}{2} |u\rangle \langle u| + \frac{1}{2} |d\rangle \langle d|. \quad (7.42)$$

This density matrix shows that spin 1 is in an uncertain quantum state: it is in the quantum state  $|u\rangle$  with a probability of  $1/2$ , and in the quantum state  $|d\rangle$  with a probability of  $1/2$ . As to why this density matrix has such an odd form, how it is obtained, and what its implication is, the answers to these questions are all beyond the scope of this book. Interested readers are referred to Landau's *Quantum Mechanics*.

# Chapter 8

## Quantum Measurement



At any given time, any person or object has a definite position and moves at a certain speed (a stationary object has a zero speed). This observation is so normal and obvious that it almost never catches our attention. Occasionally, we may notice it, for example in driving, for which case the speedometer will tell us the speed of the car and the GPS will tell us the position of the car. Obviously, the speedometer and the GPS do not affect each other, nor do the speedometer's measurement of speed and the GPS's measurement of position affect our driving or the vehicle's conditions. This is our life experience: every object has a well-defined position and speed at any given moment; we can measure the position and speed of an object simultaneously; the two measurements will not affect each other, nor will they change the state of the object.

Our daily observation is perfectly consistent with classical mechanics. In classical mechanics, an object has a well-defined position and momentum (or velocity) at any instant of time. Not only can we measure the position and momentum of an object, but we can measure them simultaneously; the outcome of the measurement is always certain, and the effect of the measurement on an object can always, at least in principle, be reduced to such an extent that we can ignore it. However, when you study classical mechanics, no textbook or teacher would mention these features, let alone emphasize them as they are simply taken for granted in classical mechanics.

For those who are not familiar with quantum mechanics, the above discussion may appear strange and meaningless. It is like when someone describes the appearance of a person with "He has two eyes, one nose, ...", you may think this person is a weirdo or trying to waste your time. But as soon as this person begins to describe the appearance of an alien, you would no longer feel strange. In this chapter, I will introduce such an "alien", quantum measurement.

In classical physics, measurement is not a part of the theory. The discussion of measurement, such as how to improve the measurement precision or to reduce noises, is limited to specific experimental techniques. In principle, the results of different measurements are always deterministic, whether they are performed simultaneously



or separately; the effects of noise and devices on the measurement can always be reduced so that be safely ignored.

In quantum mechanics, measurement is not only a technical operation, it is an intrinsic part of the theory of quantum mechanics. Let us recapitulate its basic framework (see Chap. 5). In quantum mechanics, a quantum state is represented by a vector in a Hilbert space. Hilbert spaces and their vectors are abstract, and can not be measured directly. To establish their connection with the reality, observables are introduced. An observable is mathematically represented by an equally abstract operator (or matrix), and only its eigenvalues can be directly measured. The quantum state only specifies the probability of a possible measurement outcome.

It is clear from the above description, measurement plays a unique and fundamental role in quantum mechanics. However, the concept of measurement is so subtle that it is one of the most debated topics in quantum mechanics. The general consensus on a quantum measurement is as follows. (1) If the operators representing two measurements do not commute, the two measurements cannot give definite results simultaneously. This is the famous Heisenberg's uncertainty relation. (2) A measurement leaves an impact on a quantum system, which cannot be ignored. Most debates about quantum measurement are concerned with the second point. How does a measurement affect a quantum system? What is the consequence of this effect? There are different views on this question, and I will introduce two of them: the Copenhagen interpretation and the many-worlds theory. The Copenhagen interpretation, at the heart of which is the collapse of wave function, is the most popular; the many-worlds theory is, in my opinion, the most reasonable. Unfortunately, we are not yet in a position to test experimentally which of these viewpoints is right or wrong.

## 8.1 The Uncertainty Relation

In 1927, Heisenberg discovered an important inequality

$$\Delta x \Delta p \geq \hbar/2. \quad (8.1)$$

This is the famous Heisenberg's uncertainty relation. In this inequality,  $\Delta x$  is the uncertainty of position;  $\Delta p$  is the uncertainty of momentum. It means that, if a particle has a definite position, i.e.,  $\Delta x = 0$ , then  $\Delta p = \infty$ , i.e., its momentum is completely uncertain, having equal probability to have any value. Conversely, if the particle has a definite momentum,  $\Delta p = 0$ , then  $\Delta x = \infty$ , i.e., its position is completely unknown, having equal probability to be anywhere in the universe. Heisenberg's uncertainty relation reveals a stunning fact: regardless how advanced your measurement devices are, even if you are able to eliminate all measurement noises, you still cannot simultaneously make definite predictions for a measurement of position and for a measurement of momentum. This is a bizarre yet fundamental feature of quantum measurement.

In quantum mechanics, an observable is represented by an operator (i.e., matrix) and the order of multiplication of two matrices can affect the result. For example, you can directly verify  $\hat{\sigma}_x \hat{\sigma}_y \neq \hat{\sigma}_y \hat{\sigma}_x$ . In this case, we say the operators  $\hat{\sigma}_x$  and  $\hat{\sigma}_y$  do not commute. The Heisenberg's uncertainty relation arises exactly from the non-commutativity of position and momentum operators. There is a general conclusion in quantum mechanics: if two observables  $\hat{O}_1$  and  $\hat{O}_2$  do not commute,  $\hat{O}_1 \hat{O}_2 \neq \hat{O}_2 \hat{O}_1$ , then the measurement uncertainty of  $\hat{O}_1$  and the measurement uncertainty of  $\hat{O}_2$  cannot simultaneously vanish. Since the mathematical representations of position and momentum operators involve calculus, which is beyond this book, we illustrate the relation between the non-commutativity of operators and measurement with spin. For a rigorous proof of the uncertainty relation (8.1), interested readers are referred to Landau's *Quantum Mechanics*.

We start with some simple mathematics of statistics. In particular, we shall define what is statistical uncertainty. For a variable  $w$ , suppose it has  $n$  possible values  $w_1, w_2, \dots, w_n$  with probability  $p_1, p_2, \dots, p_n$ , respectively. The average value or expectation value of this variable is defined as

$$\bar{w} = \sum_{i=1}^n p_i w_i. \quad (8.2)$$

The statistical uncertainty of this variable is then

$$\Delta w = \sqrt{\sum_{i=1}^n p_i (w_i - \bar{w})^2}. \quad (8.3)$$

Consider an example of tossing two coins, each of them has 1 on one side and 0 on the other. We are interested in the total values of the two coins after tossing. Apparently, there are three possible outcomes,  $w_1 = 2$ ,  $w_2 = 1$ , and  $w_3 = 0$ , which occur with probability  $p_1 = 1/4$ ,  $p_2 = 1/2$ , and  $p_3 = 1/4$ , respectively. As a result, the expectation value is

$$\bar{w} = \frac{1}{4}w_1 + \frac{1}{2}w_2 + \frac{1}{4}w_3 = 1. \quad (8.4)$$

The uncertainty is

$$\Delta w = \sqrt{\frac{1}{4}(w_1 - \bar{w})^2 + \frac{1}{2}(w_2 - \bar{w})^2 + \frac{1}{4}(w_3 - \bar{w})^2} = \frac{\sqrt{2}}{2}. \quad (8.5)$$

Consider another example. Suppose we measure a particle's position  $x$  with two different measurement devices. Each device performs four measurements, yielding the following results

Device 1 (m)				Device 2 (m)			
0.12	0.11	0.10	0.09	0.101	0.098	0.103	0.099

We assume that every measurement outcome occurs with equal probability. For device 1, we have the average value  $\bar{x}_1 = (0.12 + 0.11 + 0.10 + 0.09)/4 = 0.105$ , and the uncertainty

$$\Delta x_1 = \sqrt{\frac{(0.12 - \bar{x}_1)^2 + (0.11 - \bar{x}_1)^2 + (0.10 - \bar{x}_1)^2 + (0.09 - \bar{x}_1)^2}{4}} \approx 0.11. \quad (8.6)$$

For device 2, we have  $\bar{x}_2 = (0.101 + 0.098 + 0.103 + 0.099)/4 = 0.10025$ , and

$$\Delta x_2 = \sqrt{\frac{(0.101 - \bar{x}_2)^2 + (0.098 - \bar{x}_2)^2 + (0.103 - \bar{x}_2)^2 + (0.099 - \bar{x}_2)^2}{4}} \approx 0.002. \quad (8.7)$$

We see that  $\Delta x_2 \ll \Delta x_1$ , meaning the measurement precision of device 2 is higher than device 1.

Returning to quantum mechanics, we consider a spin in the following quantum state

$$|\psi\rangle = a|u\rangle + b|d\rangle. \quad (8.8)$$

We are interested in two spin operators  $\hat{\sigma}_x$  and  $\hat{\sigma}_z$ , which do not commute with each other. For the measurement of  $\hat{\sigma}_z$ , there are two possible outcomes: 1 with probability of  $|a|^2$ ;  $-1$  with probability of  $|b|^2$ . According to Eq. (8.2), its expectation value is

$$\bar{\sigma}_z = |a|^2 - |b|^2. \quad (8.9)$$

As can be straightforwardly shown, this result is consistent with the expectation value calculated for an operator in Chap. 5, i.e.,  $\bar{\sigma}_z = \langle \psi | \hat{\sigma}_z | \psi \rangle$ . According to Eq. (8.3), the uncertainty in a measurement of  $\hat{\sigma}_z$  is

$$\Delta \hat{\sigma}_z^2 = |a|^2(1 - \bar{\sigma}_z)^2 + |b|^2(-1 - \bar{\sigma}_z)^2 = 4|a|^2|b|^2. \quad (8.10)$$

We can directly verify that

$$\begin{aligned} \Delta \hat{\sigma}_z^2 &= \langle \psi | (\hat{\sigma}_z - \bar{\sigma}_z)^2 | \psi \rangle \\ &= \langle \psi | \hat{\sigma}_z^2 | \psi \rangle - 2\langle \psi | \bar{\sigma}_z \hat{\sigma}_z | \psi \rangle + \langle \psi | \bar{\sigma}_z^2 | \psi \rangle \\ &= \langle \psi | \hat{\sigma}_z^2 | \psi \rangle - 2\bar{\sigma}_z \langle \psi | \hat{\sigma}_z | \psi \rangle + \bar{\sigma}_z^2 \\ &= \langle \psi | \hat{\sigma}_z^2 | \psi \rangle - \bar{\sigma}_z^2 = 4|a|^2|b|^2. \end{aligned} \quad (8.11)$$

Therefore, not only have we obtained the expectation value and uncertainty in the measurement of  $\hat{\sigma}_z$ , we have also demonstrated that the calculation based on statistics is consistent with the calculation in quantum mechanics. For  $\hat{\sigma}_x$ , we use the prescriptions of quantum mechanics. The expectation value of the spin operator  $\hat{\sigma}_x$  is

$$\begin{aligned}\bar{\hat{\sigma}}_x &= \langle \psi | \hat{\sigma}_x | \psi \rangle = [a^* \langle u | + b^* \langle d |] \hat{\sigma}_x [a | u \rangle + b | d \rangle] \\ &= [a^* \langle u | + b^* \langle d |] [a | d \rangle + b | u \rangle] = a^* b + ab^*.\end{aligned}\quad (8.12)$$

For the uncertainty  $\Delta \hat{\sigma}_x$  in the measurement of  $\hat{\sigma}_x$ , we have

$$\Delta \hat{\sigma}_x^2 = \langle \psi | \hat{\sigma}_x^2 | \psi \rangle - \bar{\hat{\sigma}}_x^2 = 1 - (a^* b + ab^*)^2. \quad (8.13)$$

Due to the normalization condition  $|a|^2 + |b|^2 = 1$ , we choose  $a = \cos \theta$ ,  $b = e^{i\beta} \sin \theta$  and obtain

$$\Delta \hat{\sigma}_z = |\sin 2\theta|, \quad (8.14)$$

$$\Delta \hat{\sigma}_x = \sqrt{1 - \sin^2 2\theta \cos^2 \beta}. \quad (8.15)$$

Evidently, the maximum values of  $\Delta \hat{\sigma}_x$  and  $\Delta \hat{\sigma}_z$  are both 1, and the minimum values of  $\Delta \hat{\sigma}_x$  and  $\Delta \hat{\sigma}_z$  are both 0, i.e.,

$$0 \leq \Delta \hat{\sigma}_x \leq 1, \quad 0 \leq \Delta \hat{\sigma}_z \leq 1. \quad (8.16)$$

This is different from position and momentum operators, where the uncertainties  $\Delta x$  and  $\Delta p$  can be infinitely large. From Eqs. (8.14 and 8.15), we obtain an inequality

$$\Delta \hat{\sigma}_x^2 + \Delta \hat{\sigma}_z^2 = 1 + \sin^2 2\theta \sin^2 \beta \geq 1. \quad (8.17)$$

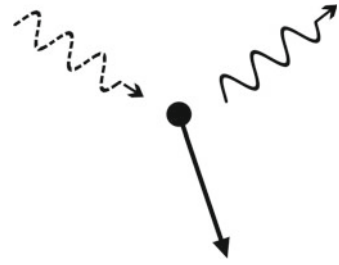
This is the Heisenberg's uncertainty relation for spin. Therefore, if the outcome of a measurement of  $\hat{\sigma}_x$  is absolutely certain,  $\Delta \hat{\sigma}_x = 0$ , then  $\Delta \hat{\sigma}_z = 1$ , meaning that the measurement of  $\hat{\sigma}_z$  is maximally uncertain, and vice versa.

To be more specific, let us consider the following spin state

$$|\psi\rangle = \frac{\sqrt{2}}{2} |u\rangle + \frac{\sqrt{2}}{2} |d\rangle. \quad (8.18)$$

We measure this spin state with the Stern-Gerlach setup (see Fig. 5.1). We repeat the experiment for 100 times, making sure that the silver atom is in the above state every time. For about 50 times, we find the silver atoms arriving at the upper spot; and for about 50 times, we find the silver atom arriving at the lower spot. It means that the measurement of  $\hat{\sigma}_z$  has the largest uncertainty. This observation is consistent with Eq. (8.11), which states  $\Delta \hat{\sigma}_z = 1$ . What will happen if we reorient the magnetic field in the Stern-Gerlach experiment to be along the  $x$  axis? Some readers may have already

**Fig. 8.1** Heisenberg's intuitive interpretation of the uncertainty relation. The solid circle represents the particle being observed, the dashed wavy line represents the incident photon, and the solid wavy line represents the emitted photon



noticed  $|\psi\rangle = |f\rangle$ , which is an eigenstate of  $\hat{\sigma}_x$ . Thus in every measurement, the outcome is the same and all the silver atoms hit the same spot on the detection screen along the positive direction of  $x$  axis. Therefore, the uncertainty of a measurement of  $\hat{\sigma}_x$  is zero. Again, this result is consistent with Eq. (8.13), for which we indeed have  $\Delta\hat{\sigma}_x = 0$ .

If a system is in an eigenstate of the operator  $\hat{O}$ ,  $|\phi_j\rangle$ , a measurement of  $\hat{O}$  of the system will produce a definite outcome, i.e., the uncertainty  $\Delta\hat{O} = 0$ . Interested readers are encouraged to prove it. The state  $|\psi\rangle$  in Eq. (8.18) with operator  $\hat{\sigma}_x$  discussed above is a special example of this general conclusion.

Why does the noncommutativity of operators result in the uncertainty relation? To answer this question, let us make a bold assumption that  $\sigma_x$  and  $\sigma_z$  share a common eigenstate  $|\phi_0\rangle$ . According to the previous analysis, if a spin is in this eigenstate, we will have  $\Delta\hat{\sigma}_x = \Delta\hat{\sigma}_z = 0$  simultaneously, which violates Heisenberg's uncertainty relation (8.17). Hence our assumption is wrong; the two observables  $\sigma_x$  and  $\sigma_z$  cannot have common eigenstates. We know that the two eigenstates of  $\sigma_x$  are  $|f\rangle$  and  $|b\rangle$ , which are indeed different from the eigenstates of  $\sigma_z$ ,  $|u\rangle$  and  $|d\rangle$ . There is a general conclusion in linear algebra: two commuting matrices share common eigenstates; two non-commuting matrices, except for some extremely special cases,<sup>1</sup> do not have common eigenstates. Detailed proofs and explanations of these two statements are beyond the scope of this book. Interested readers are referred to Dirac's *Principles of Quantum Mechanics*. With these two rigorous mathematical results, we immediately have: there is a Heisenberg's uncertainty relation between any pair of non-commuting operators; there is no uncertainty relation between any two commuting operators. For instance, two momentum operators along the different directions, such as  $\hat{p}_x$  and  $\hat{p}_y$ , are commutative, so there is no uncertainty relation between them as there are quantum states that are eigenstates of both  $\hat{p}_x$  and  $\hat{p}_y$ .

Heisenberg gave an intuitive explanation of his uncertainty principle. As shown in Fig. 8.1, one tries to measure a particle's position using light of wavelength  $\lambda$ . The accuracy of a measurement of position is approximately determined by the wave length,  $\Delta x \sim \lambda$ . During the measurement process, the photon collides with the particle. Since the momentum of a photon is  $\hbar k$  ( $k = 2\pi/\lambda$ ), the collision will cause an uncertainty in the measurement of the particle's momentum, which is approximately  $\Delta p \sim \hbar k$ . We then have

<sup>1</sup> Such as the angular momentum quantum state  $s$ , which is the common eigenstate shared by the angular momentum operators  $\hat{L}_x$ ,  $\hat{L}_y$ , and  $\hat{L}_z$ , which do not commute.

$$\Delta x \Delta p \sim 2\pi \hbar. \quad (8.19)$$

This estimate agrees very well with Eq. (8.1). Heisenberg's analysis is very profound, indicating measurement inevitably changes the measured state of a particle. We usually think it possible to eliminate all the noises in measurement devices at least in principle. For example, to measure the temperature of a bowl of water, we can insert an ordinary thermometer into the water, wait for the heat transfer between the water and the thermometer until an equilibrium is reached, and then read the temperature on the thermometer. But this method does not work if there is only 1 gram of water because the heat transfer between the thermometer and 1 gram of water will seriously affect the water temperature, resulting in a large measurement error. To avoid this issue, experimentalists can use a more sophisticated thermometer, such as contactless thermometer, to accurately measure the temperature, where the effect of the thermometer on the water is again negligible. Heisenberg's analysis shows that in a measurement of microscopic particles, the disturbance of a measuring device must be taken into account.

Due to the above analysis, Heisenberg's uncertainty relation is often interpreted as the impossibility to achieve precise measurement due to the unavoidable disturbance from the measuring device. This is wrong. The similarity between Eq. (8.19) and the Heisenberg's inequality (8.1) is coincident. Equation (8.19) is caused by the disturbance of the measuring device; Heisenberg's inequality (8.1) is a consequence of the noncommutativity of operators. They reflect different physics. Since position and momentum are non-commuting operators, one cannot find a quantum state which is an eigenfunction of both position and momentum. In other words, there does not exist a quantum state in which a particle has simultaneously a definite position and a definite momentum. As a result, measurements of position and momentum, of course, cannot simultaneously yield definite outcomes. This has nothing to do with how the measurement is performed and how much the particle is disturbed by the device. Nevertheless, Heisenberg's estimation (8.19) and the inequality (8.1) are so close that our argument may not be so convincing to you. Therefore, we shall return to spin, with which we shall make a convincing demonstration that the uncertainty relation has nothing to do with perturbations introduced by measurement.

Let us briefly review the spin measurement in the Stern-Gerlach experiment (see Fig. 5.1). To measure  $\hat{\sigma}_x$ , we need to orient the magnetic field along the  $x$  direction; to measure  $\hat{\sigma}_z$ , we need to orient the magnetic field along the  $z$  direction. If we want to measure simultaneously  $\hat{\sigma}_x$  and  $\hat{\sigma}_z$ , we need to orient the magnetic field along both the  $x$  and  $z$  directions. But it turns out the direction of the field is actually  $\mathbf{n} = \{n_x, 0, n_z\}$ , corresponding to a measurement of  $\mathbf{n} \cdot \hat{\sigma}$ . Therefore, because of practical limitations, we cannot measure  $\hat{\sigma}_x$  and  $\hat{\sigma}_z$  simultaneously in the Stern-Gerlach experiment, which has nothing to do with perturbations by the measurement. To see more clearly, we consider a special situation: the magnetic field is oriented along the  $z$ -direction and the spin state of the silver atom is fixed at  $|u\rangle$ . In this case, the silver atom flies upwards every time, producing just one spot on the screen, and the measurement becomes completely deterministic. If the perturbation from a measurement is connected to the

uncertainty relation, it is incomprehensible as to why its effect vanishes entirely in this case.

Therefore, it is a mathematical coincidence that Heisenberg's estimate (8.19) is close to the inequality (8.1). By analyzing Fig. (8.1), Heisenberg revealed that a measurement inevitably disturbs the measured system. Accordingly, Heisenberg estimated the minimum value of the perturbation (8.19). There is nothing wrong with this estimation, but we should not link it to the Heisenberg's uncertainty relation (8.1) as they are not related in physics. If they were related, it would be manifested also in the spin measurement. However, from our analysis of the spin measurement, we do not see any connection between the uncertainty relation of spin (8.17) and the perturbation in the measurement. In conclusion, Heisenberg's uncertainty relation is a consequence of the noncommutativity of operators, which is not related to the perturbation brought about by measurement.

## 8.2 Collapse of a Wave Function

In Chap. 5, we have introduced the basic theoretical framework of quantum mechanics. One of the fundamental principles is that the results of quantum measurement are probabilistic. At that time, we have deliberately avoided an important question: what happens to a quantum state after measurement? This question is very controversial with widely varied answers among physicists. We will begin with a widely accepted view that is in most textbooks: a measurement on a quantum state results in the “collapse of a wave function”. The discussion of wavefunction collapse can be found in the earliest and most prestigious monographs on quantum mechanics, *Principles of Quantum Mechanics* by Dirac and *Mathematical Foundations of Quantum Mechanics* by von Neumann (John von Neumann, 1903–1957). Both Dirac and von Neumann argued that observation and measurement cause a discontinuous change in the quantum state  $|\psi\rangle$ , i.e., the collapse of a wave function. They describe the process of collapse as follows: during a measurement of an observable  $\hat{O}$ , the quantum state  $|\psi\rangle$  is reduced with some probability to a single eigenstate of  $\hat{O}$ ,

$$|\psi\rangle \longrightarrow |\phi_j\rangle. \quad (8.20)$$

Here  $|\phi_j\rangle$  is an eigenstate of  $\hat{O}$ , and the probability of collapsing to it is  $|\langle\phi_j|\psi\rangle|^2$ .

This point of view seems to be supported by experiments. Again, let us recall the stern-Gerlach experiment (see Fig. 5.1). Suppose that a silver atom in the spin state  $|\psi\rangle = (|u\rangle + |d\rangle)/\sqrt{2}$  flies from the source to the detection screen. Before hitting the detection screen, the spin state of a silver atom is unchanged, and we cannot predict with certainty where the silver atom will hit on the screen. But once the atom hits the detection screen, it appears only in one place, the upper spot or the lower spot. If the atom lands at the upper spot, we say that the spin state of the silver atom is collapsed,

$$|\psi\rangle = (|u\rangle + |d\rangle)/\sqrt{2} \longrightarrow |u\rangle. \quad (8.21)$$

So, the collapse of the wave packet offers a reasonable interpretation of the Stern-Gerlach experiment.

But the collapse of a wave function is very puzzling. The collapse should be a physical process because it results from the interaction of two physical systems—the object and the measuring instrument. How exactly does the collapse occur? How long does it take for a wave function to collapse? Could the collapse be described by some mathematical equation? Neither Dirac nor von Neumann answered these questions; they simply described the collapse of a wave function in words without any mathematics.

In some sense dice rolling is similar to the collapse of a wave function: a dice is not in a definite state before the roll; it is in a specific state only when it stops rolling. The dice, too, “collapses” from an uncertain state to a definite state. But the “collapse” of a dice is an actual physical process. We can observe the trajectory of the dice in the air, its collision with the table, as well as its rolling on the table. A clever and competent physicist can even study this process in detail and describe it precisely (see Fig. 5.2). It is mysterious and confusing that the collapse of a wave function in quantum mechanics does not involve a physical process.

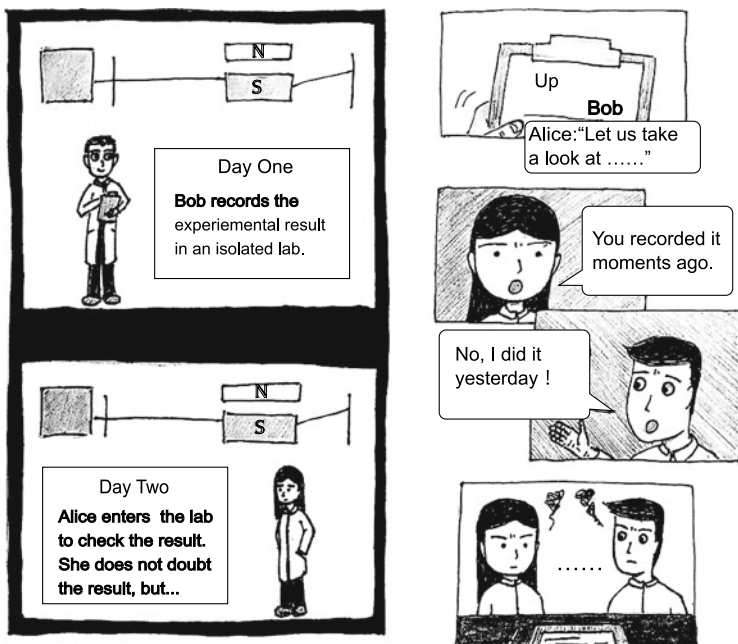
If we accept the postulate of wavefunction collapse, a quantum states evolves in time by two processes: (1) the collapse of a wave function; (2) a continuous time evolution via the Schrödinger equation (6.1) (see Chap. 6). The former is discontinuous and non-unitary; the latter is continuous and unitary. In his *Mathematical Foundations of Quantum Mechanics*, von Neumann repeatedly emphasized that the difference between these two types of changes of wave function lies in the difference of the system: the system undergoing the unitary evolution is an isolated quantum system, i.e., there is no transfer of energy or matter between the quantum system and the external world; whereas the system undergoing the collapse of a wave function interacts with measurement apparatus.

In 1956, Everett (Hugh Everett III, 1930–1982) pointed out that the postulate of wave function collapse leads to unavoidable contradiction. Let us see what it is.

Consider a laboratory that is in complete isolation from the external world, i.e., there is no energy or matter transfer between the laboratory and the external world. An experimenter, Bob, performs the Stern-Gerlach experiment in this laboratory (see Fig. 8.2). One silver atom in the spin state  $|f\rangle = (|u\rangle + |d\rangle)/\sqrt{2}$  is emitted from the particle source. Bob observes a mark in the upper part of the detection screen, and he then turns off the experiment and records this result in his notebook. The next day Alice opens the lab to check the result. She has no issue with the result, but insists that the wave function collapsed at the moment she opened the door. Bob, of course, insists that the wave function collapsed and he recorded it one day before. Let us see how this controversy emerges.

For Bob, the measured system is spin, and the measuring instrument is the apparatus such as the magnet and the detection screen. For Bob, the wave function collapses at the instant when the silver atom strikes the detection screen. But for Alice, the other experimenter, the measured system consists of all objects in the lab, including





**Fig. 8.2** The dilemma of the collapse of wave function. Alice and Bob are the two experimenters. At first, Bob's lab is completely isolated from the external world. Alice opens the door to enter Bob's lab one day after he finished his experiment

Bob; the measuring instrument is Alice's eyes. Until Alice opens the door of the laboratory, the entire laboratory is a quantum system isolated from the external world. As we have emphasized earlier, the collapse of a wave function only occurs when the measured system interacts with the measuring instrument. Therefore, for Alice, before she makes any observation, the quantum state of the whole laboratory is in a continuous evolution with a unitary matrix  $\hat{U}$

$$\hat{U}\left[\frac{1}{\sqrt{2}}(|u\rangle + |d\rangle) \otimes |\emptyset\rangle\right] = \frac{1}{\sqrt{2}}[|u\rangle \otimes |\text{Up}\rangle + |d\rangle \otimes |\text{Down}\rangle]. \quad (8.22)$$

Here  $|\emptyset\rangle$  represents no record on the notebook;  $|\text{Up}\rangle$  represents the record "spin up";  $|\text{Down}\rangle$  represents the record "spin down". So, for Alice, both "spin up" and "spin down" are possible records on the notebook. At the moment when the door is opened, Alice's eyes interact with all the objects in the lab, and the wave function collapses as

$$\frac{1}{\sqrt{2}}[|u\rangle \otimes |\text{Up}\rangle + |d\rangle \otimes |\text{Down}\rangle] \longrightarrow |u\rangle \otimes |\text{Up}\rangle. \quad (8.23)$$

This brings us to the inherent conflict mentioned earlier: Alice thinks that the record on the notebook was only determined when she opened the door; Bob, on the other hand, is absolutely certain that he recorded it down one day earlier.

Let us look at this contradiction in a different perspective. In our daily life, the interaction of a macroscopic object with another macroscopic object does not cause the collapse of a wave function. For example, when we open a notebook to check a record, our action does not cause any “collapse” or “change” of the record. The record is the same before and after we look at it. Wavefunction collapse only occurs when a microscopic system interacts with a large measuring instrument, such as a silver atom interacting with a detection screen. As the Stern-Gerlach experiment was performed one day earlier, when Alice enters the lab, all she sees is macroscopic objects, including Bob and his record. Alice’s action would not cause any collapse. In addition, for Alice, the Stern-Gerlach experiment was carried out in complete isolation. So from Alice’s point of view, no discontinuous collapse of the state of a system occurred throughout the experiment, and the whole experiment should be regarded as a continuous and deterministic unitary time evolution. After seeing “spin up” in the notebook, Alice is absolutely certain that the same result would be obtained when the experiment were repeated. Bob, on the other hand, thinks the experiment was not deterministic and there was a 50% probability the result was “spin down”. We know that Bob is correct and Alice is wrong. Alice’s misconception comes from the wavefunction collapse postulate.

In 1961, Wigner (Eugene Paul Wigner, 1902–1995) proposed a similar thought experiment,<sup>2</sup> now known as Wigner’s friend paradox. In Wigner’s thought experiment, his friend performs a quantum experiment in an isolated laboratory, where the measured object is in a superposition state  $\alpha|\psi_1\rangle + \beta|\psi_2\rangle$ . The two quantum states  $|\psi_1\rangle$  and  $|\psi_2\rangle$  predict different measurement outcomes, which trigger different perceptions in his friend. Let  $|\chi_1\rangle$  represent his friend’s perception of  $|\psi_1\rangle$ , and  $|\chi_2\rangle$  represent his perception of  $|\psi_2\rangle$ . Wigner argues that after his friend completes the measurement, the whole laboratory, including his friend, should be described by a superposition state  $\alpha|\psi_1\rangle \otimes |\chi_1\rangle + \beta|\psi_2\rangle \otimes |\chi_2\rangle$ . However, his friend thinks that the measurement caused the system to collapse into either the state  $|\psi_1\rangle \otimes |\chi_1\rangle$  or the state  $|\psi_2\rangle \otimes |\chi_2\rangle$ . Wigner and his friend’s conclusions contradict each other. Note that in this thought experiment, Wigner does not actually enter the laboratory.

In the above example, no matter how it is viewed or analyzed, whenever there are two observers (Alice and Bob, or Wigner and his friend), they arrive at contradicting conclusions or realities. Despite this inherent flaw, the postulate of wavefunction collapse remains one of the most commonly accepted, and many physicists still use it in understanding the theoretical results or explaining the experimental observations. As we saw earlier, the postulate can explain the Stern-Gerlach experiment in a simple and straightforward way. Although I do not agree to this postulate, I still use it often

---

<sup>2</sup> E. P. Wigner (1961), “Remarks on the mind-body question”, in: I. J. Good, *The Scientist Speculates* (London, Heinemann). As Everett and Wigner were both in Princeton at the time, they might had some private discussions. There is no clear evidence as to who first proposed the thought experiment; the only thing that is certain is that Everett’s study was officially published earlier.

in thinking and discussions. After all, no contradiction arises in most situations where we unconsciously regard ourselves as the only conscious observer.

If the postulate of wavefunction collapse is wrong, how do we describe the effect of a measurement on quantum states? This question remains one of the great unsolved puzzles for physicists. In the next section, we will present Everett's theory. A great physicist, when finding a flaw in the old theory, will come up with a new theory to replace the old one. Everett was one of these people, and his theory is known as the many-worlds theory.

### 8.3 Many-Worlds Theory and Schrödinger's Cat

The microscopic world described by quantum mechanics, being a radical departure from our normal experience, is mind-boggling and weird. A natural question is why we do not see or feel these amazing quantum phenomena in everyday classical world. A generally accepted answer is that Planck's constant is so small that quantum effects are negligible at the macroscopic scale. This does answer a number of questions. According to quantum mechanics, light is a collection of particles called photons, each with an energy  $h\nu$ . But in our everyday world light appears continuous, and we do not perceive individual photons. There are two reasons: (1) the energy of each photon is small because Planck's constant  $h$  is very small, and (2) ordinary light is made up of a large number of photons. As a result, our impression of light is that it is continuous. This is very similar to water. The water that we experience in daily life is made up of a lot of very small water molecules, giving us the impression that water is continuous. This is in marked contrast to a beach, which, although consisting of many grains of sand, is grainy because the grains of sand are large.

According to Heisenberg's uncertainty relation, a particle cannot simultaneously have definite momentum and position. However, because Planck's constant is very small, we usually do not feel this quantum effect. Human eyes have about 0.01 m/s speed resolution and 0.1 mm position resolution. If the mass of the object is about 1 g, for human eyes, we have approximately  $\Delta p \Delta x \sim 10^{-9}$  J·s, which is about 24 orders of magnitude larger than Planck's constant  $6.62607004 \times 10^{-34}$  J·s. As a result, any object in our everyday world appears to have simultaneously determined instantaneous position and momentum, and we do not feel any quantum effects associated with the uncertainty relation.

But there are some quantum effects or phenomena, such as the superposition principle of quantum states and quantum entanglement, are completely independent of Planck's constant. The superposition principle builds on two fundamental elements of quantum mechanics: Hilbert space and the Schrödinger equation. A Hilbert space is a linear space. Therefore, the sum of any two vectors in a Hilbert space is also a vector in this Hilbert space. This means that the superposition of any two quantum states is still a valid quantum state, which has nothing to do with Planck's constant. In addition, because the Schrödinger equation is a linear equation, a linear

superposition of two solutions of the Schrödinger equation is still a solution of the equation. This property seems to be related to Planck's constant, for the Schrödinger equation contains Planck's constant. But the Schrödinger equation is linear regardless Planck's constant is small or large. To sum up, we arrive at an interesting and profound conclusion: the superposition principle of quantum states does not rely on Planck's constant. As we discussed in Chap. 6, a consequence of the superposition principle is that a single electron can pass through both slits simultaneously in the interference experiment. Given that the superposition principle has nothing to do with Planck's constant, how come we cannot go to work by simultaneously taking two different paths?

Quantum entanglement is independent of Planck's constant as well. Let us review the origin of quantum entanglement with a system of two spins. Suppose the Hilbert space of spin 1 is  $V_1$  and the Hilbert space of spin 2 is  $V_2$ . The Hilbert space of the composite system of these two spins is their direct product  $V = V_1 \otimes V_2$ . The dimension of Hilbert space  $V$  is 4 and there are four orthonormal basis vectors  $|uu\rangle$ ,  $|ud\rangle$ ,  $|du\rangle$ ,  $|dd\rangle$ . Using this set of basis, we can construct an entangled state, such as the spin singlet state  $|S\rangle$ . Here Planck's constant is not involved throughout. For a composite quantum many-particle system, the associated Hilbert space is a direct product of the Hilbert space of individual particles in the system. This is the origin of quantum entanglement, which does not involve Planck's constant at all. It is puzzling that quantum entanglement disappears in our everyday world, where every object has its well defined position and momentum, never losing its individuality.

A macroscopic object is made up of microscopic particles—atoms and molecules. So why atoms and molecules can interfere and become entangled, but the soccer balls and glasses that they make up can not? This puzzling problem is inexplicable with Planck's constant being small, for neither interference nor entanglement is related to Planck's constant. The most widely accepted view is the Copenhagen interpretation, which is based on the postulate of wavefunction collapse.<sup>3</sup>

The Copenhagen interpretation makes a brutal assumption that the macroscopic world that we live in is classical in nature, where quantum phenomena disappear. When a macroscopic measuring instrument interacts with the measured quantum system, the wave function collapses, naturally eliminating the linear superposition (or interference) and entanglement between the macroscopic and microscopic systems. This can be clearly seen from Eq. (8.23): the left side of the equation is a linear superposition of quantum states describing an entanglement between Bob (macroscopic) and spin (microscopic); the right side of the equation contains only one component, which is a product state and has no superposition and entanglement.

---

<sup>3</sup> The Copenhagen interpretation of quantum mechanics is in fact a complete theory, including the basic principles of quantum mechanics we have introduced in Chap. 5. The vast majority of it are not controversial. Here we only discuss its controversial parts: (1) the distinction between the classical and quantum worlds; (2) the collapse of a wave function.

The Copenhagen interpretation allows simple and straightforward explanations that are in excellent agreement with existing experimental observations. However, it is full of ambiguous and heuristic assumptions. The Copenhagen interpretation asserts macroscopic objects are classical without any explanation. The Copenhagen interpretation also fails to specify a boundary between macroscopic and microscopic systems. A copper atom is a microscopic system, and we describe it with quantum mechanics. Two copper atoms, too, are microscopic and described by quantum mechanics. Ten copper atoms remain microscopic, so we continue to employ quantum mechanics. But what about a thousand copper atoms? Or 10,000 copper atoms? Are these objects macroscopic or microscopic? Are they described by quantum mechanics or classical mechanics? The answer becomes less clear. But for a small ball made up of about  $10^{23}$  copper atoms, once again, we have a clear-cut answer: when the ball is thrown, we can determine its trajectory precisely using classical mechanics. However, it is not clear at all as to where exactly is this boundary between a single copper atom, which behaves quantum mechanically, and a copper ball, which behaves classically. This is the inherent flaw of the Copenhagen interpretation. The problem is in fact even more serious than we have just described. Although a small copper ball should be treated as a macroscopic object by any standard, not all of its properties are compatible with classical mechanics. Condensed matter physicists will tell you that to understand the conductivity of small copper balls requires quantum mechanics. Condensed matter physics is the field of physics that deal with the properties of matter, such as why some materials conduct electricity and others do not, and why some materials are hard and others are brittle. Why do we need quantum mechanics to explain the conductivity of a copper ball? Because an electron has such a small mass that its Fermi temperature is much higher than the room temperature. What is the Fermi temperature? Why only electrons copper atoms are taken into account, whereas the nuclei are not considered? To address these issues is beyond the scope of this book. If you are interested in these questions, please look up textbooks on condensed matter physics books.

Another centerpiece of the Copenhagen interpretation, the postulate of wavefunction collapse, is equally ambiguous and heuristic: the collapsing process cannot be described mathematically; it leads to logical contradictions as we have shown in detail earlier.

Over the years, there have been many objections to various aspects of the Copenhagen interpretations, and many new theories have been proposed. One of them is the many-worlds theory. In my opinion, the many-worlds theory is the best and most self-consistent interpretation of quantum mechanics to date. In addition, many-worlds theory can naturally explain why there is no quantum superposition and entanglement in our daily life.

In 1956, Everett (Hugh Everett III, 1930–1982) proposed the many-worlds interpretation of quantum mechanics in his doctoral dissertation, solving the problems encountered by the Copenhagen interpretation: (1) the intrinsic logical inconsistency and (2) the ambiguous quantum-classical boundary. His theory had not been well-recognized for a long time but Everett appeared not bothered by it very much. He graduated from a not well-known university in 1953 with a degree in chemical engineering. Then Everett started his graduate studies in Princeton University, where he was initially interested in game theory. Later he got interested in quantum physics and chose Wheeler (John Archibald Wheeler, 1911–2008) as his advisor. In 1956, Everett completed his doctoral dissertation entitled “The Theory of the Universal Wave Function”, and finished his doctoral defense in 1957. He then took a job outside academics as an analyst in the Department of Defense of US, and later established several corporations. He never published any paper in physics since.



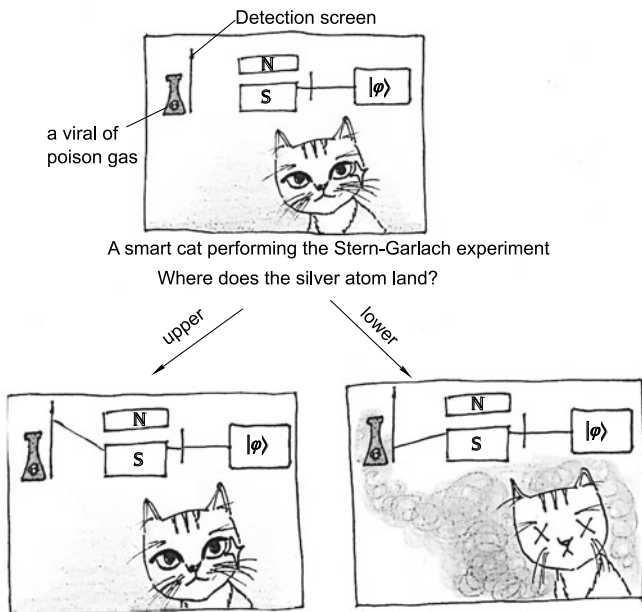
Everett (1930-1982)

The theory that Everett proposed in his doctoral dissertation is the many-worlds theory. Even with the support of his supervisor Wheeler, Everett's theory was not quickly accepted. For example, in 1959, with Wheeler's arrangement, Everett visited Copenhagen to explain his theory to Bohr. However, Bohr completely rejected the young man's new theory. He and his followers dismissed Everett's theory as “heresy”. But a good theory always finds a way to survive the initial cold shoulder. In the 1970s, DeWitt (Bryce DeWitt, 1923–2004) reintroduced Everett's theory to the physics community and named it many-worlds theory.<sup>4</sup> The many-worlds theory is now accepted by growing number of physicists, and even appeared in science fictions. Everett was in poor health in his later life due to heavy drinking and died of a heart attack at the age of 51.

Everett's doctoral dissertation has more than 100 pages and is full of mathematical formulas. Everett first formulated a quantum theory of composite systems, where he discussed systematically quantum entanglement between different parts of a composite system. However, Everett did not use the term entanglement, instead he called it canonical correlation. Based on this theory of composite system, he then established his many-worlds theory and discussed extensively how his theory was consistent with all the experiments and our daily experience. We now illustrate the essence of Everett's many-worlds theory with Schrödinger's cat.

---

<sup>4</sup> Note: Everett himself did not call his theory the many-worlds theory.



**Fig. 8.3** The many-worlds theory of Schrödinger’s cat. The clever cat is performing the Stern-Gerlach experiment. If a silver atom hits the upper screen, nothing happens; if it hits the lower screen, it triggers a chain reaction that eventually shatters the vial and releases the poison gas inside. At the end of the experiment, two worlds appear: one in which the cat is alive, and the other in which the cat is dead. Note that there is only one cat from the beginning to the end: before the experiment the cat has only one state—alive; after the experiment the cat has two states—alive or dead

Consider a laboratory that is completely isolated from the external world. In the lab, there is a clever cat performing the Stern-Gerlach experiment, and near the lower part of the detection screen, there is a vial of poison (see Fig. 8.3). When a silver atom hits the lower screen, it triggers a chain of reactions that eventually shatter the vial and release the poison. The cat fires a silver atom in the spin state  $|f\rangle = (|u\rangle + |d\rangle)/\sqrt{2}$  from the source. Before this silver atom hits the detection screen, the state of the laboratory can be described by the following quantum state,

$$|\Phi_0\rangle = \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle) \otimes |\text{live cat}\rangle. \tag{8.24}$$

When the silver atom hits the detection screen, there are two possible outcomes: if the silver atom hits the upper part, nothing happens and the cat remains alive; if the silver atom hits the lower part, it triggers a chain reaction and releases the poison, which kills the cat. According to Everett’s theory, this is when the silver atom and the cat become entangled, and the state of the lab becomes

$$|\Phi_1\rangle = \frac{1}{\sqrt{2}}(|u\rangle \otimes |\text{live cat}\rangle + |d\rangle \otimes |\text{dead cat}\rangle). \quad (8.25)$$

The two components of this wave function represent two parallel worlds: one in which the cat is alive; the other in which the cat is dead (see Fig. 8.3). These two worlds are equally real and exist in parallel. It is important to note that the lower part of Fig. 8.3 depicts the two states of the same system—the laboratory, instead of a cat splitting into two cats—one alive and one dead. This is the same as  $|f\rangle = (|u\rangle + |d\rangle)/\sqrt{2}$ , which describes one silver atom that has two states, spin-up and spin-down, instead of two silver atoms, one with spin up and the other with spin down.

Many-worlds theory interprets measurement as a process that creates entanglement between the observer (or measuring device) and the quantum system, with each component of the entangled wave function describing one of many possible realities. The wave function in Eq. (8.25) indicates that entanglement occurs between the cat and the spin. Each of its two components represents a real world: one world in which the cat is alive; the other in which the cat is dead. Since the Schrödinger equation is linear, these two worlds evolve in parallel without affecting each other with one world feels no presence of the other.

The thought experiment in Fig. 8.2 makes it apparent that the collapse postulate cannot explain the situation involving two observers in a self-consistent manner. Let us revisit this thought experiment with the many-worlds theory. We will see that the many-worlds theory does not lead to logical contradiction. For this experiment, the initial state of the system was

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle) \otimes |B^0\rangle \otimes |A^0\rangle. \quad (8.26)$$

Here  $|A^0\rangle$  and  $|B^0\rangle$ , respectively, describe the initial state of the two experimenters, Alice and Bob. Before the silver atom strikes the screen, they both have no idea what the result would be. After Bob completes the experiment, the system becomes

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}}[|u\rangle \otimes |B^u\rangle + |d\rangle \otimes |B^d\rangle] \otimes |A^0\rangle. \quad (8.27)$$

Similar to Schrödinger's cat being entangled with the silver atom, here an entanglement is generated between Bob and the silver atom.  $|B^u\rangle$  represents that Bob observed the silver atom hitting the upper screen, and  $|B^d\rangle$  represents that Bob observed the silver atom hitting the lower screen. The next day, when Alice opens the door and enters the laboratory, she sees the experimental result, and gets entangled with both the silver atom and Bob. The state of the system becomes

$$|\Psi_2\rangle = \frac{1}{\sqrt{2}}[|u\rangle \otimes |B^u\rangle \otimes |A^u\rangle + |d\rangle \otimes |B^d\rangle \otimes |A^d\rangle]. \quad (8.28)$$



Here  $|A^u\rangle$  and  $|A^d\rangle$  describe the two states of Alice: seeing “spin up” and seeing “spin down”.

We have already pointed out the weakness of the Copenhagen interpretation of this experiment. In contrast, Everett’s many-worlds theory is evidently self-consistent. Alice and Bob both agree that Bob recorded the measurement outcome one day earlier. Before Alice opens the door, there are two parallel worlds: one world in which Bob recorded “spin up”, and the other world in which Bob recorded “spin down”. Alice lives in both worlds, and her state is the same in both worlds unaware of what Bob recorded. After opening the door, there are still two worlds, but Alice’s state becomes different in these two worlds: in one world she saw “spin up” in Bob’s notebook; in the other world she saw “spin down” in Bob’s notebook. Everett’s many-worlds theory does not contain logical inconsistency. So far no one has been able to design a thought experiment that shows that many-worlds theory is logically contradictory.

In Everett’s theory, there is no distinction between the quantum and classical worlds. All the systems including silver atoms, detection screens, cats, and experimenters, regardless of how large or small they are, are treated equally as quantum objects, which can be in a superposition state or entangled with one another. Different components in a superposition state represent different parallel worlds.

In Chap. 6, we have discussed in detail the wave functions in Fig. 6.4a. These wave functions are spread out in the entire box, meaning a single ball can be simultaneously on the left and right halves of the box. In contrast, in our everyday world, we have never seen a soccer ball appear in both the left and right halves of the soccer field. Why? According to the Copenhagen interpretation, soccer balls are macroscopic classical objects, and they can not be in a superposition state. This is just a statement, not an explanation at all. The many-worlds theory instead offers a more natural and reasonable explanation. According to this theory, both the soccer ball and the whole world are quantum objects, so that it is possible to have the following quantum state

$$|\Psi_{\text{Univ}}^0\rangle = \frac{1}{\sqrt{2}} (|L\rangle + |R\rangle) \otimes |\Psi_{\text{Env}}\rangle, \quad (8.29)$$

where  $|L\rangle$  is for the state where the ball is on the left field,  $|R\rangle$  the state where the ball is on the right field, and  $|\Psi_{\text{Env}}\rangle$  the state associated with objects other than the ball. The above wave function represents the state where soccer ball can be both on the left and right halves of the field, but other objects do not feel any difference. Such a quantum state, even if successfully prepared, would only persist for a very short time. There is a significant physical difference between a soccer ball on the left and on the right. If the ball is on the left, it attracts more players to the left field; if the ball is on the right, it attracts more players to the right field. In other words, two states  $|L\rangle$  and  $|R\rangle$  induce different physical responses from the environment. Even there are no players on the field, the grass on the field will feel the difference: when the ball is at a different location, a different piece of grass will be bent. In any case, the soccer ball is a big object. When it appears at a place, its ambient environment

immediately feels its physical presence and gets entangled with it. Therefore, in a very short time, the above wave function becomes

$$|\Psi_{\text{Univ}}\rangle = \frac{1}{\sqrt{2}} (|L\rangle \otimes |\Psi_{\text{Env,L}}\rangle + |R\rangle \otimes |\Psi_{\text{Env,R}}\rangle). \quad (8.30)$$

The world is split into two: in one world, we see the ball on the left field; in the other world, we see the ball on the right field.

The electron in a hydrogen atom, being confined to a very small space, has a very different fate from the soccer ball. Assuming that the hydrogen atom is in the ground state, then the wave function of the whole system can be written as

$$|\Phi_{\text{Univ}}^0\rangle = |\phi_s\rangle \otimes |\Phi_{\text{Env}}\rangle, \quad (8.31)$$

where  $|\phi_s\rangle$  is the ground state of the hydrogen atom and  $|\Psi_{\text{Env}}\rangle$  describes objects other than the hydrogen atom. According to Fig. 6.5, the wave function  $|\phi_s\rangle$  spreads over the space: the electron shows up everywhere around the proton simultaneously. But since  $|\phi_s\rangle$  is confined in a small space (the radius is only about  $0.53 \times 10^{-10}$  m), the electron, regardless being on the left or right of the proton, does not affect other objects. Equivalently, we say that the environment cannot distinguish if an electron is on the left or right. To be specific, we illuminate a hydrogen atom with a beam of visible light. Since the wavelength of the visible light is about several hundred nanometers, which is much larger than the radius of a hydrogen atom  $0.53 \times 10^{-10}$  m, the visible light is unable to resolve the position of electrons. On the other hand, the ground state electron can be excited to the  $p$  orbital of the hydrogen, absorbing a photon from the visible light. In this case, there is a change in the environment and the whole wave function becomes

$$|\Phi_{\text{Univ}}\rangle = c_1 |\phi_s\rangle \otimes |\Phi_{\text{Env},s}\rangle + c_2 |\phi_p\rangle \otimes |\Phi_{\text{Env},p}\rangle, \quad (8.32)$$

where  $|c_2|^2$  is the excitation probability,  $|\phi_p\rangle$  is the  $p$  orbital wave function. The world is again split into two: a world with a hydrogen atom in the ground state, and a world with a hydrogen atom in the excited state. However, in both worlds, the electron position cannot be precisely determined, because both wave functions  $|\phi_s\rangle$  and  $|\phi_p\rangle$  are spread out in space (see Fig. 6.5). In fact, how to locate an electron in a hydrogen atom is a topic in the forefront of physics research.

According to Everett's theory, you can work in the office and stay at home simultaneously. All you need to do is to perform a Stern-Gerlach experiment in the office: if the silver atom is detected on the upper screen, you stay in the office and if it is detected on the lower screen, you stay at home. At the end of the experiment, you can work in the office and rest at home simultaneously, except that they happen in different worlds, and no one will notice it, including yourself as these two worlds are parallel and do not affect each other physically.

Now it is clear that the macroscopic world, too, can exhibit superposition and entanglement, but we can not feel or perceive them. We can only perceive one parallel

world and be in one parallel world, where there is no superposition and entanglement. For example, the first component of Eq. (8.28) represents a parallel world in which the quantum state is not a superposition state but rather a product state. DeWitt once wrote to Everett and asked, “Why can’t I feel parallel worlds?” Everett replied, “Can you feel the rotation of the earth?” DeWitt was convinced and became an ardent advocate of the many-worlds theory.

There is no electron or spin in our daily experience, though. So it is relatively easy to accept that an electron can travel through both slits, or that a spin can be both up and down simultaneously. This is similar to the psychology behind some people’s perception of ghosts. Because there are no ghosts in life, many people are prone to accept the existence of ghosts, no matter how ridiculous their powers and actions are. But when you say that an ordinary cat can be both alive and dead simultaneously, people immediately regard it ridiculous. We are all too familiar with cats: a live cat and a dead cat must be two different cats; one cat must be either alive or dead. We have never seen a cat that is both alive and dead. If this is why you find the many-worlds theory hard to accept, the cause is not scientific or logical, but rather psychological.

The many-worlds theory is quite stimulating: what am I doing in another parallel world? Do the alive and dead states of the same cat exist in the same space-time? Is there a superposition of space-time? It is foreseeable that many-worlds theory may inspire physicists to think more deeply about the nature of quantum and spacetime, leading us to solutions of the measurement problem, and even the unification of quantum theory and gravity.

The many-worlds theory reconciles the seeming conflict between the free will<sup>5</sup> of a human being and the determinism of physical laws. As mentioned in Chap. 3, classical mechanics and quantum mechanics have a crucial distinction. In classical mechanics, once a complete and exact specification of the initial state of a particle is provided, the future state of the particle is completely determined. In quantum mechanics, however, the outcome of any measurement is probabilistic. As a quantum particle inevitably interacts with the external world, so the future state of a quantum particle is uncertain. Since the world is made up of microscopic particles that obey quantum mechanics, naturally, we may infer that every one of us, as well as the society, has an uncertain future; any random event in the microscopic world might change our thoughts and actions at present, and therefore, change our future. This amounts to saying that every individual person has a free will. But some physicists might argue against this conclusion. They would say that the Schrödinger equation is deterministic: once the initial wave function is given, the evolution of the wave function is completely deterministic. So, if the whole universe is described by a giant wave function, then this wave function evolves deterministically according to the Schrödinger equation,<sup>6</sup> the future of the universe is thus completely determined because there are no people or other objects outside the universe to observe the

---

<sup>5</sup> There are many definitions of free will. Here free will is defined as the inability of a person to infer his previous states from his current state.

<sup>6</sup> Strictly speaking, it evolves according to the equations in quantum field theories; the Schrödinger equation is used here only for convenience.

universe. The many-worlds theory provides a solution for this problem: although the whole cosmic wave function evolves deterministically, it contains an infinite number of parallel worlds; every person can only perceive one parallel world, and cannot control which parallel world he or she will be in. The many-worlds theory allows everyone to have a free will in a deterministically evolving universe.

Both the wavefunction collapse postulate and the many-worlds theory concern a measurement on a quantum system. According to the former, a measurement collapses the quantum system into certain eigenstates; according to the latter, the measuring apparatus gets entangled with the quantum system during the measurement. So, both theories agree on at least one thing: the quantum system is changed unavoidably by the measurement. Is this effect related to Heisenberg's uncertainty relation discussed earlier? No. Because Heisenberg's uncertainty relation involves two different types of measurements, but here we are discussing the influence generated by a single measurement.

In my opinion, Everett was the first physicist who completely and utterly break free from the shackles of classical physics. In Chap. 2, we have described the revolutionary development of quantum mechanics by physicists such as Planck, Einstein, Bohr, Heisenberg, and Schrödinger. From 1900 to 1926, these pioneers, motivated by the indisputable experimental data, established quantum mechanics which was a radical departure from classical mechanics. After 1926, physicists, on the one hand, continued with the development of quantum theory, and on the other hand, began to think more deeply about the meaning of quantum mechanics and its relation to the everyday world. The former has led to the achievement in fields such as particle physics and condensed matter physics, while the latter has led to various debates, such as the debate between Einstein and Bohr, and the debate on quantum measurement. In these debates, although various viewpoints were proposed, they have a common feature: classical mechanics was indispensable for quantum theory to be logically self-consistent or quantum mechanics needs an interpretation with classical mechanics. The physicists represented by Einstein considered quantum mechanics as an approximation to some classical hidden variable theory. The Copenhagen interpretation, principally attributed to Bohr and Heisenberg, insisted that the measuring apparatus be classical. Von Neumann and Wigner went even further and argued that human consciousness played a decisive role in the measurement. Wigner's thought experiment was intended to illustrate the importance of consciousness in measurement instead of criticizing the Copenhagen interpretation. De Broglie presented a pilot wave theory, which was later developed by Bohm (David Bohm, 1917–1992) and others into Bohmian mechanics. This theory attempted to interpret quantum mechanics with classical physics. Finally, Everett bravely stepped forward and declared in his doctoral thesis that quantum mechanics is by itself self-consistent, which does not require classical mechanics at all. It is not quantum mechanics that needs to be interpreted with classical mechanics; we should do the reverse, explaining classical mechanics with quantum physics. History will remember Everett as a brave thinker, who, for the first time ever, completely abandoned concepts of classical physics in principle.

Unfortunately, Everett's many-worlds theory is still a minority view among physicists. However, the number of its supporters has been increasing. Even some of those who initially supported the Copenhagen interpretation have switched to support the many-worlds theory. On the other hand, supporters of the many-worlds theory never switch to the Copenhagen interpretation. It is a one-way road.

## 8.4 Flaws in Feynman's Argument

In Chap. 6, we discussed the double-slit interference of electrons. In the experiment, the beam of electrons diffract through the two slits and form interference pattern on the detection screen. Many famous physicists, including Einstein and Wheeler, have deeply thought through the double-slit interference experiment as it is a simple and yet spectacular demonstration of the mysterious superposition of states in quantum mechanics: a single particle can show up in two different positions like a wave. Here we discuss how to make the interference disappear by quantum measurements. These discussions will further help us understand quantum measurements.

Feynman (Richard Phillips Feynman, 1918–1988) discussed how to make the double-slit interference pattern disappear (see *The Feynman Lectures on Physics* Volume I). Let us first review Feynman's illuminating discussion. Feynman considered a variation of the double-slit interference experiment: a light source is positioned between the two slits (see Fig. 8.4). The light source emits photons continuously. If an electron passes nearby, it will collide with the photons. Since the collision between photons and electrons can be experimentally detected, it is possible to determine precisely where the collision occurs, allowing us to determine which slit an electron has passed, slit  $s_1$  or slit  $s_2$ . Of course, the experiment must be done very carefully, and the change in the electron momentum due to the collision should be minimized. We leave these details to the experimentalist. In the following discussion, we simply assume that the light source only acts to tell us which slit the electrons have passed through, and it has no other effects.



**Fig. 8.4** Feynman's double-slit interference experiment. The light bulb between the double slits schematically represents a light source, which can distinguish which slit the electrons have passed through and thus causes the interference pattern to disappear

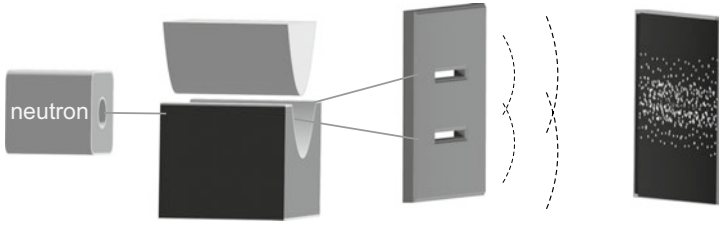
For simplicity, we reduce the intensity of the electron beam so that the electrons pass through the two slits one at a time. As in Chap. 6, we will focus on the electrons detected by detector  $d_5$  (see Fig. 6.8). Whenever an electron is detected at  $d_5$ , due to the light source, we know which slit the electron comes from. Therefore, we can divide the electrons detected by  $d_5$  into two groups: group A containing electrons passing through slit  $s_1$  and group B containing electrons passing through slit  $s_2$ . If a total of  $N$  electrons are recorded by all the detectors, there are roughly  $N/2$  passing through each of the two slits. Therefore, for detector  $d_5$ , there are approximately  $N|a_5|^2/2$  electrons in group A and  $N|b_5|^2/2$  electrons in group B. By virtue of symmetry, we have  $a_5 = b_5$ , and the number of electrons detected at  $d_5$  is approximately  $N|a_5|^2$ . If the light source is absent, as discussed in Chap. 6, the number of electrons detected at  $d_5$  is approximately  $N|a_5 + b_5|^2/2 = 2N|a_5|^2$ . This means that when we find out which slit an electron passes, the interference pattern will mysteriously disappear.

Feynman continued with his fascinating discussion on why the interference disappears. His conclusion was that photons from the light source cause a non-negligible perturbation to an electron, and that this perturbation erases the interference effect. Feynman argued that the collision between a photon and an electron perturbs the momentum of electron, and affects the interference. The magnitude of this perturbation is roughly the magnitude of the momentum of the photon  $\hbar k$  ( $k = 2\pi/\lambda$ ). In order to minimize this perturbation, we must use a photon with a very long wavelength  $\lambda$ . However, if the wavelength  $\lambda$  is so long that it far exceeds the distance between the two slits, the photon will no longer be able to distinguish which slit an electron passes, and as a result, the interference persists. Feynman's final statement is that the perturbation caused by an "efficient" measurement will make the interference disappear; if the measurement is sufficiently "gentle", the perturbation is too weak to affect the interference. Feynman's argument is reminiscent of Heisenberg's analysis of Fig. 8.1. Indeed, they are physically equivalent. Feynman's analysis is rather misleading: it seems to suggest that in order to eliminate the interference, we must effectively perturb the momentum of an electron. This is not true. In the following we shall discuss another variation of the double-slit interference experiment in which the momentum of the particle is not perturbed at all, but the interference still vanishes.

Our key revision of the double-slit interference experiment is to add a Stern-Gerlach-type magnetic field (see Fig. 8.5). In addition, to avoid the Lorentz force, we replace electrons by neutrons which carry no charge. All the neutrons are in the spin state  $|f\rangle = (|u\rangle + |d\rangle)/\sqrt{2}$ . After traveling through the non-uniform magnetic field, the neutron stream splits into two beams. The horizontal position of the two slits and the distance between the two slits are carefully tuned such that the two beams pass through the two slits, respectively. Is there any interference pattern? The answer is no. Let us see why.

The time evolution from the neutron source to the double-slit can be described as

$$|\psi_0\rangle \otimes \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle) \longrightarrow |\psi_1\rangle \otimes |u\rangle + |\psi_2\rangle \otimes |d\rangle. \quad (8.33)$$



**Fig. 8.5** Stern-Gerlach double-slit interference experiment. A neutron beam in a given spin state splits into two beams after traveling through a non-uniform magnetic field. The two beams pass through the two slits, respectively. Although it is uncertain which slit a single neutron passes, the interference vanishes. The replacement of electrons by neutrons is to avoid the Lorentz force experienced by a charged particle. In the absence of a magnetic field, the neutron beam, like the electron beam, can interfere, producing interference fringes on the screen

Here  $|\psi_0\rangle$ ,  $|\psi_1\rangle$ ,  $|\psi_2\rangle$  are similar to the wave functions in Eqs. (6.36, 6.37 and 6.38):  $|\psi_0\rangle$  describes the state of neutrons before neutrons passing the slits and  $|\psi_1\rangle$ ,  $|\psi_2\rangle$  the state right after passing the slits. The equation to the right of the arrow indicates that a neutron passing through slit  $s_1$  is in the spin up state while a neutron passing through slit  $s_2$  is in the spin down state. In other words, the spatial position of the neutron is entangled with its spin. There is no such entanglement in the double-slit interference in Chap. 6 (see Fig. 6.8). This is the key difference. After passing through the double-slit, the time evolution is again similar to the previous double-slit interference experiment. So we directly write down the wave function when it reaches the detector

$$|\psi_1\rangle \otimes |u\rangle + |\psi_2\rangle \otimes |d\rangle \longrightarrow \sum_{j=1}^9 \left[ a_j |d_j\rangle \otimes |u\rangle + b_j |d_j\rangle \otimes |d\rangle \right]. \quad (8.34)$$

The probability at the detector at  $d_j$  is

$$\begin{aligned} & \left( \langle u|a_j^* + \langle d|b_j^* \right) \left( a_j |u\rangle + b_j |d\rangle \right) \\ &= |a_j|^2 + |b_j|^2 + a_j^* b_j \langle u|d\rangle + a_j b_j^* \langle d|u\rangle = |a_j|^2 + |b_j|^2. \end{aligned} \quad (8.35)$$

As the two spin states are orthogonal,  $\langle u|d\rangle = \langle d|u\rangle = 0$ , the interference terms  $a_j^* b_j$  and  $a_j b_j^*$  vanish. Therefore, different from the double-slit experiment in Fig. 6.8, we do not observe the interference pattern in a double-slit experiment with a Stern-Gerlach magnetic field. Note that we have been sloppy in the above discussion about the normalization of the quantum states as the normalization does not play any essential role in the physics.

In this interference experiment, we do not measure the position of a neutron to see which slit it has passed. As a result, there is no perturbation to the neutron's momentum. Yet still, the interference pattern disappeared. Was the great Feynman

wrong? Strictly speaking, we cannot say that Feynman was wrong as his discussion was qualitative. But we can say for sure that Feynman's analysis did not capture the essence of the measurement problem. Feynman was limited in his analysis, failing to consider situations illustrated by the experiment in Fig. 8.5. As a result, technically, he failed to see the possibility to tell which slit a neutron passes without perturbing its momentum; conceptually, he failed to see the close relation between measurement and entanglement indicated in Eq. (8.33).

We can further improve the experiment in Fig. 8.5 by replacing its detectors. These new detector can not only response to the arrival of a single neutron but also can detect the neutron spin along the z-direction. If it detects spin up, the neutron has passed through slit  $s_1$ ; if it detects spin down, the neutron has passed through slit  $s_2$ . Thus we can detect which slit a neutron passes without perturbing the neutron momentum. Feynman apparently did not see this possibility. As far as I know, no one seems to have considered before a double-slit interference experiment combined with the Stern-Gerlach apparatus. But similar experiments are not difficult to construct. For example, similar double-slit interference experiments can be performed with light. We can use a special crystal, such as calcite, to divide a laser beam into two beams: one horizontally polarized and one vertically polarized (see Fig. 10.1 in Chap. 10), each passing through a slit. In this case, the interference pattern will disappear as well. For whatever reason, Feynman failed to see this kind of possibility.

In essence, quantum measurement is a process which generates quantum entanglement between the measuring instrument and the measured object. This is a natural physical process that does not necessarily involve human activity, nor does it require a macroscopic measuring instrument. Let us use the previous double-slit interference experiment to illustrate.

In the Stern-Gerlach double-slit experiment, when a neutron travels through the two splits, an entanglement is created between its spatial degrees of freedom and spin degrees of freedom. This entangled state is described by the following expression

$$|\psi_1\rangle \otimes |u\rangle + |\psi_2\rangle \otimes |d\rangle, \quad (8.36)$$

which is the right hand side of Eq. (8.33). This entanglement accomplishes a measurement, where the measured quantity is the neutron's position and the measuring device is the neutron's spin. After this measurement, neutrons passing through slit  $s_1$  are labeled by  $|u\rangle$ , the spin up state; neutrons passing through slit  $s_2$  are labeled by  $|d\rangle$ , the spin down state. As a result of this 'measurement', the interference disappears. It is worth emphasizing that in this measurement, neutron's spin acts as a measuring instrument, which is not macroscopic and classical from any perspective. Moreover, this measurement does not require an experimenter to make any record in a notebook.

Finally, let us look at Feynman's double-slit interference experiment from the perspective of entanglement. According to Feynman, there is a measurement process in his double-slit interference thought experiment: by detecting the collision between a photon and an electron, we can tell which slit an electron passes; as a result of this measurement, the interference pattern will disappear. In our perspective, this



measurement is essentially an entanglement between electrons and scattered photons. If we denote by  $|\text{top photon}\rangle$  the photon scattered by electrons passing through slit  $s_1$  and by  $|\text{bottom photon}\rangle$  the photon scattered by the electrons passing through slit  $s_2$ , the entangled state of electrons and photons can be written as

$$|\psi_1\rangle \otimes |\text{top photon}\rangle + |\psi_2\rangle \otimes |\text{bottom photon}\rangle. \quad (8.37)$$

This is the entanglement manifested here that makes the interference disappear, not the disturbance to the electron's momentum as argued by Feynman. If the wavelength  $\lambda$  of the photon is so long that the photon will not be able to distinguish which slit an electron passes, this means that we will not be able to tell the electron's position by detecting photons. In other words, all the scattered photons are the same, so the above equation should be changed into

$$(|\psi_1\rangle + |\psi_2\rangle) \otimes |\text{photon}\rangle. \quad (8.38)$$

This represents a product state, with no entanglement between electron's position and photon's state. This measurement, therefore, is not effective, and the interference pattern will not disappear.

To sum up, at the heart of quantum measurement is entanglement. The measuring apparatus can be microscopic, and the measurement process does not require any human intervention. This is in sharp contrast to the postulate of the wave function collapse, where the measuring apparatus must be macroscopic and the measurement outcome must have a macroscopic record. All these are essentially not necessary. In previous discussions, we saw that neutron spin not only plays the role of a measuring instrument, it also encodes the information about the measurement outcome in the microscopic state.

With this in mind, it is interesting to revisit the Stern-Gerlach experiment in Fig. 5.1, which aims to measure the spin state of a silver atom. Similar to the double-slit interference experiment in Fig. 8.5, the spin measurement is achieved through entanglement, i.e.,

$$|\psi_1\rangle \otimes |u\rangle + |\psi_2\rangle \otimes |d\rangle, \quad (8.39)$$

except that the roles of the spatial degrees of freedom and the spin degrees of freedom of a silver atom is switched in the present scenario: the measured object is atom's spin while the measuring instrument is atom's position. The spin up state  $|u\rangle$ , labeled by  $|\psi_1\rangle$ , represents a silver atom flying upward, which is recorded by the upper spot on the detection screen; the spin down state  $|d\rangle$ , labeled by  $|\psi_2\rangle$ , represents a silver atom flying downward, which is recorded by the lower spot on the detection screen.

# Chapter 9

## Quantum Computation



The concept of quantum computation was first suggested in the early 1980s. In 1980, the American physicist Benioff (Paul A. Benioff, 1930–2022) proposed an implementation of classical computer with quantum systems. In the same year, the Russian mathematician Manin (Yuri Ivanovitch Manin, 1937–2023) wrote a book entitled *Computable and Incomputable*<sup>1</sup>, in which he pointed out the essential difficulties of using classical computers to compute quantum mechanical problems (see Sect. 9.1.2). Manin suggested a quantum machine, which “use only the most general quantum principles” and whose “evolution is a unitary rotation in a finite-dimensional Hilbert space”. In 1981, Feynman independently suggested using “a suitable class of quantum machines” to “imitate any quantum system”. In 1985, Deutsch proposed the first universal, albeit abstract, model of a quantum computer, the quantum Turing machine. But their ideas had garnered not much attention for a long time.

A breakthrough occurred in 1994, when Shor (Peter Williston Shor, 1959–) discovered that a quantum computer can be much faster than a classical computer in factoring integers. In technical terms, Shor discovered a quantum algorithm for factoring integers which is exponentially faster than the most efficient known classical factoring algorithm (see Sect. 9.4 for a detailed explanation). Shor’s quantum algorithm immediately caused a sensation, giving a strong boost to the development of quantum computing. The mathematical problem of integer factorization, as simple as it may appear, is a very difficult problem in computer science: given a very large integer, especially that is the multiplication of two very large prime numbers, it takes a classical computer an exponentially long time to find its prime factors. Capitalizing on this difficulty, cryptographers devised cryptosystems which are now applied in various business activities, such as credit card transactions (see Chap. 10 for a detailed description). Shor’s algorithm shows that we could easily break these cryptosystems with a quantum computer. Shor’s algorithm is beyond the initial expectations of quantum computers. Scientists like Manin and Feynman, when proposing quantum

---

<sup>1</sup> This book is available only in Russian.

computation, intended to use it to help solve some of quantum mechanical problems. Shor's algorithm clearly demonstrates that a quantum computer can also solve familiar mathematical problems substantially faster, and has a great potential in practical applications. Inspired by this development, physicists finally began to seriously think about building a physical quantum computer, rather than just being satisfied with theoretical discussion. Soon, in 1995, physicists implemented the world's first two-bit quantum logic gate in the laboratory with trapped ions. Later on, physicists came up with many proposals for building quantum computers. Physical quantum computers with around 100 quantum bits have already been constructed. A company called D-Wave claimed to have built a quantum computer with a few hundred quantum bits, but this claim is still not widely accepted in the scientific community. For those interested in these developments, please refer to the wikipedia article "Timeline of quantum computing".

After nearly 40 years of development, quantum computing has evolved from a field attracting a small number of theorists to one of the hottest research fields. Major world powers have announced ambitious plans in the field of quantum information. Many big international companies are making major investments in quantum computing. As a result, major progresses in quantum information technologies are frequently featured in mass media.

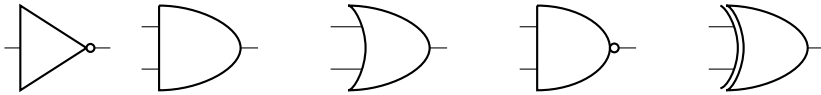
However, in my opinion, there are still many complex and difficult technical problems in the field of quantum computing, and general-purpose quantum computers with practical applications are at least 50 years away. In the following, I'll give tutorial introduction of quantum computing with the following three questions in mind: How does a quantum computer work? What advantages can a quantum computer provide over a classical computer? Why is it so hard to build physically a quantum computer? Before introducing quantum computers, let us first review the basic framework of classical computers and briefly explain why classical computers are inherently inefficient in solving quantum mechanical problems.

## 9.1 Classical Computer

### 9.1.1 Basic Framework

In everyday life we primarily use decimal numbers. In order to reduce errors, digital computers use binary numbers, made up of only 0 and 1, usually corresponding to two states of an electronic device. We can represent any number  $x$  by a sequence of binary digits

$$\dots x_n x_{n-1} \dots x_2 x_1 x_0 \cdot x_{-1} x_{-2} \dots \quad (9.1)$$



**Fig. 9.1** Basic logic gates in a classical computer. From the left to the right: NOT gate, AND gate, OR gate, NAND gate, and XOR gate

**Table 9.1** Function of classical logic gates

AND gate		NAND gate		OR gate		XOR gate	
Input	Output	Input	Output	Input	Output	Input	Output
0 0	0	0 0	1	0 0	0	0 0	0
0 1	0	0 1	1	0 1	1	0 1	1
1 0	0	1 0	1	1 0	1	1 0	1
1 1	1	1 1	0	1 1	1	1 1	0

with

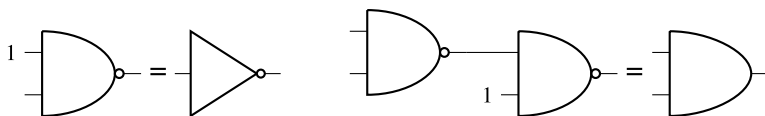
$$x = \sum_{j=-\infty}^{\infty} x_j 2^j, \tag{9.2}$$

where  $x_j$  can only be 0 or 1. For example, the binary representation of 5 is 101; the binary representation of 15 is 1111; the binary representation of 4.5 is 100.1. A complex number can be viewed being made of two real numbers, so it can also be expressed in binary digits. As such, all the numbers in the world can be expressed with two digits, 0 and 1.

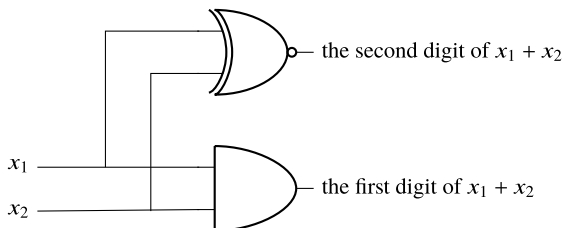
A bit, the basic unit of information storage in a conventional computer, can be in one of two states, representing 0 or 1. If your computer has 2 gigabytes<sup>2</sup> (GB) of memory, this is roughly equivalent to 16 billion bits. These bits encode documents, images, and audios stored on a computer, plus a lot of information that you can not see or hear. After receiving the instructions from you and the programs, the computer performs a series of logic gate operations, changing the state of these bits. After the operations, some bits are switched from 1 to 0, or from 0 to 1, and other bits remain unchanged. The results are the changes of images on the screen or the sound from the computer, or some other changes that you may not perceive.

We have learnt many operations on numbers in math classes, such as the familiar addition, subtraction, multiplication, and division, as well as the matrix-related calculations discussed in Chap. 4 and calculus. These operations have different levels of complexity and difficulty. But scientists have discovered that all these operations can be completed using combinations of a number of simple logic gates. Figure 9.1 shows some basic logic gates. The first logic gate is a NOT gate. Its main function is to invert the input: if the input is 1, the output is 0; if the input is 0, the output is 1. Functions of other logic gates are shown in Table 9.1.

<sup>2</sup> 1 byte = 8 bits.



**Fig. 9.2** Implementation of the NOT gate and the AND gate with NAND gates



**Fig. 9.3** Addition of two single-digit binary numbers with classical logic gates

Among the logic gates in Fig. 9.1, the NAND gate is special as all other classical logic gates can be implemented with NAND gates. The NAND gate is therefore called universal gate. Fig. 9.2 are two examples. The NOT gate is implemented by fixing one of the two inputs of the NAND gate at 1; the AND gate is implemented with two NAND gates by fixing one of the inputs at 1.

The computers that we use everyday are physical implementations of these simple and abstract mathematical concepts and operations. They consist of computer memories that store information in sequences of 1s and 0s, chips that have billions of logical gates operating on the bits under program instructions, and peripheral devices including input devices that allow you to input instructions and output devices that enable you to display information. This now ubiquitous electronic device demonstrates a profound and amazing fact: colors, sounds, symbols, and their inexhaustible combinations, all of these, can be stored as sequences of 1 and 0; with the orchestrated manipulations of logic gates, these 1s and 0s can be transformed into brilliant movies, beautiful musics, profound mathematical theorems, and amazing laws of nature.

The computing power of computers is realized through a computer program. After decades' development, computer programs are now mostly written in high-level programming languages, which are then translated by compilers into a sequence of logical operations that can be implemented by the machine. Here we write a program directly using logic gates, which calculates the addition of two single-digit binary numbers  $x_1$  and  $x_2$ . We need two bits as the input and another two bits as the output (see Fig. 9.3). In binary, both  $x_1$  and  $x_2$  are either 0 or 1. Unless  $x_1 = x_2 = 1$ , the result of the addition,  $x_1 + x_2$ , is also a single-digit binary number, which can be stored with a single bit. When  $x_1 = x_2 = 1$ ,  $x_1 + x_2 = 2$ . The binary expression of 2 is 10, so two bits are needed to record the result. Based on these considerations, we design an algorithm to accomplish this addition. As shown in Fig. 9.3, this algorithm is simple and can be implemented using a combination of AND gate and XOR gate.

For comparison, we will later design an algorithm for a quantum computer (quantum algorithm) to accomplish the same task.

### 9.1.2 The Unbearable Quantum

In the 1980s, many scientists, such as Manin and Feynman, began to realize that classical computers were fundamentally inefficient in solving quantum mechanical problems. Let us see why.

Consider a one-dimensional classical system. If this system contains only one particle, it has 2 variables, its position  $x_1$  and momentum  $p_1$ . The dimension of the system's phase space is 2. If the system contains two particles, there are 4 variables,  $x_1$ ,  $x_2$  and  $p_1$ ,  $p_2$ , so the dimension of the system's phase space is 4. If the system contains three particles, it has 6 variables,  $x_1$ ,  $x_2$ ,  $x_3$  and  $p_1$ ,  $p_2$ ,  $p_3$ , and the dimension of the system's phase space is 6, and so on. A system of  $n$  particles has  $2n$  variables and the dimension of its phase space is  $2n$ . Thus in a classical system, the number of variables and the dimension of phase space are proportional to the number of particles in the system. Suppose we wish to simulate a one-dimensional classical system with 100 particles on a classical computer. It requires a memory capacity for storing 200 variables. If each variable is represented by 4 bytes (32 bits), we need 800 bytes, which is far below the typical memory capacity of a modern computer (about  $\sim 10^9$  bytes).

As a comparison, let us consider a quantum spin system. For one spin, the dimension of the Hilbert space is 2, same as the phase space of a single classical particle; for two spins, the dimension of the Hilbert space is 4, which is still the same as the dimension of the phase space of two classical particles. This seems to suggest that the dimension of the Hilbert space of three spins is 6. But this is not the case; the dimension of the Hilbert space of three spins is 8. Here is why. According to our previous discussion, the Hilbert space of a single spin is spanned by the basis  $|u\rangle$  and  $|d\rangle$ . By direct product, there are 4 basis vectors in the Hilbert space of two spins, namely,  $|uu\rangle$ ,  $|ud\rangle$ ,  $|du\rangle$  and  $|dd\rangle$ . The basis vectors of three spins can be constructed by direct product of the two-spin basis vectors and the single-spin basis. The result is 8 basis vectors, namely,

$$|uuu\rangle, |uud\rangle, |udu\rangle, |udd\rangle, |duu\rangle, |dud\rangle, |ddu\rangle, |ddd\rangle. \quad (9.3)$$

So the dimension of the Hilbert space of three spins is 8. Following this construction procedure, the dimension of the Hilbert space of  $n$  spins is  $2^n$ . This means that the dimension of the Hilbert space of a quantum system increases exponentially with the number of spins (or particles).

Let us try to simulate a quantum system of 100 spins using a classical computer. The dimension of the Hilbert space of this system is  $2^{100}$ . Its vector has  $2^{100}$  complex components. A complex number consists of two real numbers, so the total number of variables is  $2^{101}$ . As in the classical case, we represent a variable by 4 bytes,

so the memory of a classical computer should have at least  $2^{103}$  bytes, which is equivalent to about  $10^{22}$  GB. This far exceeds the memory capacity of the largest existing computer, which is about  $10^{10}$  GB in 2022. Even in the very distant future, it is unlikely that we will ever be able to build a classical computer with such a large memory. Thus it is impossible for a classical computer to fully simulate a quantum system with 100 spins. Even worse, a 100-spin quantum system is not a large system at all. For example, let us consider the carbon fullerene  $C_{60}$ . Since a carbon atom has 4 valence electrons,  $C_{60}$  has 240 valence electrons. An electron has both spin degree of freedom and spatial degrees of freedom, so the dimension of the Hilbert space of  $C_{60}$  is already significantly larger than 100 spins. Therefore, it is tremendously difficult to simulate many-body quantum systems on a classical computer. But for quantum computers, this difficulty does not arise: 100 quantum bits naturally span a  $2^{100}$ -dimensional Hilbert space, which is big enough to accommodate all the quantum states of 100 spins.

## 9.2 Quantum Computer

The basic operations of quantum computing are not complicated. A quantum computer consists of quantum bits (or qubits)—the quantum mechanical analogue of classical bits, and quantum logic gates (or simply quantum gates)—the quantum mechanical counterparts of classical logic gates. Quantum computing is achieved through a sequence of orchestrated gate operations on qubits. Similar to a classical bit, a qubit has two states, represented by  $|0\rangle$  and  $|1\rangle$ . But unlike a classical bit, a qubit can be in a linear superposition of  $|0\rangle$  and  $|1\rangle$ , that is,

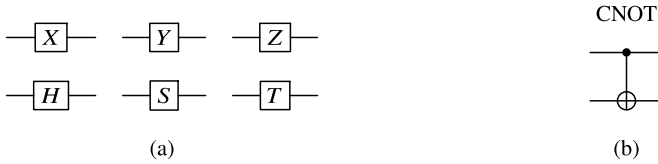
$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (9.4)$$

Many physical systems can be used for qubits. For example, the familiar spins can be used as qubits, with  $|u\rangle$  representing  $|0\rangle$  and  $|d\rangle$  representing  $|1\rangle$ .

Analogous to classical computers, the wide variety of operations on qubits can be decomposed into combinations of a number of elementary quantum gate operations. Figure 9.4 shows these basic quantum gates. They are divided into two categories: single-qubit quantum gates, which operate on a single qubit (Fig. 9.4a) and two-qubit quantum gates, which operate simultaneously on two qubits (Fig. 9.4b).

Every single qubit quantum gate can be represented by a  $2 \times 2$  unitary matrix. In Fig. 9.4a, the three quantum gates in the first row actually correspond to the familiar Pauli matrices  $\hat{\sigma}_x$ ,  $\hat{\sigma}_y$ , and  $\hat{\sigma}_z$ . The three quantum gates in the second row, from the left to the right, are Hadamard gate, phase gate, and  $\pi/8$  gate. Their corresponding unitary matrices are

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}. \quad (9.5)$$



**Fig. 9.4** Quantum gates. **a** First row: X gate, Y gate, Z gate; second row, Hadamard gate, phase gate,  $\pi/8$  gate. **b** CNOT gate. The gates in **a** are single qubit quantum gates while the CNOT gate is a two-qubit quantum gate

We use the Hadamard gate as an example to illustrate how a quantum gate operates on a qubit. Like the spin states  $|u\rangle$  and  $|d\rangle$  can be represented as column vectors, we represent  $|0\rangle$  and  $|1\rangle$  as the column vectors,

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{9.6}$$

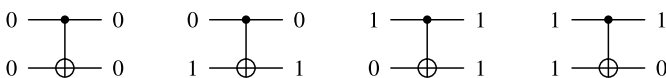
Suppose a qubit is initially in state  $|0\rangle$ . After the Hadamard gate transformation, it becomes

$$H|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \tag{9.7}$$

The qubit is now neither in the state  $|0\rangle$  nor in the state  $|1\rangle$ , but rather in their superposition. This is the most fundamental distinction between quantum computers and classical computers. In a classical computer, a bit is always in a definite state (either 0 or 1), and it remains in a definite state (either 0 or 1) after a logic gate operation. In a quantum computer, primarily due to the Hadamard gate, the superposition of 0 and 1 can occur and in fact occurs frequently.

Figure 9.4b shows a two-qubit quantum gate, the CNOT gate. The symbol of the CNOT gate contains two parallel lines, the upper one representing the control qubit, and the lower one representing the target qubit. As illustrated in Fig. 9.5, the CNOT gate functions as follows: if the control qubit is in the state  $|0\rangle$ , then the target qubit remains unchanged; if the control qubit is in the state  $|1\rangle$ , then the state of the target qubit is flipped.

The dimension of the Hilbert space of two spins is 4. Similarly, the dimension of the Hilbert space of two qubits is 4. Its four basis vectors can be chosen as  $|00\rangle$ ,  $|01\rangle$ ,  $|10\rangle$  and  $|11\rangle$ , where the qubit on the left is the control qubit and the qubit on the right is the target qubit. For example,  $|10\rangle$  represents the control qubit in the state  $|1\rangle$  and the target qubit in the state  $|0\rangle$ . We express these basis vectors as column vectors,



**Fig. 9.5** Operations of the CNOT gate



$$|00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad (9.8)$$

Accordingly, the CNOT gate is mathematically a  $4 \times 4$  matrix,

$$U_{\text{cnot}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (9.9)$$

Interested readers can verify directly the following transformations with the above matrix and column vectors

$$U_{\text{cnot}}|00\rangle = |00\rangle, \quad U_{\text{cnot}}|01\rangle = |01\rangle, \quad (9.10)$$

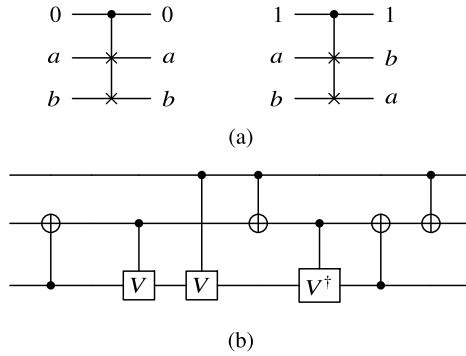
$$U_{\text{cnot}}|10\rangle = |11\rangle, \quad U_{\text{cnot}}|11\rangle = |10\rangle, \quad (9.11)$$

which, respectively, correspond to the operations of the CNOT gate in Fig. 9.5.

As we already mentioned, for a classical computer, the NAND gate is a universal gate and it can be used to realized all other logical gates. Similarly, for a quantum computer, there are also universal gates, which include the Hadamard gate, the  $\pi/8$  gate, and the CNOT gate. All unitary evolutions in a finite dimensional Hilbert space can be realized with these three quantum gates.

Although the single qubit gates and the two-qubit CNOT gates, in principle, are sufficient for achieving arbitrary unitary evolutions, three-qubit quantum gates are often used for practical convenience. We now introduce a typical three-qubit quantum gate, the Fredkin gate, which has one control bit and two target qubits. As we will see later, the Fredkin gate was in fact introduced to construct a reversible classical computer and plays a crucial role in understanding classical computing. Its operations are illustrated in Fig. 9.6a: when the control qubit is 0, the other two qubits remain unchanged; when the control qubit is 1, the other two qubits swap their states. This gate operation can also be expressed as a matrix. Let us denote the state of three qubits as  $|z_1, z_2, z_3\rangle$ , where  $z_i$  labels the  $i$ th qubit. The corresponding column vectors take the form

$$|000\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |001\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |010\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |011\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (9.12)$$



**Fig. 9.6** **a** The Fredkin gate: when the control qubit is 0, the other two qubits do not change; when the control qubit is 1, the other two qubits are swapped. **b** Realization of the Fredkin gate with a combination of two-qubit gates.  $V \equiv (1 - i)(I + iX)/2$  with  $X$  being the  $X$  gate

$$|100\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |101\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |110\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |111\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad (9.13)$$

Thus the Fredkin gate can be represented as the following matrix

$$F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (9.14)$$

Interested readers can verify that this  $8 \times 8$  matrix indeed represents the Fredkin gate.

Fredkin gates can be implemented using a sequence of two-qubit quantum gates, as shown in Fig. 9.6b. In order to understand this circuit, we first introduce the basic rules for quantum circuits. In a quantum circuit, each horizontal line represents a qubit. Therefore, the single qubit gate in Fig. 9.4 contains only one horizontal line while the two-qubit CNOT gate has two horizontal lines. These horizontal lines also represent the passage of time from the left to the right. A symbol on a horizontal line denotes a certain unitary operation on the qubit. The symbols for frequently

used single qubit operations are listed in Fig. 9.4a. There is a special symbol, the filled circle, which indicates that the qubit is a control qubit. A filled circle is always accompanied by a vertical line, which connects the control qubit to the target qubit. When the control qubit is 0, nothing happens; when the control qubit is 1, the target qubit undergoes an operation represented by the symbol at the other end of the vertical line. Here are two examples. For the CNOT gate in Fig. 9.4b, the filled circle on the top horizontal line indicates a control qubit; the vertical line connects the first and second horizontal lines, so the second horizontal line represents a target qubit; the symbol  $\oplus$  at the other end of the vertical line indicates flipping of the state of the target qubit: 0 to 1 or 1 to 0. For the Fredkin gate in Fig. 9.6a, three horizontal lines represent three qubits, the filled circle indicates that the first horizontal line is the control qubit, the vertical line from the filled circle connects the other two horizontal lines, indicating that the other two qubits are target qubits; the two symbols  $\times$  on the vertical line indicate that the states of the two target qubits are swapped when the control bit is 1.

With these rules, we can read the quantum circuit in Fig. 9.6b. The three horizontal lines represent three qubits. The first quantum operation is a CNOT gate, in which the control qubit is the third qubit and the target qubit is the second qubit. The second step is a two-qubit gate operation, in which the control qubit is the second qubit and the target qubit is the third qubit: when the second qubit is 1, the transformation  $V$  is performed on the third qubit. The third step is the same two-qubit gate operation, except that the control qubit is the first qubit. Note that in step three, the vertical line and the second horizontal line intersect, but there is no symbol at the intersection point, so the second qubit is not connected with the control qubit. The fourth step is another CNOT gate, where the first horizontal line is the control qubit and the second horizontal line denotes the target qubit. Step five is similar to step two, except that the transformation is  $V^\dagger$  instead of  $V$ . Step six and step seven are two consecutive CNOT gates. Interested readers can verify that this quantum circuit of 7 two-qubit gates indeed implements the Fredkin gate. It is interesting to consider if there is a simpler combination of single-qubit gates and two-qubit gates to implement the Fredkin gate.

All the quantum gates above have one common feature: they are unitary transformations and represented by unitary matrices. For example, by direct calculation you can verify that  $U_{\text{cnot}}^\dagger U_{\text{cnot}} = 1$ . One can prove mathematically that the combination of these gate operations is also represented by a unitary matrix. So every step in a quantum computer is a unitary transformation. This is a fundamental feature of quantum computing. Why is this feature essential? As said before, the time evolution of a quantum system is unitary, and therefore, to simulate quantum dynamics requires unitary logic gates.

Similar to a classical computer, for a quantum computer to accomplish a certain task, it needs to follow an algorithm, a set of quantum gates that operate according to a thoughtfully designed sequence. Such a quantum algorithm has three stages: (1) all qubits are initialized in a given quantum state; (2) after a sequence of quantum gate operations, the qubits reach a certain quantum state; (3) the final quantum state is measured and the result is read out. For the same problem, if an algorithm

requires fewer gate operations, it has faster performance. Once a quantum algorithm is designed, one can compare it with the corresponding classical algorithms to see which one is faster. In comparing a quantum algorithm with a classical algorithm, we compare the number of steps of gate operations in an algorithm for solving the same problem; the one with fewer steps is faster. In Sect. 9.4, we describe more specifically how to compare the speeds of different algorithms.

As an example, we introduce a simple quantum algorithm. In order to compare with the classical algorithm in Fig. 9.3, we consider the same problem of adding two single binary digits  $x_1$  and  $x_2$ . We quickly notice that  $x_1 = 1, x_2 = 0$  and  $x_1 = 0, x_2 = 1$  give the same sum, i.e. two different inputs produce the same output. This is common in classical computers, as featured in all but the NOT gate in Fig. 9.1: different inputs produce the same output. But for a quantum gate or a quantum computer, different inputs must give different outputs. Let us see why.

Suppose there is a quantum gate  $U$ , which converts two different input states  $|\phi_1\rangle$  and  $|\phi_2\rangle$  into one identical output. Let  $|\psi\rangle = |\phi_1\rangle - |\phi_2\rangle$  be the difference between these two inputs. Then we have

$$|\varphi\rangle = U|\psi\rangle = U|\phi_1\rangle - U|\phi_2\rangle = 0. \quad (9.15)$$

At the same time, as  $U$  is unitary satisfying  $U^\dagger U = 1$  and we have

$$\langle\varphi|\varphi\rangle = \langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle. \quad (9.16)$$

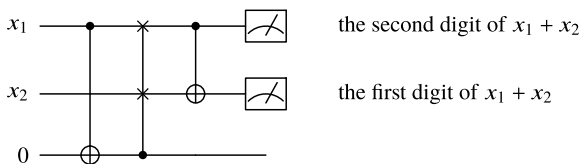
On the one hand, since  $|\phi_1\rangle$  and  $|\phi_2\rangle$  are different, we have  $|\psi\rangle \neq 0$  and therefore  $\langle\psi|\psi\rangle \neq 0$ . On the other hand, we have  $\langle\varphi|\varphi\rangle = 0$ . So, the above equality can not hold and the assumption is wrong.

For a quantum gate, it always transforms two different input states into two different outputs. Since a quantum computer operates with a sequence of quantum gates, different inputs must give different outputs on a quantum computer. As such, it seems that a quantum computer cannot even perform the simplest calculation like  $x_1 + x_2$ .

We solve this difficulty by using three qubits. We use the first two qubits to represent the two numbers  $x_1, x_2$  and the third for another single digit number  $x_3$ . The corresponding quantum states are denoted by  $|x_1, x_2, x_3\rangle$ . At the input, we always have  $|x_1, x_2, 0\rangle$  with the third qubit  $x_3$  fixed at 0. At the output, qubit  $x_1$  represents the first digit of the sum, qubit  $x_2$  the second digit of the sum, and the third is discarded as a redundant output. If the third qubit behaves differently when adding  $x_1 = 1, x_2 = 0$  and  $x_1 = 0, x_2 = 1$ , we would be able to satisfy the requirement of quantum computing—different inputs give different outputs. We find such a quantum algorithm as shown in Fig. 9.7. For adding  $x_1 = 1, x_2 = 0$  and  $x_1 = 0, x_2 = 1$ , our algorithm achieves the following transformations, respectively,

$$|010\rangle \longrightarrow |010\rangle; \quad |100\rangle \longrightarrow |011\rangle. \quad (9.17)$$

**Fig. 9.7** A quantum algorithm for adding two single binary digits  $x_1$  and  $x_2$



At the output, although the first two digits are the same, indicating that the two additions produces the same sum, the third qubit is different: 0 for  $x_1 = 0, x_2 = 1$  and 1 for  $x_1 = 1, x_2 = 0$ .

The quantum circuit of our quantum algorithm for adding two single binary digits  $x_1$  and  $x_2$  is shown in Fig. 9.7. Here are its steps:

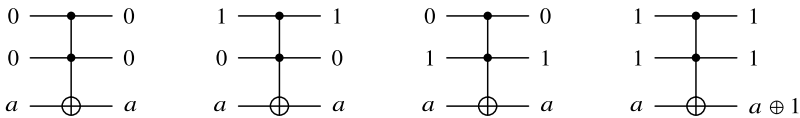
1. Initialize the three qubits as  $|x_1, x_2, 0\rangle$ ;
2. apply the CNOT gate with the first qubit  $x_1$  as the control qubit and the third qubit  $x_3$  as the target qubit;
3. apply the Fredkin gate with the third qubit  $x_3$  as the control qubit;
4. apply the CNOT gate with the first qubit  $x_1$  as the control qubit and the second qubit  $x_2$  as the target qubit;
5. measure the first two qubits  $x_1, x_2$  as the output.

Let us illustrate how the algorithm works with input  $x_1 = 1, x_2 = 1$ . In the first step, the quantum state  $|110\rangle$  is prepared. In the second step, since the first qubit is  $|1\rangle$ , after the CNOT gate, the target qubit (i.e., the third qubit) is flipped, producing  $|111\rangle$ . In the third step, after the Fredkin gate, the state remains unchanged. In the fourth step, since the first qubit is  $|1\rangle$ , after the CNOT gate, the target qubit (i.e. the second qubit) is flipped, producing  $|101\rangle$ . In step 5, the first two qubits are measured, giving the correct result 10. Mission accomplished. Interested readers can similarly verify that our algorithm can indeed achieve the input-output transformations in Eq. (9.17).

Let us now compare this quantum algorithm to the classical algorithm Fig. 9.3. The classical algorithm is significantly simpler, and it consists of two logic gates in one step. The quantum algorithm takes, apart from the input and output, at least three intermediate steps. If we require that all quantum operations be single-qubit or two-qubit gates, then we need at least eight steps (see the implementation of the Fredkin gate in Fig. 9.6). So, for this specific problem, the quantum algorithm is much slower than the classical algorithm. Thus a very important and fundamental question is: how powerful are quantum computers? We will discuss this in detail in Sect. 9.4. Interested readers are encouraged to design a simpler and faster quantum algorithm for adding  $x_1$  and  $x_2$ .

### 9.3 Reversible Classical Computers

In the previous section we have noted an important difference between classical gates and quantum gates. Most classical logic gates are irreversible: different inputs can produce the same output; by contrast, all quantum gates are reversible: different

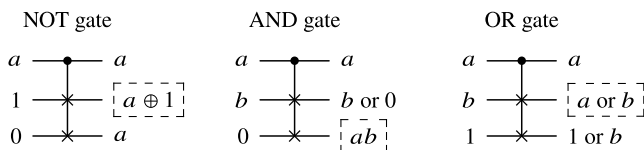


**Fig. 9.8** A Toffoli gate has two control bits. When both of the control bits are 1, the third bit is flipped; otherwise, all bits remain unchanged. Here  $\oplus$  is modulo-2 addition: add two numbers, divide the sum by 2, and the remainder is the result of modulo-2 addition

inputs always produce different outputs. This seems to imply that a classical computer is irreversible and a quantum computer is reversible, and the reversibility is the distinguishing feature between classical computers and quantum computers. This is wrong! Although the classical computers that we use in practice are irreversible, classical computers can theoretically be reversible. In the 1970s, many scientists, including Fredkin (Edward Fredkin, 1934–) and Toffoli (Tommaso Toffoli, 1943–), studied reversible classical computing. They found that the reversible classical computing was not only theoretically feasible, it was also as powerful as irreversible classical computers. Several physical implementations of reversible classical computers were later proposed but no one has been built.

Fredkin and Toffoli invented two three-bit logic gates that now bear their names. We have introduced the Fredkin gate in the previous section. Now we briefly introduce the Toffoli gate. As shown in Fig. 9.8, the Toffoli gate has two control bits and one target bit. It functions as follows: the target bit is flipped under the condition that both control bits are 1, otherwise, the target bit does not change. This is clearly a reversible logic gate: different inputs produce different outputs. The Toffoli gate can also be represented by an  $8 \times 8$  unitary matrix.

Although the Fredkin gate and the Toffoli gate are reversible logic gates, we can use them to realize all the classical logic gates in Fig. 9.1, regardless of their reversibility. Let us take the Fredkin gate as an example. Figure 9.9 depicts how to use the Fredkin gates to realize the classical NOT gate, AND gate, and OR gate. The Fredkin gate is a three-bit reversible logic gate, which has three inputs and three outputs. But the desired classic gate has one or two inputs and only one output. Thus the Fredkin gate always has redundant inputs and outputs when used to realize these classical one-bit or two-bit gates. We can fix the values of the redundant inputs, and select the output for our purpose while ignoring other outputs. In Fig. 9.9, we see that in implementing the NOT gate, the inputs of the two target bits are fixed at 1 and 0. The output important for us is the first target bit, which has been marked by a box for clarity. To implement the AND gate and OR gate, we only need to fix the input of the second target bit. Since the NAND gate, which is effectively a combination of the NOT gate and AND gate, is a universal gate, the implementations in Fig. 9.9 shows that the Fredkin gate is a universal gate, i.e., any logic function and operation in a classical computer can be implemented using only the Fredkin gates. The Toffoli gate is also a universal logic gate. Interested readers can try to use Fredkin gates to implement the NAND gate and XOR gate, and use the Toffoli gate to achieve the NOT gate, AND gate, OR gate, etc.



**Fig. 9.9** Implementation of the classical NOT gate, AND gate and OR gate with the Fredkin gate. Since the Fredkin gate is a reversible three-bit gate, it has redundant inputs and outputs. Valid inputs are denoted by variables  $a$  and  $b$ ; valid outputs are marked by dashed boxes

The above analysis has two immediate important corollaries. Firstly, a reversible classical computer is indeed feasible at least in principle; secondly, a reversible computer is equivalent to the usual irreversible computer in terms of the computational power. If you have an algorithm for an irreversible computer, you can immediately obtain an algorithm for a reversible computer, simply by replacing the logic gates with the Fredkin gates or the Toffoli gates. Both algorithms involve the same number of logic gates, except that the reversible algorithm obtained in this way requires a few more bits.

There is another important corollary from the above analysis: a reversible classical computer is a special case of a quantum computer. The reason is that a reversible classical computer is composed of either the Fredkin gates or Toffoli gates, both being unitary transformations. Combining with the second corollary above, which states that a reversible classical computer is equivalent to an irreversible classical computer, we can assert that

A classical computer is theoretically a special quantum computer.

An immediate implication of this conclusion is that theoretically a classical computer cannot be more powerful than a quantum computer just like the fastest white horse is at best as fast as the fastest horse. Although the Fredkin gate and the Toffoli gate were first proposed in the context of reversible classical computers, they are now widely used in quantum computers.

So what is the fundamental distinction between a classical computer and a quantum computer? To address this question, consider three qubits in quantum state  $|101\rangle$ . This state is a product state, where there is no entanglement between the qubits and each qubit is in a definite state,  $|0\rangle$  or  $|1\rangle$ . Since this state can also be realized with classical bits, we call it classical state. Note that there are some product states that are non-classical. For example,  $(|101\rangle + |100\rangle)/\sqrt{2}$  is a product state, but it is not a classical state for the last qubit is in a superposition state  $(|1\rangle + |0\rangle)/\sqrt{2}$ . Using the classical states, we can classify quantum gates into two categories: product quantum gates and superposition quantum gates. For product quantum gates, if the input is a classical state, the output is also a classical state; otherwise, the gate is a superposition quantum gate. It can be verified by directly checking all the possible inputs of classical states that both the Fredkin gate and Toffoli gate are product quantum gates. Because the CNOT gate can be realized with Toffoli gates (see Fig. 9.10a), the CNOT gate is also a product quantum gate. Among the quantum gates introduced earlier, only the Hadamard gate is a superposition quantum gate.



**Fig. 9.10** **a** The CNOT gate implemented with the Toffoli gate. **b** A quantum circuit capable of generating an entangled state from a classical input

A quantum computer with only product quantum gates is essentially a classical reversible computer. In this type of quantum computer, when an input is a classical state, the output of every gate operation is also a classical state. The quantum algorithm in Fig. 9.7 is essentially a reversible classical algorithm, for it only involves product quantum gates. A quantum circuit beyond a classical computer must have at least one superposition quantum gate, such as a Hadamard gate. Figure 9.10b shows a genuine quantum circuit. If the input state is a classical state  $|10\rangle$ , then it will undergo the following evolution

$$|10\rangle \xrightarrow{H} \frac{1}{\sqrt{2}}(|1\rangle \otimes (|0\rangle + |1\rangle)) \xrightarrow{\text{CNOT}} \frac{1}{\sqrt{2}}(|10\rangle + |01\rangle). \quad (9.18)$$

The final state is not only a superposition state but also an entangled state: each qubit loses its individuality and is in an indefinite quantum state. This result is fundamentally impossible in a classical computer (reversible or not). Therefore, the most fundamental distinguishing feature of a quantum computer from a classical computer is the superposition and entanglement. Quantum computers may become more powerful than classical computers because of these two unique features. Indeed, there are already quantum algorithms that are faster than classical counterparts. However, it is unclear in general how exactly superposition and entanglement contribute to the power of a quantum computer. It is highly non-trivial to find a quantum algorithm that is better than its classical counterpart.

Historically, theoretical models of a computer were regarded as purely mathematical; people thought that physics was only needed in building a computer. Only when Manin and Feynman proposed the concept of quantum computing, people began to realize that physics was involved not only in building a computer but also in constructing theoretical models of computing. This is understandable since it is hard to see why the logic gates in Fig. 9.1 for the irreversible classical computer or the Fredkin gate for the reversible classical computer are related to classical mechanics. In addition, these gates can be combined to simulate not only Newton’s equations of motion but also the Schrödinger equation. As we discussed earlier, the crucial difference of a quantum computer from a classical computer is that it has superposition and entanglement. This became apparent only when Deutsch proposed the model of quantum Turing machine. As discussed in the above, the most fundamental difference of a quantum computer from a classical computer is that the qubits can superpose and entangle. This fundamental difference can be equivalently put in another perspective: the information processed by a classical computer is cloneable whereas the information processed by a quantum computer is uncloneable.



## 9.4 How Powerful Is a Quantum Computer

Previously, we have shown that quantum computers, in principle, are at least as powerful as classical computers. How much more powerful are quantum computers than classical computers? To answer this question, we must first figure out how to compare the power or speed of two computers. The most straightforward way is to use them to solve the same problem and see which one is faster. Unfortunately, a quantum computer that can solve a useful problem has not yet been built.

Another approach is to compare the algorithms. According to the working principles of the two computers, we can design different algorithms to solve the same problem. If one algorithm requires fewer number of gate operations than the other, we say that the corresponding computer is more efficient. In the previous section, we introduced a quantum algorithm for adding two single-digit binary numbers, which has more steps than the classical algorithm. For this particular problem, apparently, the quantum computer is slower than the classical computer. But can we conclude that quantum computers are not as powerful as classical computers? No. Let us see why.

Suppose we have two computers: one called Vermilion Bird, which can only perform addition, and the other is called White Tiger, which can perform, apart from addition, subtraction, multiplication and division. We now use them to compute the function  $g(n) = 1 + 2 + 3 + \dots + n$ . Because Vermilion Bird can only perform addition, our algorithm for Vermilion Bird can only add consecutively every number from 1 to  $n$ . But for White Tiger, we can use the formula  $g(n) = n(1 + n)/2$  to compute it. When  $n \leq 2$ , obviously Vermilion Bird is faster: for  $n = 1$ , no computation is needed; for  $n = 2$ , only one more step is needed. In contrast, for either  $n = 1$  or  $n = 2$ , White Tiger needs to perform one step of addition, one step of multiplication, and one step of division. However, it is also clear that when  $n$  is large, White Tiger is much faster, and everyone will agree that White Tiger is more powerful.

As you can see from the above discussion, it is not easy to compare the power of various computers. Before classical computers were physically built, computer scientists and mathematicians have discussed abstractly about various hypothetical computing machines and how to compare their computation speeds. The famous Turing machine was a product of these discussions. In order to compare the efficiency of various computers, scientists proposed a function called time complexity,  $O(f(n))$ , where the variable  $n$  is the size of the input and  $f(n)$  is a function of  $n$ . The time complexity describes how the algorithm's running time varies when the input size  $n$  increases. Let us skip the abstract math and illustrate time complexity  $O(f(n))$  with concrete examples.

*Determine the parity of an integer.* The input is an integer, represented by  $n$  digits in binary. The size of this input is thus  $n$ . To determine the parity of this integer, we only need to look at the last digit of its binary number, regardless how large this integer is (i.e., how many digits it has). This means that the running time is independent of the input size, and the time complexity of the parity problem is  $O(1)$ .

*Random search.* Suppose you have ten untagged keys, but only one of them can open the door. To find the right key, you have to try these keys one by one. If you are very lucky, you succeed the first time; if you are very unlucky, you succeed the last time. On the average, you need 5 trials. This type of problem is called random search: you have  $N$  unlabeled objects, one of which is the desired target. In this case, the size of the input is  $N$ , and, on the average, you need to try  $N/2$  times to find your target. Computer scientists consider the time complexity for random search as  $O(N)$ . Why is it not  $O(N/2)$ ? Because the time complexity is aimed to describe how the running time varies with the size of the input size. Both  $O(N)$  and  $O(N/2)$  show that the search time is doubled when the input size is doubled. Therefore, the factor  $1/2$  is not important.

We now use the time complexity to measure how powerful a quantum computer is. As said at the beginning of this chapter, in 1994 Shor discovered a quantum algorithm for factoring integers which is exponentially faster than the best classical factoring algorithm that has been found so far. What is meant here is that Shor's quantum algorithm is exponentially faster than classical factoring algorithms in terms of time complexity. Let us take a closer look at this example. Suppose there is an integer  $N$ , encoded by  $n$  bits in binary. The time complexity of Shor's quantum algorithm is  $O(n^2 \cdot \log n \cdot \log \log n)$ . By contrast, the time complexity of the fastest classical algorithm is  $O(e^{1.9n^{1/3} \log^{2/3} n})$ . Currently the RSA cryptosystem (see Sect. 10.3 for details) uses an integer with  $n = 2048$  bits for encryption. To break such a RSA encryption requires roughly  $n^2 \cdot \log n \cdot \log \log n \approx 1.6 \times 10^8$  operations for a quantum algorithm and  $e^{1.9n^{1/3} \log^{2/3} n} \approx 6.75 \times 10^{51}$  operations for a classical algorithm. If both quantum and classical computers can perform  $10^9$  operations per second, a quantum computer can break the RSA code in less than one second. Even if the quantum computer is much slower physically, performing only  $10^6$  operations per second, it takes only about tens of minutes to break the code. By contrast, it takes a classical computer about  $2 \times 10^{35}$  years to break it, which is 25 orders of magnitude longer than the age of the universe. The difference is stunning! Unfortunately, the introduction of Shor's algorithm is beyond the scope of this book.

The time complexity of random search on a classical computer is  $O(N)$ . In 1996, Grover (Lov Kumar Grover, 1961–) proposed a quantum algorithm for random search with time complexity  $O(\sqrt{N})$ , which is substantially faster than classical algorithms: when the number of objects to be searched is quadruplicated, the running time of Grover's quantum algorithm is only doubled whereas the running time for the classical search algorithm is quadruplicated. As Grover's algorithm involves some complex mathematics, we will not describe how it works exactly, but rather give an intuitive explanation as to why quantum search is faster. Let us use  $|1\rangle, |2\rangle, \dots, |j\rangle, \dots, |N-1\rangle, |N\rangle$  to denote the  $N$  objects to be searched. In the Grover's algorithm, the quantum computer is initialized in the quantum state

$$|\Psi_0\rangle = \sum_{j=1}^N \frac{1}{\sqrt{N}} |j\rangle, \quad (9.19)$$

where the amplitude of state  $|j\rangle$  is  $1/\sqrt{N}$ . This reflects the fact that, as the desired target is not known *a priori*, every state is equally likely to appear. In contrast, in classical random search, every object has probability  $1/N$  to appear. The difference between the quantum amplitude  $1/\sqrt{N}$  and the classical probability  $1/N$  is exactly the reason why a quantum search algorithm is faster.

From the two examples above, we see that, indeed, quantum computers can be more powerful than classical computers. Unfortunately, scientists have hitherto found only a small number of quantum algorithms that are faster than their classical counterparts. One possible reason is that quantum computers are based on the laws of quantum mechanics, and the intuition that we acquire in our everyday experience is not very useful in designing quantum algorithms. A more important reason, in my opinion, is that we do not yet have a deep understanding as to why quantum computers are more powerful than classical computers. Although we provided an explanation for the efficiency of the quantum search algorithm with respect to the classical search algorithm, it does not apply to Shor's quantum algorithm. The reason why Shor's algorithm is faster is very different. Without general guidelines, it is difficult to design quantum algorithms that are faster than classical algorithms. Physicists are now trying to use various known quantum processes, such as quantum tunneling, quantum adiabatic evolution, and cooling, to assist the design of quantum algorithms. Some early progress has been reported.

There is a widespread misconception that a quantum computer is powerful because it allows for parallel computing: using the superposition of quantum states, one can simultaneously operate on multiple inputs. But wait! The output is also a superposition of many answers. In order to distinguish these answers, we have to make measurements. For one measurement only produces one outcome, we have to repeat all the calculations in order to obtain other answers. Thus, although the state superposition is an important feature that distinguishes quantum computers from classical computers, the utility of the state superposition *per se* does not make quantum computers more efficient.

Note that any quantum algorithm ends with a solution that needs to be obtained by measurement. If there are several solutions, the computation must be repeated in order to get a new solution, which is obtained through a new measurement. This statement does not rely on our interpretation of measurement. If we assume the collapse of the wave function, then the quantum computer will collapse into one of the solutions after the measurement. You will simply get the same solution if you continue to measure on this state. If we use many-worlds theory, the world splits after the measurement, and there are as many worlds as there are solutions, with only one solution in each world. No matter what you may believe, if you want to know other solutions, you have to repeat the computation and measure again.

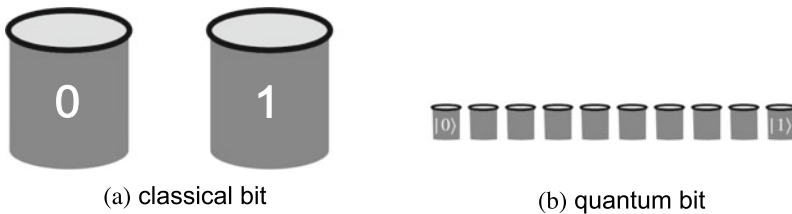
In the above, we have already shown that quantum computers can outperform classical computers for certain problems. But for many other problems, it is still not known whether quantum computers are more efficient than classical computers. For example, there is a class of very hard problems called NP-complete problems. For these well known difficult problems, quantum algorithms faster than their classical counterparts are yet to be found.

## 9.5 Technical Difficulties of Building a Quantum Computer

The concept of quantum computation was conceived in the early 1980s. After over four decades of development, scientists have constructed some primitive quantum computers in the laboratory, which are yet outperform an ordinary classical computer for any practical purpose, not even close. At the moment, there is a general consensus in the scientific community that it is difficult to build a general-purpose quantum computer that is capable of surpassing the fastest classical computer. General-purpose means that it can be applied to solve any problem. The opposite is a special-purpose quantum computer, which is able to solve one or several particular problems. Personally, I think it will take at least 50 years to build the first general-purpose quantum computer that is able to outperform classical computers. On the other hand, special-purpose quantum computers, which can perform better than classical computers for some given problems, are likely in the near future. Time, the most impartial judge, will give the final verdict.

Now humans can readily fly to the blue sky, even land vehicles on the Mars, and put billions of transistors in a nail-sized chip, but still can not build a practical quantum computer. Why is it so difficult to build a quantum computer?

Let us first review the technological development in classical computers. In modern computers, the two states of a bit (0 or 1) generally correspond to the high and low gate voltages in the field-effect transistor. A gate voltage higher than a threshold is read as a “1”; conversely, a voltage lower than the threshold is read as a “0”. Either 0 or 1 corresponds to a rather big range of voltage, and we do not need fine control or measurement of voltage in order to get 0 or 1 accurately. This is like throwing small balls into two large, deep buckets separated by a certain distance. It is easy to throw the ball into the desired bucket, but difficult for it to escape to the other bucket (see Fig. 9.11a). Even so, there is a small chance that a bit in the 0 state becomes 1 due to noise, leading to an error. To avoid and reduce errors as much as possible, classical information technology uses fault-tolerance schemes. A simple and efficient scheme is to use three bits as one bit:



**Fig. 9.11** Classical and quantum bits. A classical bit has only two states, 0 and 1; a qubit can be, apart from  $|0\rangle$  and  $|1\rangle$ , a superposition of  $|0\rangle$  and  $|1\rangle$ . **a** Manipulating a classical bit is relatively easy. It is like throwing a small ball into two large, deep buckets: it is easy to get the ball in but difficult for the ball to accidentally jump into the other bucket. **b** The manipulation of quantum bits is much more difficult. It is like throwing a small ball into many small, shallow buckets, which not only requires a high degree of precision to get the ball into a particular bucket, it is also easy for the ball to escape to another bucket

$$0 \rightarrow 000; \quad 1 \rightarrow 111. \quad (9.20)$$

It is common to call the bit on the left side logical bit and the three bits on the right side physical bits. Three physical bits in state 000 means that the logical bits are in state 0, and three physical bits in state 111 means that the logical bits are in state 1. Suppose the physical bits in state 000 become 010 due to noise. When the computer finds one of the three bits is 1 and two others are 0, it knows the one in the middle is in a wrong state and corrects it to 0. As the probability for two bits to be simultaneously wrong is very small, the error rate in computing is reduced significantly. To further lowering the error rate, we can continue increasing the number of physical bits.

Similar errors occur in quantum computers as well, actually even worse. A qubit can be, in addition to the two states,  $|0\rangle$  and  $|1\rangle$ , a superposition state  $\alpha|0\rangle + \beta|1\rangle$ , where  $\alpha$  and  $\beta$  are complex continuous variables. As the difference between various superposition states can be very small, precise manipulation of qubits is required, and is extremely challenging. The superposition states of a qubit can be regarded as many small and shallow buckets. Apparently, it is more difficult to throw a ball into a particular small bucket and small noise can move it to other buckets (see Fig. 9.11b). Let us consider a concrete example. Suppose there is a qubit in state  $|0\rangle$  and our goal is to flip the qubit to  $|1\rangle$  with an X-gate. Due to noise or imperfect manipulation, the actually realized unitary transformation is

$$\tilde{X} = \frac{1}{\sqrt{1+\epsilon^2}} \begin{pmatrix} \epsilon & 1 \\ 1 & -\epsilon \end{pmatrix}, \quad (|\epsilon| \ll 1). \quad (9.21)$$

After this unitary transformation, the qubit is in the following state

$$|1_\epsilon\rangle = \frac{1}{\sqrt{1+\epsilon^2}} (\epsilon|0\rangle + |1\rangle), \quad (9.22)$$

which is very close to  $|1\rangle$ . A classical bit has only two possible states, 0 or 1; any state close to 1 is read as 1. But for a qubit,  $|1_\epsilon\rangle$  and  $|1\rangle$  are two different states. A quantum computing process involves more than tens of thousands of gate operations. If each operation introduces a small deviation, these deviations will accumulate and eventually cause the whole operation to fail. Therefore, a more robust fault-tolerant scheme is needed for a quantum computer. Scientists find that the quantum fault-tolerant scheme requires at least 6 physical bits to implement a logical bit. This is both good news and bad news. The good news is that practical fault-tolerant schemes do exist; the bad news is that it adds to the difficulties to construct physically a quantum computer. There is a general consensus that a general-purpose quantum computer requires at least 50 qubits to outperform a classical computer, so a practical quantum computer would require at least 300 physical qubits. There is a long way to go before we have a practical quantum computer.

A bigger challenge facing quantum computers is decoherence. This is a difficulty that is unique to quantum computers and does not exist in classical computers. Decoherence means that the qubits get entangled with the environment or other devices

in the computer, losing their quantum coherence. Let us look at how this actually happens. At the level of hardware, a quantum computer consists of two parts: the qubits and the devices that implement the quantum logic gates. For convenience, we will refer to the latter as gate devices. Using gate devices, one implements quantum gate operations on qubits and change their states. Gate devices, because they perform quantum gate operations on qubits, must interact with qubits, which will necessarily generate entanglement. After a certain gate operation, the quantum computer is very likely to be in the following quantum state,

$$|QC\rangle = \frac{1}{\sqrt{1 + \epsilon^2}} \left\{ |Qubits_A\rangle \otimes |Gates_1\rangle + \epsilon |Qubits_B\rangle \otimes |Gates_2\rangle \right\}, \quad (9.23)$$

where  $|Qubits_{A,B}\rangle$  denotes the state of the qubits and  $|Gates_{1,2}\rangle$  denotes the state of the gate devices. This is an entangled state. As we discussed earlier, once two systems are in an entangled state, both systems will lose their individualities and no longer have a definite quantum state. If the quantum computer as a whole is in an entangled state like the one above, then its quantum bits do not have a definite quantum state. A specialist would say that the quantum computer undergoes decoherence and is no longer a quantum computer. Things are even worse in realistic situations, because in addition to the gate device, the qubits are disturbed by many other noises that also cause decoherence.

The central challenge in building a quantum computer is to overcome decoherence. Physicists have considered many methods to reduce decoherence, such as using certain types of qubits that are easy for manipulation and keeping them very cold. After many experiments, most experimental groups now prefer to use superconducting qubits that are based on Josephson junctions. Recently, topological quantum computing has become very popular, which aims to use topological property to protect the quantum coherence of qubits.

It is apparent that the more qubits there are, the more likely they get entangled with the environment and the easier decoherence occurs. On the other hand, in order to make quantum computers more powerful, we need to integrate more qubits. This dilemma is the biggest technical challenge facing physicists today. I am not sure whether a practical general-purpose quantum computer will ever be built. But in this process humans will certainly understand better the microscopic world and push the boundaries of the micro-manipulation technology to its limits. As Feynman said, “There’s plenty of room at the bottom.” If we do not find quantum computers at the bottom, we will find other things which may equally beneficial to human society.

# Chapter 10

## Quantum Communication

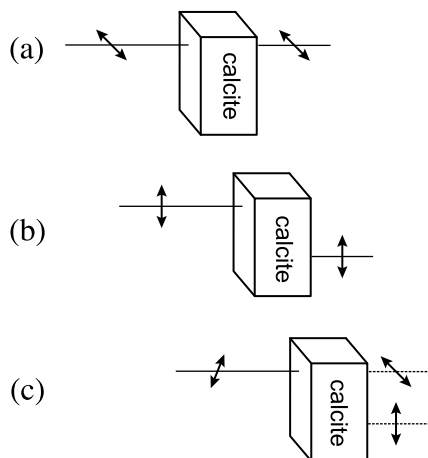


In the information age, our daily lives are becoming increasingly digitalized, and they are stored in magnetic memories, processed by computers, and transmitted via optical fibers and electromagnetic waves. Yet our world is quantum in nature. Therefore, a natural question is how to store, process and transmit the information encoded in qubits. These questions are addressed by quantum information theory. Quantum computation, as introduced in the previous chapter, explores how to process quantum information. Quantum communication is another important branch of quantum information science, where quantum entanglement has been successfully exploited to achieve quantum teleportation for transferring quantum information.

In the context of quantum computation, many different quantum systems have been proposed to physically implement qubits, such as nuclear spins, trapped ions, and quantum dots. At present, most experimental groups prefer to use superconducting qubits based on Josephson junctions. There are also significant experimental efforts devoted to the realization of topological qubits. For quantum communication, photons are the unanimous choice for qubits. There are at least three reasons: (1) photons display significant quantum effects even in a normal environment; (2) photons do not easily get entangled with other systems; (3) one can use mature technologies from classical optical communication. For the first two reasons, photons are also a serious candidate for qubits in quantum computing. Below I shall start with introducing optics (or photons) and how to use photons as qubits.

### 10.1 Polarization of Photons

In 1865, Maxwell (James Clerk Maxwell, 1831–1879) wrote down a set of universal laws for electric and magnetic fields. These are the well known Maxwell equations that we use today. After writing down the equations, Maxwell immediately realized that light is an electromagnetic wave. Conversely, one can say that every



**Fig. 10.1** The polarization states of a photon and its experimental measurement. A photon with a definite polarization state is incident from the left onto a calcite crystal. **a** If the direction of polarization is horizontal, the photon is not affected by the calcite and maintains its horizontal polarization when exiting from the right side. **b** If the direction of polarization is vertical, the path of the photon will be shifted downward by the calcite while the photon maintains the vertical polarization when exiting from the right side; **c** If the direction of polarization is  $45^\circ$ , then the photon has about 50% probability to exit with a horizontal polarization and 50% probability to exit with a vertical polarization. This experiment can be regarded as the Stern-Gerlach experiment for photons

electromagnetic wave is a specific kind of light. In 1905, inspired by Planck, Einstein proposed the theory of light quanta (i.e., photons), which says that electromagnetic waves are made up of discrete photons. The theory of light quanta was eventually confirmed in experiments, leading us to the interesting conclusion that light (or electromagnetic wave) is both wave and particle. Light and electromagnetic waves that we are exposed to in our daily life are made up of vast numbers of photons. For example, a strong mobile phone signal is about 10 nW (or  $10^{-8}$  W), which is equivalent to emitting or receiving about  $10^{16}$  photons of frequency  $\sim 900$  MHz per second. This is a million times more than the number of grains in a ton of sand. If you are using a 4G mobile phone, every bit of information you receive is carried by more than one million photons. As such, signals in conventional communication are noise-resilient with high fidelity. That we do not feel single photons is due to that each photon has no size and carries a tiny amount of energy. It is akin to our inability to see and detect water molecules in water with our own biological sensors.

In quantum communication, one bit of information is carried by a single photon. As shown in Fig. 10.1, a photon has two orthogonal polarization states, horizontal and vertical. To encode the states of a qubit, we use the horizontal polarization state to represent  $|0\rangle$  and the vertical polarization state to represent  $|1\rangle$ . A generic polarization state can be expressed as a superposition of these two states  $\alpha|0\rangle + \beta|1\rangle$ . The polarization state can be measured with a calcite crystal. If the photon is in the



horizontal polarization state  $|0\rangle$ , the photon will pass through the calcite without any change. If it is in the vertical polarization state  $|1\rangle$ , the path of the photon will be shifted downward after passing through the calcite. If the photon is in other polarization state  $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ , then after passing the calcite, the probability to detect a horizontally polarized photon is  $|\alpha|^2 = |\langle\psi|0\rangle|^2$ , and the probability to observe a downshifted vertically polarized photon is  $|\beta|^2 = |\langle\psi|1\rangle|^2$ . If it is not a single photon but a light beam made of many photons that are all in the polarization state  $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ , it will be split into two beams after passing the calcite: the upper one is in the horizontal polarization state with an intensity proportional to  $\alpha^2$ , and the lower one is in the vertical polarization state with an intensity proportional to  $\beta^2$ .

This is analogous to the Stern-Gerlach experiment that measures the spin state. Indeed, their underlying physics is essentially the same. At a deeper level, physicists find that photon’s polarization is in fact its spin, and the calcite crystal can be seen as an equivalent of “magnetic field” that distinguishes the different spin states of a photon.

In the Stern-Gerlach experiment, if we change the orientation of the magnetic field, the observable is changed. Likewise, when we rotate the calcite and change its orientation, the observable is changed. Let us examine a special case where the calcite is rotated to an angle  $45^\circ$  from the vertical. In this case, the polarization state of light that can pass calcite without being unaffected is  $|0_x\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$  while the light in the polarization state  $|1_x\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$  is shifted after passing through the calcite, albeit maintaining its polarization state. One usually refers to  $|0_x\rangle$  as the  $45^\circ$  polarization state and  $|1_x\rangle$  as the  $135^\circ$  polarization state.

In fact, the calcite can be rotated by an arbitrary angle. Yet, regardless of the orientation of the calcite, there are only two possible outcomes after photons pass through it: (1) without deflection, which is recorded as 0; (2) with deflection, which is recorded as 1. The angle of the calcite only affects the probabilities to observe the two outcomes. For example, when the calcite is placed vertically, the measurement of the polarization state  $|0\rangle$  gives 0 with probability of 100% and 1 with probability of 0%. So far we have discussed two special measurements of the photon polarization state: (1) the calcite is placed vertically; (2) the calcite rotates an angle  $45^\circ$  from the vertical. We refer to them as  $M_z$  and  $M_x$ , respectively. Since these two measurements are used in the quantum communication, we summarize the measurement results for the four polarization states  $|0\rangle, |1\rangle, |0_x\rangle, |1_x\rangle$ , respectively, in Table 10.1.

**Table 10.1** The measurement of the photon polarization state

Polarization state	$ 0\rangle$		$ 1\rangle$		$ 0_x\rangle$		$ 1_x\rangle$	
	0	1	0	1	0	1	0	1
$M_z$ (%)	100	0	0	100	50	50	50	50
$M_x$ (%)	50	50	50	50	100	0	0	100

Quantum communication involves encoding information in the polarization of photons and transferring them from one place to another. The encoded information can be either classical or quantum. For classical information stored in a sequence of 0s and 1s, the encoded photons will be in the horizontally or vertically polarized state. If quantum information is encoded, the photons can be in a superposition state of the two mutually orthogonal polarized states, or even in an entangled state. Since the polarization of individual photons and entanglement among photons are easily disturbed by various kinds of noise, there are challenging requirements on the media transferring these photons. We call a communication channel which transmits quantum information as quantum channel. A communication channel which transmits classical information is called classical channel. To maintain a quantum channel requires a lot of sophisticated and expensive instruments. Even so, there is still photon dissipation in quantum channels and photons need to be replenished at a regular rate; this is called quantum relay. Quantum relay demands the replenished photon has exactly the same polarization as its predecessor photon, making quantum relay technically challenging.

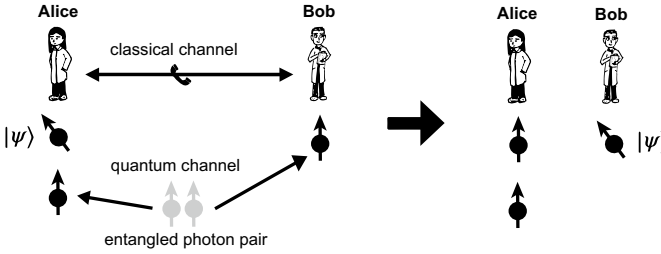
Compared to quantum communication, classical communication in our everyday world is very stable. First, in optical (or electromagnetic wave) communication, one bit of information is carried by millions of photons, instead of a single photon. Therefore, the information carried by light is very robust to small noises, ensuring a high level of fidelity. Second, state-of-the-art classical communications rely on the modulation of light frequency (or electromagnetic wave frequency) for transmission, which does not depend on the polarization of light. Whereas the frequency of light can be hardly altered, its intensity can be easily compensated through the relay. As a result, classical communication is very reliable. This is consistent with our daily experience: we almost never need to worry about noise or weak signals while using mobile phones unless we are in very remote areas.

Due to the stability problems, application of quantum communication is very limited. So far, quantum communication is mostly used for distributing cryptokeys over a long distance. It is safe to say that quantum communication will never replace classical communication in our daily life.

## 10.2 Quantum Teleportation

Quantum teleportation is a simple but magical way of quantum communication (see Fig. 10.2). It consists of transferring a quantum state from a sender to a receiver by sharing a pair of entangled photons. By convention, we refer to the sender as Alice and the receiver as Bob.

The task is that Alice wants to transfer a photon polarization state,  $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ , to Bob some distance away. There are many possible ways to accomplish this task: (1) Alice knows this polarization state, i.e., she knows the superposition coefficients  $\alpha$  and  $\beta$ . In this case, she can use a classical communication, like a phone call, to tell Bob these two coefficients, which allows Bob to prepare a photon into



**Fig. 10.2** Quantum teleportation. By using an entangled photon pair, Alice can transfer a quantum state  $|\psi\rangle$  to Bob at another place. Bob uses the quantum channel to send one photon in the entangled pair to Alice. Alice tells Bob her measurement outcome through classical communication, like a phone call. In the whole process, the quantum information  $|\psi\rangle$  is carried neither in either the quantum channel nor the classical channel

the state  $|\psi\rangle$ . (2) Alice does not know the exact polarization state being transferred. In this case, she can transmit this photon directly to Bob through a quantum channel. The first method uses only a classical channel whereas the second method uses only a quantum channel. In the following we present the third method, quantum teleportation, in which Alice and Bob use both classical and quantum channels.

Figure 10.2 schematically shows a protocol for quantum teleportation. Alice has a photon in state  $|\psi\rangle$ , and she wants to send this bit of quantum information  $|\psi\rangle$  to Bob who is some distance away. A pair of photons in the following entangled state is prepared,

$$|\gamma_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{10.1}$$

Bob generates this pair and sends one photon to Alice via a quantum channel and keeps the other to himself. Now Alice has two photons and Bob has one, they are together in the state

$$\begin{aligned} |\Phi_0\rangle &= |\psi\rangle \otimes |\gamma_{00}\rangle = \frac{1}{\sqrt{2}} \left\{ \alpha|0\rangle \otimes (|00\rangle + |11\rangle) + \beta|1\rangle \otimes (|00\rangle + |11\rangle) \right\} \\ &= \frac{1}{\sqrt{2}} \left\{ \alpha(|000\rangle + |011\rangle) + \beta(|100\rangle + |111\rangle) \right\}. \end{aligned} \tag{10.2}$$

For convenience of discussion, let us fix the notations. Hereafter we use the left two qubits represent Alice’s two photons, while the rightmost qubit represents Bob’s photon. For example,  $|101\rangle$  means Alice’s initial photon is in the state  $|1\rangle$ , the photon from the entangled pair is in the state  $|0\rangle$ , and Bob’s photon is in the state  $|1\rangle$ .

Alice next performs a CNOT gate operation on two of her photons—the first photon is the control qubit and the second photon is the target qubit. After this CNOT operation, the state of the three photons becomes

$$|\Phi_1\rangle = \frac{1}{\sqrt{2}} \left\{ \alpha(|000\rangle + |011\rangle) + \beta(|110\rangle + |101\rangle) \right\}. \quad (10.3)$$

Then Alice performs a Hadamard gate operation on the first photon and obtains

$$|\Phi_2\rangle = \frac{1}{2} \left\{ \alpha(|0\rangle + |1\rangle) \otimes (|00\rangle + |11\rangle) + \beta(|0\rangle - |1\rangle) \otimes (|10\rangle + |01\rangle) \right\}. \quad (10.4)$$

There are eight items in total here. We re-arrange them in such a way that Alice's qubits and Bob's qubit are separated with  $\otimes$

$$\begin{aligned} |\Phi_2\rangle = \frac{1}{2} \left\{ |00\rangle \otimes (\alpha|0\rangle + \beta|1\rangle) + |01\rangle \otimes (\alpha|1\rangle + \beta|0\rangle) \right. \\ \left. + |10\rangle \otimes (\alpha|0\rangle - \beta|1\rangle) + |11\rangle \otimes (\alpha|1\rangle - \beta|0\rangle) \right\}. \end{aligned} \quad (10.5)$$

Finally, Alice performs the measurement  $M_z$  on her two photons and tells Bob the result through a classical channel such as a phone call. Upon the measurement, the state  $|\Phi_2\rangle$  will collapse into one of the four possible states<sup>1</sup>. For example, it collapses to

$$|10\rangle \otimes (\alpha|0\rangle - \beta|1\rangle). \quad (10.6)$$

In this case, for the Alice, her two photons are in the state  $|10\rangle$ ; for Bob, his photon is the state  $|\psi_b\rangle = \alpha|0\rangle - \beta|1\rangle$ . To obtain  $|\psi\rangle$ , Bob only needs to perform a  $Z$  gate on his photon. In general, after knowing Alice's measurement outcome, Bob can put his photon to the state  $|\psi\rangle$  by performing a corresponding gate operation:

1. If the outcome is  $|00\rangle$ , then Bob's photon is in the state  $|\psi_b\rangle = \alpha|0\rangle + \beta|1\rangle$ . This is already the state  $|\psi\rangle$ , Bob does nothing.
2. If the outcome is  $|01\rangle$ , then Bob's photon is in the state  $|\psi_b\rangle = \alpha|1\rangle + \beta|0\rangle$ . Bob performs an  $X$  gate on his photon, and obtains  $|\psi\rangle$ .
3. If the outcome is  $|10\rangle$ , then Bob's photon is in the state  $|\psi_b\rangle = \alpha|0\rangle - \beta|1\rangle$ . Bob performs a  $Z$  gate and obtains  $|\psi\rangle$ .
4. If the outcome is  $|11\rangle$ , then Bob's photon is in the state  $|\psi_b\rangle = \alpha|1\rangle - \beta|0\rangle$ . Bob performs an  $X$  gate then a  $Z$  gate, and obtains  $|\psi\rangle$ .

Whatever the outcome of Alice's measurement, Bob can successfully prepare his photon into Alice's original photon state  $|\psi\rangle$ .

The quantum teleportation has the following important features. Firstly, it is the state  $|\psi\rangle$  carried by the photon, instead of the photon itself, that is being transferred. Secondly, it requires Alice and Bob to communicate twice in the entire process, one through the classical channel and the other through the quantum channel. We emphasize that not only the state  $|\psi\rangle$  itself but also any related information never appears in the classical channel and quantum channel. The photon pair in the quantum channel is always in the entangled state  $|\gamma_{00}\rangle$ , completely independent of the

---

<sup>1</sup> According to the many-worlds theory, it splits into four different worlds.

quantum information  $|\psi\rangle$  to be transferred. For Alice, no matter what the state  $|\psi\rangle$  is, there are always four possible measurement outcomes, each with a probability of 25%. Finally, since Alice and Bob need to transmit photons and to communicate on the measurement results, quantum teleportation can not occur faster than the speed of light. This indicates that, although quantum entanglement is non-local, the communication based on it cannot occur faster than the speed of light.

Quantum teleportation was first proposed theoretically by six physicists in 1993. The first experimental quantum teleportation was achieved by an Austrian group in 1997. Since then, laboratories world wide have been devoted to extending the teleportation distance. Now the longest distance for a ground-based quantum teleportation has exceeded 400 km. China was the first to achieve quantum teleportation from a satellite to the ground.

So far, we have described three methods for transferring quantum information, a purely classical communication method, a purely quantum communication method, and quantum teleportation. The purely classical communication method uses only classical channels, while the latter two involve quantum channels. As such, the purely classical communication method is far more stable than the other two. But every coin has two sides. Quantum channels provide a security of communication at the price of decreased stability. We describe below how a quantum channel ensures the safety of communications.

Suppose a third party, named Eve, attempts to eavesdrop on the quantum communication between Alice and Bob, i.e., to access the quantum information exchanged between Alice and Bob without being detected. If the communication between Alice and Bob is classical, where one bit of information is carried by millions of photons, Eve could simply intercept a small fraction of photons and obtain the information, without being noticed by Alice and Bob. In quantum communication, however, one bit of information is carried by a single photon. In order to eavesdrop, Eve must intercept the photon carrying the quantum information  $|\psi\rangle$  in the quantum channel. In order to avoid being detected by Alice and Bob, it is better that Eve copy  $|\psi\rangle$  to another photon and then put the original photon back into the quantum channel. But this operation is fundamentally forbidden by the no-cloning theorem. Eve has no choice but to make the measurement. After Eve's measurement, the photon becomes entangled with the measuring instrument. As a result, the photon state  $|\psi\rangle$  is changed into an eigenstate of some observable and cannot be recovered. As a result, Eve fails the mission: on the one hand, the photon state is changed, which can be easily detected by Alice and Bob; on the other hand, Eve only obtains limited amount of information about  $|\psi\rangle$ . If the communication between Alice and Bob is by quantum teleportation, it is even more difficult for Eve to eavesdrop. This is because when Eve intercepts the photon in the quantum channel, what is intercepted is an entangled photon, which does not contain any information about the quantum state  $|\psi\rangle$ . The information in the classical channel is about the measurement results, which do not provide any specific information about  $|\psi\rangle$ , either. As a result, Eve cannot obtain any information about  $|\psi\rangle$  either by intercepting photons in the quantum channel and listening to the classical communication.

**Table 10.2** Comparison of three different methods of transferring quantum information

Method	Purely classical	Purely quantum	Quantum teleportation
Knowledge of the state	Yes	No	No
Classical channel	Yes	No	Yes
Quantum channel	No	Yes	Yes
Security level	Low	High	High
Stability	High	Low	Medium

The comparison of the above three state transfer methods, the purely classical communication method, the purely quantum communication method, and quantum teleportation, is summarized in Table 10.2.

### 10.3 Classic Encryption

To avoid our phone call being overheard, usually we try to find a place, which is far away from everyone, or speak a hard-to-understand dialect. To protect against an email hack, you can set up a long and hard-to-guess password. These simple tricks satisfy our basic need for protecting the privacy and confidentiality in our daily lives. But for a professional spy, these tricks are useless. Counterintelligence agencies can easily tap your phone and go to your service provider with a search warrant to access your email. Spies encrypt their messages systematically to ensure the confidentiality of their communications. Being “systematic” is important. If the spy is not required to communicate with the headquarters during the mission, except reporting the result of the mission, then he only needs a simple code to signal success or failure. In this case, the other side can not do much to break the code. In general, however, the spy has to stay in long term contact with the headquarter, reporting information and unpredictable situations. In this case, the spy can only use systematic encryption to encrypt the report and then transfer it back to the headquarters. Since it is systematically encrypted, there is a pattern. Counterintelligence agencies can hire many clever experts to break your codes by analyzing the patterns of your communications. Here is a simple example. Suppose you communicate in English. A simple encryption method is to convert letter a into b, b into c, c into d, and so on. After intercepting your communication, counterintelligence agencies analyze the frequency at which a letter appears in your correspondence. Soon, they will find that b appears almost as frequent in your correspondence as a in newspaper articles, c appears almost frequent as b, and so on. Thus your code is broken. Nowadays, information encryption is no longer limited to espionage and military, but is also used in our daily lives. Whenever you make a purchase online, your account and purchase information are encrypted before being transferred to the credit card companies or any payment company.

**Table 10.3** Encrypted communication with Vernam cipher

Message	Letter	q	u	a	n	t	u	m
	ASCII	113	117	97	110	116	117	109
Random key		014	013	000	031	000	012	010
Encrypted message	Letter	c	h	a	O	t	i	c
	ASCII	099	104	097	079	116	105	099

After decades of research, it is found that Vernam cipher (also called one-time pad), which is invented by Vernam (Gilbert Sandford Vernam, 1890–1960) in 1917, is an unbreakable encryption technique. Let us briefly illustrate how Vernam cipher works. Alice and Bob want to encrypt their communication, and they decide to use Vernam cipher. They randomly generate a long key and each of them keeps a copy. Now Alice wants to send the word “quantum” to Bob. To encrypt it, Alice first converts the word into the familiar ASCII code {113, 117, 97, 110, 116, 117, 109}; then she subtracts from them the first seven numbers in her copy of the key {014, 013, 000, 031, 000, 012, 010}, which resulted in a new string of numbers {099, 104, 097, 079, 116, 105, 099}; finally, she sends these numbers to Bob using a public communication channel. When Bob receives these numbers, he adds them with the first 7 numbers in his copy of the same key and then converts them back to letters by referring to ASCII codes. The whole process is shown in Table 10.3.

If Eve is eavesdropping, she will obtain the string of numbers {099, 104, 097, 079, 116, 105, 099}, which, according to ASCII codes, translates to “chaOtic”. Of course, Eve knows that this is not Alice’s real message, and that the capital letter “O” is like Alice is challenging her, “Welcome to crack my code!” Eve would try all out to figure out what the numbers mean. As Alice and Bob is using the Vernam cipher, where the key is randomly generated, any piece of the key once used will be discarded. Therefore, there is no pattern in the key, and the only way for Eve to decipher successfully Alice’s message is to guess. For the above example, the probability for Eve to correctly guess all the seven 3-digit numbers right in the key is  $10^{-21}$ . To have a feeling for how small this number is, imagine you have a grain of sand marked with some symbol. You accidentally drop it on a beach, which is about one-kilometer long. The probability that you find the marked sand is at least 10 million times higher than the probability for Eve to make a correct guess.

While Vernam cipher and other similar encryption methods are widely used in military and espionage, they are not useful for commercial purpose. If a credit card company use Vernam cipher to encrypt their customers’ credit card information, the company needs to assign a different key to each customer. As a customer is encouraged to use the credit card frequently, the key would have to be very long, and at the same time be kept safely by the customer. This is practically impossible for a normal person who has no rigorous training in espionage.

Modern cryptosystem for commercial use is called public-key cryptography. It works like this: the credit card company generates a pair of keys—a public key and a private key. The company keeps the private key securely while the public key is openly distributed, which is available to everyone (including criminals). When a customer makes a purchase with his credit card, the shop uses the public key to encrypt his credit card and purchase information and transfers it back to the credit card company through the public network. Such an encrypted information is eventually decoded by the credit card company with the securely-kept private key. The crucial feature in such a public-key cryptography is that the encryption with the public key is a one-way process: one can easily encrypt information with the public key but no one can easily decode the encrypted information with the public key. Only the credit card company who has the private key can easily decode it. This is similar to the mailboxes on the streets: on the one hand, it is easy to put your envelope in through the narrow slot but difficulty to get it out from the same slot; on the other hand, the mailman who has the key to mailbox can easily open the mailbox door and take the envelope out.

RSA is a widely used public-key cryptosystem. The acronym “RSA” comes from the names of its inventor Rivest (Ron Rivest, 1947–), Shamir (Adi Shamir, 1952–) and Adleman (Leonard Adleman, 1945–). The key idea of RSA is to create a pair of keys, the public key  $(N, e)$  and the private key  $(N, d)$ . Here  $N$  is a product of two large prime numbers; both  $d$  and  $e$  are two positive integers, which are chosen in connection with these two prime numbers. The public key  $(N, e)$  is announced and widely distributed; the private key is securely kept by the company. The security of RSA relies on the difficulty to obtain these two large primes by factoring  $N$ . RSA only requires the credit card company to secure the two prime numbers and the private key  $d$ . The credit card company can even destroy the two prime numbers after the key is created and keep only the private key  $d$ . The customer does not even need to know the public key (in fact, most of customers do not even know the existence of such a public key). When a purchase is made with a credit card, the POS (point of sale) machine at the cashier encrypts the purchase information with the public key stored in the machine and transfers it to the credit card company.

Now let us see if quantum communication can be useful in the context of encryption.

## 10.4 Quantum Key Distribution

We have already mentioned that the encryption technique called Vernam cipher is unbreakable. However, it has a shortcoming: the key is as long as the message. A spy working for a long time away from the headquarter has to keep a very long key. Once the key is lost, the spy has to go back to the headquarter to get a new key. Quantum communication offers a secure way to generate and distribute the key over distance. In the 1980s, inspired by the early work of Wiesner (Stephen J. Wiesner, 1942–), Bennett (Charles Henry Bennett, 1943–) and Brassard (Gilles Brassard, 1955–) proposed the first feasible quantum key distribution protocol, which is now



called BB84. Other similar protocols have been proposed. All these protocols, albeit vary in details, rely on similar principles and steps. Alice and Bob use this kind of protocols to randomly generate a Vernam cipher and then use it to encrypt their classical communication.

In the BB84 protocol, Alice uses quantum teleportation to transfer a sequence of photon polarization states to Bob. Alice publicly announces that these polarization states are chosen from the following four,

$$|\varphi_{00}\rangle = |0\rangle, \quad (10.7)$$

$$|\varphi_{10}\rangle = |1\rangle, \quad (10.8)$$

$$|\varphi_{01}\rangle = |0_x\rangle = (|0\rangle + |1\rangle)/\sqrt{2}, \quad (10.9)$$

$$|\varphi_{11}\rangle = |1_x\rangle = (|0\rangle - |1\rangle)/\sqrt{2}. \quad (10.10)$$

But exactly which polarization state is transmitted each time is random and confidential. After Bob receives these polarization states, he measures them randomly in two ways (either  $M_z$  or  $M_x$ , see Sect. 10.1). Then he discusses and compares the measurement results with Alice through the classical channel. Finally, a series of good bits are chosen as the key. Note the subscript of the polarization state in the above equation: this binary notation is crucial for understanding step 4 of the BB84 scheme below.

We illustrate the BB84 protocol with the following example, and the goal is to create a short binary key.

1. Alice randomly generates two 9-bit binary numbers  $a$  and  $b$ , where  $a$  is always kept secret and  $b$  is temporarily kept secret. We use  $a_1, a_2, \dots, a_9$  and  $b_1, b_2, \dots, b_9$  to denote the digits of  $a$  and  $b$ , respectively. For example, in Table 10.4,  $a_2 = 0$ ,  $b_7 = 1$ .
2. Using the quantum teleportation, Alice sends a sequence of polarization states  $|\varphi_{a_k b_k}\rangle$  to Bob according to the digits of  $a$  and  $b$ . For example, in Table 10.4,  $a_1 = 1, b_1 = 0$ , so the first polarization state Alice sent to Bob is  $|\varphi_{10}\rangle = |1\rangle$ . In this way, Alice sends a sequence of 9 polarization states to Bob according to this table.
3. Bob generates a random 9-bit binary number  $b'$ , which decides how he measures the photon polarization states: if  $b'_k = 0$ , he performs  $M_z$  measurement; if  $b'_k = 1$ , he performs  $M_x$  measurement. In Table 10.4,  $b'_1 = 1$ , so Bob performs  $M_x$  measurement on the first photon state;  $b'_2 = 0$ , so Bob performs  $M_z$  measurement on the second photon state, and so on. According to Table 10.1, Bob records the measurement outcomes and obtain a sequence of 0s and 1s, which make up another 9-bit binary number  $a'$ .
4. Alice announces  $b$ , and Bob compares it with  $b'$ . If  $b_k = b'_k$ ,  $a'_k$  is retained, otherwise  $a'_k$  is abandoned. Afterwards, Bob tells Alice the values of  $k$  at which  $b_k = b'_k$  through a public classical communication channel. Then Alice retains the corresponding  $a_k$ . Kind of magically, for these  $k$ 's, we always have  $a_k = a'_k$ . The retained  $a_k$  (or  $a'_k$ ) is the key. For the example shown in Table 10.4, the retained  $a_k$  and  $a'_k$  (in bold font) are the same 4-bit binary number 0100.

**Table 10.4** BB84 quantum key distribution scheme

a	1	0	1	1	1	0	0	0	1
b	0	0*	1	0*	0	1*	1	0*	1
Polarization state	$ \varphi_{10}\rangle$	$ \varphi_{00}\rangle$	$ \varphi_{11}\rangle$	$ \varphi_{10}\rangle$	$ \varphi_{10}\rangle$	$ \varphi_{01}\rangle$	$ \varphi_{01}\rangle$	$ \varphi_{00}\rangle$	$ \varphi_{11}\rangle$
b'	1	0*	0	0*	1	1*	0	0*	0
Measurement	$M_x$	$M_z$	$M_z$	$M_z$	$M_x$	$M_x$	$M_z$	$M_z$	$M_z$
a'	1	0	0	1	0	0	0	0	1

\*marks where  $b_k = b'_k$ . The corresponding  $a_k$  and  $a'_k$  are labeled in bold

Let us analyze the last step, i.e., the fourth step. Due to the clever numbering of the four polarization states in Eq. (10.10), the measurement has a definite outcome when  $b_k = b'_k$  and the outcome  $a'_k$  is always the same as  $a_k$ . As an example, in Table 10.4,  $b_6 = b'_6 = 1$ , so Bob carries out measurement  $M_x$ . Because the corresponding polarization state is  $|\varphi_{01}\rangle = |0_x\rangle$ , the outcome is 100% 0 (see Table 10.1), i.e.,  $a'_6 = 0$ , which is the same as  $a_6 = 0$ . When  $b_k \neq b'_k$ , the probability to obtain  $a_k = a'_k$  is 50%. In a miraculous way, Bob is able to partially know  $a$  by quantum measurements through a combination of quantum and classical communication with Alice, even though he knows nothing *a priori* about  $a$ .

It is clear that the encryption key generated with the BB84 protocol is random and is known only to Alice and Bob. Anyone, for example, Eve, who is interested in this key, can not learn anything about the key by monitoring the communication between Alice and Bob. The number  $a$ , where the encryption key comes from, is kept secret by Alice. The quantum communication is done via the quantum teleportation; Eve has no chance to gain any knowledge of the polarization states being transferred. By listening to the classical communication between Alice and Bob, Eve would know the number  $b$  and which binary digits of  $a$  are kept. The former has nothing to do with  $a$ ; the latter is meaningless if you do not know  $a$ . In short, BB84 is a very secure protocol to distribute a key.

When we introduced the BB84 scheme above, we assumed that  $a$  and  $b$  have only 9 digits, and we ended up with a 4-digit encryption key. Obviously,  $a, b$  can be any positive integers. In general, in order to get an  $n$ -bit binary key,  $a, b$  is chosen to be a binary number with  $4n + \delta$  bits. Here  $\delta$  is usually a large number, the value of which is case dependent. Why choose  $4n + \delta$  bits? It is for the reason of security and noise-resistance. When Alice uses the quantum teleportation to transfer photon states to Bob, noise in the quantum channel, or Eve's eavesdropping, may partially or completely destroy the entanglement between the photon pair, which can result in Bob getting a different photon polarization state from Alice's. As a consequence, even if  $b_k = b'_k$ , there may not be  $a_k = a'_k$ . In order to assess how much damage the noise or Eve's eavesdropping actually cause, Alice and Bob could randomly select another  $n$  numbers from the remaining  $2n$   $a_k$  and  $a'_k$ , and compare them through the classical channel. If they agree, the other half  $a_k$  and  $a'_k$  are retained; otherwise they are abandoned and the whole process starts over.

As can be seen from the above description, the quantum key distribution, which simply creates and distributes Vernam ciphers, does not affect the RSA scheme that

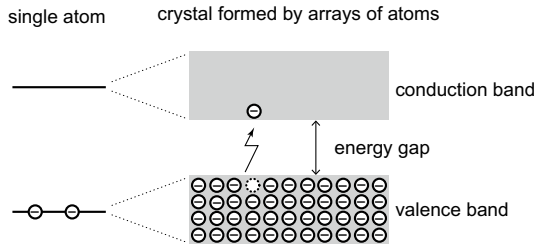
is widely used for commercial use. The advantage of the quantum key distribution is that it allows users in separate places to create and share an encryption key without any close physical contact.

## 10.5 Future Quantum Technologies

In Chap. 1, we have classified quantum technologies into two groups, implicit quantum technology and explicit quantum technology. Implicit quantum technology, in principle, can be realized by classical technologies; in contrast, explicit quantum technology, in principle, is infeasible with any classical technology. Chip technology is a typical implicit quantum technology; quantum computer is a typical explicit quantum technology. These two types of technologies are not competing, but rather complementary and mutually beneficial. Modern classical computers and classical communications rely on many implicit quantum technologies, and they will continue to develop and will never be replaced by quantum computers and quantum communications. History is a great guide for future. Let us look back, trying to extrapolate from history and, hopefully, catch a glimpse of future quantum technologies.

We review two examples of well-developed quantum technologies, semiconductor technology and magnetic resonance imaging (MRI) technology. We first take a look at semiconductor technology. As is well known, metals conduct electricity, but gemstones, like diamond, do not. Physicists find that the conductive nature of these materials can only be explained by quantum mechanics. As said before, energy levels of the electron in a hydrogen atom are discrete. i.e. there are “gaps” between the energy levels. Other atoms have similar discrete energy levels. When arrays of these atoms form a crystal, these discrete energy levels are broadened into energy bands. There, some “gaps” disappear whereas some “gaps” are retained, which are called energy gap in crystalline materials. Because electrons are fermions, there is a limit on the number of electrons that can fill an energy band. Starting from the lowest energy band, electrons in a crystal fill the energy bands one by one until there are no more electrons. For a conductor such as a metal, the electrons with the highest-energy only partially fill an energy band. In this case, the electrons can easily participate in the conductivity. For an insulator, the electrons will fill all the energy bands below a certain energy gap. Therefore, electrons must acquire energy to overcome this energy gap in order to be conductive (see Fig. 10.3). Physicists have further discovered semiconductors, whose electrical conductivity falls between that of a conductor and an insulator. A semiconductor also has an energy gap, but the gap is relatively small. The conducting properties of a semiconductor can be easily altered with various methods, switching quickly between the conductive and insulating. Based on this unique property of semiconductors, physicists invented the transistor in 1947, which is the beginning of modern semiconductor technology. Now a single computer chip consists of billions of transistors.

MRI technology is physically based on the resonant interaction of spin and light (or electromagnetic radiation) in an external magnetic field. This technique is now



**Fig. 10.3** Energy bands of an insulator or a semiconductor. Individual atom has discrete energy levels. When they form a crystal, these energy levels are broadened into energy bands. There are two types of energy bands: if the electrons in the band are involved in conducting electricity, the band is called the conduction band; if the electrons in the band are not involved in conducting electricity, the band is called the valence band. In a semiconductor, the energy gap does not exceed 3 eV

widely used for medical diagnosis. As hydrogen atoms are naturally abundant in human tissues, particularly in water and fat, MRI uses the spin of hydrogen nuclei, which consists of a single proton. In a magnetic field, the two nuclear spin states along the direction of the magnetic field have different energies,  $E_+ = \mu_b B$  and  $E_- = -\mu_b B$ , where  $\mu_b$  is the proton magnetic moment, and  $B$  is the strength of magnetic field (see Sect. 6.4). The nuclear spin with energy  $E_-$  can transit to the energy level  $E_+$  by absorbing a photon. Conversely, the nuclear spin with  $E_+$  can emit a photon and transit to the energy level  $E_-$ . Both the absorbed and emitted photons have a frequency  $\nu = (E_+ - E_-)/h$ . This is the physics behind MRI technology. As the proton magnetic moment is very small, MRI requires a very strong magnetic field, so that the absorbed or emitted photons correspond to ordinary radio waves, which can be detected with state of the art techniques. During the diagnosis, the states of nuclear spins are controlled by external instruments, which emit electromagnetic waves to excite the nuclear spins; the excited nuclear spins then emit electromagnetic signals, which are detected by a small antenna nearby. In order to detect the positions of the electromagnetic signals, a nonuniform magnetic field is used, i.e.  $B$  varies in space. As a result, the frequencies  $\nu$  of emitted signals at different places are different, allowing the antenna to locate the nuclear spins. The environment surrounding a nuclear spin induces spin relaxation; different environment exhibits different relaxation behavior. Therefore, different contrasts may be generated between tissues in MRI, based on the relaxation properties of the nuclear spins.

The two technologies described above harness different quantum effects. But all these effects are single-particle physics in nature, and do not depend on the coherence of quantum states. By solving the Schrödinger equation, physicists know that an electron in a periodic lattice exhibits a band structure and an energy gap. There are many electrons in a semiconductor material and there are interactions among electrons, which affect the quantum state (or wave function) of electrons to some extent. However, the effect of interaction does not change the energy band structure of the material and has limited influence on the function of materials. Defects and

impurities in semiconductor materials can also affect the electron wave function, but their effect is not substantial as long as their amount is small. In addition, the function of a semiconductor material is only related to how the energy bands are filled by electrons, which does not relate to the coherence of the electron wave function. In MRI technology, effects of interactions between nuclear spins are very weak and negligible; therefore, it is only necessary to manipulate the single-spin quantum state. The resonance of a single spin under electromagnetic excitation is in principle a coherent quantum phenomenon, but the surrounding environment of nuclear spins causes decoherence, which is the relaxation phenomenon mentioned earlier. MRI techniques make clever use of decoherence to create images of human tissues.

The semiconductor and MRI technologies are just typical examples of all current mature implicit quantum technologies. They are based on single-particle quantum effects, and rely on the manipulation of single-particle wave functions (or quantum states); most of them do not pertain to the coherence of these single-particle quantum states. The most sophisticated quantum technology so far, which relies on the coherence of single-particle wave functions, is the laser technology. Although there are vast numbers of photons in a laser beam, there is no entanglement between these photons. All photons in the beam are in the same quantum state, whose polarization, spatial distribution, and phase can be precisely controlled in modern laser technology. The world's most precise clock builds on the laser coherence.

In contrast, quantum information technology requires a precise control of a quantum many-particle state (or wave functions) while maintaining its coherence. Previously, we have seen that the quantum teleportation involves precise manipulations of the polarization states of three entangled photons. It is similar for quantum computing, where many qubits are entangled and have to remain coherent while being operated on by quantum logic gates. As such, we can also classify quantum technologies into two groups from a different perspective: single-particle quantum technology and many-body quantum technology. Single-particle quantum technology still has lots of room for improvement. But eventually, a journey toward many-body quantum technologies is inevitable and its ultimate goal is to build a useful quantum computer as it represents an ability to precisely manipulate every details of a many-body wave function and to accurately control every step of its evolution in a Hilbert space. To achieve this goal may be the biggest and most complex technological challenges that we human have ever faced, more difficult than controlling nuclear fusion. In this long marathon, each challenge that we overcome will be a small incremental success along the way. Eventually, there will be a giant leap—when quantum computing becomes a reality.

## Further Reading

For readers interested in continuing to explore the quantum world, here is a list of books and papers that you may find useful and interesting.

### • General

1. Wilczek, F. (2016). *A beautiful question*. Penguin. This is a popular science book. You can read the section on quantum physics directly. There is an appendix at the end of the book with easy-to-understand explanations of various technical terms in physics, which can be used as a toolkit.
2. Susskind, L. (2014). *Quantum mechanics*. Basic Books. This book also requires only rudimentary knowledge of mathematics and physics, similar to this book .
3. Feynman, R. *Lectures on physics*. Addison-Wesley (the part on quantum physics). This is the lecture notes of Feynman's class for the first-year undergraduates at Caltech, so it is not too demanding on mathematics.
4. Dirac, P. A. M. (1958). *The principles of quantum mechanics*. Oxford University Press. It was written by Dirac for professionals, but there is discussion at the beginning of the book that does not involve advanced mathematics.
5. von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton University Press. In this book, von Neumann pointed out for the first time that the quantum world lives in Hilbert spaces. He also discussed in detail a theory of quantum measurement, at the core of which is the collapse of a wave packet, setting the standard for all subsequent quantum mechanics textbooks on quantum measurement. This book is for professionals.
6. Messiah, A. (1999). *Quantum mechanics*. Dover. This is a comprehensive textbook for professionals.

## • Quantum History

Here is a list of books that I have gathered historical materials for Chap. 2.

1. Kragh, H. *Quantum generations*. Princeton University Press.
2. Cassidy, D. D. (2009). *Beyond uncertainty: Heisenberg, quantum physics, and the bomb*. Bellevue Literary Press.
3. Wali, K. C. (2009). *Satyendra Nath Bose: His life and times*. World Scientific.
4. Farmelo, G. (2009). *The Strangest man: The hidden life of Paul Dirac, mystic of the atom*. Basic Books.
5. Moore, W. J. (1989). *Schrödinger: Life and thought*. Cambridge University Press.

## • Classical Mechanics

If you are not familiar with classical mechanics, here are two books to start.

1. Susskind, L. (2014). *Classical mechanics*. Basic Books. This book requires only rudimentary knowledge of mathematics and physics, similar to this book .
2. Feynman, R. (1963). *Lectures on physics*. Addison-Wesley (the part on classical physics). This is the lecture notes of Feynman’s class for the first-year undergraduates at Caltech, so it is not too demanding on mathematics.

## • Quantum Computing

1. Deutsch, D. (1997). *The fabric of reality*. Penguin Books. This is a popular science book. There is a beautiful discussion on computation in general and quantum computation in particular. It also describes vividly the many-worlds theory to the general public.
2. Nielsen, M., & Chuang, I. (2000). *Quantum computation and quantum information*. Cambridge. This is a comprehensive book on quantum information and has been widely regarded as a “bible” in this field. It is written for professionals. However, its Chap. 1 is accessible to non-professionals. And its Sect. 2.1 on linear algebra is readily accessible to many and can be used as a reference for linear algebra introduced in this book.
3. Wu, B. (2021). *Classical computer, quantum computer, and the Gödel’s theorem*. arXiv:2106.05189 (2021); collected in Wilczek, F. (2022). *50 years of theoretical physics* (pp.281–290). World Scientific. In this article, I point out that the fundamental difference between classical information and quantum information is that the former is cloneable and the latter is uncloneable. Any object (man-made machine or brain) that processes classical information is a classical computer and any object that processes quantum information is a quantum computer.

## • Papers

Here is a list of important papers that are discussed and mentioned in this book.

1. Planck, M. (1900). Über eine Verbesserung der Wienschen Spektralgleichung [English translation: On an Improvement of Wien's equation for the spectrum], *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 2, 202–204; Ueber das Gesetz der Energieverteilung in Normalspectrum [English translation: On the law of distribution of energy in the normal spectrum]. *Annalen der Physik*, 4, 553–563 (1901). In the first paper, Planck obtained the correct law of black-body radiation without introducing the concept of quanta, which was introduced in the second paper by Planck. Planck also defined two fundamental constants, the Planck constant and Boltzmann constant and found their values in the second paper for the first time.
2. Einstein, A. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt [English translation: Concerning an heuristic point of view toward the emission and transformation of light]. *Annalen der Physik*, 17, 132–148 (1905). In this paper, Einstein proposed the concept of photon (or light quanta) and used it to explain the photoelectric effect.
3. Bohr, N. (1913). On the constitution of atoms and molecules. *Philosophical Magazine*, 26, 1–24; vol. 26: 476–502 (1913). In these two papers, Bohr proposed his model of a hydrogen atom, the first quantum model of atoms.
4. Bose, S. N. (1924). Plancks Gesetz und Lichtquantenhypothese [English translation: Planck's law and light quantum hypothesis] *Zeitschrift für Physik*, 26, 178–181. In this paper, Bose pointed out that photons are indistinguishable from each other.
5. Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen [English translation: On quantum theoretical reinterpretation of kinematic and mechanical relations]. *Zeitschrift für Physik*, 33, 879–893. The matrix mechanics was first proposed in this paper.
6. de Broglie, L.-V. (1925). Recherches sur la théorie des quanta [English translation: Researches on the quantum theory]. Thesis, Paris, 1924, *Annales de Physique*, (10)3, 22. In this paper, de Broglie proposed that an electron is also a wave.
7. Schrödinger, E. (1926). Quantisierung als Eigenwertproblem (Erste Mitteilung) [English translation: Quantisation as an eigenvalue problem (first communication)]. *Annalen der Physik*, 79, 361–376; Quantisierung als eigenwertproblem (Zweite Mitteilung) [English translation: Quantisation as an eigenvalue problem (second communication)]. *Annalen der Physik*, 79, 489–527 (1926); Quantisierung als eigenwertproblem (Dritte Mitteilung: Störungstheorie, mit Anwendung auf den Starkeffekt der Balmerlinien) [English translation: Quantisation as an eigenvalue problem (third communication: perturbation theory, with application to the Stark effect of Balmer lines)]. *Annalen der Physik*, 80, 437–490 (1926); Quantisierung als eigenwertproblem (Vierte Mitteilung) [English translation: Quantisation as an eigenvalue problem (fourth communication)]. *Annalen der Physik*, 81, 109–139 (1926). The foundation of the wave mechanics was laid in these four papers.
8. Gerlach, W., & Stern, O. (1922). Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld [English translation: the experimental proof



- of the directional quantization in the magnetic field]. *Zeitschrift für Physik*, 9, 349–352. The Stern-Gerlach experiment was first reported in this paper.
9. Park, J. (1982). The concept of transition in quantum mechanics. *Foundations of Physics*, 1, 23–33. Wootters, W., & Zurek, W. (1982). A single quantum cannot be cloned. *Nature*, 299, 802–803. These are two earliest and independent proofs of the no-cloning theorem.
  10. Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik [English translation: On the descriptive content of quantum theoretical kinematics and mechanics. *Zeitschrift für Physik*, 43, 172–198. Kennard, E. H. (1927). Zur Quantenmechanik einfacher Bewegungstypen [English translation: On the quantum mechanics of simple types of motion]. *Zeitschrift für Physik*, 44, 326–352. The Heisenberg uncertainty relation was first proposed and proved in these two papers.
  11. Bell, J. S. (1964). On the Einstein Podolsky Rosen Paradox. *Physics*, 1, 195–200. The proof of Bell’s inequality.
  12. Everett, H. (1973). The theory of the universal wave function, collected in *The many-worlds interpretation of quantum mechanics*. Princeton University Press. B. S. DeWitt & N. Graham (Eds.), pp. 3–140; “Relative State” of formulation of quantum mechanics. *Reviews of Modern Physics*, 29, 454–462; Wu, B. (2021). Everett’s theory of the universal wave function. *The European Physical Journal H*, 46, 7. The first paper is Everett’s long thesis, where the many-worlds theory is systematically presented and clearly elucidated. The second paper is a short version of the long thesis with many key elements left out. The third paper is also a short version the long thesis but with only the detailed mathematical derivation left out.
  13. Manin, Y. I. (1980). Vychislimoe i nevychislimoe [Computable and noncomputable] (in Russian). *Sov. Radio*. 13–15; Feynman, R. Simulating physics with computers. *International Journal of Theoretical Physics*, 21, 467–488.
  14. Deutsch, D. (1985). Quantum theory, the Church-turing principle and the universal quantum computer. *Proceedings of the Royal Society of London A*, 400: 97–117; Quantum computational networks. *Proceedings of the Royal Society of London. Series A*, 425, 73–90 (1989).
  15. Bennett, C. H., Brassard, G., Crépeau, C., Jozsa, R., Peres, A., & Wootters, W. K. (1993). Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70, 1895–1899. Quantum teleportation was first proposed in this paper.
  16. Bennett, C. H., & Brassard, G. (1984). Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* (Vol. 175, p. 8). The quantum key distribution protocol BB84 was proposed here.
  17. Shor, P. W. (1994). Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (pp. 124–134). IEEE Computer Society Press.

# Index

## A

Addition of complex numbers, 43  
Addition of imaginary numbers, 42  
Adleman, 174  
Angular momentum, 65  
Antiparticle, 29  
Argument of a complex number, 42, 44

## B

Balmer, 14  
Basis, 49  
Basis vector, 49  
BB84, 175  
Bell, 107  
Bell's inequality, 5, 71, 108, 111–113  
Benioff, 143  
Bennett, 174  
Big Bang, 91  
Binary numbers, 144  
Bit, 145  
Black body, 10  
Black-body radiation, 10  
Bohm, 137  
Bohmian mechanics, 137  
Bohr, 17, 23, 25, 29, 30, 131, 137  
Bohr's model of hydrogen atom, 17, 18, 27, 29  
Bohr-Sommerfeld quantization rule, 39  
Bohr-Sommerfeld theory, 18, 38  
Bohr-Van Leeuwen theorem, 17  
Boltzmann's constant, 11  
Born, 19, 23, 25, 29, 30  
Bose, 20, 30  
Bose-Einstein condensation, 21  
Boson, 24, 66  
Brassard, 174

## C

Canonical correlation, 131  
Carbon fullerene, 148  
Classical channel, 168  
Classical communication, 6, 168  
Classical computer, 6, 144  
Classical interference, 99  
Classical probability, 68, 69, 71, 113  
Classical state, 156  
Classical technology, 6  
CNOT gate, 149, 156  
Collapse of a wave function, 124, 125, 142  
Column vector, 46  
Commutator, 76  
Complex conjugate, 44  
Complex number, 42  
Complex plane, 42  
Copenhagen interpretation, 118, 129, 130, 137  
Correlation at distance, 5  
Correspondence principle, 18  
Counting rod, 41

## D

De Broglie, 26, 29, 137  
Decoherence, 86, 162, 163, 179  
Degenerate, 60  
Degree of freedom, 66  
Density matrix, 116  
Destructive quantum interference, 99  
Deutsch, 143  
DeWitt, 131, 136  
Diagonal element, 57  
Diagonal matrix, 58  
Dice, 64, 68, 70, 125  
Dirac, 19, 23, 24, 29, 66, 81, 124

Dirac notation, 48  
 Dirac's equation, 29  
 Direct product, 102, 61, 147  
 Division of complex numbers, 43  
 Dot product, 45, 47  
 Double-slit interference, 96  
 Double-spin operator, 104, 107  
 Double-spin Stern-Gerlach experiment, 105, 108  
 Dynamics, 81

**E**

Eigenenergy, 87  
 Eigenfunction, 39, 90  
 Eigenstate, 60  
 Eigenvector, 60  
 Einstein, 12, 13, 21, 26, 28, 70, 137, 138  
 Electronic orbital, 91  
 Electron radius, 92  
 Energy band, 177  
 Energy eigenstate, 87  
 Energy gap, 177  
 Energy level, 18, 87  
 Entangled state, 104, 105  
 Euler, 42  
 Everett, 125, 128, 131  
 Expectation value, 77, 119, 121  
 Explicit quantum technology, 6, 177

**F**

Fermi, 22, 23, 29  
 Fermion, 24, 66  
 Fermi temperature, 130  
 Feynman, 138, 143, 147  
 Fredkin, 155  
 Fredkin gate, 150, 151  
 Free fall, 31  
 Free will, 136

**G**

Galileo, 31  
 Geiger, 16  
 General-purpose quantum computer, 161  
 Gerlach, 63  
 Ground state wave function, 91  
 Grover, 159  
 Grover's algorithm, 159

**H**

Hadamard gate, 148, 156

Hamilton, 85  
 Hamiltonian, 34, 85  
 Hamiltonian operator, 82, 84, 87  
 Harmonic oscillator, 37  
 Heisenberg, 24, 29, 81, 122  
 Heisenberg equation, 81, 87  
 Heisenberg's uncertainty relation, 3, 77, 90, 118, 119, 121, 124, 128, 137  
 Hermitian conjugate, 58  
 Hermitian matrix, 58, 60  
 Hero of Alexandria, 42  
 Hidden variable, 70  
 Hidden variable theory, 70, 113, 137  
 Hilbert space, 47, 48  
 Hippasus, 42  
 Hydrogen atom, 1, 14, 24, 91, 92, 135  
 Hydrogen molecule, 92, 93

**I**

Identical particle, 4, 20  
 Identity matrix, 58  
 Imaginary number, 42  
 Implicit quantum technology, 6–8, 177  
 Inner product, 47, 49, 51, 103  
 Interference term, 98, 140  
 Intrinsic angular momentum, 65  
 Inverse matrix, 58  
 Irrational number, 41  
 Isoenergetic line, 34

**J**

Jordan, 25

**K**

Kelvin, 14  
 Kinetic energy, 33  
 Kirchhoff, 9  
 Kramers, 18

**L**

Landau, 119  
 Langevin, 26  
 Light quanta, 12  
 Lightyear, 1  
 Linear operator, 56  
 Linear space, 46  
 Linear transformation, 56  
 Logical bit, 162  
 Logic gate, 145  
 Lorentz, 12, 13  
 Loss of individuality, 5, 101, 114

**M**

Mach, 14, 22  
 Magnetic moment, 88  
 Manin, 143, 147  
 Many-body quantum technology, 179  
 Many-worlds theory, 118, 128, 134  
 Marsden, 16  
 Matrix, 53  
 Matrix equation, 87  
 Matrix multiplication, 54, 55  
 Maxwell, 165  
 Mixed state, 116  
 Modulus of a complex number, 42, 44  
 Momentum, 33  
 Momentum operator, 79, 84  
 MRI technology, 88, 177  
 Multiplication of complex numbers, 43  
 Multiplication of imaginary numbers, 43  
 Multiplication of matrix and vector, 53

**N**

Nernst, 13  
 Newton, 32  
 Newtonian mechanics, 85  
 Newton's second law, 81, 87, 94  
 No-cloning theorem, 94, 95, 171  
 Noncommutativity, 55–57  
 Noncommutativity of operators, 119, 122–124  
 Nonlocal correlation, 5, 101, 106, 111, 112  
 Normalization, 60  
 Normalization condition, 67, 78  
 NP-complete problem, 160

**O**

Observable, 72, 78  
 Off-diagonal matrix element, 57  
 One-time pad, 173  
 Operator, 72  
 Orbital angular momentum, 65  
 Orthogonal, 49  
 Orthonormal basis, 49

**P**

Parity of an integer, 158  
 Partial differential equation, 82  
 Pauli, 22, 23, 29  
 Pauli exclusion principle, 23  
 Pauli matrix, 59  
 Perpendicular, 49  
 Phase, 44

Phase gate, 148  
 Phase space, 34, 67  
 Photoelectric effect, 13, 26  
 Physical bit, 162  
 $\pi/8$  gate, 148, 150  
 Planck, 1, 2, 9, 13, 28  
 Planck cell, 40  
 Planck's constant, 11, 82, 128  
 Planck's law of black-body radiation, 10, 13, 21  
 Polarization state of photon, 166  
 Position operator, 79  
 Positron, 29  
 Potential energy, 33  
 Probabilistic interpretation, 78  
 Product quantum gate, 156  
 Product state, 103, 114  
 Pythagorean, 42

**Q**

Quantized orbit, 17, 27, 39, 91  
 Quantum, 1, 3, 6  
 Quantum algorithm, 147, 152  
 Quantum bit, 67, 148  
 Quantum channel, 168  
 Quantum circuit, 151  
 Quantum-classical correspondence, 82, 87  
 Quantum cloning, 94, 95  
 Quantum communication, 6, 86, 165, 168  
 Quantum computation, 86  
 Quantum computer, 6, 148, 156, 158  
 Quantum dot, 2  
 Quantum entanglement, 5, 101, 128, 129, 131  
 Quantum gate, 148  
 Quantum indistinguishability, 20  
 Quantum information technology, 86  
 Quantum interference, 98, 99  
 Quantum key distribution, 174  
 Quantum logic gate, 148  
 Quantum measurement, 117, 118, 141  
 Quantum probability, 68, 69, 71, 113  
 Quantum randomness, 4  
 Quantum relay, 168  
 Quantum state, 68, 77  
 Quantum teleportation, 168, 171, 179  
 Quantum theory, 11, 77  
 Quantum Turing machine, 143  
 Qubit, 67, 86, 148, 165

**R**

Radius of a hydrogen atom, 92

Radius of a proton, 92  
 Random search, 159  
 Rational number, 41  
 Reversibility, 155  
 Reversible classical computer, 155  
 Rigidity of wave function, 92  
 Rivest, 174  
 Row vector, 46  
 RSA public-key cryptosystem, 174  
 Rutherford, 16  
 Rydberg, 14  
 Rydberg formula, 15

**S**

Scalar, 45  
 Scalar product, 45  
 Schrödinger, 27, 30, 81, 91  
 Schrödinger equation, 28, 39, 81, 84  
 Schrödinger's cat, 3, 131  
 Shamir, 174  
 Shear transformation, 53  
 Shor, 143  
 Shor's algorithm, 143, 159  
 Single-particle quantum technology, 179  
 Single qubit quantum gate, 148  
 Solvay conference, 14  
 Sommerfeld, 18, 22, 24, 29, 30  
 Special relativity, 30, 66  
 Specific heat of solids, 13  
 Spin, 64, 65  
 Spin precession frequency, 89  
 Spin resonance absorption, 88  
 Spin singlet state, 105, 106, 112  
 Spin state, 66  
 Standing wave, 27, 91  
 Statistical uncertainty, 119  
 Stern, 63  
 Stern-Gerlach double-slit experiment, 140  
 Stern-Gerlach experiment, 63, 70, 123–125, 132, 167  
 Summation symbol, 52  
 Superposition principle, 3, 93, 128  
 Superposition quantum gate, 156  
 Susskind, 108  
 Symmetric matrix, 58

**T**

Thomson, 15  
 Thomson's model of atom, 15  
 Time complexity, 158  
 Toffoli, 155  
 Toffoli gate, 155

Transistor, 6  
 Transpose, 47, 48, 58  
 Two dark clouds, 14  
 Two-qubit quantum gate, 148

**U**

Unitary evolution, 84, 86  
 Unitary matrix, 59  
 Universal gate, 146, 150, 155

**V**

Vector, 45, 46, 49  
 Vector space, 46  
 Vernam, 173  
 Vernam cipher, 173  
 Von Lindemann, 24  
 Von Neumann, 124, 125, 137

**W**

Watson, 30  
 Wave function, 82, 83, 91  
 Wave-particle duality, 27, 82  
 Weyl, 27  
 Wheeler, 131, 138  
 Wien, 10, 25  
 Wien's law, 12  
 Wiesner, 174  
 Wigner, 127, 137  
 Wigner's friend paradox, 127

**Y**

Young, 99

**Z**

Zero-point energy, 90  
 Zero-point vibration, 3