Donald C. Chang

On the Wave Nature of Matter

A New Approach to Reconciling Quantum Mechanics and Relativity

Springer

On the Wave Nature of Matter

Donald C. Chang

On the Wave Nature of Matter

A New Approach to Reconciling Quantum Mechanics and Relativity

Springer

Donald C. Chang
Division of Life Science
Hong Kong University of Science
and Technology
Kowloon, Hong Kong

ISBN 978-3-031-48776-7 ISBN 978-3-031-48777-4 (eBook)
https://doi.org/10.1007/978-3-031-48777-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

I would like to dedicate this book to my beloved mother, K. Chen

Preface

What is this book about? This book is to introduce a new approach for explaining the foundation of quantum physics and to resolve the conflict between quantum mechanics and relativity. At present, there are many *deep questions* in modern physics that have not been satisfactorily resolved. For example,

- **What is the physical basis of wave-particle duality?** The quantum particle is supposed to behave like a point mass, why does it sometimes behave like a wave?
- **What is the origin of matter?** How can particles be created in the vacuum? How can energy be converted into mass?
- **What is the meaning of matter wave?** The movement of quantum particles can be described accurately using the quantum wave equations (e.g., Schrödinger equation). However, it is not clear what is the physical meaning of the quantum wave function. Does it represent a physical wave or a wave of probability?
- **What is the physical meaning of *mass*?** Could a photon have mass? Why does the moving mass of a particle change with speed?
- **How to resolve the conflict between quantum physics and relativity? What is the physical property of the *vacuum*? Is the vacuum empty?** There is a serious conflict on the view of the vacuum between quantum mechanics and relativity. The vacuum in quantum mechanics is not empty; it is just the ground state of the quantum system. However, the vacuum in special relativity must be an empty space; otherwise, the vacuum will provide a universal reference frame to tell which inertial frame is stationary and which one is in motion.

To resolve these deep questions, one needs to take a brave new approach that is fundamentally different from the conventional thinking. In the past one hundred years, physicists have been using a "particle" approach to explain the quantum world. In this traditional view, our material world is made up of point-like particles called "fermions". This "particle" view, however, has great difficulty in explaining the observation of *wave-particle duality*. How can a quantum particle (such as an electron or neutron) behave like a light wave in a diffraction experiment (or in a double-slit experiment)? Furthermore, how can a point-mass-like particle be created from nowhere or annihilated in the vacuum?

These observations give us a hint: Could the quantum particle be a quantized excitation wave, just like the photon? This hint stimulated us to think: Can this *wave* approach help us to overcome the previous difficulties? Could one use a *wave* hypothesis to resolve some of the outstanding questions encountered in modern physics? Over the past decade, we have conducted a series of investigations to explore this possibility. The results have been highly encouraging.

This book is a comprehensive review of our work in using the wave approach to understand the foundation of quantum physics. Our hypothesis is called the "*Quantum Wave Model*", in which we propose that the vacuum is a dielectric medium according to the Maxwell theory, and the quantum particles are quantized excitation waves of the vacuum medium. Thus, matter in our universe is really made of waves.

Based on this idea, one can easily explain the physical basis of wave-particle duality. At the microscopic level, the quantum particle is a wave; but at the macroscopic level, the wave packet behaves like a particle. Using this model, the known quantum wave equations, including the Klein-Gordon equation, the Dirac equation, and the Schrödinger equation can be directly derived based on vacuum excitation. Furthermore, this model provides a clear physical meaning of *energy*, *momentum,* and *mass* based on the geometrical properties of the vacuum. Some well-known "*relativistic effects*" can also emerge naturally from this model. Thus, this model can provide a comprehensive explanation for most of the existing mysteries in quantum physics.

The model discussed in this book is not a *theory of everything*. However, it can provide useful hints for the future studies of particle physics or cosmology. The final part of this book (**Part VI**) is an in-depth discussion of the open questions and remaining challenges in this regard.

This book is written for people curious in science. Its purpose is to provide a big picture view. It discusses the foundations of quantum theory and relativity in a way that undergraduate physics students and advanced research physicists can understand. The focus of this book is on physical concepts and logical thinking. Readers do not need advanced mathematical training to enjoy this book.

This book is written on a semi-technical level. Some of the more technical discussions are presented in the ***Appendix***. For readers who are not mathematically inclined, they can skip reading the derivation of equations. The results of most mathematical derivation are summarized in the ***Chapter Summary*** given at the end of each chapter.

Acknowledgment. I would like to thank people who gave me help in this book project. I am grateful for the encouragement and support from many friends and colleagues, particularly Profs. Harold E. Rorschach, John A. Wheeler, K. K. Phuah, Zhen Cao, Yuk L. Yung, and Edmund Chiang. I thank Ms. Lan Fu for her great contribution in the manuscript preparation work. The early draft of this book was previewed by Drs. Avery Ching and Chantal Roth; their feedbacks are very helpful. Finally, I would like to thank Dr. Zachary Evenson and Mrs. V. B. Saranyaa of Springer/Nature for their valuable supports in making the publication of this book possible.

Clear Water Bay, Hong Kong
October 2023

Donald C. Chang

Contents

Chapter 1
Introduction: The Particle World *Versus* the Wave World

Our material world is basically composed of matter and radiation. According to the current quantum theory, both matter and radiation are made up of “particles”. These particles are quantum objects, which behave in a very strange way. In classical physics, a particle is a corpuscular object (a rigid point mass); it behaves like a tiny billiard ball. The quantum particle, however, behaves very differently from the classical particle. It behaves as both a corpuscular object and a wave (e.g., “matter wave”). Over the past one hundred years, physicists have developed a highly successful theory called “quantum mechanics” to describe the interactions between these quantum particles (e.g., electrons and photons). However, they cannot explain why the quantum particle behaves differently from the classical particles. This has become a big mystery.

Philosophically speaking, a physical object might be viewed as a particle or a wave. For example, several hundred years ago, Newton and Huygens had an active debate on the physical nature of light; they showed that the behavior of light can be explained either based on a particle perspective or on a wave perspective. In the nineteenth century, it was clearly demonstrated that light is an electromagnetic wave. But at the beginning of the twentieth century, it was shown by Planck and Einstein that light can also be considered as a stream of particles (called “photon”). This strange property of light is called “wave-particle duality”.

Even more interestingly, it was discovered later that not only massless particles such as photons can exhibit the property of wave-particle duality; massive particles (such as electrons or neutrons) can also exhibit both corpuscular and wave properties. This new understanding was first proposed by de Broglie and later verified in a series of diffraction experiments [1–3]. Nowadays, many physical events at the atomic level are found to be associated with wave behavior [4–9].

In view of the discovery of wave-particle duality, one could interpret the quantum world from a particle view or from a wave view. But historically, the development of quantum physics was mainly based on the particle view. In fact, such an approach has achieved great success in the past one hundred years. However, the particle theory of

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_1

quantum physics also encountered some major conceptual challenges, which even the most brilliant physicists do not see any way out. (For details, see below.) In this book, we would like to explore a new approach by interpreting the quantum world from a wave point of view. We found this new approach can indeed overcome some of the major conceptual limitations of the current quantum theory. **The objective of this book is to give a detailed explanation on how one can use the wave model to resolve the quantum mysteries encountered today.**

1.1 The Current Quantum Theory is a Particle Theory

In the quantum physics theory used today, it is basically taking a particle view. Most physicists regard the quantum system as a mechanical system composed of particles, each of which behaves like a point mass. For example, we know the building blocks of matter are atoms. In most textbooks, an atom is depicted as a miniature solar system (see Fig. 1.1), where electrons move around the nucleus like tiny planets. In this picture, not only the nucleus is viewed as a tiny massive particle, each electron is also regarded as a corpuscular object.

Today, the branch of physics that studies the interaction between sub-atomic particles is called "particle physics". Its mainstream theory, the so-called "Standard Model", proposes that all matters in the world are made of a sub-set of particles called "fermions", and the interacting forces between these fermions are carried by another sub-set of particles called "bosons". For example, the atom is made up of protons, neutrons, and electrons, which are fermions. The particle carrying electromagnetic interactions is supposed to be the photon, which is a boson. In fact, the force-mediating bosons could be very massive. For example, the particles carrying the weak force (W± and Z particles) and the particles carrying the strong force (gluons) are very heavy particles.

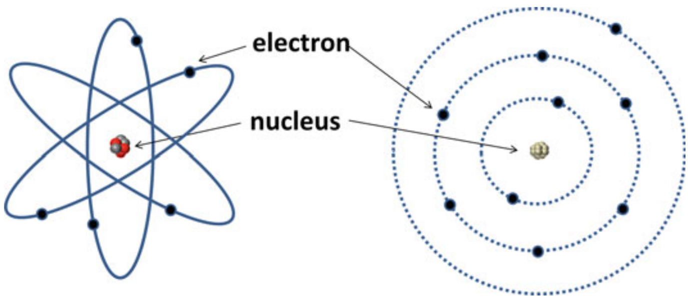


Fig. 1.1 Particle view of an atom. Traditionally, an atom is depicted as a miniature solar system, where electrons move around the nucleus like tiny planets. In this picture, each electron is regarded as a corpuscular object

1.2 What is the Problem with the Current View of Quantum Physics? Why Do We Need a Paradigm Shift?

The quantum theory used today is generally referred to as the "quantum field theory", which is a combination of quantum mechanics, special relativity, and the classical field theory. Quantum mechanics, of course, is a highly successful theory. It provides the basis for the development of many branches of modern physics, including atomic physics, molecular physics, condensed matter physics, etc. In fact, most of the modern technology we use today, including all electronic devices, computers, mobile phones, and communication networks, all depends on quantum mechanics and quantum electrodynamics.

Then, why are we not satisfied with it?

That is because there are serious fundamental problems with the current version of quantum theory. These problems include:

(1) After more than one hundred years of development, the physical foundation of quantum mechanics is still not well understood; it remains a deep mystery.
(2) There is a serious conflict between the fundamental assumptions of quantum mechanics and that of relativity.
(3) There are still many unanswered fundamental questions in the current quantum theory. It is not clear that the current model can take us to a deeper understanding of our physical world.

In the following, let us review these problems one by one.

1.2.1 Lack of Understanding on the Physical Basis of Quantum Mechanics

Today, there is still a lot of mystery in quantum mechanics. First, the current theory cannot explain why a quantum particle has both corpuscular and wave properties (wave-particle duality). Second, unlike the equations of motion in classical physics, which are based on well-established physical laws, the derivations of quantum wave equations (the Schrödinger equation and the Dirac equation) were based mainly on conjectures; the strongest justification is that these wave equations can lead to results consistent with experiments [6, 10–12]. This is not very satisfactory. In the physics tradition, we always want to know the physical basis behind a working theory.

Finally, there is serious confusion about the physical meaning of the quantum wave function. Is it a real physical wave or just a statistical parameter that gives the probability of finding the particle? There was a famous debate between Bohr and Einstein about this question almost one hundred years ago. We still do not have a clear answer today [13].

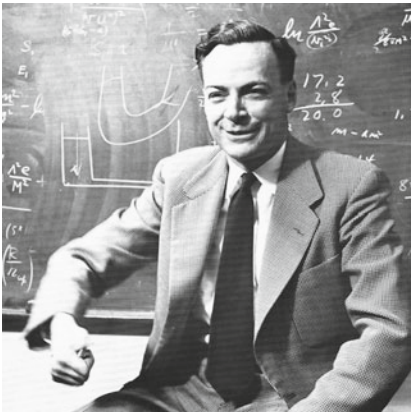


Fig. 1.2 Richard Feynman. Richard Feynman (1918–1988) was a famous theoretical physicist in the US. He made important contributions in the development of quantum electrodynamics. He invented the technique of "Feynman diagram" for analyzing the interaction of particles. Feynman was awarded the Nobel Prize in Physics in 1965. He was a scientist of great personality. "The Feynman Lectures on Physics" was based on his lectures given at Caltech, which has become a widely used textbook for many top universities in the United States. According to a 1999 poll conducted by the British magazine "Physics World", Feynman was named one of the ten greatest physicists of all time. Photo Credit: The Big T (Yearbook of California Institute of Technology); Wikimedia Commons, Public domain

In the past one hundred years, many leading physicists have been aware of these problems. For example, Richard Feynman (see Fig. 1.2), a very well-known physicist in the twentieth century, had a famous quote: "*I think I can safely say that nobody understands quantum mechanics*". [14]

In his famous book entitled "*The Feynman Lectures on Physics*", he further elaborated why he thought quantum physics is beyond human understanding:

> Because atomic behavior is so unlike ordinary experience, it is very difficult to get used to and it appears peculiar and mysterious to everyone, both to the novice and to the experienced physicist. Even the experts do not understand it the way they would like to, and it is perfectly reasonable that they should not, because all of direct, human experience and of human intuition applies to large objects. We know how large objects will act, but things on a small scale just do not act that way. So we have to learn about them in a sort of abstract or imaginative fashion and not by connection with our direct experience. [15]

Feynman was totally pessimistic about any hope for physicists to understand the foundation of quantum mechanics. Feynman was by no means the only pessimistic physicist in this regard; many leading physicists had expressed similar views and admitted that the foundation of quantum mechanics is very mysterious. For example, Roger Penrose, a mathematical physicist in Cambridge University who won the Nobel Prize of physics in 2020, once wrote:

> I should begin by expressing my general attitude to present-day quantum theory, by which I mean standard non-relativistic quantum mechanics. The theory has, indeed, two powerful bodies of fact in its favour, and only one thing against it. First, in its favour are all the marvellous agreements that the theory has had with every experimental result to date. Second, and to me almost as important, it is a theory of astonishing and profound mathematical beauty. The one thing that can be said against it is that it makes absolutely no sense! [16]

1.2.2 Fundamental Conflict Between Quantum Mechanics and Relativity

The current quantum theory is a combination of quantum mechanics, special relativity, and classical field theory. There seems to be a serious contradiction between these theories on the concept of vacuum. More specifically, the vacuum in quantum mechanics is not empty; it is just the ground state of the quantum system. However, the vacuum in special relativity must be an empty space; otherwise, the vacuum will provide a universal reference frame to tell which inertial frame is stationary and which one is in motion. This would defeat the principle of relativity. This is a serious discrepancy. Yet, this problem has been overlooked entirely in the development of the current quantum theory.

1.2.3 Important Questions that the Current Quantum Theory Cannot Resolve

Furthermore, there are many more important questions about fundamental physics that the current quantum theory cannot resolve. For example,

- **Where do the particles come from?** How can matter be created from energy? In most textbooks, it is frequently stated that energy-mass convertibility is based on special relativity. However, many careful literature reviews had shown that this is not the case (see Chap. 11). Furthermore, nothing is known about the physical mechanism that allows energy to be converted into matter and vice versa.
- According to the current quantum field theory, the creation of particles in the vacuum is based on a process that converts virtual particles into real particles. However, **it is not clear why virtual particles can pre-exist in the vacuum**. If one thinks the vacuum is an empty space (as it is assumed in special relativity), how can something come out from nothing?
- In the current version of quantum field theory, it is hypothesized that a quantum particle is an excitation of its own quantum field. It is not clear that, **between particle and field, which one is more fundamental?** Furthermore, we know there are many types of quantum particles; **why does nature need so many different fields?**

Apparently, the current version of quantum theory has serious limitations; it is not clear that it can take us to a deeper understanding of our physical world.

1.2.4 The Particle Physics Establishment Had Given up Hopes to Resolve the Fundamental Issues

To make things worse, not only we do not understand these fundamental issues, the physics community today has given up hope to understand such issues; they are not interested in this kind of research. For example, Sean Carroll, a physicist in Caltech who is also a popular scientific writer, stated explicitly in the *Prologue* of his recent book [17]: "*You might think …that the quest to understand quantum mechanics would be the single biggest goal in all of physics. Millions of dollars of grant money would flow to researchers in quantum foundations, the brightest minds would flock to the problem, … Sadly, no. Not only is the quest to make sense of quantum mechanics not considered a high-status specialty within modern physics; in many quarters it's considered barely respectable at all, if not actively disparaged. Most physics departments have nobody working on the problem, and those who choose to do so are looked upon with suspicion*".

Today, the physics establishment teaches only theories that fit the current views. In most physics textbooks, they try to shy away from discussing any discrepancy within the established theories. Most physicists treat quantum mechanics as a useful tool, but do not wish to be bothered with finding out where this tool comes from. There was a famous quote about the common attitude of today's physicists when they are asked by their students to explain quantum physics. Their answer could be: "*Shut up and calculate!*".[1]

This is not a very satisfying situation. To overcome the problems facing quantum theory today, perhaps we should explore the use of alternative approaches to develop quantum theory. In other words, should we be considering a paradigm shift? Would it be more effective to use the wave view rather than the particle view to explain physical events at the sub-atomic level?

1.3 The Basic Idea of the Quantum Wave Model

We agree with Feynman that the microscopic world is indeed vastly different from the macroscopic world. However, we believe it is still possible for humans to understand the physical mechanisms in the microscopic world if we are willing to carefully study nature. Our proposal is that physicists should give up the traditional

[1] This quote was often attributed to Feynman. But according to David Mermin, he was the first one to make this quote. He wrote in *Physics Today* in 1989: *"If I were forced to sum up in one sentence what the Copenhagen interpretation says to me, it would be "Shut up and calculate!"*.

concept of regarding the sub-atomic particle (such as an electron) as a corpuscular object. Instead, one needs to realize that the electron is more like a quantized excitation wave instead of a tiny billiard ball (see Fig. 1.3). The motion of the quantum particle is governed by wave mechanics which should be developed based on the wave properties of the vacuum medium.

Unlike the particle view used in mainstream quantum physics today, we proposed that all sub-atomic particles are quantized excitation waves of the vacuum medium. Thus, not only the photon is a quantized excitation wave, the electron is also an excitation wave in the microscopic view; it behaves like a particle only in the macroscopic view (see Fig. 1.3). In fact, in our model, sub-atomic particles like proton and neutron are also quantized excitation waves. Thus, the atom is composed of quantized excitation waves instead of corpuscular particles. This implies that the entire material world is composed of quantized excitation waves of the vacuum medium.

We call this new theory "***the quantum wave model***"; it is based on three simple hypotheses:

(1) **The vacuum is a wave medium.**
(2) **The quantum particle is a quantized excitation wave of the vacuum medium.**
(3) **Different types of quantum particles are represented by different excitation modes of the vacuum medium.**

Here, the quantum particles include all sorts of "elementary particles", such as photons, electrons/positrons, muons, and neutrinos. Hadrons, like protons and neutrons, are slightly more complicated because they are composite particles. We

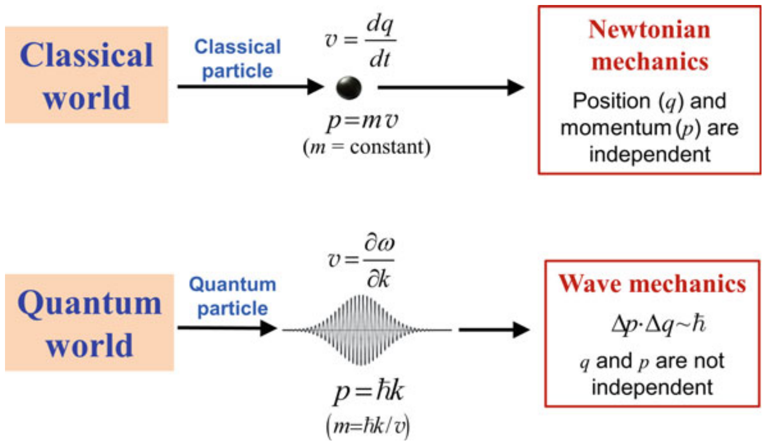


Fig. 1.3 Classical view versus the quantum view of a particle. The classical particle is generally thought to be a corpuscular object, just like a tiny billiard ball; the position and momentum of this particle can be independently defined. In the case of a quantum particle, it is more like a quantized excitation wave instead of a tiny billiard ball; its position and momentum cannot be independently determined

believe they are also quantized excitation waves, but their wave structures could be highly complicated (see Chap. 17).

From this hypothesis, it is predicted that the equation of motion for a quantum particle is essentially determined by the physical properties of the vacuum medium. Thus, we expect that the wave equation of photon and the quantum wave equation of electron can both be derived based on the same wave excitation mechanism of the vacuum medium (see Part 2).

1.3.1 Justification for the Hypotheses of the Quantum Wave Model

Is there any supporting evidence for the above hypotheses? Yes, our model is based on well-established experimental facts, for example:

1. The concept of vacuum being a medium is a well-based idea in physics. Before the twentieth century, it was widely believed that the vacuum is filled with an aether-like medium. This understanding was based on the studies of light and electromagnetism. Although the original aether hypothesis was later disfavored, there was strong evidence indicating that vacuum is not an empty space (see **Appendix A**).
2. We know at least one quantum particle, i.e., the photon, is a quantized physical wave.
3. The phenomena of particle-wave duality are well demonstrated in experiments. Particularly, the diffraction experiment indicated that the electron behaves very similarly to a photon. Thus, the electron must also be a quantized physical wave.
4. A very important finding in modern physics is that particles can be created or annihilated in the vacuum. This fact suggests that quantum particles could be excitation waves of the vacuum medium.
5. In the experimental study of particle physics, it is frequently observed that one type of particle can be converted into other types of particles during collision or decay. These observations suggest that different particles could be different excitation modes of the vacuum medium.
6. It is well known that all particles have the same traveling speed limit. That is, no particle can travel faster than the speed of light, which is c. Since the speed of wave traveling in a medium is determined entirely by the physical properties of the wave medium, the fact that the speed limits for all particles are the same suggests that all quantum particles could be excitation waves of the same medium.

These points will be discussed in detail in the following chapters of this book.

1.4 How Can the Quantum Wave Model Help to Resolve the Problems Encountered in the Current Quantum Theory?

Using this quantum wave model, we will show that it is possible to resolve the "mystery" of quantum mechanics. First, **the quantum wave model can directly explain the phenomenon of wave-particle duality**. According to our hypothesis, photons and electrons are both excitation waves of the vacuum; this explains why they behave similarly in the double-slit or diffraction experiments.

Second, by using the wave approach instead of the particle approach, one can **explain more easily the physical basis of particle creation/annihilation**. If particles are excitation waves of the vacuum, the vacuum will be free of particle when it is at its resting state. Then, when the vacuum is excited with an energy stimulus, particles will be created due to the generation of new waves. Similarly, in the process of wave-wave interactions, some waves may be destroyed in order to create new waves. This may explain why particles can be converted into different types during interactions.

Third, there is a clear advantage of a **conceptual unification between light wave and matter wave**. As we will show later, from the Maxwell theory and Helmholtz decomposition, one can **derive not only the wave equation of photons, but also the quantum wave equations for electrons** (e.g., the Dirac equation and the Schrödinger equation). The only assumption here is that both the matter wave and the radiation wave are excitation waves of the vacuum medium. (For details, see Chaps. 7, 8 and 9).

Finally, to further demonstrate the usefulness of the wave approach, **we will use the quantum wave model to address a number of deep questions in fundamental physics**. For example:

- What is the physical basis of Heisenberg's Uncertainty Principle? (See Chap. 3)
- What is the physical property of the vacuum? Can the vacuum satisfy the physical requirements of being a wave medium? (See Chap. 6)
- Can one use the wave excitation mechanism in the vacuum to derive the quantum wave equations for a massive particle (such as an electron)? (See Chaps. 8 and 9)
- How can a wave have mass? (See Chap. 11)
- Why can energy and mass be converted between each other? (See Chap. 11).
- Why is mass speed-dependent? (See Chap. 12)
- Why do all particles have the same traveling speed limit? (See Chap. 12)
- How can a wave behave like a particle in the macroscopic world? (See Chap. 11)
- If matter is composed of waves, how can the waves generate gravity? (See Chaps. 12 and 15)

In the following chapters of this book, we will try to answer these questions in detail one by one. These discussions are based on a series of research papers published by the author in recent years [18–28].

References

1. de Broglie, L. (1924). *Recherches sur la théorie des quanta.*
2. Davisson, C. J., & Germer, L. H. (1927). The scattering of electrons by a single crystal of nickel. *Nature, 119*, 558–560.
3. Thomson, G. P., & Reid, A. (1927). Diffraction of cathode rays by a thin film. *Nature, 119*, 890.
4. Bach, R., Pope, D., Liou, S.-H., & Batelaan, H. (2013). Controlled double-slit electron diffraction. *New Journal of Physics, 15*, 033018.
5. Chang, D. C. (2005). *On the wave nature of matter*. https://doi.org/10.48550/arXiv.physics/0505010
6. Hey, T., & Walters, P. (2003). *The New Quantum Universe*. Cambridge University Press.
7. Estermann, I., & Stern, O. (1930). Beugung von molekularstrahlen. *Zeitschrift für Physik, 61*, 95–125.
8. Von Halban, J., & Preiswerk, P. (1936). Preuve expérimentale de la diffraction des neutrons. *CR Academy Science Paris,* ***203***, 73–75.
9. Carnal, O., & Mlynek, J. (1991). Young's double-slit experiment with atoms: A simple atom interferometer. *Physical Review Letters, 66*, 2689–2692.
10. Slater, J. C. (1969). *Concepts and Development of Quantum Physics.* Dover Publications.
11. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Chapter 16 The dependence of amplitudes on position. In *The Feynman Lectures on Physics* (Vol. 3). Addison-Wesley.
12. Dirac, P. A. M. (1981). *The Principles of Quantum Mechanics* (4th ed.). Clarendon Press.
13. Skibba, R. (2018). Einstein, Bohr and the war over quantum theory. *Nature, 555*, 582–584.
14. Feynman, R. P. (1965). *The Character of Physical Law.* MIT Press.
15. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Chapter 37 Quantum behaviour. In *The Feynman Lectures on Physics* (Vol. 1). Addison-Wesley.
16. Penrose, R. (1986). Gravity and state vector reduction. In R. Penrose, & C. J. Isham (Eds.) *Quantum Concepts in Space and Time* (pp. 1–129). Oxford University Press
17. Carroll, S. M. (2019). *Something Deeply Hidden: Quantum Worlds and the Emergence of Spacetime.* Dutton.
18. Chang, D. C. (2004). *What is rest mass in the wave-particle duality? A proposed model.* arXivpreprintphysics/0404044
19. Chang, D. C. (2013). A classical approach to the modeling of quantum mass. *Journal of Modern Physics, 4*, 21–30.
20. Chang, D. C., & Lee, Y.-K. (2015). Study on the physical basis of wave-particle duality: Modelling the vacuum as a continuous mechanical medium. *Journal of Modern Physics, 6*, 1058–1070.
21. Chang, D. C. (2017). *On the wave nature of matter: A transition from classical mechanics to quantum mechanics.* https://doi.org/10.48550/arXiv.physics/0505010
22. Chang, D. C. (2017). Physical interpretation of the Planck's constant based on the Maxwell theory. *Chinese Physics B, 26*, 040301.
23. Chang, D. C. (2017). Is there a resting frame in the universe? A proposed experimental test based on a precise measurement of particle mass. *The European Physical Journal Plus, 132*, 140.
24. Chang, D. C. (2018). A quantum mechanical interpretation of gravitational redshift of electromagnetic wave. *Optik, 174*, 636–641.
25. Chang, D. C. (2020). A quantum interpretation of the physical basis of mass–energy equivalence. *Modern Physics Letters B, 34*, 2030002.
26. Chang, D. C. (2021). Review on the physical basis of wave–particle duality: Conceptual connection between quantum mechanics and the Maxwell theory. *Modern Physics Letters B, 35*, 2130004.

27. Chang, D. C., Wang, Y., Wong, M. K.-Y., & Chan, K.-L. (2021). What is the Origin of Our Material World? (in Chinese) (1st ed.). China Youth Publishing House.
28. Chang, D. C. (2023). A quantum view of photon gravity: The gravitational mass of photon and its implications on previous experimental tests of general relativity. *Modern Physics Letters B* 2250179

Part I
The Physical Basis of Wave-Particle Duality

Chapter 2
The Birth of Quantum Mechanics: Arriving of the Photon Concept

Before the twentieth century, the concepts of *particle* and *wave* are entirely independent. A particle is supposed to be a mechanical object like a point mass, while the wave is a spread-out oscillation of a medium. The concepts of "particle" and "wave" thus are totally different.

At the end of the nineteenth century, physicists discovered that this conventional thinking may not be entirely true. This was mainly due to the study of blackbody radiation by a German scientist, Max Planck.

As we all know, light is a form of radiation. In classical physics, the distribution of energy in the radiation wave is supposed to be continuous. However, Planck discovered that the energy of light cannot be continuous. Light is transmitted in an undividable package of energy, which was later called a "*quantum*". Such behavior is very much like that of a particle. So, Planck's finding suggested that light could have both *particle* and *wave* properties.

2.1 Is Light a Wave or a Particle? How Do We Know that Light is a Wave?

In fact, even before Planck's study, there had been a historical debate on the physical nature of light. The modern study of light probably started in the seventeenth century. There were two major contributors in studying the physics of light: One is Isaac Newton and the other is Christiaan Huygens. In Newton's view, light is composed of particles or corpuscular objects. His major argument was that light always travels in straight lines, while waves will bend around obstacles. Another argument is that Newton's theory could predict the reflection of light. When light is reflected on a smooth surface, the incoming angle and the outgoing angle is exactly the same. This is just like bouncing a ball. Because Newton was such an influential physicist at that

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_2

time, his theory regarding light as particles played a dominant role in the seventeenth century.

2.1.1 The Double-Slit Experiment

Around the same time, another well-known physicist Huygens proposed a different model of light. He proposed that light was emitted in all directions as a series of waves in a medium called the luminiferous aether [1, 2]. The debate between the particle view of light and the wave view of light carried on for the next two centuries. Finally, in the beginning of the nineteenth century, the wave nature of light was clearly demonstrated in the results of the double-slit interference experiment. This experiment was conducted by Thomas Young in 1801 [3, 4]. The design of this experiment was very simple: Let a beam of light pass through a partition with two small slits, and then project it onto a screen behind the partition. By looking at the interference results on the screen, one can tell whether the light is a particle or a wave.

What is the experimental principle of this double-slit interference experiment? If a light beam is composed of individual particles (like a train of tiny bullets), when it passes through a slit, it will just appear as a bright line on the screen. When another beam of light passes through another slit, it will also appear as a bright line behind the slit. So, if light is composed of particles, the light beam passing through the double-slit can only leave two bright lines on the screen behind the double-slit. (See Fig. 2.1).

Now, if light is a wave, it will exhibit an interference phenomenon (see Fig. 2.2). A series of alternating light and dark bands will appear on the screen. This is called "interference fringes". The principle of the double-slit interference experiment is very simple. When Thomas Young conducted this light double-slit experiment, he

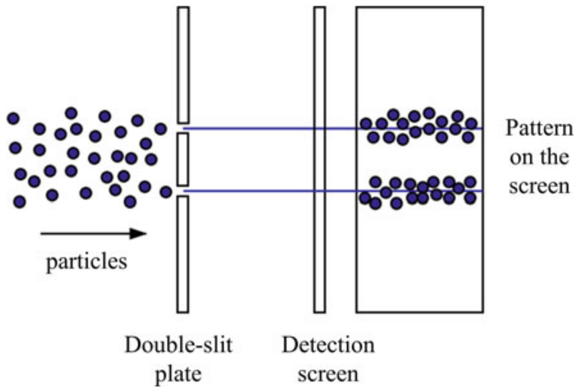


Fig. 2.1 A double-slit experiment using mechanical particles. According to the double-slit interference experiment designed by Thomas Young, if light were made of particles like steel balls, they would not interfere, so there would only be two bright lines on the detection screen. Image Credit: Inductive load, Wikimedia Commons; Public domain

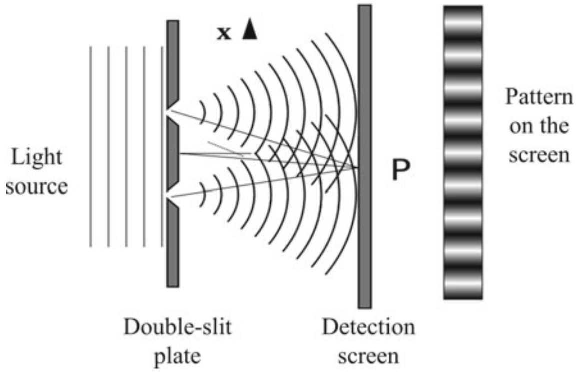


Fig. 2.2 A double-slit experiment with light waves gives a pattern of interference fringes. If light is a wave, it will exhibit an interference phenomenon in the double-slit experiment. Image Credit: Original figure from Wikimedia Commons; CC BY-SA 3.0

observed that light indeed produced interference fringes after passing through the double-slit, so he demonstrated that light was a kind of wave [3, 4].

2.1.2 The Bragg Diffraction Experiment

Another strong evidence to indicate that light is a wave is the result of the Bragg diffraction experiment. Following the double-slit experiment, physicists later discovered one can generate an interference pattern from the refraction of light from a crystal surface. This is called the Bragg diffraction. The Bragg diffraction was proposed by Lawrence Bragg and his father Henry Bragg in 1913 to explain their discovery that crystalline solids produced peculiar patterns of reflected X-rays [5]. They found that these crystals produced intense peaks of reflected radiation at certain specific wavelengths and incident angles.

The principle of Bragg diffraction is very simple. It is based on the summation of two light beams diffracted from two neighboring rows of atoms in a crystal. Depending on the diffraction angle, the combination of these neighboring light beams could be constructive or destructive based on their phase difference. And thus, an alternating light-and-dark pattern could be observed in the diffracted plate. Constructive interference occurs when the path difference between two adjacent reflecting light beams equals to an integer multiple of the wavelength of the radiation; this is called the "Bragg condition". Since the diffraction of X-ray can be fully described by the Bragg diffraction, it was convincingly concluded that X-ray must be a wave.

2.1.3 *Maxwell and Hertz Showed that Light is a Kind of Electromagnetic Wave*

The most convincing argument for light being a wave comes from the work of Maxwell and Hertz. In the middle of nineteenth century, Maxwell worked out a very influential treatise on the study of electromagnetic waves [6]. In his theory, he predicted that light is a kind of electromagnetic radiation, one of the evidence he cited is that based on his mathematical calculation, the speed of the propagation of the electromagnetic wave is identical to the speed of light that was experimentally determined at that time. A few years later, Hertz experimentally confirmed Maxwell's theory by building an experimental device that can generate and detect radio waves between two location points [7]. He demonstrated very convincingly that the electromagnetic radiation wave behaves exactly like light. So, from then on, people clearly recognize that light is a wave.

2.2 The Discovery of Light Wave Behaving like a Particle

2.2.1 *Quantization of Light*

At the end of the nineteenth century, there was a remarkable conceptual change regarding the physical nature of light. As we discussed above, light is clearly a wave in the classical view. But with the emerging of quantum physics, people started to realize that light is composed of discrete packages of energy, each of which could be regarded as a particle. (This particle-like quantum of energy is now called "photon").

The birth of quantum mechanics is commonly attributed to the discovery of Max Planck (see Fig. 2.3). In order to explain black-body radiation, Planck postulated that the radiation energy is transmitted in package (called "quantum") [8]. Einstein later studied the photoelectric effect and also came up with the conclusion that the energy of light is not transferred continuously as in a classical wave, but in small "packets". Einstein's explanation for these observations was that light itself is quantized. The size of these "packets" of energy was the same as Planck's "energy element", which was found to be proportional to the frequency ν of the radiation wave. This relation is now called the Planck's relation:

$$E = h\nu.$$

The constant "h" here is called the "Planck's constant". When the Planck's relation was first proposed in 1900, h was thought to be just a proportional constant between the minimal increment of energy E (of a hypothetical linear oscillator in a black body radiation cavity) and the frequency of its associated electromagnetic wave [8]. Later in 1905, in a study of photoelectric effect, the value E was theoretically associated

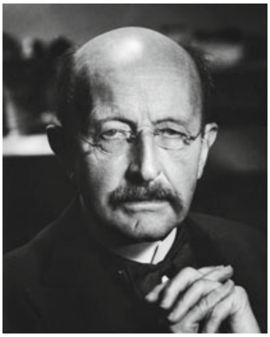


Fig. 2.3 **Max Planck**. Max Planck (1858–1947) was a German physicist. He is most famous for the Planck relation published in 1900, which proposed the concept of "quantum". He was awarded the 1918 Nobel Prize in Physics. Credit: Archives of the Max Planck Society, Berlin

by Einstein with quantum energy of the electromagnetic wave itself [9]. The light quantum behaved in some aspects like a particle, which was later named the "photon".

2.3 How Did Planck Derive the Planck's Relation?

How did Planck develop his revolutionary idea? It was closely related with the study of black-body radiation. In 1894, Planck was commissioned by electric companies to improve the efficiency of lightbulbs. At that time, it was known that the intensity and color of the light emitted by a black body (a perfect absorber, also known as a cavity radiator) is closely related to its temperature. But no theoretical treatment was found to agree with experimental data. Before Planck's work, Wilhelm Wien in 1896 had proposed a model to describe the spectrum of thermal radiation [10, 11],

$$u_\nu = \frac{8\pi h\nu^3}{c^3}\exp\frac{-h\nu}{kT}, \tag{2.1}$$

where h is an empirical fitting constant. The Wien's law correctly predicted the radiation behavior at high frequencies but failed at low frequencies. In 1900, a British physicist John Rayleigh proposed another model to describe thermal radiation which was later developed to become the Rayleigh–Jeans law [12, 13],

$$u_\nu d\nu = \frac{8\pi\nu^2 kT d\nu}{c^3}. \tag{2.2}$$

This model was able to fit the data well at low frequencies but failed badly at high frequency (known as "UV catastrophe") (See Fig. 2.4).

Planck realized that if he can interpolate the Wein's Law with Rayleigh-Jeans Law, he might be able to find a correct theory to describe the black-body radiation. Planck's original argument was very complicated, which was hidden in cumbersome formulism of thermodynamics and statistical mechanics [11]. A more comprehensible treatment of Planck's argument was put forward by Debye in 1910 [11,

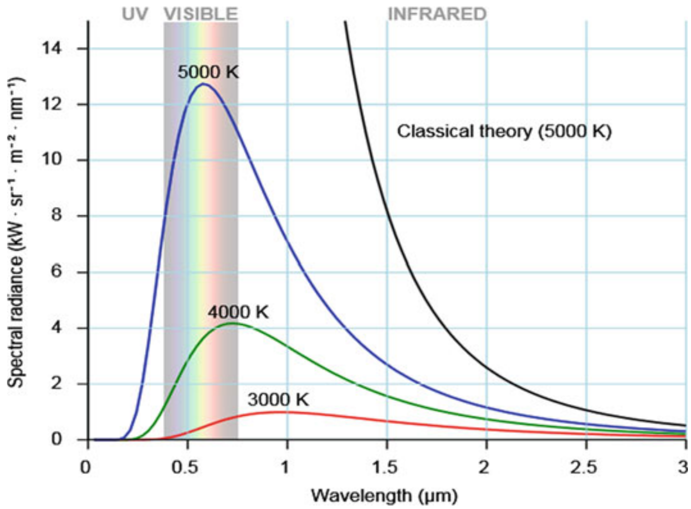


Fig. 2.4 **The UV catastrophe**. The black-line is the theoretical result based on the classical theory of Rayleigh-Jeans law. The color lines are experimental results. The classical theory fails at short wavelength. Credit: Darth Kule, Wikimedia Commons; Public domain

14]. In the following, we will briefly review Planck's derivation of h using Debye's cleaned-up version as outlined by John Slater [11].

Planck's theory was basically to treat the emitter in the black-body radiation as a linear oscillator,

$$E = T + V = \frac{p^2}{2m} + \frac{m\omega^2}{2}q^2. \tag{2.3}$$

Here, E, T and V represent the total energy, kinetic energy and potential energy, respectively; q is the generalized coordinate and p is the generalized momentum; m and ω are the mass and frequency of the oscillator. If one maps the energy distribution in the phase space, one will find that the contour of a constant energy is an ellipse (Fig. 2.5a), since Eq. (2.3) can be reduced into

$$\frac{p^2}{a^2} + \frac{q^2}{b^2} = 1, \tag{2.4}$$

where $a = \sqrt{2mE}$, $b = \sqrt{2E/m\omega^2}$. The area enclosed by this ellipse is known to be

$$\pi ab = \frac{2\pi E}{\omega} = \frac{E}{\nu}. \tag{2.5}$$

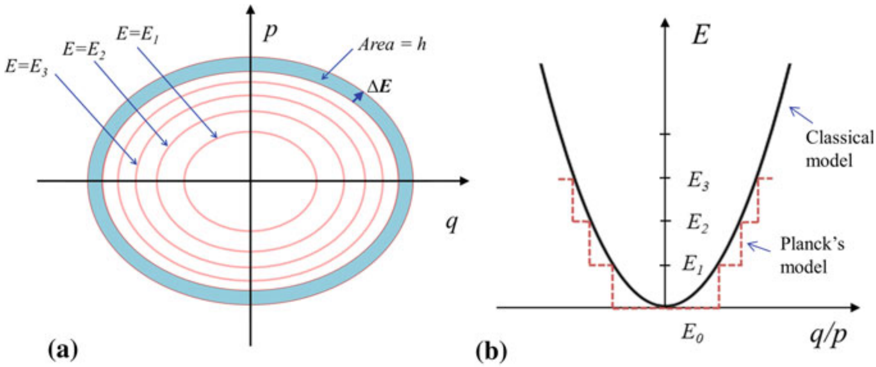


Fig. 2.5 Modeling the emitter in black-body radiation as a linear oscillator. **a** Energy distribution in phase space. Here, q is the generalized coordinate and p is the generalized momentum. The energy distribution is given by Eq. (2.3). The contour of a constant energy is an ellipse. When energy is increased from E to $E + \Delta E$, the corresponding incremental area in the phase space is $\Delta A = \Delta E/\nu$ (shadowed area). **b** Planck's step-wise energy distribution model for emitter in black-body radiation. The smooth curve was assumed to become a step-wise function with a constant jump of magnitude ΔE (dotted lines). Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, Chin. Phys. B **26**, 040301 (2017)

Then, when the energy is increased from E to $E + \Delta E$, the corresponding incremental area in the phase space is (see Fig. 2.5a)

$$\Delta \mathrm{A} = \Delta E/\nu. \tag{2.6}$$

Planck then made the following formal assumptions: *The energy distribution of the emitter in reality does not follow a smooth curve as described in Eq. (2.3). Instead, it is a step-wise function with a constant jump of magnitude ΔE.* (See Fig. 2.5b.) The reason that ΔE is constant is because nature somehow partitions the phase space of the emitter in constant incremental areas (ΔA), which can be called "*h*". (i.e., $\Delta \mathrm{A} = h$, see Fig. 2.5a). According to Eq. (2.6), Planck's assumptions implied that

$$\Delta E = h\nu. \tag{2.7}$$

(This is now called the "Planck's relation"). With the above assumption, one can calculate the average energy of a black-body radiation emitter oscillating at frequency ν using the Boltzmann distribution, that is,

$$\text{Average Energy} = \frac{0 + h\nu\exp(-h\nu/kT) + 2h\nu\exp(-2h\nu/kT) + 3h\nu\exp(-3h\nu/kT) + \ldots}{1 + \exp(-h\nu/kT) + \exp(-2h\nu/kT) + \exp(-3h\nu/kT) + \ldots} \tag{2.8}$$

For simplicity, let us denote $\exp(-h\nu/kT) \equiv x$,

$$\text{Average Energy } = \frac{h\nu x(1 + 2x + 3x^2 + \ldots)}{1 + x + x^2 + x^3 + \ldots}.$$

We know $1+x+x^2+x^3+\ldots = \frac{1}{1-x}$, and $1+2x+3x^2+\ldots = \frac{d}{dx}(1+x+x^2+x^3+\ldots)$. Therefore,

$$\text{Average Energy } = \frac{h\nu x}{1 - x} = \frac{h\nu \exp(-h\nu/kT)}{1 - \exp(-h\nu/kT)} = \frac{h\nu}{\exp(h\nu/kT) - 1}. \tag{2.9}$$

Since we know the number of energy states per unit volume and per unit frequency range in the radiation field is $8\pi \nu^2/c^3$, the energy distribution in the radiator then is [11]

$$u_\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{\exp(h\nu/kT) - 1}. \tag{2.10}$$

This equation is now known as the "Planck's law". It fitted the experimental data of black-body radiation as reported by Lummer and Pringsheim very well. One can see easily that at low frequency, Planck's law can be reduced to become the Rayleigh-Jeans law. And, at high frequency, Planck's law becomes Wien's law.

From the above review, one can see that in Planck's original model, h can be interpreted as a hypothetical constant incremental area in the phase space of the emitter. If the value of h is very small, this model will become the classical equipartition model as used in the derivation of Rayleigh-Jeans Law. So the basic difference between the quantum model and the classical model of black-body radiation is essentially whether the magnitude of h is significantly large or not. Based on experimental observations, Planck estimated that the value of h was 6.55×10^{-34} J s [8], which is within 1.2% of the present value (6.62607×10^{-34} J s) [15]. This value is proven sufficiently large so that we cannot ignore the quantum effects in radiation.

2.4 Further Evidence Supporting the Idea of Photon

2.4.1 The Photo-Electric Effect

Although Planck was the first one to suggest that light energy was emitted in packets, Einstein (see Fig. 2.6) was the first person to realize that light energy is transmitted and absorbed in packets (now called "photons") [9]. Such quantum behavior of photon was demonstrated in his analysis of the photoelectric effect. Einstein's idea was very straight forward. Based on Planck's result, he simply assumed that the energy of a photon is $E = h\nu$. When light is shined on an object, some of the electrons within that solid object will absorb part of the light energy. Einstein assumed that the energy of

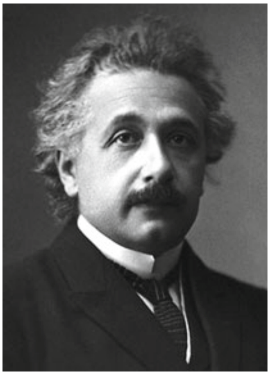


Fig. 2.6 Albert Einstein. Albert Einstein (1879–1955) was probably the most famous physicist in modern times. He made important contributions in many fields, including photoelectric effect, special relativity, general relativity, statistical mechanics, and more. He was awarded the Nobel Prize in Physics in 1921 for his work on the photoelectric effect. As a legendary figure, Einstein was a social activist. He was influential both in political thinking and in philosophy. Photo Credit: Nobel Prize in Physics photograph 1921; Wikimedia Commons, Public domain

the entire photon is absorbed by one electron. Such energy can be used to (a) knock out the electron from the solid; and (b) provide the excited electron with kinetic energy. Thus,

$$\text{Kinetic energy of the electron} = h\nu - W, \tag{2.11}$$

where the "work function" W is the amount of work required to move an electron from the solid to its surface. This idea can be tested by conducting a simple experiment. The kinetic energy of the excited electron can be measured by surrounding the solid with a conducting cage with a negative potential (V). When eV equals to the kinetic energy of the excited electron, no electron can leave the cage. In this situation,

$$eV = h\nu - W. \tag{2.12}$$

This relation predicted by Einstein was shown to agree well with experiments [16, 17].

The study of the photoelectric effect was an important demonstration of the quantum nature of light. Einstein, however, did not try to examine the physical meaning of h in his study. He simply used the Planck's relation to obtain the value of the energy quanta.

2.4.2 The Compton Scattering

Another evidence showing that light could behave like a corpuscular object is the Compton scattering experiment [18]. This experiment was carried out by Arthur Compton in 1923 at Washington University. The Compton scattering experiment basically studied the inelastic scattering of light by an electron. The wavelength of the scattered light is different from that of the incident radiation. In Compton's original experiment, he used X ray as the light source and treated the electrons as free particles after the scattering. In his experiment, he measured mainly two things: (1) Change of the wavelength between the incident light and the scattered light; and (2) the angle between incoming light and the scattered light. He explained his experimental data using a model which treated the X ray photon as a classical particle.

In the Compton scattering experiment, the photon and electron are treated as corpuscular objects [18]. Compton used classical mechanics to explain the interaction between the incoming X ray photon and the recoil of the targeted electron. He worked out a simple formula called the "Compton scattering". The result of his experiment was highly consistent with his calculated result. The success of the Compton scattering experiment convinced physicists that light can be treated as a stream of particle-like objects called "photon"; and, the energy of the photon is proportional to the frequency of the light wave as predicted by the Planck's relation.

From then on, the *photon* was regarded as an *elementary particle* in the mainstream physics community.

2.5 Chapter Summary

- Is light a wave or a particle? How do we know that light is a wave? The wave nature of light was clearly demonstrated in the double-slit interference experiment and the Bragg diffraction experiment.
- The most convincing argument for light being a wave comes from the work of Maxwell and Hertz. In the middle of nineteenth century, Maxwell theoretically demonstrated that light is a kind of electromagnetic radiation. Later, Hertz experimentally demonstrated that the electromagnetic radiation wave behaves exactly like light.
- At the end of the nineteenth century, it was discovered that light could behave like particles. This discovery was due to the work of Max Planck. In order to explain the spectrum of black-body radiation, Planck postulated that the radiation energy is transmitted in package ("quantum"). The size of the quantum energy was proportional to the frequency ν of the radiation wave.
- Einstein later studied the photoelectric effect and found that the energy of light is not transferred continuously as in a classical wave, but in small "quantum", as predicted in the Planck's relation.

- From then on, people started to realize that light is composed of discrete packages of energy, each of which could be regarded as a particle. This particle-like energy quantum is now called "photon".
- Another piece of evidence showing that light could behave like a corpuscular object is the scattering experiment carried out by Arthur Compton in 1923. Compton studied the inelastic scattering of light by an electron. He found that both the photon and electron can be treated as corpuscular objects.

References

1. Huygens, C. (1690). *Traité de La Lumière*. Pieter van der Aa.
2. Whittaker, E. (1951). *A History of the Theories of Aether and Electricity*. Thomas Nelson and Sons Ltd.
3. Young, T. (1802). II. The Bakerian lecture. On the theory of light and colours. *Phil. Trans. R. Soc., 92*, 12–48.
4. This Month in Physics History. (2008). May 1801: Thomas Young and the nature of light. *APS News, 17* (2008)
5. Bragg, W. H., & Bragg, W. L. (1913). The reflection of X-rays by crystals. *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences, 88*, 428–438.
6. Maxwell, J. C. (1873). *A Treatise on Electricity and Magnetism* (Vol. 1, 1st ed.). Oxford University Press.
7. Hertz, H. (1893). *Electric Waves: Being Researches on the Propagation of Electric Action with Finite Velocity Through Space*. Macmillan.
8. Planck, M. (1900). Zur theorie des gesetzes der energieverteilung im normalspectrum. *Verhandl. Dtsch. Phys. Ges., 2*, 237–245.
9. Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik, 17*, 132–147.
10. Wien, W. (1897). XXX. On the division of energy in the emission-spectrum of a black body. *Philosophical Magazine, 43*, 214–220.
11. Slater, J. C. (1969). *Concepts and Development of Quantum Physics*. Dover Publications.
12. Rayleigh, L. (1900). LIII. Remarks upon the law of complete radiation. *Philosophical Magazine, 49*, 539–540.
13. Jeans, J. H. (1905). XI. On the partition of energy between matter and Æther. *Philosophical Magazine, 10*, 91–98.
14. Debye, P. (1910). Der Wahrscheinlichkeitsbegriff in der Theorie der Strahlung. *Annalen der Physik, 338*, 1427–1434.
15. Steiner, R., Newell, D., & Williams, E. (2005). Details of the 1998 watt balance experiment determining the Planck constant. *Journal of Research of the National Institute of Standards and Technology, 110*, 1.
16. Millikan, R. A. (1914). A direct determination of "h." *Physical Review, 4*, 73–75.
17. Millikan, R. A. (1916). A Direct photoelectric determination of Planck's "h." *Physical Review, 7*, 355–388.
18. Compton, A. H. (1923). A quantum theory of the scattering of X-rays by light elements. *Physical Review, 21*, 483–502.

Chapter 3
Derivation of the Planck's Relation, the de Broglie Relation, and Heisenberg's Uncertainty Principle Based on the Maxwell Theory

As we summarized in the last chapter, the arrival of quantum physics began with the discovery that the energy of light appears to be quantized. Based on the study of black-body radiation, Max Planck found that the energy of light is not transmitted continuously as in a classical wave, but in small "packets" [1]. Later, Einstein studied the photoelectric effect and inferred that the absorption of light by an atom is also quantized. From then on, the idea of energy quantization of light is widely accepted. It is generally believed that light (a radiation wave) is carried by a stream of particles ("photons"). The energy of each photon is proportional to the frequency ν of the radiation wave as given by the Planck's relation

$$E = h\nu$$

The physical basis of this relation, however, remains unclear. Although Planck was able to derive the Planck's relation using a simple "step-wise oscillator energy" model (see Chap. 2), he realized that his original proposal was somewhat arbitrary. Also, he was not able to identify the physical meaning of h based on any well-established physical laws. In this chapter, we will show that those deficiency can be overcome. We believe that there should be a connection between quantum mechanics and classical electrodynamics. All quantum relations, including the Planck's relation, the de Broglie relation, and Heisenberg's Uncertainty Principle, can find a root in the classical Maxwell theory.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_3

3.1 Why is Light Quantized? What is the Physical Meaning of the Planck's Constant?

3.1.1 Planck Was not Satisfied with His Original Derivation

At first, Planck thought that the idea of energy quantization was only "*a purely formal assumption … actually I did not think much about it…*" [2]. He was fully aware that this assumption was not compatible with classical physics. Thus, Planck spent the rest of his life trying to grasp the meaning of energy quanta. Planck was concerned that his black-body radiation theory had a number of problems, including:

- He did not directly calculate the radiation energy of the electromagnetic wave emitted from the black body. Instead, his model only calculated the energy distribution of the radiation emitter which was modeled as a linear oscillator. Such modeling needs to be justified.
- He assumed that the *energy distribution of the emitter does not follow a smooth curve as described by the linear oscillator; instead, it is a step-wise function with a constant jump of magnitude* ΔE. This assumption of partitioning the phase space (of the oscillator) in equal incremental area was somewhat arbitrary.
- There was a lack of understanding on the physical basis of the Planck's constant, h. He could not derive the physical meaning of this constant based on first principle.

Planck spent subsequent years trying to justify his theory on better physical grounds but was not successful [3]. "*My unavailing attempts to somehow reintegrate the action quantum into classical theory extended over several years and caused me much trouble*" [4].

So, although the Planck's relation was a great success, Planck was not satisfied with the physical meaning of h as he derived it. The other major contributors of quantum physics did not give much help in this regard. For example, the study of the photoelectric effect was an important demonstration of the quantum nature of light; Einstein, however, did not try to examine the physical meaning of h in his study. He simply used the Planck's relation to obtain the value of the energy quanta. Like Einstein, de Broglie did not try to re-examine the physical origin of the Planck's constant.

3.2 Derivation of the Planck's Relation Based on the Maxwell Theory

3.2.1 Energy and Momentum of the Electromagnetic Wave

The Planck's relation was originally postulated based on phenomenological considerations rather than first principles. In order to have a better understanding of this relation, we believe that it should be derived based on a more solid physical foundation, such as the Maxwell's theory [5]. In this chapter, we will uncover the physical meaning of h by treating the photon as a wave packet of electromagnetic radiation and directly calculating the total energy and momentum contained within the wave packet.[1] More explicitly,

- We will regard the photon as a wave packet which is made up of an oscillating electromagnetic field.
- The energy (E) and momentum ($\boldsymbol{p}$) of the electromagnetic field contained within the wave packet can be calculated based on the Maxwell theory.
- We will examine whether E is proportional to the oscillating frequency ν. If yes, the proportional constant will be identified as the Planck's constant.

The energy density of an electromagnetic field is known to be [6]

$$U = \frac{1}{2}\left(\varepsilon \mathbf{E}^2 + \frac{1}{\mu}\mathbf{B}^2\right), \tag{3.1}$$

where ε and μ are the dielectric permittivity and magnetic permeability of the vacuum; $\mathbf{E}$ and $\mathbf{B}$ are electric field and magnetic induction, respectively. According to the Maxwell's theory, $\mathbf{E}$ and $\mathbf{B}$ can be derived from the scalar potential Φ and the vector potential $\mathbf{A}$:

$$\begin{cases} \mathbf{B} = \nabla \times \mathbf{A} & (3.2) \\ \mathbf{E} = -\nabla\Phi - \dfrac{\partial \mathbf{A}}{\partial t}. & (3.3) \end{cases}$$

In electromagnetic radiation, the vector potential $\mathbf{A}$ obeys the wave equation

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0. \tag{3.4}$$

In order to calculate the energy density of an electromagnetic wave, let us choose a simple system in which the wave is traveling along the z-axis and the vector potential is along the x-axis (see Fig. 3.1), i.e., $\mathbf{A} = A_x \hat{x}$, and

[1] This chapter is based on our previous publication: D. C. Chang, Chin. Phys. B 26, 040301 (2017).

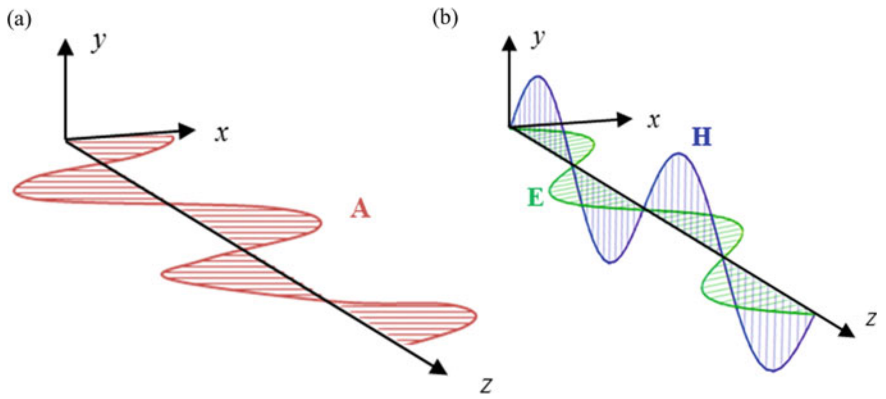


Fig. 3.1. A simplified model for the propagation of electromagnetic radiation. **a** Vector potential **A** oscillates along the x-axis and the wave is traveling along the z-axis. **b** The electric field **E** oscillates in x direction; the magnetic field **H** is perpendicular to **E** and oscillates in y direction. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Chin. Phys. B* **26**, 040301 (2017)

$$A_x = A_0 e^{i(kz-\omega t)}. \tag{3.5}$$

Since there is no embedded charge within the vacuum, $\Phi = 0$. Thus, Eqs. (3.2) and (3.3) become

$$\begin{cases} \mathbf{B} = \left(\dfrac{\partial A_z}{\partial y} - \dfrac{\partial A_y}{\partial z}\right)\hat{x} + \left(\dfrac{\partial A_x}{\partial z} - \dfrac{\partial A_z}{\partial x}\right)\hat{y} + \left(\dfrac{\partial A_y}{\partial x} - \dfrac{\partial A_x}{\partial y}\right)\hat{z} = \dfrac{\partial A_x}{\partial z}\hat{y} & (3.6) \\ \mathbf{E} = -\dfrac{\partial \mathbf{A}}{\partial t} = -\dfrac{\partial A_x}{\partial t}\hat{x}. & (3.7) \end{cases}$$

Substituting Eqs. (3.6) and (3.7) into Eq. (3.1), we have

$$U = \frac{1}{2}\left[\varepsilon\left|\frac{\partial A_x}{\partial t}\right|^2 + \frac{1}{\mu}\left|\frac{\partial A_x}{\partial z}\right|^2\right]. \tag{3.8}$$

This relation suggests that A_x is playing the role of the "basic field" in wave propagation (For a more detailed discussion about the concept of *basic field*, please see **Appendix B**). This point can be easily seen by comparing Eq. (3.8) with the energy density equation in a one-dimensional stretched string (see Fig. 3.2), which is

$$U = \frac{1}{2}\rho\left(\frac{\partial \phi}{\partial t}\right)^2 + \frac{1}{2}F_1\left(\frac{\partial \phi}{\partial z}\right)^2. \tag{3.9}$$

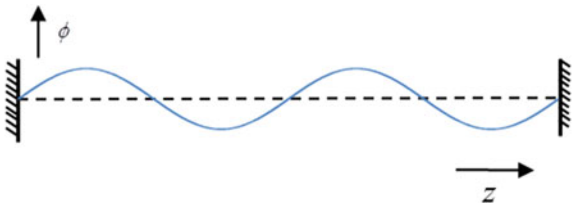


Fig. 3.2 **Wave propagation in a 1-D stretched string**. The wave is traveling along the z-axis and ϕ is the local displacement of the string. Credit: This figure is reproduced from an earlier publication of the author: D. C. *Chang, Chin. Phys. B* **26**, 040301 (2017)

Here, ρ is the mass density of the string, and F_1 is the tension of the string [7]. One can immediately see that, in the electromagnetic system, A_x appears to play the role of a propagating field, just like the displacement ϕ in the stretched string.

Recall that the speed of light $c = \omega/k = \sqrt{1/\varepsilon\mu}$, one can directly calculate the energy density of the electromagnetic system from Eqs. (3.5) and (3.8),

$$U = \frac{1}{2}\left[\varepsilon(\omega A_0)^2 + \frac{1}{\mu}(kA_0)^2\right] = \varepsilon\omega^2 A_0^2. \tag{3.10}$$

3.3 Calculating the Energy Contained Within a Wave Packet Based on Fourier Transform

To find the physical meaning of the Planck's constant, we need to calculate the total energy $\langle U \rangle$ contained within the wave packet representing one single photon. This energy can be obtained directly by integrating the energy density described in Eq. (3.10) over the entire volume of the wave packet:

$$\langle U \rangle = \iiint\limits_{v} U(x, y, z)\, dx\, dy\, dz. \tag{3.11}$$

In order to carry out this integration, one must know the structure of a photon. In the literature, a photon is usually described by Eq. (3.5). This, however, is not strictly correct since it represents a continuous wave, which spreads over the entire space and time. The photon should have a limited size along its trajectory (z-axis) and in the transverse plane (xy plane). It should be a wave packet, which is constructed by superposition of multiple wave components. Figure 3.3 shows three basic types of traveling waves: (a) *A continuous wave*. The wave frequency is a fixed constant. (b) *A wave packet with limited spread on the space and time dimensions* ($\Delta\omega$ is very small). (c) *A wave packet with a narrow spread over space and time* ($\Delta\omega$ is very large). What does a photon look like? Since a coherent light (such as a laser) has

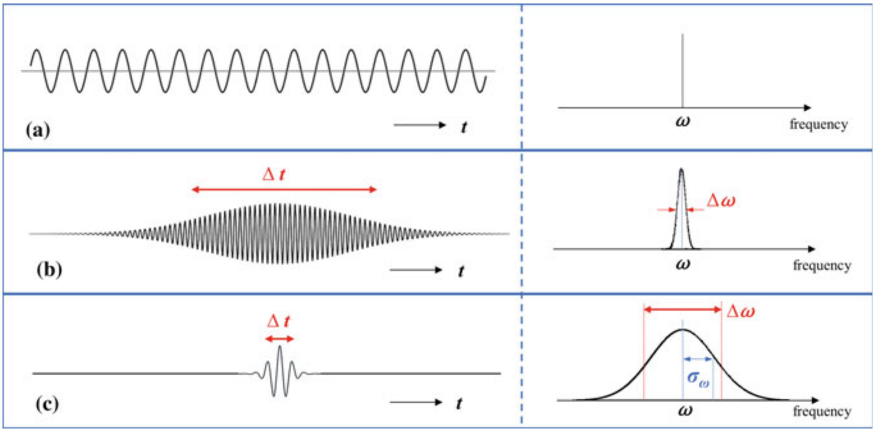


Fig. 3.3 Three basic types of traveling wave. The left-panel is plotted in time domain; the right-panel is plotted in frequency domain. **a** A plane wave; its frequency ω is a constant. **b** A wave packet with narrow linewidth; $\Delta\omega$ is very small in comparison with its average frequency ω. The half-width of the wave packet in the time domain is Δt. **c** A wave packet with large linewidth. Here, we also show the standard deviation σ_ω of the Gaussian function. Credit: This figure is reproduced from an earlier publication of the author: D. C. *Chang, Chin. Phys. B* **26**, 040301 (2017)

very narrow linewidth, the wave packet of a photon must be similar to that shown in Fig. 3.3b. Such a wave function can be written as

$$A_{wp} = A_0(x,\ y,\ z,\ t)\mathrm{e}^{i(kz-\omega t)}, \tag{3.12}$$

where $A_0(x,\ y,\ z,\ t)$ describes the envelope of the wave packet. Since the photon travels in a straight line along the z-axis, we can assume that it has axis symmetry, i.e.,

$$A_0(x,\ y,\ z,\ t)\ = A_T(r,\theta)A_L(z-ct), \tag{3.13}$$

where A_T and A_L are the transverse and longitudinal component, respectively. Equation (3.12) then becomes

$$A_{wp} = A_T(r,\theta)\ \underbrace{A_L(z-ct)\mathrm{e}^{i(kz-\omega t)}}_{A_{\text{path}}}. \tag{3.14}$$

A_{path} can be constructed by superposition of plane waves with frequency slightly different from the central frequency (ω), i.e.,

$$A_{\text{path}}(z,t) \approx \int_0^\infty g(\omega')\ \mathrm{e}^{i(kz-\omega' t)}d\omega', \tag{3.15}$$

where $g(\omega')$ is the frequency distribution function, which can be assumed to follow a Gaussian distribution with a standard deviation σ_ω, i.e.,

$$g(\omega') \sim \mathrm{e}^{-\left(\omega'-\omega\right)^2/2(\sigma_\omega)^2}. \tag{3.16}$$

It is well known that the Fourier transform of which will also give a Gaussian distribution in the time domain. That is,

$$A_L(z-ct) \sim \mathrm{e}^{-(z-ct)^2/2c^2(\sigma_t)^2}, \tag{3.17}$$

and

$$\sigma_\omega \cdot \sigma_t = 1. \tag{3.18}$$

Once we know the envelope function in the time domain, we can easily obtain the envelope function in the spatial domain along the z-axis,

$$A_L(z-ct) \sim \mathrm{e}^{-(z-ct)^2/2(\sigma_z)^2}, \tag{3.19}$$

where $\sigma_z = c \cdot \sigma_t$. Based on Eqs. (3.10) and (3.11), we can now calculate the total energy of the wave packet:

$$\begin{aligned}\langle U\rangle &= \iiint\limits_{v} (\varepsilon\omega^2 A_0^2)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\ &= \varepsilon\omega^2 \underbrace{\int_0^\infty \int_0^{2\pi} A_T^2(r,\theta)\, r\mathrm{d}\theta\mathrm{d}r}_{\rho_T} \underbrace{\int_{-\infty}^{\infty} A_L^2(z-ct)\,\mathrm{d}z}_{\rho_L}\end{aligned} \tag{3.20}$$

ρ_T and ρ_L in Eq. (3.20) can be calculated separately:

$$\rho_L = \int_{-\infty}^{\infty} \mathrm{e}^{-(z-ct)^2/(\sigma_z)^2}\mathrm{d}z = \sqrt{\pi}(\sigma_z). \tag{3.21}$$

Recall that $\sigma_z = c\ \sigma_t$ and $\sigma_\omega \cdot \sigma_t = 1$,

$$\sigma_z = c \cdot \sigma_t = c\,\frac{1}{\sigma_\omega}. \tag{3.22}$$

σ_ω is known to be related to the linewidth (or half-width, $\Delta\omega$) of the photon,

$$\Delta\omega = 2\sqrt{2\ln 2}\ \sigma_\omega = 2.355\ \sigma_\omega. \tag{3.23}$$

In most transmitting media, the linewidth of a wave is proportional to the frequency ω. This ratio is defined as the "Q factor",

$$Q = \frac{\omega}{\Delta\omega}. \tag{3.24}$$

The value of the Q factor is determined by the properties of the transmitting medium. Combining Eqs. (3.22), (3.23) and (3.24), we have

$$\sigma_z = c\,\frac{2.355}{\Delta\omega} = \frac{2.355\,cQ}{\omega}.$$

Substituting this into Eq. (3.21), and recall $c = \sqrt{1/\varepsilon\mu}$, we have

$$\langle U\rangle = 2.355\sqrt{\pi}\,\varepsilon\omega^2\frac{cQ}{\omega}\rho_{_T} = 2.355\;Q\sqrt{\frac{\pi\varepsilon}{\mu}}\rho_{_T}\omega. \tag{3.25}$$

3.3.1 Determination of the Planck's Constant

Next, we need to calculate the value of $\rho_{_T}$. Its value can be easily determined if one knows the functional form of A_T. As we had discussed earlier, the size of the wave packet in the transverse plane cannot be infinite. The simplest way to model A_T is to assume that it has a constant value up to a cut-off radius (r_0). A_T then vanishes when $r > r_0$. A more reasonable model, however, is to assume that A_T follows a bell-shaped Gaussian distribution (see Fig. 3.4), i.e.,

$$A_T(r,\theta) = a\mathrm{e}^{-r^2/2\sigma^2}, \tag{3.26}$$

where a is the amplitude of the envelope function. From Eqs. (3.20) and (3.26),

$$\begin{aligned}\rho_{_T} &= \int_0^{\infty}\int_0^{2\pi} a^2 e^{-r^2/\sigma^2} r\,\mathrm{d}\theta\,\mathrm{d}r = 2\pi a^2\int_0^{\infty}\mathrm{e}^{-r^2/\sigma^2} r\,\mathrm{d}r\\ &= \pi\sigma^2 a^2.\end{aligned} \tag{3.27}$$

This is closely related to the area of integrating the transverse component A_T along the x-axis, which we can call it "ζ" (see Fig. 3.4b),

$$\zeta \equiv \int_{-\infty}^{\infty} A_T\,\mathrm{d}x = \int_{-\infty}^{\infty} a\mathrm{e}^{-x^2/2\sigma^2}\mathrm{d}x = \sqrt{2\pi}\;a\sigma. \tag{3.28}$$

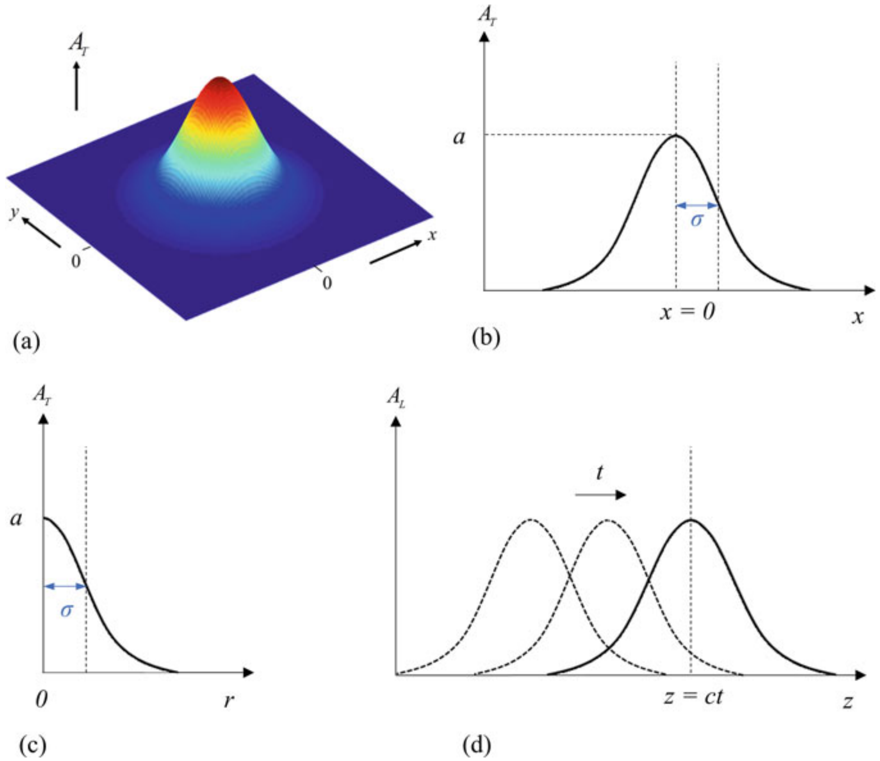


Fig. 3.4 A Gaussian distribution model of an electromagnetic wave packet. **a** A two-dimensional plot of the transverse component of the envelope function, A_T. **b** The cross-section plot of A_T along the x-axis; σ is the standard deviation; a is the wave amplitude at the peak. **c** Variation of A_T along the radius dimension. **d** A plot of the longitudinal component of the envelope function, A_L, along the z-axis. The wave packet moves at the speed c. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Chin. Phys. B* **26**, 040301 (2017)

Combining Eqs. (3.27) and (3.28), we have

$$\rho_T = \frac{1}{2}\zeta^2. \tag{3.29}$$

Substituting this result into Eq. (3.25), and recall $\omega = 2\pi\nu$, we get

$$\langle U \rangle = 2.355\, Q\sqrt{\frac{\varepsilon\pi}{\mu}}\frac{1}{2}\zeta^2\omega = \left(13.113\, Q\sqrt{\frac{\varepsilon}{\mu}}\zeta^2\right)\nu. \tag{3.30}$$

Since $\langle U \rangle$ represents the total electromagnetic energy of a single photon, Eq. (3.30) is identical to the Planck's relation $E = h\nu$, and the Planck's constant h can be identified as

$$h = 13.113\, Q\sqrt{\frac{\varepsilon}{\mu}}\zeta^2. \tag{3.31}$$

We will show later that ζ^2 has a clear meaning in quantum physics and should have a fixed cut-off value for a photon.

3.4 Derivation of the de Broglie Relation: Total Momentum Carried in a Wave Packet

By treating the photon as a wave packet of electromagnetic radiation, we can also calculate the total momentum contained within the wave packet. It is well known that the energy flow of the electromagnetic field can be described by the Poynting vector $\mathbf{S}$: [6]

$$\mathbf{S} = \varepsilon c^2 \mathbf{E} \times \mathbf{B}. \tag{3.32}$$

Since it can be shown that in a radiation system, $|\mathbf{E}| = c|\mathbf{B}|$, then,

$$\mathbf{S} = \varepsilon c^2 \mathbf{E} \times \mathbf{B} = \varepsilon c |\mathbf{E}|^2 \hat{\mathbf{z}}. \tag{3.33}$$

From Eq. (3.10), we know $\varepsilon|\mathbf{E}|^2 = \varepsilon\omega^2 A_0^2 = U$, Eq. (3.33) becomes

$$\mathbf{S} = cU\,\hat{\mathbf{z}}. \tag{3.33A}$$

The total energy flux of a wave packet of electromagnetic radiation then is

$$\langle \mathbf{S} \rangle = \iiint\limits_{v} \mathbf{S}\, \mathrm{d}x \mathrm{d}y \mathrm{d}z = \iiint\limits_{v} (cU)\, \mathrm{d}x \mathrm{d}y \mathrm{d}z\, \hat{\mathbf{z}} = c\langle U \rangle\, \hat{\mathbf{z}} \tag{3.34}$$

For an electromagnetic wave, the momentum density ($\mathbf{g}$) is known to be related to the Poynting vector $\mathbf{S}$ by [6]

$$\mathbf{g} = \frac{1}{c^2}\mathbf{S}. \tag{3.35}$$

Thus, the total momentum of a wave packet is

$$\langle \mathbf{g} \rangle = \iiint\limits_{v} \mathbf{g}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = \frac{1}{c^2} \iiint\limits_{v} \mathbf{S}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = \frac{1}{c^2}\langle \mathbf{S} \rangle. \tag{3.36}$$

Substituting Eq. (3.34) into Eq. (3.36), and using Eq. (3.30), we have.

$$\langle \mathbf{g} \rangle = \frac{1}{c} \langle U \rangle \, \hat{\mathbf{z}} = 13.113 \, Q \sqrt{\frac{\varepsilon}{\mu}} \zeta^2 \, \frac{\nu}{c} \, \hat{\mathbf{z}}. \tag{3.37}$$

Previously, we have already identified the value of h from Eq. (3.31). Recall that the wave vector $k = 2\pi/\lambda = 2\pi\nu/c$, Eq. (3.37) becomes

$$\langle \mathbf{g} \rangle = h \, \frac{\nu}{c} \, \hat{\mathbf{z}} = \frac{h}{2\pi} k \, \hat{\mathbf{z}} = \hbar k \, \hat{\mathbf{z}}. \tag{3.38}$$

Equation (3.38) shows that the total momentum of a photon is proportional to its wave vector $\boldsymbol{k}$. Equation (3.38) is identical to the de Broglie relation [8],

$$\boldsymbol{p} = \hbar \boldsymbol{k}. \tag{3.39}$$

Therefore, the Planck's constant derived by us not only satisfies the Planck's relation, it also satisfies the de Broglie relation.

3.5 Derivation of Heisenberg's Uncertainty Principle

One of the most important relations in quantum mechanics is the "*Uncertainty Principle*" proposed by the German physicist, Werner Heisenberg (see Fig. 3.5). Based on the wave packet model, not only can we derive the Planck's relation and de Broglie's relation, but we can also easily derive Heisenberg's Uncertainty Principle.

Fig. 3.5 Werner Heisenberg. Werner Heisenberg (1901–1976) was a German theoretical physicist. He introduced the use of matrix into quantum mechanics and developed a set of quantum mechanical equations. During his work at the University of Copenhagen with Bohr, he published the famous "Heisenberg Uncertainty Principle" (1927). In 1932, he was awarded the Nobel Prize in Physics for his research on quantum mechanics. Credit: AIP Emilio Segrè Visual Archives, Segrè Collection

If one accepts that a photon is a wave packet of oscillating electromagnetic field, which follows a Gaussian distribution along the particle trajectory, the half-width of the wave packet in the time domain can be directly determined from the linewidth of the radiation wave. From the condition of Fourier transform, we know

$$\sigma_\omega \cdot \sigma_t = 1, \tag{3.18}$$

where σ_ω and σ_t are the standard deviations in the frequency domain and the time domain. Since we know the half-width of the wave packet is $\Delta t = 2.355\ \sigma_t$ and the linewidth of the oscillation frequency is $\Delta\omega = 2.355\ \sigma_\omega$, Equation (3.18) implies $\Delta\omega \cdot \Delta t = (2.355)^2$. From the Planck's relation, $E = \hbar\omega$, we have

$$\Delta E \cdot \Delta t = \hbar\Delta\omega \cdot \Delta t = \frac{h}{2\pi}(2.355)^2 = 0.8827\ h. \tag{3.40}$$

This suggests that the product of linewidths in the energy and time domains for a single photon is very close to h. Such a result agrees with Heisenberg's conjecture that

$$\Delta E \cdot \Delta t \approx h. \tag{3.41}$$

Thus, Heisenberg's Uncertainty Principle can be interpreted as a direct result of the fact that a photon is a wave packet which follows a Gaussian distribution.

Using this wave packet model, we can also easily obtain the Uncertainty Principle between Δp and Δz. Recall from the de Broglie relation, $\Delta\ p = \hbar\ \Delta k = \hbar(2.355\sigma_k)$. The half-width of the wave packet is $\Delta z = 2.355\ \sigma_z$. From the conditions of the Fourier transform,

$$\sigma_k \cdot \sigma_z = 1. \tag{3.42}$$

Then, the above relations give

$$\Delta p \cdot \Delta z = \hbar\ \Delta k \cdot \Delta z = \frac{h}{2\pi}(2.355)^2 \sigma_k\ \sigma_z = 0.8827\ h. \tag{3.43}$$

This agrees with the conjecture of Heisenberg that $\Delta p \cdot \Delta z \approx h$ [9].

3.6 The Principle of All-Or-None: Physical Meaning of the Planck's Constant as Derived from the Maxwell Theory

In the foregoing sections, we demonstrated that one can directly calculate the energy of a photon based on the Maxwell's theory. Based on this result, the Planck's constant is given by

$$h = 13.113\, Q\sqrt{\frac{\varepsilon}{\mu}}\zeta^2. \tag{3.31}$$

Apparently, the Planck's constant is dependent on the physical properties of the vacuum, e.g., the dielectric permittivity ε and magnetic permeability μ. The quality factor Q is also a property of the vacuum, since it is dependent on the transmitting medium. At this point, we do not know enough about the detailed properties of the vacuum to directly calculate Q. But the value of Q can be determined by experiment. One can use an optical device to directly measure the linewidth of a photon with known frequency. In the literature, there were already some hints about the value of Q. For example, it was reported that a solid-state dye laser (at 590 nm) could have a linewidth around 350 MHz [10]. This suggests that the Q factor is about 1.45×10^6. Our work may motivate more accurate measurement of Q in the future.

The remaining problem is to consider whether ζ^2 can be regarded as a constant and what does that mean. From Eq. (3.28), ζ is defined as the integrated area of the vector potential at the center of the wave packet. The requirement for ζ being a constant means that: *Regardless of the oscillating frequency, in order to generate a sustainable oscillating wave, nature requires a fixed amount of disturbance in the electromagnetic field*.

This situation may be understood using an analogy; it is somewhat similar to the generation of a nerve impulse. We know that a neuron can transmit a signal to its downstream target along its nerve fiber (called "axon"). This signal is called an "action potential" [11]. It is well known that the generation of action potential has the property of "**all-or-none**". That means, when the stimulus to the axon is below a threshold, no action potential can be generated. But when the stimulus is higher than the threshold, a full size action potential will be generated. This action potential will propagate along the axon with a constant amplitude (about 100 mV) [12]. In another word, one cannot generate an action potential with arbitrarily small amplitude. And, no matter how large is the stimulus; one cannot generate an action potential which is much larger than 100 mV. That is why people called it "all-or-none".

As it turns out, this principle of *all-or-none* is applied in multiple aspects of nature. Not only the transmission of a nerve impulse is all-or-none, the transmission of the electromagnetic radiation is also all-or-none. The radiation energy is apparently transmitted in small packets (photon), each of which has a limited size and is not sub-dividable. Either one can generate a complete photon or generate no photon at all. In another word, if the energy of the electromagnetic field is smaller than a critical

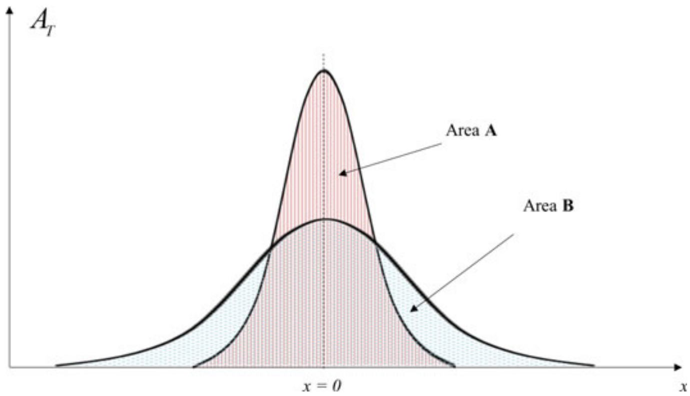


Fig. 3.6 Integrated area of the transverse component A_T along the x-axis (denoted by "ζ") is a constant during wave propagation. **A** and **B** represent the transverse components of two different wave packets with different oscillation amplitudes (a) and widths (σ). Nature requires that their integrated areas ($\sqrt{2\pi}\ a\sigma$) to be the same, i.e., Area A = Area B. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Chin. Phys. B* **26**, 040301 (2017)

value, it will not be able to trigger a transmissible excitation wave traveling as a wave packet. Instead, the energy will just dissipate in the surrounding.

This requirement of "*all-or-none*" means that ζ should have a fixed cut-off value; it cannot be arbitrarily small. Thus, although the size of the wave packet is not fixed, the total amount of disturbance in the electromagnetic field (as measured by ζ) is fixed (see Fig. 3.6). If the diameter of the wave packet is very small, the oscillation amplitude of the electromagnetic field within the wave packet must be large enough to make ζ reach the threshold value. Alternatively, if the oscillation amplitude of the wave packet is small, the size of the wave packet must be large enough to compensate it so that the integrated area reaches the threshold value.

Besides the above considerations, there is another reason suggesting that ζ^2 should be a constant. Recall that ζ^2 was defined by

$$\rho_T = \int_0^\infty \int_0^{2\pi} |A_T(r,\theta)|^2 r \mathrm{d}\theta \mathrm{d}r = \frac{1}{2}\zeta^2. \tag{3.29}$$

As we pointed out earlier, the vector potential (**A**) in an electromagnetic radiation system is equivalent to the amplitude of an oscillation wave in a one-dimensional vibrating string. Thus, if one wants to write down the wave function of a photon (ϕ), one can guess that ϕ must be related to **A** (with a normalizing factor). Since the absolute square of the wave function $|\phi|^2$ is usually interpreted as the probability of finding the particle, this suggests that the absolute square of the vector potential of the wave packet, $|A_{wp}|^2$, is proportional to the probability of finding the photon. Recall that $|A_{wp}|^2 = |A_0|^2 = |A_T|^2$ at the center of the wave packet, where $z - ct = 0$, one can interpret

$$\rho_T = \int_0^{\infty}\int_0^{2\pi} |A_T(r,\theta)|^2 r d\theta dr \sim \int_0^{\infty}\int_0^{2\pi} |\phi|^2 r d\theta dr$$

as a measure of "the total probability of finding the photon at the center of the wave packet". So long as the wave packet represents a single photon, this probability should remain constant as the wave packet travels along the trajectory of the photon. Thus, the requirement of ζ^2 being a constant essentially means that, in an optical measurement, *the probability of finding the photon at the center of the wave packet is always the same.*

In conclusion, we showed that the energy and momentum of a photon can be determined by treating the photon as a wave packet of electromagnetic radiation. We found the energy of a photon is indeed proportional to its oscillation frequency, as postulated in the Planck's relation. Using the Maxwell theory, one can explicitly derive the Planck's constant *h*. Furthermore, by treating the photon as a quantized wave packet of electromagnetic wave, one can also derive the de Broglie relation and Heisenberg's Uncertainly Principle (see Table 3.1)

Therefore, the observation of radiation transmitted in discrete energy quanta does not mean that the photon is a corpuscular object. Instead, the phenomenon of energy quantization only means that the photon is a wave packet which has a critical size; the wave packet cannot be arbitrarily small. And, emission and transmission of radiation energy must follow the principle of *all-or-none*.

Table 3.1 Derivation of the quantum relations based on the Maxwell theory. Here, the quantum particle is modeled as a quantized wave packet of electromagnetic wave

Name of the quantum relation	**Relation for a quantum particle**	**Equivalent relations for a wave packet**	**Physical basis of the quantum relation**
Planck's relation	$E = h\nu$	Total energy of a wave packet $\langle U\rangle = \left(13.113\, Q\sqrt{\frac{\varepsilon}{\mu}}\zeta^2\right)\nu$	The Planck's constant is identified as $h = 13.113\, Q\sqrt{\frac{\varepsilon}{\mu}}\zeta^2$
de Broglie relation	$\boldsymbol{p} = \hbar\boldsymbol{k}$	Total momentum of a wave packet $\langle \mathbf{g}\rangle = 13.113\,\mathrm{Q}\sqrt{\frac{\varepsilon}{\mu}}\zeta^2\,\frac{\boldsymbol{k}}{2\pi}$	The Planck's constant is identified as $h = 13.113\, Q\sqrt{\frac{\varepsilon}{\mu}}\zeta^2$
Heisenberg's Uncertainty Principle	$\Delta E \cdot \Delta t \approx h$ $\Delta p \cdot \Delta z \approx h$	$\sigma_\omega \cdot \sigma_t = 1$	This relation is based on the requirement of Fourier transform

3.7 Chapter Summary

- In this chapter, we showed that there is a logical transition from classical physics to quantum mechanics.
- More specifically, the fundamental quantum relations, including the *Planck's relation*, the *de Broglie relation*, and *Heisenberg's Uncertainty Principle*, can all be derived based on the classical Maxwell theory.
- From these derivations, one can identify the physical meaning of the *Planck's constant*, which is shown to be dependent on the physical properties of the vacuum.
- The quantization of light is a manifestation of the *principle of all-or-none* in the transmission of the electromagnetic radiation; it means that the wave packet has a minimum size and is not sub-dividable. When the vacuum is excited, it can either generate a complete wave packet (a photon) or generate no transmittable wave at all.
- Thus, the photon is a *wave* in nature; it behaves like a *particle* only in the sense that it carries a package of energy which is not sub-dividable. Furthermore, its momentum is also quantized following the de Broglie relation.
- Heisenberg's Uncertainty Principle can be interpreted as a direct result of the fact that the quantum particle (such as a photon) is a wave packet, the energy content of which follows a Gaussian distribution.

References

1. Planck, M. (1900). Zur theorie des gesetzes der energieverteilung im normalspectrum. *Verhandl. Dtsch. Phys. Ges., 2*, 237–245.
2. Longair, M. S. (1984).*Theoretical Concepts in Physics: An Alternative View of Theoretical Reasoning in Physics*. Cambridge University Press.
3. Slater, J. C. (1969). *Concepts and Development of Quantum Physics*. Dover Publications.
4. Strickland, J. (2011). *Weird Scientists—the Creators of Quantum Physics*. Lulu.com.
5. Chang, D. C. (2017). Physical interpretation of the Planck's constant based on the Maxwell theory. *Chinese Physics B, 26*, 040301.
6. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). *The Feynman Lectures on Physics* (Vol. 2). Addison-Wesley.
7. Chang, D. C., & Lee, Y.-K. (2015). Study on the physical basis of wave-particle duality: Modelling the vacuum as a continuous mechanical medium. *Journal of Modern Physics, 6*, 1058–1070.
8. de Broglie, L. (1924). Recherches sur la théorie des quanta
9. Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Physik, 43*, 172–198.
10. Duarte, F. J. (1999). Multiple-prism grating solid-state dye laser oscillator: Optimized architecture. *Applied Optics, 38*, 6347.
11. Kuffler, S. W., Nicholls, J. G., & Martin, A. R. (1984). *From Neuron to Brain: A Cellular Approach to the Function of the Nervous System* (2nd ed.). Sinauer Associates.
12. Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology, 117*, 500–544.

Chapter 4
The Merging of the Particle and Wave Concepts: Evidence Suggesting that the Sub-atomic Particle is a Quantized Excitation Wave

As we mentioned above, one of the most amazing discoveries in quantum physics is that a quantum particle can behave both like a corpuscular object and a wave. How did this concept of *wave-particle duality* develop? It was based on both theoretical predictions and experimental verifications. In the nineteenth century, it was well known that light is an electromagnetic wave. However, it was discovered later that light can have the properties of particle too. As we mentioned in the last two chapters, this new understanding was mainly due to the works of three scientists, namely: Max Planck's work on black-body radiation [1], Einstein's photoelectric effect [2], and Compton's X-ray scattering experiment [3]. Their work demonstrated that the light wave could behave like a particle. The next surprising discovery in quantum physics is that the opposite can also be true; namely, a particle can behave like a wave!

We know the electron is a particle with rest mass. Surprisingly, it was found in some experiments that the electron can sometimes behave like a wave. This creates a problem that has troubled physicists for a long time. This problem is generally known as the "*wave-particle duality*".

4.1 The Discovery of Massive Particle Behaving Like a Wave

4.1.1 The Revolutionary Idea of de Broglie

In 1924, a physics PhD student in France named Louis de Broglie (see Fig. 4.1) came up with an interesting idea [4]. He thought that since **light waves** have the properties of **particles,** can **particles** also have the properties of **waves**? At that time, it was already known that a photon has momentum, which is

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_4

Fig. 4.1 Louis de Broglie. Louis de Broglie (1891–1987) was a French physicist. In 1924, he put forward a bold hypothesis in his doctoral dissertation; that is, electrons also have the wave nature as photons, and the momentum of the particle is proportional to its wave number. His hypothesis was later confirmed experimentally. This won him the Nobel Prize in Physics in 1929. Photo Credit: Wikimedia Commons; Public domain

$$p = \hbar k, \tag{4.1}$$

(where p is momentum, k is wavenumber, $k = 2\pi/\lambda$). The left-hand side of this equation describes the physical property of a particle (momentum), while the right-hand side of this equation describes the physical property of a wave (wavenumber). De Broglie put forward a bold hypothesis in his doctoral dissertation: he proposed that not only photons obey $p = \hbar k$, particles with mass (such as electrons) also obey this relationship.

What is the basis of this hypothesis? This is based on a bold conjecture of de Broglie. He was mainly inspired by Einstein's work on the photoelectric effect. He said: "*After long reflection in solitude and meditation, I suddenly had the idea, during the year 1923, that the discovery made by Einstein in 1905 should be generalized by extending it to all material particles and notably to electrons*" [5]. Thus, in de Broglie's mind, radiation waves and matter waves could have similar physical properties.

De Broglie's hypothesis was too innovative at the time. A professor who reviewed his doctoral dissertation, Paul Langevin, was not sure if that was reasonable. After reading de Broglie's doctoral dissertation, he was very hesitant to let de Broglie pass. So he wrote a letter to Einstein, along with de Broglie's thesis, to seek for his opinion. Einstein wrote a short message to Langevin, giving a positive evaluation of the thesis. Because of that, de Broglie was able to successfully get his doctorate degree. [6]

4.1.2 Confirmation of the de Broglie Relation Using Bragg's Diffraction Experiments

When de Broglie proposed his hypothesis in 1924, it attracted some attention in the European physics community; there were some interests to test his hypothesis experimentally. If de Broglie's hypothesis is correct, electrons would have wave properties and will exhibit a diffraction pattern similar to light.

The early twentieth century was a golden age of new discoveries in physics. Several years before de Broglie put forward his hypothesis, some experimental physicists had already started doing diffraction experiments using electrons. In 1921, a physicist at the Western Electric laboratory (later becomes the "Bell Labs") in the United States, Clinton Davisson, had conducted diffraction experiments by using electrons to hit different metal surfaces [7]. The original purpose of the experiment was to gain a better understanding of the structure of atoms. Inspired by Rutherford's 1911 experiment in striking gold foil with alpha particles, Davisson thought that using electrons for bombardment could yield more information because the electron is much smaller than alpha particles. He placed a piece of metal in a vacuum tube, shot it with an electron beam, and used a movable electron detector to record the spread of the diffracted electrons. However, the results he got only supported the atomic model of Rutherford; he had no new discoveries.

In late 1924, Germer joined Davisson's group and they resumed the diffraction experiments. The progress of their experiment was still very slow. Later, an accident in their equipment gave them an unexpected inspiration. They found out that if they switched their metal sample to single-crystal nickel, they could obtain much clearer diffraction experimental results than before [7].

In the summer of 1926, Davisson took a vacation to Britain. During this trip, Davisson stopped by Oxford to attend an academic conference. At that time, Europe was the center of physics. After de Broglie put forward the hypothesis of matter waves in 1924, Schrödinger in 1926 had just published the wave equation of electrons (Schrödinger equation). Being far away in America, Davisson didn't know much about these new developments. Davisson attended a lecture by Max Born at the Oxford conference in which it was mentioned that de Broglie's theory of matter waves could be tested experimentally. One of the experiments cited by Max Born was the diffraction experiment Davisson did in 1923. This surprised Davisson. After this meeting, Davisson returned to his laboratory and decided to change the purpose of his experiment to verifying the wave theory of electrons. Through rigorous and meticulous work, Davisson and Germer finally found clear electron diffraction results. Its characterization is similar to that obtained by diffraction using X-rays, which fully agrees with de Broglie's relation. They published their experimental results in *Nature* in 1927, clearly demonstrating the wave properties of electrons and thus supporting the theory of matter waves as proposed by de Broglie [8].

One month after Davisson and Germer's article was published, a British physicist George Thomson also published the results of their experiments verifying the properties of electron waves in *Nature* [9]. They shot electrons to a sheet of metal, behind which there was a detection plate. They found that the detection plate showed a pattern of diffraction with co-centric rings. This image was very similar to using X-rays to shine on a sheet of metal.

Davisson and Thomson shared the 1937 Nobel Prize in Physics for their discoveries of electron diffraction and thus confirming that electrons have wave properties. Later, it was demonstrated that many massive particles besides electrons could also exhibit wave properties. These particles include neutrons, helium, neon, argon, etc. [10–15].

4.1.3 Double-Slit Experiment for a Single Electron

Although the diffraction experiment of electrons has confirmed the wave nature of electrons, a more direct experiment to show that electrons are waves would be the double-slit interference experiment. Feynman had proposed such a thought experiment in his "*Feynman Lectures on Physics*" [16]. This double-slit interference experiment can clearly demonstrate that an electron does not behave like a classical particle (i.e., a point mass).

Feynman's electron double-slit interference experiment was inspired by the double-slit experiment of light (see Fig. 4.2). Previously, the wave nature of light was demonstrated based on the results of the double-slit interference experiment of light as proposed by Thomas Young. Now, if one uses an electron beam instead of a light beam to conduct the double-slit interference experiment, one can easily determine if the electron has wave properties or not. For example, if an interference pattern is produced on the screen behind the double-slit, one will know that the electron is a wave.

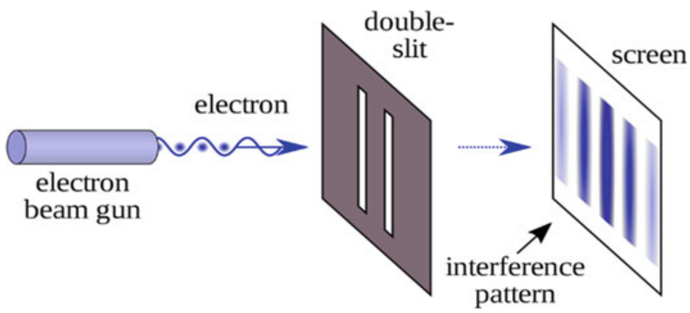


Fig. 4.2 A double-slit experiment using electrons yields a pattern of interference fringes. From the known wave-particle duality property of electrons, Feynman proposed that if a double-slit experiment was performed with electrons, the pattern of interference fringes would be obtained. Image Credit: Original: NekoJaNekoJa, Vector: Johannes Kalliauer, Wikimedia Commons; CC BY-SA 4.0

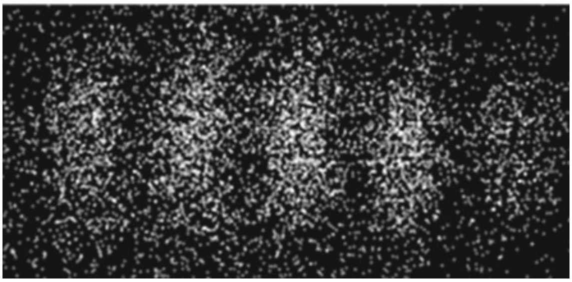


Fig. 4.3 Pattern of interference fringes obtained from the electron double-slit experiment. Experimental results of an electron double-slit experiment. The two slits were separated by a distance of 272 nm. Electrons accelerated to 600 eV were fired at the double slit. Behind the double-slit was a detector plate to record where the electron landed. The image above is a cumulative image of 6235 electrons; it shows a clear interference pattern. Image credit: R. Bach et al. *New J. Phys.*, Vol. 15. (2013) CC BY-SA 3.0

At the time when Feynman proposed the electron double-slit interference experiment (in the 1960s), it was technically difficult to do. Since the de Broglie wavelength of electrons is much shorter than that of photons of ordinary visible light, one must narrow the slits to a very small size. In recent years, with the development of semiconductor chips, scientists can use micro-fabrication technology to do the nano-scale engraving work. Now, the double-slit interference experiment for electron can be easily demonstrated [17] (see Fig. 4.3).

The double-slit experiment not only demonstrated that the electron behaves like a wave, the most amazing thing about this experiment is that the generation of such interference fringes does not come from the mutual interference between different electrons, but an electron itself can interfere with itself. How do we know that? When doing a double-slit interference experiment, the researcher can reduce the density of electrons shooting at the double-slit to less than one electron per second. One can still observe the interference fringes on the screen behind the double-slit (see Fig. 4.3).

If we treat the electron as a particle (like a tiny bullet), an electron can only pass through one of the two slits. There should be no interference, since it is impossible for the electron to pass through the other slit at the same time. The observed interference phenomenon implies that a single electron must pass through both slits at the same time. How can the electron do that? This really puzzled scientists.

So, this double-slit experiment demonstrated a very strange phenomenon; in one way, the electron seems to behave like a particle (i.e., producing a single spot at the detector); but in another way, it can't be an individual particle; it must behave like a wave (i.e., producing an interference pattern). This has really puzzled scientists for a long time.

Therefore, Feynman made the following conclusion in his *Lectures on Physics*: "*And no one has figured a way out of this puzzle. So at the present time we must limit ourselves to computing probabilities. We say "at the present time," but we suspect* *very strongly that it is something that will be with us forever—that it is impossible to beat that puzzle—that this is the way nature really is*" [16].

4.2 How to Explain Wave-Particle Duality? The Statistical Interpretation of the Copenhagen School

When de Broglie first proposed the wave-particle duality of electrons in 1924, most scientists were skeptical about it. But with the results of Davisson and Thomson's electron diffraction experiments, it became widely accepted that electrons have the properties of waves. In 1926, Schrödinger published his famous quantum wave equation for the electron, which is highly successful in explaining the structure of atoms and many other microscopic properties of physical systems [18]. The wave function (ψ) in the Schrödinger equation is called the "matter wave". Then, it became a big challenge for physicists to explain what the physical meaning of the matter wave is. How can one connect the concept of "matter wave" with the presence of the electron?

In 1927, a group of physicists working at the Copenhagen University (including Bohr and Heisenberg) proposed an explanation for the physical meaning of matter waves [19, 20]. They suggested that the electron itself is a point-mass-like particle, but its distribution could appear in the form of a wave. Therefore, the so-called "**matter wave**" is not a real wave, but a "**probability wave**"; it mainly reflects the probability of the particle appearing at a certain time and location.

According to this statistical point of view, the wave function (ψ) in the Schrödinger equation is directly related to the probability of an electron appearing at certain time and position. More specifically, they argue that *the square of the absolute value of the wave function represents the probability of a particular electron appearing in a certain space–time position*:

The probability of an electron appearing at a certain space–time $= |\psi(x, t)|^2$.

Here x is the position and t is the time. The leader of this theory was Bohr (see Fig. 4.4). In 1918, he founded an Institute for Theoretical Physics at the University of Copenhagen (later called the "Niels Bohr Institute"). The institute began operating in 1921, and Bohr was awarded the Nobel Prize in Physics in 1922. His institute attracted many quantum physicists to visit and work there. Therefore, their interpretation of matter waves was known as the "*Copenhagen Interpretation*".

4.2.1 Debates on the Probabilistic Interpretations

The Copenhagen interpretation subsequently became the mainstream theory in quantum mechanics. In fact, the interpretation of the Copenhagen School was later given a philosophical extension. It seemed to imply that the microscopic world is a

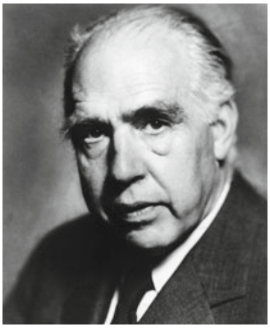


Fig. 4.4 **Niels Bohr.** Niels Bohr (1885–1962) was a Danish physicist. In 1911, he went to England to study with Rutherford. After returning to Denmark, Bohr developed his quantum model of an atom. In 1922, he won the Nobel Prize in Physics for his research on atomic structure and quantum theory. Photo Credit: Photograph by A. B. Lagrelius and Westphal, courtesy AIP Emilio Segrè Visual Archives, W.F. Meggers Gallery of Nobel Laureates Collection

probabilistic world; everything is uncertain. In theory, a microscopic system can have several possible states; the quantum state is just a superposition of these different states. The result obtained from a quantum mechanical calculation simply gives the probabilities of finding the various states. The physical state of a system can only be determined when one conducts an experiment to measure a certain physical property of the system.

The Copenhagen School basically looked at electrons from a particle point of view. While their interpretation of the quantum wave function succeeded in giving a plausible explanation for matter waves, not all scientists accepted it. Some prominent physicists have expressed doubts about the probabilistic world implied by this Copenhagen interpretation [21]. Some scientists believe that matter waves should be real waves with physical properties, not just an elusive probability.

Einstein had always been skeptical of the concept of probability waves. He has a famous saying "*God does not throw dice*" [22], which is a criticism of the Copenhagen interpretation. Another well-known physicist, Schrödinger, was also not satisfied with the statistical probability explanation advocated by the Copenhagen School. As the inventor of the quantum wave equation for electrons, he has always been interested in what the physical nature of matter waves is. In the 1930s, he had several exchanges with Einstein on this topic. In 1935, Einstein proposed a thought experiment using "gunpowder" as an example in a letter to Schrödinger [6]. Imagine there is a stack of gunpowder, some of which are unstable and could cause an explosion. If we use a quantum mechanical equation to represent this state, according to the Copenhagen interpretation, the state represented by the wave function is just a superposition of several probabilities. So, for this pile of gunpowder, does its wave function include both the part of the gunpowder that hasn't exploded and the part of the gunpowder that has already exploded? This is impossible in reality. "Because in reality gunpowder cannot exist in an intermediate state between exploded and unexploded" [6].

Quantum mechanics has been continuously developed and applied in modern times; its success is beyond doubt. But controversy over the interpretation of quantum wave functions has not stopped. Is the matter wave a real physical wave or just a probabilistic wave? What is the physical meaning of the quantum wave function? Even today, scientists are still trying to come up with different ideas to answer those questions.

4.3 Evidence Suggesting that the Electron is a Physical Wave

4.3.1 Why Do We Think Elementary Particles Are Waves?

At present, we know at least one elementary particle (photon) is a physical wave. From the Maxwell theory, we know the photon is an oscillation of the EM field; it is a quantized wave packet. For visible light waves, their wavelength is from 400 to 700 nm. Since the size of a wave packet must span over multiple wavelengths, this means the width of a photon is quite large (>10 microns). The photon is clearly not a point-like object.

Is the electron a point-like object or a wave? In many ways, an electron behaves just like a photon. Such similarities between electrons and photons had been recognized by many physicists. For example, according to Richard Feynman: "*…electrons behave just like light. The quantum behavior of atomic objects … is the same for all, they are all 'particle waves,'…so what we learn about the properties of electrons… will apply also to all 'particles,' including photons of light*" [16].

A large amount of experimental evidence had indicated that the electron has wave properties. For example, as we have discussed in the above, electrons can be diffracted by a crystal following the Bragg's Law [8]. If the electron is a point-like object, it can only bounce from one atom in the crystal and thus should not form an interference pattern following the Bragg's Law. Furthermore, it is known that a single electron can pass through two near-by slits to give an interference pattern [17]. How can a point-like electron do that? If the electron is really a corpuscular object, it can only pass one of the slits at one time and thus could never form an interference pattern.

According to the Bohr-Sommerfeld quantization condition, the length of the electron orbit within an atom must equal to an integral number of electron wavelength [21]. This suggests that the electron inside an atom is a resonating wave. Furthermore, if the electron is a point-like object, the atom will be vastly empty. How can an atom behave like a hard sphere? Also, how can the point-like electrons hold atoms together to form molecules? Can a "probability wave" form chemical bonds between atoms and hold them together?

The only logical explanation is that the electron must be a wave in nature. It should be a physical wave instead of probability wave. In fact, it is this wave property of the electron that allows us to build an electron microscope [23–25]. If the electron is a corpuscular object and its wave property is only associated with a "probability" of its distribution, how can one explain the principle of a transmitting electron microscope?

Another strong reason for us to believe that the electron is a wave is the fact that an electron can be created or annihilated in the vacuum. It is well known in quantum electrodynamics that an electron–positron pair can be created by an energetic photon. Conversely, an electron and a positron can be annihilated to give a photon. It is very difficult to explain these observations, if one regards the electron as a massive point object. On the other hand, such observations can be explained very easily if one believes that the electrons (and the positrons) are excitation waves of the vacuum.

Finally, not only de Broglie's relation, all quantum relations of photon, including Planck's relation and Heisenberg's uncertainty principles, are also shown to be applicable to massive particles. This implies that particles with or without masses are the same thing; that is, they are quantized physical waves!

4.4 Hints from the Collider Experiments: How Can Particles Be Created from Nowhere?

In modern physics, many sub-atomic particles (including electrons, neutrons, and alpha particles) were first discovered through the study of different types of radiation. Later, many new particles were discovered in collider experiments. Starting from the 1930s, scientists started to use accelerators to accelerate particles to very high speeds and bombard them with a target [26]. Later, these accelerated particles were directed to collide with each other. Using some highly sophisticated detectors, the collision products can be analyzed in detail. Occasionally, some new particles could be discovered from these collision products (see Fig. 4.5).

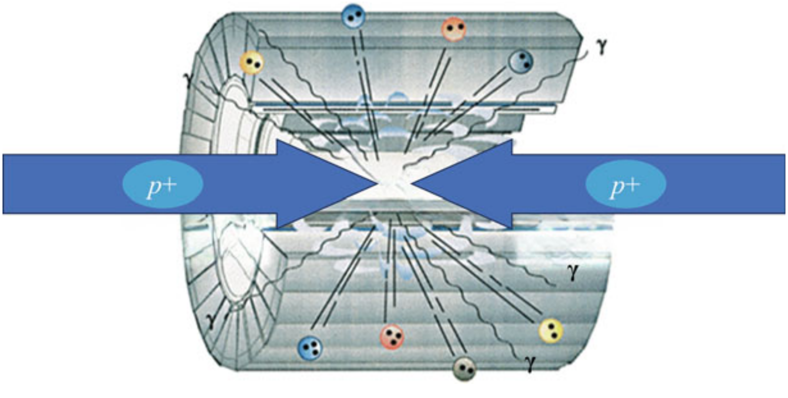


Fig. 4.5 Schematic representation of the collider experiment. Scientists use accelerators to accelerate particles to very high speeds and then direct them to collide with each other. Special detectors are used to analyze the material produced after the collision. In this example, the colliding particles are protons

The original idea behind these collider experiments was very simple. It was designed to find out what the particle is made of. Suppose an alien ("*superman*") comes to Earth for the first time and he sees cars running on the road, but he doesn't know what cars are made of. He could use his super power to accelerate two cars to a high speed, and then make them collide with each other. After the cars crashed, many parts will fall out. By analyzing the fallout parts, the alien could roughly figure out what are the basic materials that make up cars on Earth. The experimental particle physicists had a similar idea; they used particle colliders to smash one type of particle against another type of particle, and hope to find out the particle's building blocks from the debris. In the past few decades, they had indeed found a lot of useful experimental results from the products of particle collisions, allowing us to know the composition of some particles (see Fig. 4.6).

However, this method of analyzing parts by smashing an object into pieces has practical limits. In many more recent particle collision experiments, the interpretation of the results is far more complicated than the above-mentioned car collision experiment. In fact, we found a very strange phenomenon, that is, after two particles collide, it will produce many new species of particles whose rest masses are much larger than the original particles. This is as if that, when the alien mentioned above collided two cars, he found that the objects coming out were not car parts; instead, the fallout objects are very big and far heavier than the cars. For example, such fallout objects could be heavy tanks, fighter jets, etc. (see Fig. 4.7).

Particle physicists today are facing the same problem. Why did particle collision produce particles of new species with far greater masses than the original particles? Where do these new particles come from?

To answer this question, physicists basically have two choices: (1) the new particles already exist in the vacuum beforehand; (2) all particles are excitation waves of the vacuum; when energy is applied to the vacuum, new waves can be generated.

Traditionally, most particle physicists seem to favor the first explanation. For example, the first theory to predict the existence of anti-particles for electrons is Dirac's electron theory, which assumed the existence of an infinite number of negative-energy electrons in space [27, 28]. (This scenario is called the "Dirac ocean" of negative-energy electrons.) When a photon (such as a gamma ray) hits this "ocean", it excites a negative-energy electron to become a positive-energy electron, which can move freely. This creates a hole in the negative-energy state. This hole would behave like an anti-particle of the electron (i.e., positron). In other words, the positron is just a hole in the Dirac's ocean of the negative-energy electron. Particle physicists of the past have relied on this theory to explain why photons can create a pair of electron and positron in the vacuum. The same theory can be used to explain how new particles are created in the collider experiment.

This theory, however, has certain problems. First, the vacuum in this theory is very complicated; the vacuum must be filled with an infinite number of negative-energy electrons. These negative-energy electrons cannot be directly observed.

Second, there are not only electrons in our world, but also many other elementary particles. If one uses Dirac's negative-energy electron ocean theory to explain the origin of all these particles, then the vacuum must be filled with an infinite number

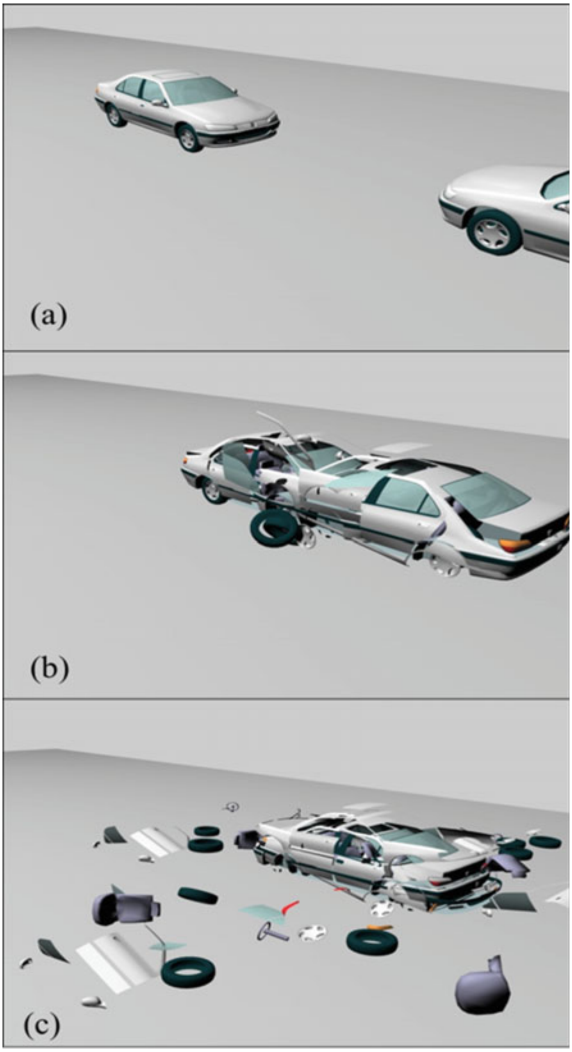


Fig. 4.6 Metaphor of the principle of the collider experiment. The analogy in the design of the particle collider experiment is similar to that of colliding two cars to learn what parts a car is made of

of all kinds of negative-energy elementary particles, including all kinds of fermions. The vacuum becomes a very crowded place. Furthermore, the Dirac theory cannot explain why bosons have anti-particles; the bosons do not obey Pauli's exclusion principle, and thus, their energy levels cannot be fully occupied.

To avoid these problems, particle physicists later developed the quantum field theory, which provides a new explanation for the creation of particles in the vacuum. They propose that there are countless "virtual particles" in the vacuum in a very short time scale [29]. Usually, you can't see or touch these virtual particles, but when the vacuum is excited, some of these virtual particles could transform into real particles

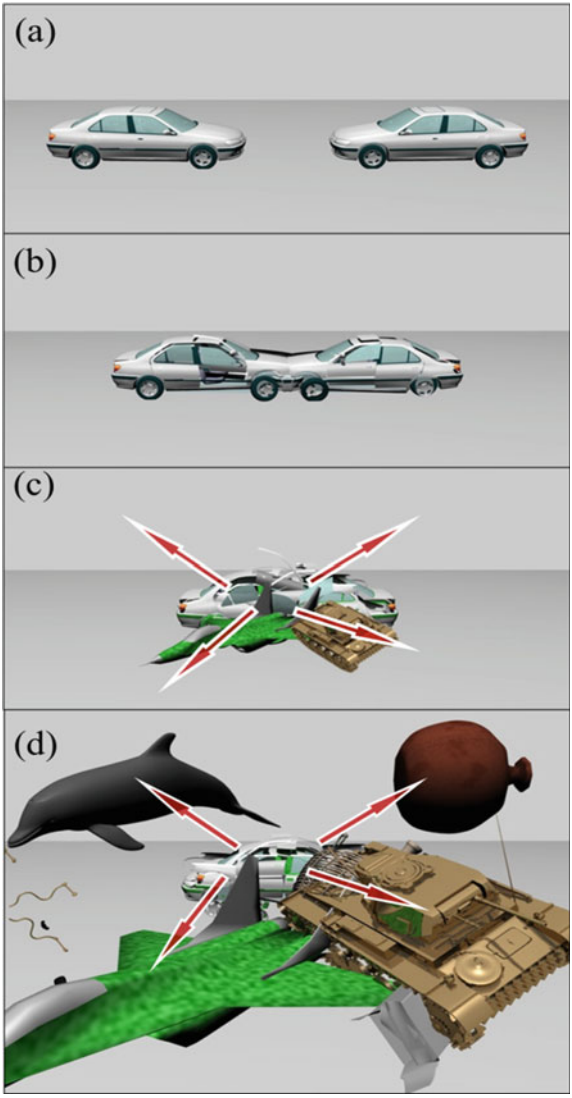


Fig. 4.7 Puzzling result of the collider experiment. Surprising results of the collider experiment. In the particle collision experiment, when the colliding particles are accelerated to very high energy, the collision will produce a variety of new species of particles whose masses are much larger than the original particles. This is like when two cars collide against each other, objects that appear after the collision are not car parts, but new objects much heavier than the original car. It is impossible to explain such a finding using the conventional particle view

and appear in space. However, it remains unexplained why nature should be filled with unlimited numbers of various kinds of virtual particles.

By comparison, the conceptual picture of the wave model is far simpler. If we regard all particles (including particles and anti-particles of different mass) as different excitation waves of the vacuum, there is no need to assume the pre-existence of an infinite number of particles occupying the negative-energy states (or virtual particles) [30]. Furthermore, from the wave model, one can see that different species of particles are just different excitation modes of the same vacuum. This can simply explain why particles can be converted from one species to another.

4.5 The Idea of Solitons

In the quantum wave model discussed above, the quantum particle is regarded as an excitation wave of the vacuum medium; it behaves like a “particle” only at a macroscopic view. The idea that a wave can behave like a particle is not totally new. For example, in the field of particle physics, many people are engaged in the research of solitons. Their aim is to develop various physical or mathematical models that could generate some specific “waves” that could mimic traveling “particles”.

The idea of solitons was inspired by the observations that a wave sometimes could behave as a non-dissipating traveling object. For example, in a certain river or water channel, one can observe some of the tidal waves that could travel in a long distance without being dispersed. Similarly, in some of the topological mechanical systems, one can also observe traveling waves that are commonly referred to as “solitons”.

The soliton phenomenon was first described in 1834 by a Scottish engineer, John Russell, who observed a solitary wave in the Union Canal in Scotland [31]. He later reproduced a similar phenomenon in a wave tank. In modern days, the soliton phenomenon has inspired many physicists to come up with specific mathematical or physical models to mimic a particle with a wave packet. The crucial step is to propose a specific form of nonlinear wave equation that can counter-act the dispersion effect of the wave medium, so that the shape of the wave packet could be maintained when the wave packet travels a long distance.

In most cases, solitons are the solutions of a widespread class of weakly nonlinear dispersive partial differential equations describing a chosen physical system. In many physical models, one could develop a wave equation that has soliton solutions. Some known cases include the Korteweg-De Vries equation, the nonlinear Schrödinger equation, and the sine–Gordon equation [32–35].

Another way to maintain the soliton stability is to impose certain topological constraints, such that the solution of the wave equation is limited in such a form that the wave packet cannot be widely dispersed. In this class of solitons, the constraints arise mainly because the solution of the wave equations must obey a set of specific boundary conditions.

We believe the research on solitons and the quantum wave model discussed here could share the same philosophical approach. Both of their works are based on the thinking that the physical nature of the quantum particle is a wave. The challenge now is to find out what are the physical properties of the wave medium and what are the mechanisms for making the excitation wave in a non-disperse state.

4.6 Chapter Summary

- In the early 1920s, it was discovered that a massive particle (e.g., an electron) can behave like a wave. This discovery was mainly due to a bold hypothesis put forward by de Broglie. He proposed that not only waves can have particle properties. Particles with mass might also have wave properties.
- This hypothesis was soon confirmed in experiments. Davisson and Germer reported that the electron diffraction results were similar to that obtained by diffraction using X-rays. Their findings suggested that electrons have wave properties as proposed by de Broglie.
- The double-slit experiment also demonstrated that the electron behaves like a wave. The most amazing thing about this experiment is that the generation of the interference fringes does not come from interference between different electrons, but an electron itself can interfere with itself.
- How to explain *wave-particle duality*? In 1927, a group of physicists working at Copenhagen University proposed an explanation. They suggested that the electron itself is a point-mass-like particle, but its distribution could appear in the form of a wave. Therefore, the so-called "*matter wave*" is not a real wave, but a "probability wave"; it mainly reflects the probability of the particle appearing at a certain time and location.
- Many physicists did not agree with the statistical interpretation of the Copenhagen school. A large amount of experimental evidence had indicated that the electron has wave properties (including the electron diffraction experiment and the double-slit experiment). Besides that, according to the Bohr-Sommerfeld quantization condition, the length of the electron orbit within an atom must equal an integral number of electron wavelength, suggesting that the electron inside an atom must behave like a resonating wave.
- Another strong reason to believe that the electron is a physical wave is the fact that an electron can be created or annihilated in the vacuum. It is very difficult to explain these observations if one regards the electron as a massive point object. On the other hand, such observations can be explained very easily if one believes that the electron is an excitation wave of the vacuum.

References

1. Planck, M. (1900). Zur theorie des gesetzes der energieverteilung im normal spectrum. *Verhandl. Dtsch. Phys. Ges., 2*, 237–245.
2. Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik, 17*, 132–147.
3. Compton, A. H. (1923). A quantum theory of the scattering of X-rays by light elements. *Physical Review, 21*, 483–502.
4. de Broglie, L. (1925). Recherches sur la théorie des Quanta. *Annalen der Physik, 10*, 22–128.

5. This Month in Physics History. (2010). October 18, 1933: Louis de Broglie elected to academy. *APS News, 19*
6. Isaacson, W. (2007). *Einstein: His Life and Universe*. Simon & Schuster.
7. Gehrenbeck, R. K. (1978). Electron diffraction: Fifty years ago. *Physics Today, 31*, 34–41.
8. Davisson, C. J., & Germer, L. H. (1927). The scattering of electrons by a single crystal of nickel. *Nature, 119*, 558–560.
9. Thomson, G. P., & Reid, A. (1927). Diffraction of cathode rays by a thin film. *Nature, 119*, 890.
10. Halban, H. J., & Preiswerk, P. (1936). Preuve expérimentale de la diffraction des neutrons," *CR Académie des Sciences, 203*, 73–75
11. Estermann, I., & Stern, O. (1930). Beugung von Molekularstrahlen. *Zeitschrift für Physik, 61*, 95–125.
12. Carnal, O., & Mlynek, J. (1991). Young's double-slit experiment with atoms: A simple atom interferometer. *Physical Review Letters, 66*, 2689–2692.
13. Schöllkopf, W., & Toennies, J. P. (1994). Nondestructive mass selection of small van der Waals Clusters. *Science, 266*, 1345–1348.
14. Bordé, Ch. J., Courtier, N., du Burck, F., Goncharov, A. N., & Gorlicki, M. (1994). Molecular interferometry experiments. *Physics Letters A, 188*, 187–197.
15. Arndt, M., Nairz, O., Vos-Andreae, J., Keller, C., Van der Zouw, G., & Zeilinger, A. (1999). Wave–particle duality of C60 molecules. *Nature, 401*, 680–682.
16. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Quantum behaviour. In *The Feynman Lectures on Physics* (Vol. 1). Addison-Wesley.
17. Bach, R., Pope, D., Liou, S.-H., & Batelaan, H. (2013). Controlled double-slit electron diffraction. *New Journal of Physics, 15*, 033018.
18. Schrödinger, E. (1926). An Undulatory theory of the mechanics of atoms and molecules. *Physical Review, 28*, 1049–1070.
19. Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. *Nature, 121*, 580–590.
20. Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Physik, 43*, 172–198.
21. Messiah, A. (1965). *Quantum Mechanics* (Vol. 1). Wiley.
22. Einstein, A., & Born, M. (1971). *The Born-Einstein Letters*. Walker and Company.
23. Knoll, M., & Ruska, E. (1932). Das Elektronenmikroskop. *Zeitschrift für Physik, 78*, 318–339.
24. von Ardenne, M. (1938). Das Elektronen-Rastermikroskop: Theoretische Grundlagen. *Z. Physik, 109*, 553–572.
25. Egerton, R. F. (2016). *Physical Principles of Electron Microscopy* (2nd ed.). Springer.
26. Veltman, M. (2003). *Facts and Mysteries in Elementary Particle Physics*. World Scientific Pub.
27. Dirac, P. A. M. (1928). The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences, 117*, 610.
28. Dirac, P. A. M. (1934). Discussion of the infinite distribution of electrons in the theory of the positron. *Mathematical Proceedings of the Cambridge Philosophical Society, 30*, 150–163.
29. Hey, T., & Walters, P. (2003). *The New Quantum Universe*. Cambridge University Press.
30. Chang, D. C. (2021). Review on the physical basis of wave–particle duality: Conceptual connection between quantum mechanics and the Maxwell theory. *Modern Physics Letters B, 35*, 2130004.
31. Russell, S. (1845). *Report on Waves: Made to the Meetings of the British Association in 1842–43*. Richard and John E. Taylor.
32. Korteweg, D. J., & de Vries, G. (1895). XLI. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 39*, 422–443.
33. Chiao, R. Y., Garmire, E., & Townes, C. H. (1964). Self-trapping of optical beams. *Physical Review Letters, 13*, 479–482.

34. Coleman, S. (1975). Quantum sine-Gordon equation as the massive Thirring model. *Phys. Rev. D, 11*, 2088–2097.
35. Bogoliubov, N. M., Korepin, V. E., & Izergin, A. G. (1985). Structure of the vacuum in the quantum sine-Gordon model. *Physics Letters B, 159*, 345–347.

Part II
Wave Excitation in the Vacuum: What are the Physical Properties of Matter Wave?

Chapter 5
The Mechanism of Wave Excitation and the Physical Nature of the Vacuum Medium

As stated above, we believe that a quantum particle (such as an electron) is an excitation wave in the microscopic view; it behaves like a point object only in the macroscopic view. We will show in the following that the quantum wave equation for the electron can be naturally derived if we treat the electron as an excitation wave of the vacuum. Our approach is to regard quantum mechanics as a natural extension of classical physics. We will first review the physical basis of wave excitation in a mechanical system (such as a flexible string or an elastic solid). We will show that the equation of motion for the excitation wave is entirely dependent on the physical properties of the medium. Then, we will analyze the wave mechanism in the vacuum system. The key questions are: What are the physical properties of the vacuum medium? Can we derive the equation of motion based on the vacuum's physical properties? Finally, can this approach lead us to derive the correct quantum wave equations for the matter wave?

5.1 Useful Analogy: Wave Propagation in a Classical Mechanical System

To study the physical meaning of the wave function, let us first review what happens in a classical mechanical system. Here, the basic requirement for generating a wave motion involves two types of forces: (a) ***Inertial force***, which is related to the change of momentum and thus the kinetic energy; (b) ***Restoring force***, which is related to the change of the potential energy. The inertial force is basically described by *Newton's second Law*: $F = ma$. The restoring force is generally described by *Hooke's Law*: $F = -\kappa x$. It is the interaction between these two forces that generates an oscillation in a mechanical system. In the following, we will briefly review three different cases as examples. From these examples, one can see clearly the physical meaning of the wave function.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_5

5.1.1 Wave in a Harmonic Oscillator

The simplest example of wave generation in a mechanical system is the one-dimensional harmonic oscillator. One can easily set up the equation of motion by equalizing the inertial force with the restoring force. Using Newton's Law, we know

$$F = ma = m\frac{d^2x}{dt^2}. \tag{5.1}$$

Using Hooke's Law, we have

$$F = -\kappa x. \tag{5.2}$$

Combining Eqs. (5.1) and (5.2), we have the equation of motion for the harmonic oscillator,

$$m\frac{d^2x}{dt^2} = -\kappa x. \tag{5.3}$$

The most general solution for Eq. (5.3) is

$$x = x_0 e^{i\omega t}, \tag{5.4}$$

where $\omega = \sqrt{\kappa/m}$. In this case, the wave function apparently represents the displacement in the harmonic oscillator.

5.1.2 Wave Propagation in a One-Dimensional String

In the above example, the solution of the wave equation is not a moving wave. Thus, it is not a proper example to demonstrate the generation of a propagating wave. The simplest example to demonstrate wave propagation in a mechanical system is a one-dimensional stretched string.

This string can be modeled as a string of beads, in which the mass of one segment of the string (Δz) is lumped together to become a bead. (See Fig. 5.1a). Each pair of neighboring beads is then connected by a massless string. The beads can undergo harmonic oscillation. The wave propagating along the string is generated by coupling the harmonic oscillation of neighboring beads.

Using such a simplified model, one can easily describe the wave propagating mechanism. First, the inertial motion of each bead is governed by the Newton's law. Second, the restoring force is governed by the Hooke's law, which is applied to the string connecting two neighboring beads. These two forces interact with each other to allow the wave to travel along the string.

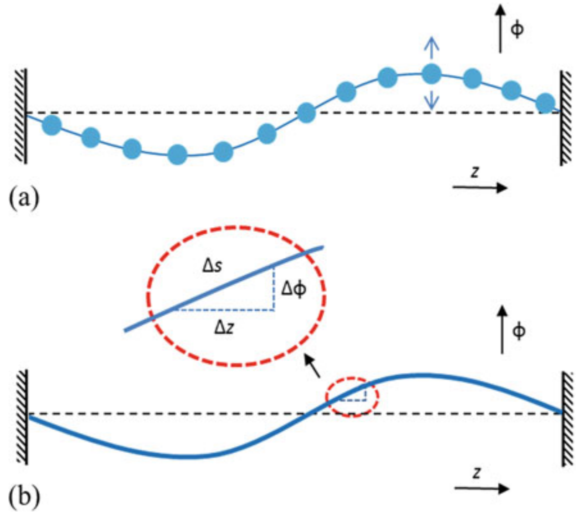


Fig. 5.1 Wave propagation in a 1-D continuum system (a stretched string). **a** The wave propagation on a string can be modelled as coupled harmonic oscillations of a string of beads. **b** In the more refined model, each "bead" becomes a tiny segment of the string. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, arXiv preprint physics/0505010v2 (2017)

In this system, it is more convenient to derive the equation of motion using the Lagrangian formulation. (See **Appendix B**). Instead of describing the system using the (position and velocity) coordinates of each bead, one can describe the motion of the string using a "basic field" concept.

From Newton's Law, we know the kinetic energy of a small segment of the string (length of Δz) is $\Delta T = \frac{1}{2}(\rho\,\Delta z)v^2$, where ρ is the mass density of the string. Let us denote the vertical displacement of the string as ϕ, which is the *basic field* of this system; the kinetic energy is

$$\Delta T = \frac{1}{2}(\rho \Delta z)\left(\frac{\partial \phi}{\partial t}\right)^2. \tag{5.5}$$

Using Hooke's Law, one can easily calculate the potential energy of the same segment. As shown in Fig. 5.1b, the length of the string over Δz is stretched to Δs during the wave motion. The potential energy for this segment is

$$\Delta V = F_1(\Delta s - \Delta z)$$

where F_1 is the tension of the stretched string between the two end points. From the inset in Fig. 5.1b, we can see

$$\Delta s = \sqrt{\Delta z^2 + \Delta\phi^2} = \Delta z\left[1 + \frac{\Delta\phi^2}{\Delta z^2}\right]^{\frac{1}{2}} = \Delta z\left[1 + \frac{1}{2}\left(\frac{\Delta\phi}{\Delta z}\right)^2 + \cdots\right].$$

When $\Delta\phi$ is small, we can ignore the higher order terms. So,

$$\Delta V = \frac{F_1}{2}\left(\frac{\partial \phi}{\partial z}\right)^2 \Delta z. \tag{5.6}$$

From Eqs. (5.5) and (5.6), we can obtain the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\rho\left(\frac{\partial \phi}{\partial t}\right)^2 - \frac{1}{2}F_1\left(\frac{\partial \phi}{\partial z}\right)^2. \tag{5.7}$$

By applying the Euler–Lagrange equation, we can easily obtain the equation of motion

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{1}{c_1^2}\frac{\partial^2 \phi}{\partial t^2} = 0 \tag{5.8}$$

where $c_1 = \sqrt{F_1/\rho}$. The general solution for Eq. (5.8) is

$$\phi = \phi_0\, e^{i(z - c_1 t)}. \tag{5.9}$$

In this case, the wave function apparently represents the transverse displacement of the string.

5.2 Wave Propagation in a 3-Dimensional Elastic Solid

In the above example, the wave function is a direct representation of the movement of the wave medium. In a more complex mechanical system, however, it is sometimes more convenient to use the *potential functions* (which indirectly describes the movement of the wave medium) as the wave function. Such a process can be accomplished with the use of the Helmholtz decomposition theorem (see **Appendix C**).

As an example, let us consider the mechanism of wave propagation in a 3-dimensional elastic solid. Here, we will denote the *displacement* and *velocity* fields of a *differential solid element* (ΔV) as r_{i} and u_i ($i = 1, 2, 3$). From Newton's Law, the time derivative of the momentum (density $\rho \times$ velocity vector u_i) is equal to the surface/traction force T_i and the body force f_i applied [1, 2],

$$\frac{\mathrm{d}}{\mathrm{d}t}(\rho\Delta V)u_i = \rho f_i\, \Delta V + T_i\, \Delta S. \tag{5.10}$$

In this system, the generalized Hooke's law is no longer a linear function. The traction force T_i is known to be related to the second-order stress tensor σ_{ij}, i.e., $T_i = \sigma_{ij}\, n_j$, where $\mathbf{n} = \{n_i\}$ is the normal unit vector with respect to the surface S [1]. It can be shown that [3]

$$T_i \, \Delta S = \sigma_{ij} n_j \, \Delta S = \frac{\partial \sigma_{ij}}{\partial x_j} \, \Delta V. \tag{5.11}$$

Here we have adopted the Einstein summation convention; repeated indices represent a summation over the three axes ($i, j, k = 1, 2, 3$). Combining Eqs. (5.10) and (5.11), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(\rho u_i)\Delta V = \frac{\partial \sigma_{ij}}{\partial x_j}\Delta V + \rho f_i \Delta V. \tag{5.12}$$

Recall that $\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + u_k \frac{\partial}{\partial x_k}$, Eq. (5.12) can be simplified as:

$$\rho \frac{\partial u_i}{\partial t} + \rho u_k \frac{\partial u_i}{\partial x_k} = \frac{\partial \sigma_{ij}}{\partial x_j} + \rho f_i. \tag{5.13}$$

If the deformation of the elastic body is small, we can neglect the second-order terms. Recall that the velocity vector u_i is the time derivative of the displacement r_i, we have

$$\frac{\partial \sigma_{ij}}{\partial x_j} + \rho f_i = \rho \frac{\partial u_i}{\partial t} = \rho \frac{\partial^2 r_i}{\partial t^2}. \tag{5.14}$$

If the material in the solid is linear, isotropic and the deformation is small, the strain tensor e_{ij} can be related to the average of the deformation gradient $r_{i,j}$ [4],

$$e_{ij} = \frac{1}{2}(r_{i,j} + r_{j,i}), \tag{5.15}$$

where $r_{i,j} = \frac{\partial r_i}{\partial x_j}$. In addition, the stress tensor for isotropic material is known to be related to the strain tensor according to the generalized Hook's law with only two constants: [1, 4]

$$\sigma_{ij} = \lambda e_{kk} \delta_{ij} + 2\mu e_{ij} = \lambda r_{k,k} \delta_{ij} + \mu(r_{i,j} + r_{j,i}), \tag{5.16}$$

where λ and μ are *Lamé's first parameter* and *Lamé's second parameter*. μ is also called the "*shear modulus*" [5]. Substituting Eq. (5.16) into Eq. (5.14), we can obtain the following equation (called the "*Navier equation*"), which is the fundamental equation for describing wave propagation in solid mechanics [4]:

$$(\lambda + \mu) r_{j,ji} + \mu r_{i,jj} + \rho f_i = \rho \ddot{r}_i. \tag{5.17}$$

Since there is no external force applied to the elastic solid, the body force $\boldsymbol{f}_i$ is zero here. Using the vector notation of the gradient and divergence operators, we can re-write the tensor equation of Eq. (5.17) as follows:

$$(\lambda + \mu)\nabla(\nabla \cdot \mathbf{r}) + \mu\nabla^2\mathbf{r} = \rho\frac{\partial^2\mathbf{r}}{\partial t^2}. \tag{5.18}$$

5.2.1 Application of the Helmholtz Decomposition Theorem on the Wave Motion of an Elastic Solid

Unlike the one-dimensional string, the wave motion in an elastic solid is more complicated, since **r** can take on different oscillating modes. The above equation could be simplified using the Helmholtz decomposition theorem [6], that is, the vector **r** can be decomposed into a curl-free component ϕ and a divergence-free component $\boldsymbol{\psi}$: (For details, see **Appendix C**).

$$\mathbf{r} = -\nabla\phi + \nabla \times \boldsymbol{\psi}, \tag{5.19}$$

where $\nabla \cdot \boldsymbol{\psi} = 0$. Substituting Eq. (5.19) into Eq. (5.18), we have

$$(\lambda + \mu)\nabla(\nabla \cdot [-\nabla\phi + \nabla \times \boldsymbol{\psi}]) + \mu\,\nabla^2[-\nabla\phi + \nabla \times \boldsymbol{\psi}]$$
$$= \rho\frac{\partial^2}{\partial t^2}[-\nabla\phi + \nabla \times \boldsymbol{\psi}].$$

This equation can be rearranged to become

$$\nabla\left\{(\lambda + 2\mu)\nabla^2\phi - \rho\frac{\partial^2\phi}{\partial t^2}\right\} = \nabla \times \left\{\mu\nabla^2\boldsymbol{\psi} - \rho\frac{\partial^2\boldsymbol{\psi}}{\partial t^2}\right\}. \tag{5.20}$$

Since ϕ and $\boldsymbol{\psi}$ are independent from each other, the simplest way to satisfy Eq. (5.20) is to assume that each bracketed term is equal to zero. Therefore, Eq. (5.20) implies that we can have two uncoupled wave equations:

$$\begin{cases} (\lambda + 2\mu)\nabla^2\phi - \rho\dfrac{\partial^2\phi}{\partial t^2} = 0 & (5.21) \\[2ex] \mu\nabla^2\boldsymbol{\psi} - \rho\dfrac{\partial^2\boldsymbol{\psi}}{\partial t^2} = 0. & (5.22) \end{cases}$$

Re-arranging the coefficients of Eqs. (5.21) and (5.22), we have

$$\begin{cases} \nabla^2\phi - \dfrac{1}{c_p^2}\dfrac{\partial^2\phi}{\partial t^2} = 0 & (5.23) \\ \\ \nabla^2\boldsymbol{\psi} - \dfrac{1}{c_s^2}\dfrac{\partial^2\boldsymbol{\psi}}{\partial t^2} = 0, & (5.24) \end{cases}$$

where $c_p = \sqrt{(\lambda + 2\mu)/\rho}$ is the velocity of the dilational wave (also called "primary wave") and $c_s = \sqrt{\mu/\rho}$ is the velocity of the transverse wave (also called "distortional wave/secondary wave/shear wave").

These wave equations have a similar form as the vibrating string as shown in Eq. (5.8). The meanings of the wave functions, however, are slightly different. In the case of the 1-D string, the wave function represents a transverse displacement of the string. But in the case of an elastic solid, the wave functions ϕ and $\boldsymbol{\psi}$ do not directly represent the displacement of the solid element. Instead, they represent potential functions, the derivatives of which are related to different modes of displacement according to the Helmholtz decomposition.

5.3 Mechanism of Wave Excitation in the Vacuum Medium

Now, let us examine the mechanism of wave propagation in the vacuum system. In the above, we reviewed the wave propagation mechanisms in the mechanical system. It is clear that the mechanical waves are clearly physical waves, i.e., their wave function represents a physical movement of its wave medium. In quantum mechanics, particles with mass are known to exhibit some wave properties. The wave associated with an electron is called "matter wave" [7]. It has been debated on whether this matter wave is a physical wave or not. In the early parts of this book, we had presented experimental evidence indicating that the electron is an excitation wave of the vacuum; so, the matter wave should be a physical wave.

From the above section, it is clear that the equation of motion for the excitation wave in a mechanical system is entirely determined by the physical properties of the medium. Now, if the matter wave of an electron is an excitation wave of the vacuum system, what are the physical properties of the vacuum medium? Can we derive the proper equation of motion for the matter wave based on the vacuum's physical properties?

5.3.1 How does Wave Propagate in the Vacuum?

Based on the above discussion, it appears that one might model the vacuum as an elastic solid. Such attempts had been made previously and some interesting results were reported [3, 8–10]. We, however, decided to take a different approach. This decision was based on two considerations. First, the hint from the photon. Second, the lack of proper mechanical properties in the vacuum. Unlike the elastic solid, Newton's law and Hooke's law cannot be applied in the vacuum medium. Since the vacuum has no rest mass, there is no inertial force. Furthermore, there is no restoring force because there is no mechanical coupling between the neighboring volume elements of the vacuum. Thus, it is extremely difficult to set up a wave equation like the Navier equation in the vacuum.

Then, the vacuum system must use a different mechanism to generate a propagating wave. To find out such a mechanism, we followed the hint of a photon. It is well known that electromagnetic wave can propagate in the vacuum. By studying how the electromagnetic wave is driven, we will be able to gain some insight about the wave propagation mechanism in the vacuum system.

For the electro-magnetic wave, the wave regenerating mechanism is entirely based on the Maxwell's equations, more specifically, the Ampere's law and the Faraday's law.

$$\begin{cases} \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} \quad \text{Faraday's Law} & (5.25) \\[2ex] \nabla \times \mathbf{H} = \mathbf{J} + \dfrac{\partial \mathbf{D}}{\partial t} \quad \text{Ampère's Law.} & (5.26) \end{cases}$$

In the above, **E** is the electric field and **H** is the magnetic field; **B** and **D** are magnetic induction and electric displacement, respectively; and

$$\begin{cases} \mathbf{B} = \mu \mathbf{H} & (5.27) \\ \mathbf{D} = \varepsilon \mathbf{E}, & (5.28) \end{cases}$$

where ε and μ are the dielectric permittivity and magnetic permeability. In the vacuum, the external current $\mathbf{J} = 0$, and $\rho_e = 0$; also $\mu = \mu_0, \varepsilon = \varepsilon_0$. We can re-write the first two Maxwell's equations as

$$\begin{cases} \nabla \times \mathbf{E} = -\mu_0 \dfrac{\partial \mathbf{H}}{\partial t} & (5.25a) \\[2ex] \nabla \times \mathbf{H} = \varepsilon_0 \dfrac{\partial \mathbf{E}}{\partial t}. & (5.26a) \end{cases}$$

This shows that **E** and **H** can cross interact with each other in the vacuum. When **E** changes with time, it generate a change in curl **H** in the vicinity; and when **H** changes with time, it would induce **E** around it to change. This cross interaction allows the electro-magnetic wave to propagate in the vacuum.

Hence, it seems that there is a close analogy between the wave-generating mechanism in the mechanical system and that in the electro-magnetic system. In the mechanical system, it is the interactions between the general coordinate (q) and the

Table 5.1 Corresponding relations in mechanical system and electro-magnetic system. Wave propagation depends on a cross-interaction between two physical parameters

Mechanical system (Newton's Law & Hooke's Law)	**Electro-magnetic system** (Maxwell's equations: Faraday's law and Ampere's law)
$\frac{\mathrm{d}p}{\mathrm{d}t} = -kq$	$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu}\nabla \times \mathbf{E}$
$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{1}{m}p$	$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon}\nabla \times \mathbf{H}$

general momentum (p) as described in Newton's Law and Hooke's Law that generates an oscillating wave. The electro-magnetic system, on the other hand, relies on the cross interactions between **E** and **H** to generate a propagating wave. This comparison can be seen more clearly in Table 5.1.

Once we know the foundation of the wave generating mechanism, it is not difficult to set up the wave equation. From Eqs. (5.25a) and (5.26a), one can easily derive the wave equation for an electro-magnetic wave, i.e.,

$$\nabla^2 \mathbf{E} - \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \tag{5.29}$$

or

$$\nabla^2 \mathbf{H} - \frac{1}{c^2}\frac{\partial^2 \mathbf{H}}{\partial t^2} = 0,$$

where $c = 1\big/\sqrt{\mu_0 \varepsilon_0}$ is the speed of light.

5.4 What is the Physical Nature of the Vacuum? The *Aether* Hypothesis

From the above discussions, we have some ideas about the wave generating mechanism in the vacuum for the electromagnetic wave. However, it is not clear whether the same mechanism works for the matter wave. First, we have not yet identified the basic field for the matter wave; we do not think the electric or magnetic fields are the proper basic field. Second, we know very little about the physical nature of the vacuum medium. Unlike those cases in the mechanical systems (such as the stretched string or elastic solid), it is not possible to associate any wave function with the local movement of the wave medium in the vacuum system.

The physical nature of the vacuum is a controversial topic in the study of physics. Historically, the perception of vacuum changed with time. In classical mechanics, the vacuum is regarded as an empty space, with nothing inside it. This view is based on the fact that a mechanical object free of applied force can move in straight line at a constant speed. There is nothing in the vacuum that can impede such a motion. But at later time, with the discovery of the electro-magnetic field, many physicists started to

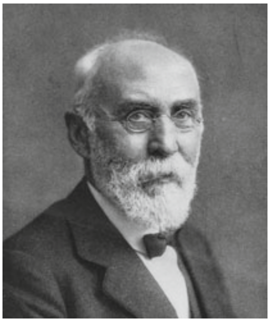


Fig. 5.2 **Hendrik Lorentz** (1853–1928) was a famous theoretical physicist in the Netherlands. He was a major contributor to the aether hypothesis. He derived the "Lorentz transformation" to connect the space–time coordinates in two inertial frames. He was awarded the Nobel Prize in Physics in 1902 for his work explaining magnetic effects in atomic spectroscopy. Photo Credit: AIP Emilio Segrè Visual Archives, Lande Collection

assume that there must be a medium that carries the electro-magnetic radiation. This medium is called "aether", which is supposed to occupy all space between matter [11]. The basic thinking was that, in order for any wave to propagate, a medium is required. Since it had been demonstrated that light is a wave, there must be a carrying medium for it.

The *Aether hypothesis* was a widely accepted physical theory during the eighteenth and nineteenth century. (For details, see **Appendix A**). Many well-known physicists and mathematicians, including Faraday, Helmholtz, Maxwell, Stokes, Cauchy, Poisson, Gauss, Riemann, and Lorentz (see Fig. 5.2), had made contributions to such a theory [11].

The idea of aether was clearly summarized by Maxwell at the end of his famous book: *A Treatise on Electricity and Magnetism, Vol. 1*:

> We have seen that the mathematical expressions for electrodynamic action led, in the mind of Gauss, to the conviction that a theory of the propagation of electric action in time would be found to be the very key-stone of electrodynamics. Now we are unable to conceive of propagation in time, except either as the flight of a material substance through space, or as the **propagation of a condition of motion or stress in a medium already existing in space**….
>
> But in all of these theories the question naturally occurs:—If something is transmitted from one particle to another at a distance, what is its condition after it has left the one particle and before it has reached the other? If this something is the potential energy of the two particles, as in Neumann's theory, how are we to conceive this energy as existing in a point of space, coinciding neither with the one particle nor with the other? In fact, whenever energy is transmitted from one body to another in time, **there must be a medium or substance** in which the energy exists after it leaves one body and before it reaches the other, for energy, as Torricelli remarked, 'is a quintessence of so subtile a nature that it cannot be contained in any vessel except the inmost substance of material things'. **Hence all these theories lead to the conception of a medium in which the propagation takes place**, and if we admit this medium as an hypothesis, I think it ought to occupy a prominent place in our investigations. [12]

The aether hypothesis, however, was later disfavored. That was because it had many serious problems. The most severe one was its inconsistency with experimental observations. In late nineteenth century, many scientists attempted to use optical interferometers to detect the movement between aether and the Earth. The most famous one was the Michelson-Morley experiment [13]. None of such measurements was able to detect any movement between Earth and the hypothetical aether.

Furthermore, the aether hypothesis was later thought to be unnecessary. In 1905, Einstein proposed the Special Theory of Relativity (STR) and showed that one can explain the null results easily without the assumption of aether [14]. So, this aether hypothesis was totally abandoned in the early twentieth century.

5.5 Evidence Indicating that the Vacuum is Not an Empty Space

The disfavoring of the aether hypothesis, however, does not imply that the vacuum is proven to be an empty space. In the more recent development of modern physics, including the emerging works of quantum mechanics, particle physics and cosmology, the view of regarding the vacuum as an empty space is clearly unacceptable. For example, in quantum electrodynamics, every oscillation mode of radiation is supposed to have a zero-point energy [7]. Such energy is assumed to be a part of the vacuum system. In fact, in the quantum field theory, the vacuum is just regarded as the ground state of the quantum system [15]. An empty vacuum is not consistent with the current theories of quantum physics. "*The quantum theory asserts that a vacuum, even the most perfect vacuum devoid of any matter, is not really empty. Rather, the quantum vacuum can be depicted as a sea of continuously appearing and disappearing [pairs of] particles that manifest themselves in the apparent jostling of particles that is quite distinct from their thermal motions. These particles are 'virtual', as opposed to real, particles. …At any given instant, the vacuum is full of such virtual pairs*". [16]

In the Standard Model of cosmology today, the vacuum has even more complicated features; the presence of our entire universe is supposed to come from the quantum fluctuations of the vacuum [17].

Finally, the idea of a non-empty vacuum is well supported in experiments. For example, the interaction between the vacuum and the electron was demonstrated in the famous experiment of Lamb shift [18]. In the Dirac theory, the energy levels of $^2S_{1/2}$ and $^2P_{1/2}$ of a hydrogen atom are degenerate [19]. There should be no energy shift between them. But if the vacuum is not empty, the vacuum energy fluctuations can interact with the electrons in different orbitals. It would cause a very small energy shift. Such a shift was detected by Lamb and Retherford in 1947 [18]. In fact, the experimental value of this "Lamb shift" agreed well with the calculation of H. A. Bethe, who applied Kramers' idea of mass renormalization to account for the interactions between a free electron and the radiation field [19]. The Lamb shift has since played a significant role in demonstrating the importance of including vacuum energy fluctuations in theoretical calculations.

In addition to the Lamb shift, non-empty properties of the vacuum were also demonstrated in several other types of experiments, including spontaneous emission, vacuum polarization and the Casimir effect [20–22]. Today, we can no longer treat the vacuum as emptiness, although its physical properties are still not well understood or agreed upon. In fact, understanding the physical properties of the vacuum is a very important topic in modern physics today; it is currently under active investigation [3, 23–26].

In this work, we will simply regard the vacuum as a wave medium, the excitation wave of which will appear as elementary particles. This vacuum medium can be regarded as a modernized quantum model of the classical aether hypothesis. The major differences between this vacuum medium and the original aether concept are that,

(1) Not only radiation wave (photons) is the excitation of the vacuum, matter waves (quantum waves representing massive particles) are also excitation waves of the same vacuum.
(2) The excitation wave of the vacuum medium is quantized, so that the emission and absorption of the excitation wave obeys the *principle of all-or-none*.
(3) The excitation wave representing a massive particle satisfies the same quantum wave relations of a photon, including the Planck's relation, the de Broglie relation, and Heisenberg's Uncertainty principle.

The physical properties of this *vacuum medium* will be discussed in detail in the next Chapter.

5.6 Chapter Summary

- We propose that a quantum particle (such as an electron) is a quantized excitation wave of the vacuum. To understand the wave excitation mechanism, we start by reviewing the mechanism of wave excitation in a classical mechanical system (such as a flexible string or an elastic solid). It is evident that the equation of motion for the excitation wave is entirely determined by the physical properties of the medium.
- **To generate a wave motion in a mechanical system, two types of forces are needed**: (a) *Inertial force*, which is related to the change of momentum and thus the kinetic energy; (b) *Restoring force*, which is related to the change of the potential energy. The inertial force is basically described by *Newton's second Law*. The restoring force is described by a generalized *Hooke's Law*. Interaction between these two forces can generate an oscillation in a mechanical system.
- In the case of the 1-D string, the wave function represents a transverse displacement of the string. But in the case of a 3-D elastic solid, the wave functions ϕ and $\boldsymbol{\psi}$ do not directly represent the displacement of the solid element. Instead, **they**

represent potential functions, the derivatives of which are related to different modes of displacement according to the *Helmholtz decomposition theorem.*

- Unlike mechanical systems, **the vacuum system uses a different mechanism to generate a propagating wave**. In the mechanical system, it is the connection between the general coordinate (q) and the general momentum (p) as described in Newton's Law and Hooke's Law that generates an oscillating wave. The vacuum system, on the other hand, relies on the cross interactions between the electric field **E** and magnetic field **H** to generate a propagating electro-magnetic wave.
- The physical nature of the vacuum is a controversial topic in physics. In classical mechanics, the vacuum was regarded as an empty space. But with the discovery of the electro-magnetic field, physicists realized that there must be a medium for carrying the electro-magnetic radiation. This medium was called "**aether**", which is supposed to occupy all space between matters [11]. The *Aether hypothesis* was a widely accepted physical theory during the eighteenth and nineteenth century. (For details, see **Appendix A**). Many well-known physicists and mathematicians, including Maxwell, Faraday, Lorentz, Gauss, and Riemann, had made contributions to such a theory.
- In the fundamental studies of modern physics, including quantum field theory, particle physics, and cosmology, the vacuum is not regarded as an empty space. In fact, **the idea of a non-empty vacuum is well supported in experiments**. For example, the interaction between vacuum and electron was clearly demonstrated in the famous experiment of Lamb shift [18]. In addition, non-empty properties of vacuum were also demonstrated in several other types of experiments, including spontaneous emission, vacuum polarization, and the Casimir effect.

References

1. Reismann, H., & Pawlik, P. S. (1980). *Elasticity, Theory and Applications.* Wiley.
2. Graff, K. F. (1991. *Wave Motion in Elastic Solids.* Clarendon Press.
3. Chang, D. C., & Lee, Y.-K. (2015). Study on the physical basis of wave-particle duality: Modelling the vacuum as a continuous mechanical medium. *Journal of Modern Physics, 6*, 1058–1070.
4. Fung, Y. C. (1977). *A First Course in Continuum Mechanics.* Prentice-Hall.
5. Salencon, J. (2001). *Handbook of Continuum Mechanics: General Concepts, Thermoelasticity.* Springer Science & Business Media.
6. Arfken, G. B., & Weber, H. J. (1995). *Mathematical Methods for Physicists* (2nd ed.). Cambridge University Press.
7. Messiah, A. (1965). *Quantum Mechanics* (Vol. 1). Wiley.
8. Danielewski, M., Sapa, L., & Roth, C. (2023). Quaternion quantum mechanics II: Resolving the problems of gravity and imaginary numbers. *Symmetry, 15*, 1672.
9. Close, R. A. (2011). Exact description of rotational waves in an elastic solid. *Advances in Applied Clifford Algebras, 21*, 273–281.
10. Schmelzer, I. (2012). A generalization of the Lorentz ether to gravity with general-relativistic limit. *Advances in Applied Clifford Algebras, 22*, 203–242.
11. Whittaker, E. (1951). *A History of the Theories of Aether and Electricity*. Thomas Nelson and Sons Ltd.

12. Maxwell, J. C. (1873). *A Treatise on Electricity and Magnetism* (1st ed., Vol. 1). Oxford University Press.
13. Michelson, A. A., & Morley, E. W. (1887). On the relative motion of the Earth and the luminiferous ether. *American Journal of Science, 34*, 333–345.
14. Einstein, A. (1905). Zur elektrodynamik bewegter körper. *Annalen der Physik, 322*, 891–921.
15. Ryder, L. H. (1996). *Quantum Field Theory*. Cambridge University Press.
16. Silk, J. (2005). *On the Shores of the Unknown: A Short History of the Universe*. Cambridge University Press.
17. Guth, A. H., & Kaiser, D. I. (2005). Inflationary cosmology: Exploring the universe from the smallest to the largest scales. *Science, 307*, 884–890.
18. Lamb, W. E., & Retherford, R. C. (1947). Fine structure of the hydrogen atom by a microwave method. *Physical Review, 72*, 241–243.
19. Bethe, H. A. (1947). The electromagnetic shift of energy levels. *Physical Review, 72*, 339–341.
20. Goy, P., Raimond, J. M., Gross, M., & Haroche, S. (1983). Observation of cavity-enhanced single-atom spontaneous emission. *Physical Review Letters, 50*, 1903–1906.
21. TOPAZ Collaboration, Levine, I., Koltick, D., Howell, B., Shibata, E., Fujimoto, J., Tauchi, T., Abe, K., Abe, T., Adachi, I., Adachi, K., & Aoki, M. (1997). Measurement of the electromagnetic coupling at large momentum transfer. *Physical Review Letters, 78*, 424–427.
22. D'Hoker, E., & Sikivie, P. (1993). Casimir forces between beads on strings. *Physical Review Letters, 71*, 1136–1139.
23. Moskalenko, A. S., & Ralph, T. C. (2019). Correlations detected in a quantum vacuum. *Nature, 568*, 178–179.
24. Benea-Chelmus, I.-C., Settembrini, F. F., Scalari, G., & Faist, J. (2019). Electric field correlation measurements on the electromagnetic vacuum state. *Nature, 568*, 202–206.
25. Oi, D. K. L., Potoček, V., & Jeffers, J. (2013). Nondemolition measurement of the vacuum state or its complement. *Physical Review Letters, 110*, 210504.
26. Boi, L. (2011). *The Quantum Vacuum: A Scientific and Philosophical Concept, from Electrodynamics to String Theory and the Geometry of the Microscopic World*. Johns Hopkins University Press.

Chapter 6
The Vacuum is a Dielectric Medium According to the Maxwell Theory; Its Basic Field is the Electric Vector Potential Z

At present, our knowledge about the quantum vacuum is still limited. In the nineteenth century, it was widely believed that the space between matters is filled with a medium called "*aether*". The *aether hypothesis*, however, had some serious difficulties and thus became disfavored later (For a more detailed discussion of the aether hypothesis, please see **Appendix A**). Nevertheless, the disfavoring of the aether hypothesis does not mean that the vacuum is an empty space. Many recent studies suggested that the vacuum must have specific physical properties. Our proposal of the vacuum being a wave medium is not a simple revival of the classical *aether hypothesis*. Instead, we are aiming to develop a new theory of *quantum vacuum* which will accommodate the major features of the Maxwell theory, the special theory of relativity, and the key concepts of quantum physics.

There are fundamental differences between our model of quantum vacuum and the traditional *aether hypothesis*. First, the *aether* was a hypothetical medium filling only the space between matters. Our concept of *vacuum*, on the other hand, is a pre-existing medium that fills the entire space of the universe. Second, the *aether* was assumed to be a medium for carrying the electromagnetic waves only. In this work, we propose that all particles found in nature (with or without rest mass) are excitation waves of the same *quantum vacuum*. Finally, our quantum vacuum hypothesis can avoid the known problems of the *aether hypothesis* (For details, see **Appendix A**).

6.1 Physical Nature of the Vacuum: Implications from the Maxwell Theory

A key challenge in this work is to determine the physical properties of the quantum vacuum. We discovered an important hint from the work of a great physicist, James Maxwell (see Fig. 6.1). That is, the Maxwell theory appears to have a hidden assumption; namely, *the vacuum should behave like a dielectric medium* (see below). This

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_6

Fig. 6.1 James Clerk Maxwell. James Clerk Maxwell (1831–1879) was a Scottish physicist who developed the modern theory of electromagnetism. The Maxwell theory provides the foundation of electrodynamics and has very profound influence in physics and engineering. Einstein described Maxwell's work as the "most profound and the most fruitful that physics has experienced since the time of Newton". Photo Credit: digitized from an engraving by G. J. Stodart from a photograph by Fergus of Greenock; Wikimedia Commons; Public domain

is a very important discovery which had been overlooked by most physicists today. Thus, let us have a careful review of Maxwell's electromagnetic theory in detail.

6.1.1 Implication of Maxwell's Introduction of the Electric Displacement Concept

It is well known that a major contribution of Maxwell was his introduction of the *electric displacement* **D** into electrodynamics [1–3]. His original purpose to do so was to make Ampère's Law consistent with the condition of charge conservation [4]. In the nineteenth century, the original Ampère's Law was

$$\nabla \times \mathbf{H} = \mathbf{J}, \tag{6.1}$$

where **J** is the electric current density and **H** is the magnetic field surrounding it. This equation had a problem; it is not consistent with the condition of *conservation of charge*,

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho_e}{\partial t}. \tag{6.2}$$

In order to fix this problem, Maxwell proposed to add a new term $\partial \mathbf{D} \big/ \partial t$ into the right-hand side of Eq. (6.1), where **D** is called "*electric displacement*" which is proportional to the electric field **E**, i.e.,

$$\mathbf{D} = \varepsilon \mathbf{E}. \tag{6.3}$$

His argument was that: When an external electric field is applied to a *dielectric medium*, it will cause a displacement of the *dielectric charges* [1, 4]. The time derivative of these displaced charges would produce a displacement current $(\partial \mathbf{D} / \partial t)$, which should be included into the Ampère's Law, i.e.,

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}. \tag{6.4}$$

Now, the revised Ampère's Law (Eq. 6.4) will be consistent with the condition of charge conservation (i.e., Eq. 6.2). In the literature, **D** sometimes is called "electric induction", since it mirrors the magnetic induction **B**, which is proportional to the magnetic field **H**, i.e.,

$$\mathbf{B} = \mu \mathbf{H}. \tag{6.5}$$

(In the above equations, ε and μ are the dielectric permittivity and magnetic permeability, respectively. In the vacuum, $\mu = \mu_0$, $\varepsilon = \varepsilon_0$)

This revised Ampère's Law was included in the final form of the Maxwell's equations. Later, Oliver Heaviside used vector calculus to further simplify them into four equations [5], which are

$$\begin{cases} \nabla \times \mathbf{H} = \mathbf{J} + \dfrac{\partial \mathbf{D}}{\partial t} & \text{(Ampère's Law)} \qquad (6.4) \\[2ex] \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} & \text{(Faraday's Law)} \qquad (6.6) \\ \nabla \cdot \mathbf{D} = \rho_e & \text{(Coulomb's Law or Gauss's Law)} \qquad (6.7) \\ \nabla \cdot \mathbf{B} = 0 & \text{(Gauss' Law for Magnetism)} \qquad (6.8) \end{cases}$$

6.1.2 Maxwell's Theory of Light Propagation Implied That the Vacuum is a Dielectric Medium

When Maxwell used his equations to construct the theory of light propagation, he assumed that the external current vanishes in the vacuum, (i.e., $\mathbf{J} = 0$), but the displacement current $(\partial \mathbf{D} / \partial t)$ does not vanish. This can be justified only if *the vacuum is a dielectric medium*! If the vacuum is an empty space, there is no dielectric charge in it; then **D** should automatically equal to zero.

In most standard electrodynamics textbooks used today, it is simply stated that $\mathbf{D} = \varepsilon_o \mathbf{E}$ in the vacuum, but offers no explanation why **D** should not vanish in the vacuum. This negligence is probably because the physical nature of the vacuum is not well understood and people tried to avoid discussing it.

With the hidden assumption that the vacuum is a dielectric medium, Maxwell hypothesized that there is a non-zero displacement current in the vacuum, and Ampere's Law now becomes

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}. \tag{6.9}$$

Then, one can relate the variations between **E** and **H** in the vacuum by using $\mathbf{D} = \varepsilon_o \mathbf{E}$, that is

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_o} \nabla \times \mathbf{H}. \tag{6.10}$$

Using $\mathbf{B} = \mu_0 \mathbf{H}$, Faraday's Law becomes

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu_o} \nabla \times \mathbf{E}. \tag{6.11}$$

By combining the above two equations, one can easily derive the wave equations for the electromagnetic wave in the vacuum; they are

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \tag{6.12}$$

or,

$$\nabla^2 \mathbf{H} - \frac{1}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0, \tag{6.13}$$

where $c = 1 \Big/ \sqrt{\mu_0 \varepsilon_0}$ is the speed of light.

From this review, it is clear that the Maxwell theory of light propagation required that *the vacuum must behave like a dielectric medium*. If the vacuum is an empty space, **D** must automatically equal to zero. It will then be impossible to derive the wave equation of light.

6.2 Structure of the Vacuum Medium According to Maxwell's Hypothesis

Hence, according to Maxwell theory of electromagnetism and light propagation, *the vacuum must behave like a dielectric medium*. Conceptually, such a dielectric medium could be like what is shown in Fig. 6.2. Namely, the dielectric medium is most likely to be composed of two types of constituents with opposite charges.

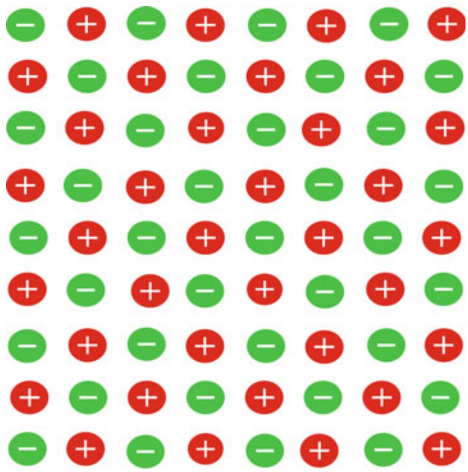


Fig. 6.2 Vacuum as dielectric medium. An important implication of the Maxwell theory of electromagnetic radiation is that the vacuum is a medium composed of dielectric charges

Here, we should point out that, the dielectric medium making up the vacuum must be very different from ordinary dielectric materials, which is composed of electrons and ions. The vacuum cannot be composed of massive particles. First, the vacuum has no rest mass and thus cannot contain any material with non-zero rest mass. Second, the vacuum must be composed of very tiny materials. If particles are excitation waves of the vacuum medium, the components making up the vacuum must be far smaller than the wavelength of any quantum particles. Hence, one can only assume that **the vacuum is filled with some sort of *primordial dielectric charges*, which are massless, highly refined, isotropic, and uniformly distributed**. (The detailed nature of the vacuum will be a very important topic for future studies of physics.) Here, we propose that the dielectric medium of the vacuum is composed of two types of components: (1) a negatively charged "*n-type medium*" (n-med) and (2) a positively charged "*p-type medium*" (p-med). Each of these medium is composed of very refined *primordial dielectric charges* (see Fig. 6.3).

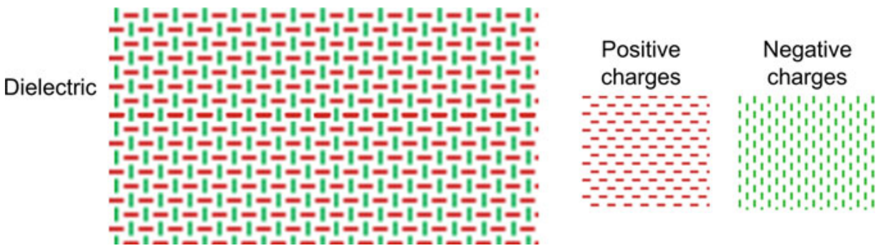


Fig. 6.3 Vacuum is a dielectric medium composed of two types of components. In a simplified model, the vacuum can be thought of as the superposition of two oppositely charged mediums, the negatively charged "n-type medium" (n-med), and the positively charged "p-type medium" (p-med)

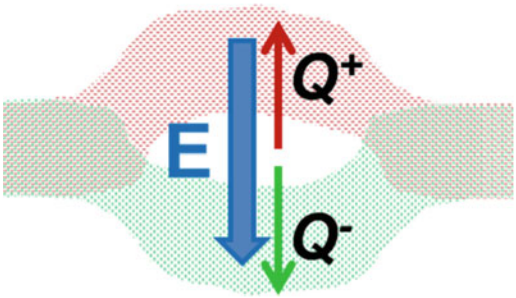


Fig. 6.4 An electric field will emerge at the location where the n-type medium and the p-type medium are separated. The vacuum medium is electrically neutral in the absence of any electric field; its positive charges and negative charges are evenly distributed. When the vacuum is excited, the charges start to separate. The distribution of the n-type medium and the p-type medium is no longer entirely overlapping. Then, an electric field will emerge at the location where the charges are separated

In the absence of any electric field, the vacuum is free of displacement charge, and the *n-type medium* and the *p-type medium* are evenly distributed; there is no net charge in the vacuum medium. But when the vacuum is under a disturbed state, such as the application of an external electric field or the passing of an electromagnetic radiation wave, the charges start to separate. The distribution of the *n-type* medium and the *p-type* medium is no longer entirely overlapping. Then, an electric field will emerge at the location where the *n-type* medium and the *p-type* medium are separated (see Fig. 6.4). The magnitude of the electric field is proportional to the "charge displacement" (i.e., $\mathbf{D} = \varepsilon\mathbf{E}$).

6.3 What is the *Basic Field* of the Vacuum Excitation Wave?

Based on the above discussions, we can formally hypothesize that: *Matter waves and radiation waves are both excitation waves of the vacuum; different types of free particles are represented by different excitation modes of this medium.* If this is the case, the wave function of a free particle must represent a local movement of the vacuum medium; such a movement can be characterized by the variation of a "*basic field*". As we had discussed in the last chapter, this field is not the same as the "*classical field*" (such as the *electric field* or the *magnetic field*). The physical meaning of our basic field is more closed to the "*quantum field*" used in the quantum field theory today (For details, please see **Appendix B**).

6.3.1 What is Its Basic Field of the Photon?

The simplest quantum particle associated with the vacuum excitation wave is the photon. From classical electrodynamics, we know the photon is an electromagnetic wave. Indeed, from Eqs. (6.12) and (6.13), one can see that the wave equation of a photon directly describes the variation of **E** or **H**. Thus, one may think that the *basic field* for a photon is either the electric or magnetic field. However, this thinking is not correct, because there are certain additional requirements for the *basic field*. In the quantum field theory, the Lagrangian density of an excitation wave is known to be composed of *the quadratic terms of the first derivatives of the basic field*, so that one can apply the Euler-Lagrange equation on the Lagrangian density to obtain the wave equation [6]. For example, for a one-dimensional string, the Lagrangian density is known to be

$$\mathcal{L} = \frac{1}{2}\rho\left(\frac{\partial\phi}{\partial t}\right)^2 - \frac{1}{2}F_1\left(\frac{\partial\phi}{\partial z}\right)^2. \tag{6.14}$$

Here, ϕ is the *basic field* which represents the vertical displacement of the string, ρ is the mass density of the string, and F_1 is the tension of the stretched string (see Fig. 6.5). Since the Lagrangian density of the electromagnetic field in the vacuum is

$$\mathcal{L} = \frac{1}{2}\left(\varepsilon_o\mathbf{E}^2 - \mu_o\mathbf{H}^2\right); \tag{6.15}$$

E or **H** does not appear to be suitable for playing the role of a *basic field*.

Then, what else can play the role of a *basic field* for the photon? According to the Maxwell theory, **E** and **H** can be derived from the scalar potential Φ and the vector potential **A**. One may guess that these potential functions might play the role of the *basic field*. It is well known that the electric and magnetic fields are derivatives of Φ and **A** [7], such that

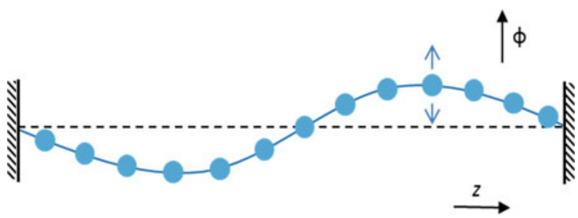


Fig. 6.5 Wave propagation in a 1-D continuum system (a stretched string). The wave propagation on a string can be modeled as coupled harmonic oscillations of a string of beads. Here, we denote the vertical displacement of the string as ϕ. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, arXiv preprint physics/0505010v2 (2017)

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{6.16}$$

$$\mathbf{E} = -\nabla\Phi - \frac{\partial \mathbf{A}}{\partial t}. \tag{6.17}$$

In the vacuum, the free charge density $\rho_e = 0$ and thus one can set $\nabla\Phi = 0$. Equation (6.17) becomes

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}. \tag{6.18}$$

Substituting Eqs. (6.15) and (6.18) into Eq. (6.10), and using the Coulomb gauge condition $\nabla \cdot \mathbf{A} = 0$, one can easily derive the wave equation

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0, \tag{6.19}$$

where $c = 1\big/\sqrt{\mu_0 \varepsilon_0}$ is the speed of light. Since the wave function of this equation is $\mathbf{A}$, it may suggest that the vector potential $\mathbf{A}$ can play the role of *basic field* for the photon. Indeed, in the standard treatment of the quantum field theory today, $\mathbf{A}$ is commonly regarded as the *quantum field* of the photon. The Lagrangian density of the vacuum associated with a photon is written as [6]

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu},$$

where

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu,$$

(Here, we use the notation of *contravariant and covariant four vectors*, μ or ν equals 0, 1, 2, 3 to represent the *four-dimensional time–space*; the *Einstein summation convention* is also applied here.)

A more careful examination of the wave excitation mechanism, however, suggests that the *vector potential* A is not the most appropriate *basic field* [8]. In the following section, we will show that, it is the *electric vector potential*, instead of the *magnetic vector potential*, which can be justified to represent the displacement of the vacuum medium during wave excitation.

6.3.2 *Origin of the Concept of Vector Potential: The Theorem of Helmholtz Decomposition*

In order to determine whether $\mathbf{A}$ is a proper measure of the vacuum displacement, let us first review the original meaning of $\mathbf{A}$. In the Maxwell theory, the vector potential $\mathbf{A}$ was defined from the magnetic flux $\mathbf{B}$ based on the relation

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{6.16}$$

According to the literature, Maxwell proposed Eq. (6.16) because he wanted to interpret Faraday's concept of "*electro-tonic state*" using the concept of *vector potential* introduced by Thomson [9]. However, from a mathematical point of view, one may recognize that Eq. (6.16) could be originated from the theorem of Helmholtz decomposition. According to this theorem, *any sufficiently smooth vector field in a three-dimensional space can be resolved into the sum of an irrotational (curl-free) vector field and a solenoidal (divergence-free) vector field*. (For details, see **Appendix C**). For example, when one studies the motion of sound waves in an elastic solid, one can use the Helmholtz decomposition to separate the longitudinal wave movement from the transverse wave movement. That is, the displacement of a volume element ($\mathbf{r}$) in the elastic solid can be decomposed as:

$$\mathbf{r} = -\nabla\phi + \nabla \times \boldsymbol{\psi}, \tag{6.20}$$

where ϕ is called the "*scalar potential*", and $\boldsymbol{\psi}$ is called the "*vector potential*". The *curl-free component* of a vector field is often referred to as the "*longitudinal component*" and the *divergence-free component* is referred to as the "*transverse component*". [10]

One may recognize that Eq. (6.16) is just an analogy of Eq. (6.20). Since $\mathbf{B}$ is a vector field, it can be decomposed based on the Helmholtz theorem. From experimental observation, we know the magnetic field is divergence-free [see Eq. (6.8)]. Thus, its *curl-free* term must be zero. One can then automatically obtain $\mathbf{B} = \nabla \times \mathbf{A}$. Here, $\mathbf{A}$ plays the role of $\boldsymbol{\psi}$ in Eq. (6.20), while $\mathbf{B}$ is the counter part of $\mathbf{r}$ in the same equation.

So, from a mathematical consideration, one could regard the vector potential $\mathbf{A}$ as a *basic field*. However, if one considers the problem from a physical point of view, one may come to a different conclusion. In the following section, we will explain why the *vector potential* $\mathbf{A}$ is not the most appropriate *basic field* representing the wave movement in the vacuum.

6.4 The Excitation Wave of the Vacuum is Characterized by the Variation of the *Electric Vector Potential Z*

From the analogy of sound wave transmission in an elastic solid, we know the excitation wave is carried by the displacement of a volume element [which is represented by $\mathbf{r}$ in Eq. (6.20)]. In the case of wave transmission in the vacuum, what could play the role of the displacement of a volume element? Clearly, ***B** could not play the role of being a displacement of the vacuum medium, since that would imply that the vacuum is composed of magnetic monopoles*. Up to now, there is no experimental evidence to support the existence of magnetic monopoles in our universe. On the other hand, as we pointed out earlier, the vacuum behaves as a dielectric medium according to the Maxwell theory. Thus, *the physical parameter representing the displacement of the vacuum medium should be the electric displacement **D** instead of the magnetic flux **B***.

In another word, if we want to model wave propagation in the vacuum in analogy to wave propagation in a physical medium, $\mathbf{D}$ should be the counter part of $\mathbf{r}$. Thus, we may apply the Helmholtz decomposition theorem to decompose $\mathbf{D}$ into a *curl-free* component and a *divergence-free* component, i.e.,

$$\mathbf{D} = -\nabla\varphi + \nabla \times \mathbf{Z}, \tag{6.21}$$

where $\nabla \cdot \mathbf{Z} = 0$. We may call $\mathbf{Z}$ the "*electric vector potential*". (From now on, $\mathbf{A}$ will be referred to as the "*magnetic vector potential*".) In the vacuum, there is no free charge, $\rho_e = 0$. Thus, $\nabla \cdot \mathbf{D} = -\nabla^2\varphi = 0$. This can be satisfied by choosing $\nabla\varphi = 0$. Equation (6.21) then becomes

$$\mathbf{D} = \nabla \times \mathbf{Z}. \tag{6.22}$$

From this relation, it is clear that the dynamic change of $\mathbf{Z}$ is a measure of the variation of the electric displacement $\mathbf{D}$, which in turn represents the local movement of the vacuum medium.

It can be shown that the *electric vector potential* $\mathbf{Z}$ can easily satisfy the mathematical requirement of being a *basic field*. From Eq. (6.22) and the relation $\mathbf{D} = \varepsilon_o\mathbf{E}$, we see

$$\mathbf{E} = \frac{1}{\varepsilon_0}\nabla \times \mathbf{Z}. \tag{6.23}$$

From the revised Ampère's Law and Eq. (6.21), we can get

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} = \frac{\partial(\nabla \times \mathbf{Z})}{\partial t} = \nabla \times \frac{\partial \mathbf{Z}}{\partial t}.$$

This implies

$$\mathbf{H} = \frac{\partial \mathbf{Z}}{\partial t}. \tag{6.24}$$

Thus, one can see that ***the electric field E is a spatial derivative of the electric vector potential Z***, while ***the magnetic field H is a time derivative of Z***. Substituting the above two equations into Eq. (6.15), we have

$$\mathcal{L} = \frac{1}{2}\left[\frac{1}{\varepsilon_0}|\nabla \times \mathbf{Z}|^2 - \mu_0\left|\frac{\partial \mathbf{Z}}{\partial t}\right|^2\right]. \tag{6.25}$$

This Lagrangian density has a similar form as that of a one-dimensional string if one equates $\mathbf{Z}$ with the wave amplitude ϕ. Thus, we may identify *the basic field for the excitation wave of the vacuum as the electric vector potential (*$\mathbf{Z}$*).*

6.4.1 Mechanism of Wave Propagation in the Vacuum as Driven by Z

In the Maxwell theory, the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ can cross-interact with each other. In fact, such cross-interactions are responsible for generating the electromagnetic radiation waves. From our analysis given in the preceding section, we found $\mathbf{Z}$ and $\mathbf{A}$ can also cross-interact with each other. In fact, such cross-interactions are fully capable to generate propagating waves in the vacuum.

By substituting Eqs. (6.23) into Faraday's Law, i.e., Eq. (6.6), we have

$$-\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \mathbf{E} = \frac{1}{\varepsilon_0}(\nabla \times \nabla \times \mathbf{Z}) = -\frac{\partial(\nabla \times \mathbf{A})}{\partial t}. \tag{6.26}$$

This implies

$$\nabla \times \mathbf{Z} = -\varepsilon_0 \frac{\partial \mathbf{A}}{\partial t}. \tag{6.27}$$

Also, by combining Eq. (6.24) and $\mathbf{H} = \frac{1}{\mu_0}\mathbf{B} = \frac{1}{\mu_0}(\nabla \times \mathbf{A})$, we can obtain

$$\nabla \times \mathbf{A} = \mu_0 \frac{\partial \mathbf{Z}}{\partial t}. \tag{6.28}$$

From the above two equations, one can clearly see that the *vector potentials* $\mathbf{Z}$ and $\mathbf{A}$ are cross-interacting with each other. Such cross-interactions can generate a transverse wave characterized by the variation of $\mathbf{Z}$. Applying a curl operation on Eq. (6.27) and combining the result with Eq. (6.28), we have

$$\nabla \times (\nabla \times \mathbf{Z}) = -\varepsilon_0 \frac{\partial(\nabla \times \mathbf{A})}{\partial t} = -\varepsilon_0 \mu_0 \frac{\partial^2 \mathbf{Z}}{\partial t^2}. \tag{6.29}$$

Since $\nabla \times (\nabla \times \mathbf{Z}) = \nabla(\nabla \cdot \mathbf{Z}) - \nabla^2 \mathbf{Z}$ and $\nabla \cdot \mathbf{Z} = 0$, Eq. (6.29) becomes

$$\nabla^2 \mathbf{Z} - \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2} = 0, \tag{6.30}$$

where $c = \sqrt{1/\varepsilon_0 \mu_0}$ is the speed of light. Thus, the variation of **Z** follows a wave equation that is similar to the wave equation of light. This supports our proposal that **Z** can represent the motion of the vacuum medium during wave excitation.

6.5 Comparison Between Wave Excitations in the Mechanical System and the Vacuum Medium

From the above discussions, one can see that there is a close analogy between the wave generating mechanisms in a mechanical medium and those in the vacuum. Table 6.1 is a detailed comparison between different types of wave generating mechanisms. When a mechanical medium (such as an elastic solid) is undergoing an excitation, its strain is represented by a displacement vector **r**, and its stress (local force/field) is a tensor related to the strain via the generalized Hooke's Law. Similarly, when the vacuum medium is undergoing an excitation, it produces an electric displacement **D**; the local electric field **E** is related to **D** through a relation analogous to Hooke's Law, $\mathbf{E} = \frac{1}{\varepsilon_0}\mathbf{D}$. Furthermore, the displacement vector **r** can be decomposed into a curl-free component ϕ and a divergence-free component $\boldsymbol{\psi}$**.** Similarly, the electric displacement vector **D** can also be decomposed into a curl-free component $\nabla\varphi$ and a divergence-free component $\nabla \times \mathbf{Z}$**.** The only difference is that here $\nabla\varphi = 0$ in the vacuum. This means that, unlike the elastic solid, the vacuum medium has no longitudinal wave; it can only generate transverse waves.

From Table 6.1, one can see that, regardless of the nature of the wave medium, wave propagation requires a cross-interacting mechanism. For wave propagation in an elastic solid, the cross-interaction is mediated through two coupling equations, i.e., the Newton's Law and the generalized Hooke's Law. For the vacuum medium, the coupling equations appear to be the Ampere's Law (as modified by Maxwell) and Faraday's Law. But at a deeper level, we find the coupling equations are more like the cross-interactions between the electric vector potential **Z** and the magnetic vector potential **A**, i.e., $\frac{\partial \mathbf{Z}}{\partial t} = \frac{1}{\mu_0}\nabla \times \mathbf{A}$ and $\frac{\partial \mathbf{A}}{\partial t} = -\frac{1}{\varepsilon_0}\nabla \times \mathbf{Z}$. It is their cross-interactions that generated the excitation wave.

Another interesting point one can see from Table 6.1 is that, no matter for mechanical medium or vacuum medium, all wave equations appear to have the same form. They are highly symmetrical. For later reference, we may call these wave equations

Table 6.1 Comparison of wave generating mechanisms between the mechanical medium and the vacuum medium

	Mechanical medium *(Elastic solid)*	**Vacuum medium**	
		Electric component	*Magnetic component*
Strain	$\mathbf{r}$	Electric displacement $\mathbf{D}$	Magnetic flux $\mathbf{B}$
Stress	$f \propto \mathbf{r}$ $\sigma_{ij} = \lambda r_{k,k}\delta_{ij} + \mu(r_{i,j} + r_{j,i})$	Electric field $\mathbf{E} = \frac{1}{\varepsilon_0}\mathbf{D}$	Magnetic field $\mathbf{H} = \frac{1}{\mu_0}\mathbf{B}$
Helmholtz decomposition	$\mathbf{r} = -\nabla\phi + \nabla \times \boldsymbol{\psi}$	$\mathbf{D} = \nabla \times \mathbf{Z}$	$\mathbf{B} = \nabla \times \mathbf{A}$
Potential function	Dilational wave ϕ and transverse wave $\boldsymbol{\psi}$	Electric vector potential $\mathbf{Z}$	Magnetic vector potential $\mathbf{A}$
Cross-interacting mechanism: **Coupling equation**	Newton's Law Generalized Hooke's Law	$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_0}\nabla \times \mathbf{H}$ Ampere's Law (as modified by Maxwell)	$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu_0}\nabla \times \mathbf{E}$ Faraday's Law
		$\frac{\partial \mathbf{Z}}{\partial t} = \frac{1}{\mu_0}\nabla \times \mathbf{A}$	$\frac{\partial \mathbf{A}}{\partial t} = -\frac{1}{\varepsilon_0}\nabla \times \mathbf{Z}$
Wave equation (transverse)	$\nabla^2\boldsymbol{\psi} - \frac{1}{c_s^2}\frac{\partial^2\boldsymbol{\psi}}{\partial t^2} = 0$ where $c_s = \sqrt{\mu/\rho}$	$\nabla^2\mathbf{Z} - \frac{1}{c^2}\frac{\partial^2\mathbf{Z}}{\partial t^2} = 0$ where $c = 1/\sqrt{\mu_0\varepsilon_0}$	$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = 0$ where $c = 1/\sqrt{\mu_0\varepsilon_0}$
Wave equation (longitudinal)	$\nabla^2\varphi - \frac{1}{c_p^2}\frac{\partial^2\varphi}{\partial t^2} = 0$ where $c_p = \sqrt{(\lambda + 2\mu)/\rho}$	Not applicable	Not applicable

"*four-dimensional Laplace equation*". We may point out that, although the wave equations in different systems look the same, their wave functions represent very different *basic fields*. Such fields are associated with different measurements of the medium displacement.

6.6 Chapter Summary

- We discovered that the Maxwell theory had a hidden assumption; namely, the vacuum should behave like a dielectric medium. A major contribution of Maxwell was his introduction of the *electric displacement* **D** into electrodynamics. Maxwell's theory of light propagation required that **the vacuum must behave**

like a dielectric medium. If the vacuum is an empty space, **D** must automatically equal to zero. It will then be impossible to derive the wave equation of light.

- Then, one can only assume that **the vacuum is filled with some sort of *primordial dielectric charges*,** which are massless, highly refined, isotropic, and uniformly distributed. Here, we propose that the dielectric medium of the vacuum is composed of two types of components: (1) a negatively charged "*n-type medium*" and (2) a positively charged "*p-type medium"*. Each of these media is composed of very refined *primordial dielectric charges.*
- The wave function of a free particle represents a local movement of the vacuum medium; such a movement can be characterized by the variation of a "*basic field*". Since the vacuum behaves as a dielectric medium, the physical parameter representing the displacement of the vacuum medium should be the *electric displacement* **D** instead of the *magnetic flux* **B**.
- By applying the Helmholtz decomposition theorem to decompose **D,** one can show that $\mathbf{D} = \nabla \times \mathbf{Z}$, where **Z** is a newly defined parameter called the "*electric vector potential*". The dynamic change of **Z** is a measure of the variation of the electric displacement **D**. One can identify **Z** as the "*basic field*" of the vacuum, since it represents the local movement of the vacuum medium.
- Based on our analysis, one can see that there is a close analogy between the wave generating mechanisms in a mechanical medium (such as an elastic solid) and those in the vacuum. But unlike the elastic solid, the vacuum medium has no longitudinal wave; it can only generate transverse waves.

References

1. Maxwell, J. C. (1862). On physical lines of force (Part 3). *Philosophical Magazine, 23*, 12–24.
2. Maxwell, J. C. (1865). A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society, 155*, 459–512.
3. Maxwell, J. C. (1873). *A Treatise on Electricity and Magnetism* (1st ed., Vol. 1). Oxford University Press.
4. Longair, M. S. (1984). *Theoretical Concepts in Physics: An Alternative View of Theoretical Reasoning in Physics*. Cambridge University Press.
5. Hunt, B. J. (2012). Oliver Heaviside: A first-rate oddity. *Physics Today, 65*, 48–54.
6. Cottingham, W. N., & Greenwood, D. A. (1998). *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press.
7. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). *The Feynman Lectures on Physics* (Vol. 2). Addison-Wesley.
8. Chang, D. C. (2021). Review on the physical basis of wave–particle duality: Conceptual connection between quantum mechanics and the Maxwell theory. *Modern Physics Letters B, 35*, 2130004.
9. Yang, C. N. (2014). The conceptual origins of Maxwell's equations and gauge theory. *Physics Today, 67*, 45–51.
10. Salencon, J. (2001). *Handbook of Continuum Mechanics: General Concepts, Thermoelasticity*. Springer Science & Business Media.

Part III
Derivation of the Quantum Wave Equations and the Physical Meaning of the Quantum Wave Function

Chapter 7
Derivation of the Quantum Wave Equations Based on Wave Excitation in the Vacuum

In the last chapter, we showed that based on the Maxwell theory, the quantum vacuum is found to behave like a dielectric medium. The excitation wave of the vacuum is carried by a newly defined *electric vector potential,* **Z**, the variation of which characterizes the local movement of *the electric displacement* **D**. Using the Helmholtz decomposition theorem, one can show that the *electric field* **E** is associated with the curl of **Z**, while the *magnetic field* **H** is associated with the time derivative of **Z**. Therefore, **Z** is a direct measure of the dynamic changes of the local electric and magnetic fields of the vacuum.

In wave mechanics, the excitation wave is often described by an oscillation of a *basic field*, which represents the movement of the wave medium. Since we believe the matter wave is an excitation wave of the vacuum, we expect that the *basic field* for the matter wave should be the *electric vector potential* (**Z**). In the following, we will show that this is indeed the case.

Furthermore, it can be shown that the known quantum wave equations (for particles with or without mass) can be derived directly from the wave equation of the vacuum. In this chapter, we will show that both the *wave equation of a photon* and the *Klein–Gordon equation* can be derived directly from the wave equation of **Z**.[1] Most interestingly, such derivation also suggests a way to connect the concept of "*rest mass*" to a wave property.

[1] This chapter is based on our previous publication: D. C. Chang, *Mod. Phys. Lett. B*, 35, 2130004 (2021).

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_7

7.1 The Wave Equation of the Quantum Vacuum

7.1.1 Identifying Z as the Wave Function of the Excitation Wave in the Vacuum

Our proposal to identify the *electric vector potential* **Z** as the *basic field* of the vacuum excitation waves is mainly based on the following considerations:

First, **Z is a direct measure of the displacement of the vacuum medium**. From the Maxwell theory, we showed that the quantum vacuum behaves like a dielectric medium. During wave propagation, the motion of this vacuum medium is represented by the dynamic change of the electric displacement **D**, which is characterized by the vector potential **Z** based on the Helmholtz theorem. Thus, **Z** is a direct measure of the motion of the vacuum medium.

Furthermore, **Z satisfies the mathematical requirement of being a basic field**. If one uses the Lagrangian formulism to derive a wave equation for the vacuum, the wave function should represent a *basic field*, which must satisfy certain criteria. One of these criteria is that the Lagrangian density for the system is composed of quadratic terms of the first derivatives of the *basic field*. **Z** can clearly satisfy this requirement. In fact, this can be demonstrated clearly using the 4-vector notation [1]. We can write $Z^{\mu} = (0,\ \mathbf{Z})$ and define a tensor

$$K^{\mu\nu} \equiv \partial^{\mu} Z^{\nu} - \partial^{\nu} Z^{\mu} \tag{7.1}$$

to construct a symmetrical Lagrangian density,

$$\mathcal{L} = a K_{\mu\nu} K^{\mu\nu}. \tag{7.2}$$

As shown in **Appendix B**, one can use this Lagrangian density and the Hamilton's principle to derive the excitation wave equation of the vacuum, which is

$$\partial_{\mu} \partial^{\mu} \mathbf{Z} = 0. \tag{7.3}$$

This equation is identical to the wave equation of the vacuum as derived from the Maxwell theory (see Eq. (6.30) in Chap. 6), i.e.,

$$\nabla^2 \mathbf{Z} - \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2} = 0. \tag{7.4}$$

In the following sections, we will show that the dynamic change of **Z** not only can give the correct wave equation for a photon, but it can also give the quantum wave equation of a massive particle.

7.1.2 Connecting Z with the Quantum Wave Function of a Particle

In this model, we propose that both the radiation wave and matter wave are excitation waves of the same quantum vacuum. Then, not only **Z** can represent the wave function of light, it can also represent the wave function of the matter wave. To connect the *vector potential* **Z** with the quantum wave function (ψ) of a particle, we can consider a simple case in which **Z** is polarized in a fixed direction, that is,

$$\mathbf{Z}(\boldsymbol{x}, t) = \boldsymbol{\epsilon_k}\, \psi(\boldsymbol{x}, t), \tag{7.5}$$

where $\boldsymbol{\epsilon_k}$ is a *polarization factor* which specifies the orientation of **Z**. Let us denote the position vector as $\boldsymbol{x} = (x_1, x_2, x_3)$ and choose the axis x_3 as parallel to the motion of the particle, i.e., $\boldsymbol{x}_3 \| \boldsymbol{k}$ (see Fig. 7.1a). Since **Z** is a transverse wave, $\boldsymbol{\epsilon_k}$ can be written as

$$\boldsymbol{\epsilon_k} = a_1 \mathrm{e}^{i\theta_1} \hat{\boldsymbol{x}}_1 + a_2 \mathrm{e}^{i\theta_2} \hat{\boldsymbol{x}}_2, \tag{7.6}$$

where a_1, a_2 and θ_1, θ_2 are amplitudes and phase angles; they are fitting constants (see Fig. 7.1b). By substituting Eq. (7.5) into Eq. (7.4), we have

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0. \tag{7.7}$$

From now on, we will call this the "***Wave Equation of the Vacuum*** (WEV)".

7.2 The Wave Equation of a Photon Based on the Dynamic Change of *Z*

One can easily recognize that the above wave equation is identical in form with the common wave equation of a photon. The simplest solution of this wave equation is a plane wave

$$\psi_{\hat{k}}(\boldsymbol{x}, t) \propto \mathrm{e}^{i\,(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)}, \tag{7.8}$$

where $\boldsymbol{k}$ and ω are the wave vector and frequency, respectively. By substituting Eq. (7.8) into Eq. (7.7), one can see its dispersion relation is

$$\omega = ck. \tag{7.9}$$

By using the Planck's relation and the de Broglie relation, the above equation becomes

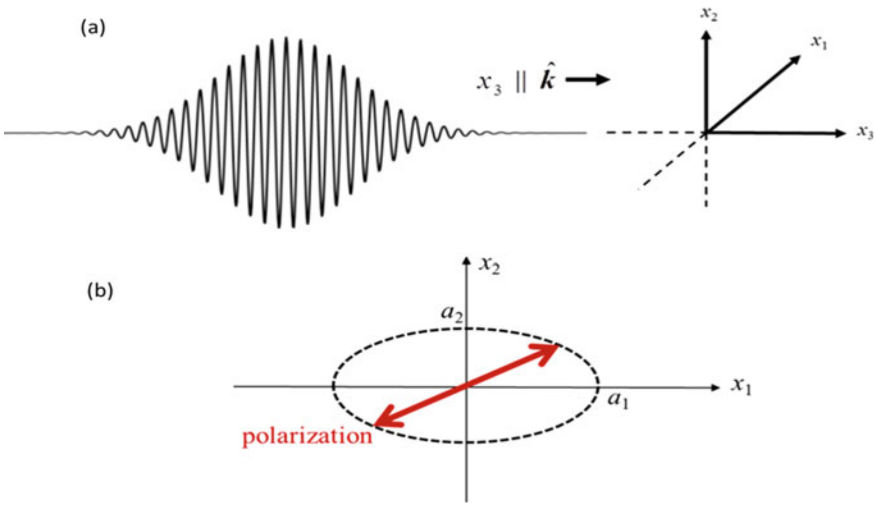


Fig. 7.1 **Excitation wave in the vacuum**. **a** The wave function represents the electric vector potential $\boldsymbol{Z}$, which is a transverse wave moving in the direction of $\boldsymbol{k}$. The amplitude of $\boldsymbol{Z}$ oscillates along the particle pathway and is described by the quantum wave function $\psi(\boldsymbol{x}, t)$. **b** $\mathbf{Z}$ can be a polarized vector. The orientation of its polarization factor $\boldsymbol{\epsilon_k}$ is shown here by the red arrow. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Mod. Phys. Lett. B* **35**, 2130004 (2021)

$$E = cp,$$

which is identical to the known *energy–momentum relation* of light. Thus, one can identify the plane wave solution of the WEV as the wave function of a photon.

7.3 Deriving the Wave Equation of a Massive Particle

One may note that the plane wave solution is only the simplest solution of WEV. The general solutions of the WEV are more complicated. Could some of these complex solutions of the WEV represent the matter wave of certain massive particles? The quantum mechanical wave equations for some massive particles are already known. For example, the quantum wave equation for a scalar particle is the *Klein–Gordon equation*, while the quantum wave equation for the electron is the *Dirac equation*. Can one derive these quantum wave equations based on the wave equation of $\mathbf{Z}$? In the following, we will show that this is indeed possible.

7.3.1 Physical Nature of the Wave Function Representing a Massive Particle

Since a massive particle behaves like a point object in the macroscopic scale, we expect that it should be found mainly along its trajectory. That means its wave function is concentrated near the center of its pathway. In another word, the wave function should vary not only along the coordinate parallel to its trajectory (i.e., $\hat{\boldsymbol{k}} \cdot \boldsymbol{x}$), but also in the transverse plane ($\hat{\boldsymbol{k}} \times \boldsymbol{x}$). Therefore, the wave function representing a free particle can be written in the following form,

$$\psi_{\hat{\boldsymbol{k}}}(\boldsymbol{x}, t) = \psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right), \tag{7.10}$$

where ψ_T is the transverse component of the wave function, while ψ_{path} is the longitudinal component of the wave function that describes the motion of the traveling wave along the particle's trajectory. Substituting Eq. (7.10) into Eq. (7.7), and recall that $\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x},\ t\right)$ is a function of (x_3), while $\psi_T\left(\hat{\boldsymbol{k}} \times\ \boldsymbol{x}\right)$ is a function of (x_1,x_2), they are independent from each other. Then, one can use the technique of separation of variables to convert the wave equation of the vacuum into two simultaneous equations. More specifically, by substituting Eq. (7.10) into Eq. (7.7), one has

$$\nabla^2\psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right) = 0.$$

Then,

$$\begin{aligned}&\left[\nabla^2\psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)\right]\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right) + \psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)\left[\nabla^2\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)\right]\\ &\quad - \left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)\right]\psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right) = 0.\end{aligned}$$

Divide the above equation by $\psi_T\left(\hat{\boldsymbol{k}} \times\ \boldsymbol{x}\right)\ \psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x},\ t\right)$, one will get

$$\begin{aligned}&\frac{1}{\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)}\left[\nabla^2\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)\right]\\ &\qquad = -\frac{\nabla^2\psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)}{\psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)}.\end{aligned}$$

Since the left-hand side of this equation is a function of $\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)$ while the right-hand side is a function of $\hat{\boldsymbol{k}} \times \boldsymbol{x}$, the two sides can be equal only when they are equal to a constant, which we may denote it as ℓ^2. Then, from each side of the above equation, we can obtain the following simultaneous equations

$$\begin{cases} \left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\psi_{path}\left(\hat{\boldsymbol{k}}\cdot\boldsymbol{x}, t\right) = \ell^2\,\psi_{path}\left(\hat{\boldsymbol{k}}\cdot\boldsymbol{x}, t\right) & (7.11) \\ \nabla^2\psi_T\left(\hat{\boldsymbol{k}}\times\boldsymbol{x}\right) = -\ell^2\psi_T\left(\hat{\boldsymbol{k}}\times\boldsymbol{x}\right), & (7.12) \end{cases}$$

The above equations can be solved separately. The solution of Eq. (7.11) is a plane wave

$$\psi_{path}\left(\hat{\boldsymbol{k}}\cdot\boldsymbol{x}, t\right) \propto e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}, \tag{7.13}$$

where $\boldsymbol{k} = k\hat{\boldsymbol{k}}$ and

$$\omega^2 = \left(k^2 + \ell^2\right)c^2. \tag{7.14}$$

The solution of Eq. (7.12) is

$$\psi_T\left(\hat{\boldsymbol{k}}\times\boldsymbol{x}\right) \propto J_n(\ell r)e^{\pm in\theta}, \tag{7.15}$$

where J_n is Bessel function of the first kind; r and θ are the radius and the azimuthal angle in the transverse (x_1, x_2) plane. To simplify our notation, let us define the direction of particle movement as the z-axis, i.e., $\hat{\boldsymbol{k}}\|\hat{x}_3\|\hat{\boldsymbol{z}}$. Then, $\boldsymbol{k}\cdot\boldsymbol{x} = k\,z$. . From Eqs. (7.13) and (7.15), the wave function shown in Eq. (7.10) becomes

$$\psi_{\hat{\boldsymbol{k}}}(\boldsymbol{x}, t) = a_k J_n(\ell r)\,e^{\pm in\theta}e^{i\,(kz-\omega t)}, \tag{7.16}$$

(where a_k is a normalizing constant). As expected, the wave function of a free particle behaves like a traveling wave along the direction of its trajectory. Because of the phase factor $e^{\pm\,in\theta}$ and the Bessel function $J_n(\ell r)$, $\psi_{\hat{\boldsymbol{k}}}$ propagates in a helical fashion and decreases in an oscillating manner in the transverse direction. It behaves almost like a vortex (see Fig. 7.2) .

A very important outcome in our solution of the wave equation of the vacuum (WEV) is the restriction on the parameter n in Eq. (7.15). Mathematically, in the "*Bessel function of the first kind*" (J_n), **the parameter *n* can be either an integer or a half-integer**. When n is an integer, the phase factor $e^{\pm\,in\theta}$ is a normal rotating function. It is straightforward that θ and $\theta + 2\pi$ give the same point. But if n is a half-integer, such as $n = ½$, the rotation of the particle is more complicated, since θ and $\theta + 2\pi$ do not give the same point. The phase factor must rotate two cycles in order to return to its original point. For example, the full cycle of the transverse wave function for a particle with $n = ½$ is plotted in Fig. 7.3.

Since the wave function (having an integer n) and the wave function (having a half-integer n) have very different topological features, it is reasonable to suspect

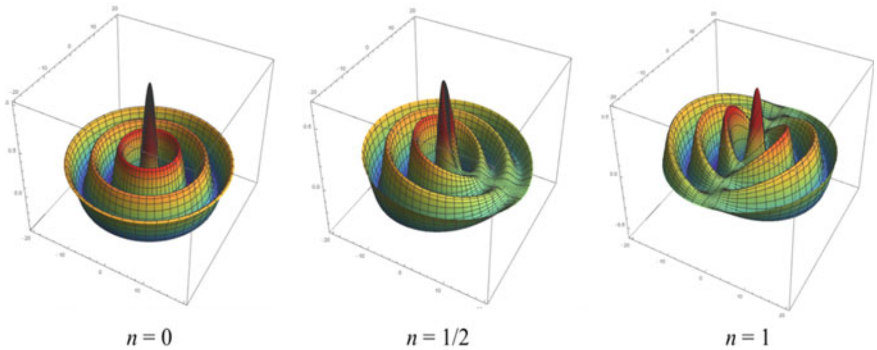


Fig. 7.2 A 3-D plot of the transverse wave function $\psi_T\left(\hat{k} \times x\right)$ in the *(x_1, x_2)* plane. Here $n = 0$, ½, 1. The magnitude of the wave function is shown in the z-axis and displayed in pseudo-color. (Here, the wave function is plotted with θ rotating for one revolution, i.e., 2π). Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, arXiv preprint physics/ 0505010v2 (2017)

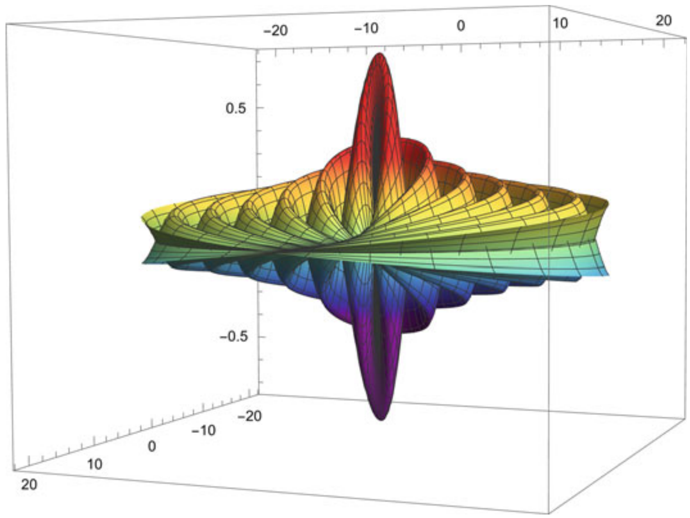


Fig. 7.3 A 3-D plot of the transverse wave function $\psi_T\left(\hat{k} \times x\right)$ *in the (x_1, x_2)* plan with $n =$ ½. The magnitude of the wave function is shown in the z-axis and displayed in pseudo-color. (Here, the wave function is plotted with θ rotating for two revolutions, i.e., 4π)

that these two types of wave functions may represent particles of different kinds. Indeed, from Eq. (7.16), it is clear that the parameter n specifies mainly the rotation of the wave function along the particle trajectory; **one may identify n as the "*spin*" of the particle**. If this interpretation is correct, then particles represented by the wave function having an integer n would be "*bosons*", and particles represented by the wave function having a half-integer n would be "*fermions*". **The mathematical**

condition of our solution thus predicts that all particles in nature must be either *bosons* or *fermions*; there is no possibility of having a third kind of particles (such as having a spin of 1/3 or other fractional numbers).

7.4 Identifying the Physical Meaning of Parameters Within the Wave Function

The wave function shown in Eq. (7.16) contains four parameters, ω, $\boldsymbol{k}$, ℓ, and n. From the discussion above, we know n could be related to the *spin* of the particle. What are the physical meanings of the other parameters? We can get some idea by comparing this wave function with that of a photon. In the case of a photon wave function (Eq. (7.8)), it is well known that ω and $\boldsymbol{k}$ are related to the energy (E) and momentum ($\boldsymbol{p}$) of the particle, as described by the Planck's relation and the de Broglie relation. If we compare the traveling wave component of Eq. (7.16) with Eq. (7.8), it is easy to see that ω and $\boldsymbol{k}$ should have the same meanings, i.e., $E = \hbar\,\omega$ and $\boldsymbol{p} = \hbar\,\boldsymbol{k}$. (In fact, according to the quantum wave model, both the Planck's relation and de Broglie's relation hold for a wave function representing a free particle with mass) [2].

Now, what is the physical meaning of ℓ? Our recent work suggested that ℓ is related to the rest mass of the particle [3]. This can be easily shown. From the Planck's relation and de Broglie's relation, Eq. (7.14) becomes

$$E^2 = c^2\left(p^2 + \hbar^2\ell^2\right). \tag{7.17}$$

$$E = \sqrt{c^2p^2 + c^2\hbar^2\ell^2}. \tag{7.17a}$$

Recall that the particle velocity (v) is determined by the group velocity of the wave packet [4], $v = \dfrac{\partial\omega}{\partial k} = \dfrac{\partial E}{\partial p}$, and the particle mass m is defined by $p = mv$ in the classical limit, one can use these relations to solve Eq. (7.17a) and obtain

$$m = \frac{\hbar\ell/c}{\left(1 - v^2/c^2\right)^{1/2}}. \tag{7.18}$$

We know at $v = 0$, m equals the rest mass m_0. Equation (7.18) then becomes

$$m_0 = \frac{\hbar\ell}{c}. \tag{7.19}$$

Substituting this into Eq. (7.17), one can see that the dispersion relation now leads to

$$E^2 = p^2c^2 + m_0^2c^4. \tag{7.20}$$

Thus, by identifying the wave parameter ℓ with the rest mass m_o, one can naturally obtain the *energy–momentum relation* of a free particle. By combining Eq. (7.18) and (7.19), one can also see that the mass of a particle is speed dependent

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{7.21}$$

By combining Eqs. (7.20) and (7.21), and $p = mv$, one can further obtain

$$E = mc^2. \tag{7.22}$$

This result suggests that the famous *mass-energy conversion relation* is really a consequence of the fact that the quantum particle is an excitation wave of the vacuum medium.

Our finding that **the wave parameter ℓ is associated with the rest mass** is not totally surprising. Since both momentum and energy of a free particle are known to be connected with "wave vector/wave number" (inverse of wavelength) in the spatial and temporal dimensions, it is reasonable to speculate that the rest mass may be connected with some sort of "intrinsic wave number" too. This is indeed the case. The asymptotic form of the Bessel function is known to be

$$J_n(\ell r) \to \left(\frac{2}{\pi \ell r}\right)^{1/2} \cos\left(\ell r - \frac{2n+1}{4}\pi\right). \tag{7.23}$$

Thus, ℓ can be regarded as the "transverse wave number" of the free particle. In fact, from Eqs. (7.23) and (7.19), one can easily see that the wavelength of this transverse oscillation is

$$\lambda_T = \frac{2\pi}{\ell} = \frac{h}{m_0 c}, \tag{7.24}$$

which is identical to the so-called "Compton wavelength" (λ_c) of the particle [5].

Our finding that *the rest mass is associated with the oscillation periodicity in the transverse direction* appears to make very good sense. It is closely parallel to the Planck's relation and the de Broglie relation, which show that the energy and momentum are related to the periodicity of oscillation of the vacuum medium. More specifically, E is shown to be related to the periodicity of oscillation in the time dimension, while $\boldsymbol{p}$ is related to the periodicity of oscillation in the spatial dimension along the direction of the particle trajectory. In essence, these results suggest that *energy, momentum, and mass have very similar physical meanings; all of them are related to the oscillation periodicity in different dimensions of space–time.*

7.5 Derivation of the Klein–Gordon Equation from the Wave Equation of the Vacuum

A well-known quantum wave equation for a massive particle is the Klein–Gordon equation [4, 6, 7]. Using the results obtained in the last section, one can easily derive the Klein–Gordon equation directly from the wave equation of the vacuum. As shown in the above, the wave function of a massive particle can be written as a product of a longitudinal component and a transverse component,

$$\psi_{\hat{\boldsymbol{k}}}(\boldsymbol{x},t) = \psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right). \tag{7.10}$$

Using the technique of "separation of variables", we showed that the WEV for the massive particle can be separated into two coupled equations, i.e., Eqs. (7.11) and (7.12). The equation of motion for the longitudinal component is

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right) = \ell^2\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right). \tag{7.11}$$

As shown in the last section, the wave parameter ℓ is found to be connected with the rest mass m_o, such that $m_0 = \hbar\ell/c$. Then, we can re-write Eq. (7.11) as

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\psi_{path} - \left(\frac{m_0 c}{\hbar}\right)^2\psi_{path} = 0. \tag{7.25}$$

This equation is identical to the "Klein–Gordon equation" [4]

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\varphi - \left(\frac{m_0 c}{\hbar}\right)^2\varphi = 0, \tag{7.26}$$

if we identify its wave function φ with ψ_{path}. This means that the Klein–Gordon equation can indeed be derived from the WEV. From this derivation, it is also clear that *the wave function of the Klein–Gordon equation describes only the longitudinal component of the matter wave*, i.e., the motion of the particle along its trajectory $\left(\psi_{path}\right)$.

7.6 Chapter Summary

- We proposed that both the radiation wave and matter wave are excitation waves of the same quantum vacuum, the wave function of which is represented by the *vector potential* **Z.** The dynamic change of **Z** is described by the *wave equation of the vacuum*, which can be derived using the Maxwell theory.

- The quantum wave equation of a photon can be derived based on the dynamic change of **Z**. The simplest solution of this wave equation is a plane wave.
- For a massive particle, its wave function not only varies along the coordinate parallel to its trajectory, but also in the transverse plane. Thus, the wave function can be written as the product of a transverse wave function (ψ_T) and a longitudinal wave function $\left(\psi_{path}\right)$, and the latter describes the motion of the traveling wave along the particle's trajectory.
- One can then solve the *wave equation of the vacuum* by using the technique of separation of variables. The wave function of the free particle behaves like a traveling wave along the direction of its trajectory; its transverse component, however, oscillates following the Bessel function $J_n(\ell r)$. Thus, the wave function representing a massive particle propagates in a helical fashion and behaves like a vortex.
- The parameter n in the Bessel function can either be an *integer* or *half-integer*. One can identify n as the *spin* of the particle. Then, particles represented by the wave function having an integer n would be "*bosons*", while particles represented by the wave function having a half-integer n would be "*fermions*". **The mathematical condition of our solution thus predicts that all particles in nature must be either *bosons* or *fermions*.**
- The wave function contains four parameters, ω, $\boldsymbol{k}$, ℓ, and n. Besides n being identified with the *spin*, ω and $\boldsymbol{k}$ are identified with the particle's *energy* and *momentum*, based on the Planck's relation and the de Broglie relation. The wave parameter ℓ is identified with the rest mass. From these results, the so-called "*relativistic energy–momentum relation*" can be directly obtained from the *dispersion relation* of the particle wave function.
- The Klein–Gordon equation can be derived directly from the *wave equation of the vacuum*. This derivation required that the wave function of the Klein–Gordon equation describes only the longitudinal component of the matter wave, i.e., the motion of the particle along its trajectory $\left(\psi_{path}\right)$.

References

1. Cottingham, W. N., & Greenwood, D. A. (1998). *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press.
2. Chang, D. C. (2016). *What is the physical meaning of mass in view of wave-particle duality?*. arXiv preprint physics/0404044v2.
3. Chang, D. C. (2013). A classical approach to the modeling of quantum mass. *Journal of Modern Physics, 4*, 21–30.
4. Messiah, A. (1965). *Quantum Mechanics* (Vol. 1). Wiley.
5. Shankland, R. (1961). *Atomic and Nuclear Physics* (2nd ed.). Newyork: MacMillan.
6. Gordon, W. (1926). Der Comptoneffekt nach der Schrödingerschen Theorie. *Zeitschrift für Physik, 40*, 117–133.
7. Klein, O. (1927). Elektrodynamik und Wellenmechanik vom Standpunkt des Korrespondenzprinzips. *Zeitschrift für Physik, 41*, 407–442.

Chapter 8
Derivation of the Dirac Equation from the Wave Equation of the Vacuum

In the last chapter, we showed that the wave equation of a photon can be easily derived from the wave equation of the vacuum. This is no surprise. The more interesting finding is that, the quantum wave equation for a *massive particle* can also be derived from the wave equation of the vacuum medium. This equation is called the "*Klein–Gordon equation*", which is known to describe the movement of a quantum particle with *spin* = 0 (i.e., a boson). Since matter is made of atoms and the physicochemical property of an atom is determined by the state of the electrons, the more important quantum wave equation would be the quantum wave equation for an electron (e.g., the *Schrödinger equation* or the *Dirac equation*). Could such quantum wave equation for *spin* = ½ particle be derived from the wave equation of the vacuum? In the following, we will show that this is indeed the case. That is, not only the Klein–Gordon equation, but the Dirac equation can also be derived directly from the wave equation of the vacuum.

8.1 Derivation of the Quantum Wave Equation for an Electron

The Schrödinger equation was the first quantum wave equation to describe the wave function of an electron. But it was designed for a non-relativistic case. Thus, another theoretical physicist, Paul Dirac (see Fig. 8.1), decided to develop a new quantum wave equation to cover the relativistic situation. His approach was trying to linearize the quantum wave equation and to develop a quantum commutation relation in analogy to classical mechanics [1–3]. The following is a concise review of Dirac's original approach in deriving his famous equation. Then, we will show how to derive the Dirac equation based on the quantum wave model. One will see that our derivation based on the wave model is more straightforward.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_8

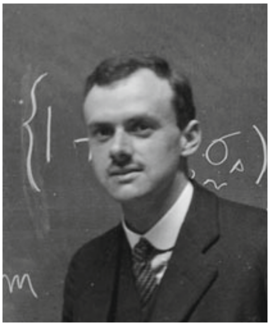


Fig. 8.1 Paul Dirac. Paul Dirac (1902–1984) was an English theoretical physicist. He was a major contributor to the development of quantum mechanics. He derived the first "relativistic quantum mechanical equation" (the "Dirac equation") and predicted the existence of anti-particles. This work earned him the 1933 Nobel Prize in Physics. Photo Credit: AIP Emilio Segrè Visual Archives, Gift of Mrs. Zemansky

8.1.1 How did Dirac Derive his Equation Originally?

In the early twentieth century, physicists were mainly trained in classical physics, and thus, their view of the sub-atomic particles was heavily influenced by the classical mechanical concepts. This was no exception for Dirac. So, when he started to develop his quantum theory of the electron, he naturally regarded the electron as a point mass. He based his development of the quantum theory on an analogy with classical mechanics [1]. Dirac's derivation of his quantum wave equation involved the following steps.

First, Dirac tried to identify the energy operator p_0 and the momentum operators p_1, p_2, p_3 based on an argument that the commutation relations between *p's* and *q's* are similar between the quantum conditions and the classical conditions. (Here, q is the canonical coordinate and p is the momentum.) That is, he assumed that in quantum mechanics, the p's and q's should obey the commutation relations:

$$\begin{cases} q_r q_s - q_s q_r = 0, \\ p_r p_s - p_s p_r = 0, \\ q_r p_s - p_s q_r = i\hbar\delta_{rs} \end{cases}$$

Dirac called these commutation relations the ***fundamental quantum conditions***.

Then, Dirac argued that, since the linear operators $-i\hbar\partial/\partial q_r$ and q's satisfy the same commutation relations as the *fundamental quantum conditions* between the *p's* and *q's*, he identified $p_0 = i\hbar\partial/\partial x_0$ and $p_r = -i\hbar\partial/\partial x_r$, where $r = 1, 2, 3$.

One may notice that these identifications are similar to what Schrödinger used earlier in his development of the quantum wave equation.

Next, he deduced from quite general arguments that the relativistic quantum wave equation must be linear in the operator $\partial/\partial t$ and $\partial/\partial x_r$, that is, the correct wave equation was assumed to be in the form of

$$\{p_0 - \alpha_1 p_1 - \alpha_2 p_2 - \alpha_3 p_3 - \beta\}\psi = 0, \tag{8.1}$$

where p_0 is the energy operator and p_1, p_2, p_3 are momentum operators, the α's and β are coefficients independent of the p's, the value of which will be determined later. Here, he used the four-vector notation (x_0, x_1, x_2, x_3), which corresponds to (t, x, y, z).

Finally, to identify the α's and β, Dirac argued that the linear wave equation proposed by him (i.e., Eq. (8.1) should be consistent with the *relativistic energy–momentum relation*, that is, the wave equation for the electron should obey

$$\left\{p_0^2 - m^2c^2 - p_1^2 - p_2^2 - p_3^2\right\}\psi = 0. \tag{8.2}$$

Thus, by multiplying the Eq. (8.1) by the operator $\{p_0 + \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta\}$ on the left, he obtained

$$\{p_0 + \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta\}\{p_0 - \alpha_1 p_1 - \alpha_2 p_2 - \alpha_3 p_3 - \beta\}\psi = 0. \tag{8.3}$$

By comparing this equation with (8.2), he could identify the values of the α's and β. As it turns out, the α's and β are shown to be associated with a vector $\boldsymbol{\sigma}$ represented by 4×4 matrix. Dirac interpreted $\boldsymbol{\sigma}$ as the origin of spin of the particle. The Eq. (8.1) is subsequently called the "Dirac equation".

The Dirac equation is a highly successful equation that provides a theoretical basis for calculating the quantum effects of electrons. However, one may see that, Dirac's original derivation was very complicated and involved a number of theoretical assumptions and conjectures, some of which could be debatable. Such complication is probably because it is very difficult to develop a quantum wave equation based on the classical mechanical view. In the following, we will show that the Dirac equation can be derived more naturally based on the quantum wave model. In fact, one can see that the Dirac equation is a direct consequence of the wave equation of the vacuum (WEV).

8.2 Derivation of the Dirac Equation Based on the Quantum Wave Model

In the quantum wave model, we hypothesize that all particles are excitation waves of the vacuum medium, thus, they should obey the same vacuum wave equation regardless of mass and spin. In the last chapter, we showed that the wave equation of the vacuum (WEV) is Eq. (7.7). Different particles are expected to be represented by its different solutions. In fact, we showed that both the wave equation of a photon and the wave equation of a massive particle (the *Klein–Gordon equation*) can be derived directly from the wave equation of the vacuum. Now, since the WEV is a general wave equation for all particles, it is not only valid for scalar particles, but it should also be valid for spin = ½ particles.

This means that one should be able to derive the wave equations for an electron from WEV. In the following, we will show that the WEV can indeed lead to the Dirac equation.

8.2.1 To Derive the Dirac Equation by Factorizing the Klein–Gordon Equation

In order to derive the Dirac equation from the WEV, we make use of three considerations:

(a) We hypothesize that, like the Klein–Gordon equation, the Dirac equation may represent only the longitudinal component of the matter wave. That is, the wave function of the Dirac equation is connected to the ψ_{path} component of $\psi_{\hat{\boldsymbol{k}}}$.

(b) If this is the case, one may be able to show that the Dirac equation is a special case of the Klein–Gordon equation. Since we already knew the Klein–Gordon equation is a special case of the WEV, then, the solution of the Dirac equation should automatically satisfy the WEV.

(c) The wave function representing a matter wave is not necessarily a scalar function. Depending on the spin of the particle, the wave functions may have different mathematical forms. Some can be scalars, while others can be vectors or matrices. For particles with spin $= 0$, the solution can simply be a scalar function. In the case of an electron, the situation could be more complicated. Since it has a spin = ½, its full wave function (as shown in Eq. (7.16) cannot be a single-valued function in respect to θ. Thus, it may be more appropriate to use a matrix to represent it.

If we denote ψ as a special solution of the Klein–Gordon equation that represents a free electron, ψ should satisfy

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\psi - \left(\frac{mc}{\hbar}\right)^2\psi = 0. \tag{8.4}$$

To simplify the mathematical calculation, let us use the natural unit (i.e., $\hbar = 1$, $c = 1$) for the above equation,

$$-\frac{\partial^2\psi}{\partial t^2} + \nabla^2\psi - m^2\psi = 0. \tag{8.4a}$$

The left-hand side of this equation can be decomposed into the product of two factors

$$\left(i\frac{\partial}{\partial t} - i\alpha\cdot\nabla + \beta m\right)\left(i\frac{\partial}{\partial t} + i\alpha\cdot\nabla - \beta m\right)\psi = 0. \tag{8.5}$$

Then, the above equation can be decoupled into two independent equations,

$$\begin{cases} \left(i\frac{\partial}{\partial t} + i\alpha\cdot\nabla - \beta m\right)\psi = 0 & (8.6) \\ \left(i\frac{\partial}{\partial t} - i\alpha\cdot\nabla + \beta m\right)\psi = 0. & (8.7) \end{cases}$$

Equation (8.5) can be rewritten explicitly as

$$\left[-\frac{\partial^2}{\partial t^2} + (\alpha\cdot\nabla)^2 + im\beta(\alpha\cdot\nabla) + im(\alpha\cdot\nabla)\beta - \beta^2 m^2\right]\psi = 0. \tag{8.5a}$$

One can see that, in order to make Eq. (8.5a) equal to Eq. (8.4a), the parameter α and β must satisfy the following conditions:

$$\begin{cases} \beta^2 = 1 \\ \alpha\beta + \beta\alpha = 0 \\ \alpha_i\alpha_j = 1, \text{if } i = j \\ \alpha_i\alpha_j + \alpha_j\alpha_i = 0, \ \text{if } i \neq j. \end{cases} \tag{8.8}$$

These conditions post a restriction on the possible mathematical form of the parameters α and β. If α and β are ordinary numbers, it is impossible for them to satisfy Eq. (8.8). In order to satisfy those conditions, both α and β must be treated as a 4×4 matrix. It can be shown that the conditions Eq. (8.8) can indeed be satisfied if one defines the matrices α and β as [4]

$$\alpha_k = \begin{pmatrix} \mathbf{0} & \sigma_k \\ \sigma_k & \mathbf{0} \end{pmatrix}, \beta = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & -I \end{pmatrix},$$

where the various σ_k are 2×2 Pauli spin matrices, where I is the unity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and $$\mathbf{0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Substituting the above values of α and β into Eq. (8.6), this equation now becomes the "Dirac equation".

From the above analysis, it is clear that, if one wants to make the Dirac equation to satisfy the Klein–Gordon equation, α and β must be treated as 4×4 matrices [5]. In other words, the Dirac equation must be treated as 4×4 matrix equation. In the literature, the Dirac equation can also be expressed using a set of 4×4 γ_μ matrices [4], which are given by

$$\gamma_k = \begin{pmatrix} \mathbf{0} & -i\sigma_k \\ i\sigma_k & \mathbf{0} \end{pmatrix}, \quad \gamma_4 = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & -I \end{pmatrix}.$$

More explicitly, these matrices look like

$$\gamma_3 = \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix}, \quad \gamma_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \text{etc.}$$

These 4×4 matrices γ_μ are called the *gamma matrices* or the *Dirac matrices*. Using these *gamma matrices,* the linearized equation (Eq. (8.6)) now can be rewritten as

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + \frac{mc}{\hbar}\right)\psi = 0, \tag{8.9}$$

where $\mu = 1, 2, 3, 4$. This is another common form of the "Dirac equation" [4].

A major advantage of the Dirac equation is that it naturally brought in the spin concept through the incorporation of Pauli's spin matrices. It can be shown that, by making use of Pauli's two-component theory, the spin of the particle can be included in the energy of the electron [4].

8.3 Physical Meaning of the Dirac Wave Function

It is now clear that the Dirac equation is a special case of the Klein–Gordon equation, which in turn is a special case of the WEV. Thus, any solution of the Dirac equation should also satisfy the WEV. This means that all electron wave functions obtained from the Dirac equation should automatically satisfy the wave equation of the vacuum [5].

Since the Dirac equation is a matrix equation, the wave function ψ must become a four-component column matrix (which is called a "spinor").

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} \tag{8.10}$$

where

$$\psi_L = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \psi_R = \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix}. \tag{8.11}$$

In current literature, the four-component column matrix ψ is called the "*Dirac spinor*" [4], ψ_L is called the *left-hand spinor* and ψ_R is called the *right-hand spinor*. The Dirac equation has been shown to be highly successful in describing the motion of an electron. In fact, the modern quantum field theory of electrons is almost entirely based on the Dirac equation. From the above review, one can see clearly that the Dirac equation is a derivative of the Klein–Gordon equation. But since the Klein–Gordon equation itself is a derivative of the WEV, it is obvious that the Dirac equation is also a derivative of the WEV. Thus, the wave function representing an electron is just a special solution of the WEV of the matter wave.

Why is the wave function representing an electron a spinor instead of a scalar? It is probably because it is a particle with $n = 1/2$. This makes the wave function not returning to the original values when θ is rotated by 2π. That means the wave function cannot be a single-valued function. It should have at least two values: (a) ψ ($\theta = 0$ to 2π); (b) ψ ($\theta = 2\pi$ to 4π). This implies that ψ can be a column matrix with at least two components. But since the phase for the wave equation can take on $\pm in\theta$, the wave function can rotate either clockwise or counter-clockwise. The wave function could be a mixture of a left-hand matrix and a right-hand matrix. This may explain why the wave function representing a free electron should have four components as shown in Eq. (8.10).

From the above analysis, one can see that, the wave functions of different types of particles are represented by different solutions of the vacuum wave equation (see Table 8.1). For example, the photon can be represented by a plane wave; the scalar particle can be represented by a cylindrically symmetrical vortex wave, while the electron can be represented by a 4-column matrix (spinor).

Table 8.1 Properties of the wave function for different free particles. These wave functions represent different solutions of the wave equation of the vacuum

Particle type	**Rest mass**	**Spin**	**Mathematical form of ψ**	**Wave equation**
Photon	No	1	Vector	Wave equation of light
Scalar particle	Yes	0	Scalar	Klein–Gordon equation
Electron / Lepton	Yes	½	Spinor	Dirac equation

8.4 Dirac's "Hole Theory" and the Prediction of Anti-Particle

Today, the Dirac equation has become the foundation of the quantum field theory of electrons; it is used to describe all spin-1/2 massive particles, called "Dirac particles", such as electrons and all leptons. The Dirac equation is also credited with predicting the existence of "anti-particles". Such a "great discovery", however, involved some bold interpretations.

When Dirac first derived his equation, he encountered a big problem. His equation not only gives positive-energy solutions, it can also generate negative-energy solutions [6]. How to interpret the negative-energy solutions of his equation becomes a problem. Mathematically speaking, there seems to be no reason for him to reject the negative-energy solutions. Then, any electron occupying a positive-energy eigenstate would decay into a negative-energy state of successively lower energy. That means the electron can keep on emitting photons and we cannot have stable atoms to construct our world.

To cope with this problem, Dirac introduced a very bold hypothesis, known as the "hole theory" [6, 7]. He argued that all the negative-energy electron eigenstates are occupied. Because of the Pauli exclusion principle, any additional electron would be forced to occupy a positive-energy eigenstate. In essence, he assumed that the vacuum is filled with an infinite number of negative-energy electrons. This "sea" of negative-energy electrons is called the "Dirac sea" [8].

Dirac further reasoned that, if one of these negative-energy electrons is knocked off by a gamma ray, the knock-off electron would gain energy to become a positive-energy electron; the unfilled vacancy—called a hole—would behave like a positively charged particle, with the same mass as the electron [9]. This hole therefore behaves like an anti-particle of the electron (called "positron"). (See Fig. 8.2). The particle positron was experimentally discovered by Carl Anderson in 1932 [10].

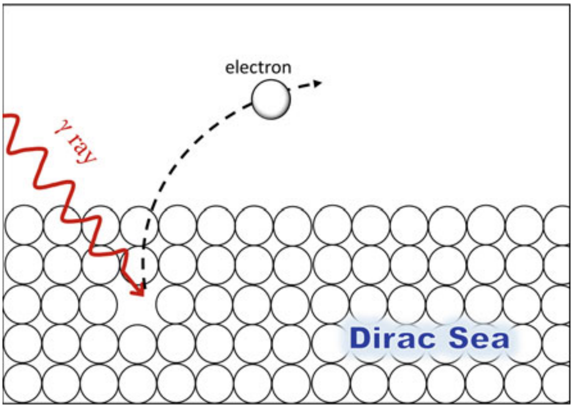


Fig. 8.2 The Dirac sea of negative-energy electrons. In Dirac's theory, the vacuum is pre-filled by a sea of negative-energy electrons. An energetic gamma ray can kick out one of the negative-energy electrons to become a positive-energy electron. The hole left behind can be regarded as the anti-particle of the electron, i.e., the positron

Although Dirac was given big credit for predicting the existence of the anti-particle, the assumption of an infinite "*Dirac sea*" of negative-energy electrons is troublesome. There is just no experimental evidence that our vacuum is filled with an infinite number of negative-energy electrons. This assumption appears to be *ad hoc* and arbitrary.

In fact, when Dirac first proposed his hole theory, most physicists did not take it seriously. For example, W. Pauli had expressed criticism on Dirac's hole theory: "*Recently Dirac attempted the explanation, already discussed by Oppenheimer, of identifying the holes with antielectrons, particles of charge* $+|e|$ *and the electron mass. Likewise, in addition to protons, there must be antiprotons. The experimental absence of such particles is then traced back to a special initial state in which only one of the two kinds of particles is present. We see that this already appears to be unsatisfactory because the laws of nature in this theory with respect to electrons and antielectrons are exactly symmetrical. Thus* γ*-ray photons (at least two in order to satisfy the laws of conservation of energy and momentum) must be able to transform, by themselves, into an electron and an antielectron. We do not believe, therefore, that this explanation can be seriously considered*" [11]. Since *electrons* and *positrons* are symmetrical in the view of nature, based on Dirac's argument, one could also assume that the vacuum is filled with a "*Dirac sea*" of infinite number of *negative-energy positrons*, and *electron* is the hole left behind when a negative-energy positron is excited. Thus, Dirac's hole theory would have its own trouble.

Such trouble can be totally avoided in our quantum wave model. In our derivation of the Dirac equation based on the wave view, there was no assumption of any pre-existing negative-energy electron sea in the vacuum. The Dirac equation only provides the quantum wave equation governing the movement of the excitation wave with ½ spin. The negative-energy solution means that the frequency has a negative sign, i.e., $\omega = -|\omega|$. That could just indicate that the wave packet is moving backward. This model does not require the assumption of a "Dirac sea" of negative-energy particles.

8.5 Chapter Summary

- Dirac's original derivation was very complicated and involved a number of conjectures. He regarded the electron as a point mass and developed his quantum theory on an analogy with classical mechanics. First, he identified the energy operator and the momentum operator, based on a hypothetical "***fundamental quantum condition***". Next, he deduced from quite general arguments that the relativistic quantum wave equation must be linear in the operator $\partial/\partial t$ and $\partial/\partial x_r$. Finally, he arrived at a linearized wave equation by assuming that the quantum wave function of an electron is a *spinor*.
- We show that the Dirac equation can be derived more naturally (and more easily) based on the quantum wave model. In fact, one can see that the Dirac equation is a direct consequence of the wave equation of the vacuum.

- Since the electron is a massive particle, according to our model, the Dirac equation should be a special case of the Klein–Gordon equation. Its wave function represents only the longitudinal component of the matter wave. In the case of an electron, its wave function can be represented by a 4-component column matrix. Then, it is easy to see that the Dirac equation is a linearized wave equation which can be obtained directly by factorizing the Klein–Gordon equation.
- A major advantage of the Dirac equation is that it naturally brought in the *spin* concept through the incorporation of Pauli's *spin matrices*.
- In current literature, the 4-component column matrix ψ is called the "*Dirac spinor*", which is made up of two sub-components: ψ_L is called the *left-hand spinor* and ψ_R is called the *right-hand spinor*. In our model, the wave functions of different types of particles are represented by different solutions of the vacuum wave equation (see Table 8.1). For example, the photon can be represented by a plane wave; the scalar particle can be represented by a cylindrically symmetrical vortex wave, while the electron can be represented by a 4-column matrix (*spinor*).
- The Dirac equation is credited with predicting the existence of "*anti-particle*". Such a "great discovery", however, involved some bold interpretations. Dirac assumed that the vacuum is filled with an infinite number of negative-energy electrons (called the "*Dirac sea*"). If one of these negative-energy electrons is knocked off by a gamma ray, the knock-off electron would gain energy to become a positive-energy electron; the unfilled vacancy—called a *hole*—would behave like a positively charged *anti-particle* of electron (called "*positron*").

References

1. Dirac, P. A. M. (1981). *The Principles of Quantum Mechanics* (4th ed.). Clarendon Press.
2. Dirac, P. A. M. (1928). The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences, 117*, 610.
3. Dirac, P. A. M. (1928). The quantum theory of the electron. Part II. *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences, 118*, 351–361.
4. Sakurai, J. J. (1973). *Advanced Quantum Mechanics.* Addison-Wesley.
5. Chang, D. C. (2021). Review on the physical basis of wave–particle duality: Conceptual connection between quantum mechanics and the Maxwell theory. *Modern Physics Letters B, 35*, 2130004.
6. Dirac, P. A. M. (1930). A theory of electrons and protons. *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences, 126*, 360–365.
7. Dirac, P. A. M. (1934). Discussion of the infinite distribution of electrons in the theory of the positron. *Mathematical Proceedings of the Cambridge Philosophical Society, 30*, 150–163.
8. Dodd, J. E. (1984). *The ideas of particle physics: An introduction for scientists.* Cambridge University Press.
9. Dirac, P. A. M. (1931). Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences, 133*, 60–72.
10. Anderson, C. D. (1933). The positive electron. *Physical Review, 43*, 491–494.
11. Pauli, W. (1933). Die allgemeinen Prinzipien der Wellenmechanik. In H. Bethe, F. Hund, N. F. Mott, W. Pauli, A. Rubinowicz, G. Wentzel, & A. Smekal, (eds.) *Handbuch der Physik. Quantentheorie*, (pp. 83–272) Springer Berlin Heidelberg.

Chapter 9
Derivation of the Schrödinger Equation: What is the Physical Meaning of Its Wave Function?

From the previous chapter, one can see that the quantum wave equation of the electron (the Dirac equation) is really a derivative of the wave equation of the vacuum, which describes the motion of the *electric vector potential* **Z**. Once the Dirac equation is derived, one can further derive the Schrödinger equation from the mechanism of vacuum excitation. It has been shown that, under the condition where the Coulombic energy and the kinetic energy of the electron are much smaller than the electron's resting energy (mc^2), the Dirac equation can be reduced into the Schrödinger equation [1]. Thus, the work shown in the above not only demonstrated that the Klein–Gordon equation and the Dirac equation can be derived from the wave equation of **Z**, it suggested that the Schrödinger equation can also be derived from the wave equation of **Z** as well.

From a historical perspective, however, the Schrödinger equation was derived before the Dirac equation. Thus, we will not try to derive the Schrödinger equation from the Dirac equation. Instead, we will trace the derivation of the Schrödinger equation based on the Klein–Gordon equation. Furthermore, since the wave function in the Dirac equation is a spinor, while the wave functions in the Schrödinger equation and Klein–Gordon equation are both scalar, it is more natural to connect the last two equations together. In this way, it is easier to see the physical meaning of the wave function.

9.1 Derivation of the Schrödinger Equation Based on the Quantum Wave Model

The Schrödinger equation is probably the most important equation in quantum mechanics. It is almost like the quantum counterpart of Newton's law in classical mechanics. The Schrödinger equation provides a way to calculate the wave function

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_9

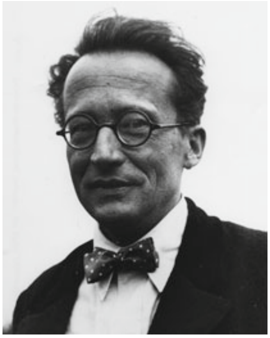


Fig. 9.1 Erwin Schrödinger. Erwin Schrödinger (1887–1961) was an Austrian physicist. He was famous for the development of the "Schrödinger equation" in quantum mechanics. He was awarded the Nobel Prize in Physics for this work in 1933, together with Paul Dirac. Photo Credit: Photograph by Francis Simon, courtesy of AIP Emilio Segrè Visual Archives

of a system and describe how it changes dynamically in time. The Schrödinger equation is used very widely nowadays. In fact, it is the major tool for most calculations conducted in atomic physics, molecular physics, quantum optics, condensed matter physics, etc. Most of the modern technologies used today can be attributed to the development of the Schrödinger equation.

Historically, the Schrödinger equation was derived by Erwin Schrödinger (See Fig. 9.1) based on his conjecture in a classical mechanical view, instead of from first principle [2–6]. According to Richard Feynman, the key step in Schrödinger's derivation of his equation was purely conjecture. "*Where did we get that from? Nowhere. It's not possible to derive it from anything you know. It came out of the mind of Schrödinger, invented in his struggle to find an understanding of the experimental observations of the real world*" [7].

In this chapter, we will show that the Schrödinger equation can be derived more naturally from the quantum wave model.

9.1.1 Development of the Correspondence Rules

In order to derive the Schrödinger equation, one needs to first introduce the "*correspondence rules*" (see Table 9.1). Historically, these rules were developed based on the Hamiltonian framework in classical mechanics [8, 9]. We, however, think it is more natural to develop the quantum wave equation based on the analogy between the wave mechanisms of radiation wave (photon) and matter wave (massive particle) [10]. In this work, we proposed that both the photon and massive particles are quantized excitation waves of the vacuum; their wave function along the trajectory path is in the form of

$$\psi_{\hat{k}} \propto e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}. \tag{9.1}$$

Based on the Planck's relation and the de Broglie relation, it can be shown that the energy (E) and momentum ($\boldsymbol{p}$) of the particle could be connected with the following operations on the wave function

$$\left(i\hbar\frac{\partial}{\partial t}\right)\psi_{\hat{k}} = \hbar\omega\psi_{\hat{k}} = E\psi_{\hat{k}}, \tag{9.2}$$

and

$$\left(\frac{\hbar}{i}\nabla\right)\psi_{\hat{k}} = \hbar\boldsymbol{k}\psi_{\hat{k}} = \boldsymbol{p}\,\psi_{\hat{k}}. \tag{9.3}$$

These relations suggest that, if one wants to associate the particle property E and $\boldsymbol{p}$ with operators on the wave function, the proper *correspondence rules* should be

$$E \to i\hbar\frac{\partial}{\partial t}. \tag{9.4}$$

$$\boldsymbol{p} \to \frac{\hbar}{i}\nabla. \tag{9.5}$$

Using these correspondence rules, one can develop the wave equation of a particle based on its energy–momentum relation. For example, one can easily derive the Klein–Gordon equation based on the energy–momentum relation of a massive particle, i.e.,

$$E^2 = p^2c^2 + m^2c^4. \tag{9.6}$$

By applying the correspondence rules, one can obtain

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\phi - \left(\frac{mc}{\hbar}\right)^2\phi = 0. \tag{9.7}$$

Here, one should note that the Klein–Gordon equation only describes the path component of the excitation wave. (See Chap. 7).

Table 9.1 Correspondence rules for both light wave and matter wave

Wave function	**Physical law**	**Operation on the wave function**	**Correspondence rule**
Light wave (photon) $\psi_{\hat{k}} \propto e^{i\,(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\, t)}$	$E = \hbar\,\omega$	$E\,\psi_{\hat{k}} = \left(i\hbar\frac{\partial}{\partial t}\right)\psi_{\hat{k}}$	$E \to i\hbar\frac{\partial}{\partial t}$
	$\boldsymbol{p} = \hbar\,\boldsymbol{k}$	$\boldsymbol{p}\,\psi_{\hat{k}} = \left(\frac{\hbar}{i}\nabla\right)\psi_{\hat{k}}$	$\boldsymbol{p} \to \frac{\hbar}{i}\nabla$
Matter wave (path component) $\psi_{path}\left(\hat{\boldsymbol{k}}\cdot\boldsymbol{x},\ t\right) \propto e^{i\,(\boldsymbol{k}\cdot\boldsymbol{x}-\omega\, t)}$	$E = \hbar\,\omega$	$E\,\psi_{path} = \left(i\hbar\frac{\partial}{\partial t}\right)\psi_{path}$	$E \to i\hbar\frac{\partial}{\partial t}$
	$\boldsymbol{p} = \hbar\,\boldsymbol{k}$	$\boldsymbol{p}\,\psi_{path} = \left(\frac{\hbar}{i}\nabla\right)\psi_{path}$	$\boldsymbol{p} \to \frac{\hbar}{i}\nabla$

9.1.2 *Construction of the Schrödinger Equation Based on the Klein–Gordon Equation*

Once people knew how to use the *correspondence rules* to derive the Klein–Gordon equation, one can derive the Schrödinger equation in a similar way. As pointed out earlier, the Klein–Gordon equation was based on the energy–momentum relation

$$E^2 = c^2\left(p^2 + m^2c^2\right). \tag{9.6}$$

For a free particle traveling at a speed far less than the speed of light, $v << c$, $p << mc$, the above relation becomes

$$E = mc^2\left(1 + \frac{p^2}{m^2c^2}\right)^{1/2} = mc^2\left(1 + \frac{1}{2}\frac{p^2}{m^2c^2} + \ldots\right) \approx mc^2 + \frac{p^2}{2m}. \tag{9.8}$$

Using the *correspondence rules*, one can obtain the wave equation for the particle, i.e.,

$$i\hbar\frac{\partial\phi}{\partial t} = \left(mc^2 + \frac{1}{2m}(-i\hbar\nabla)^2\right)\phi. \tag{9.9}$$

This was the wave equation developed by Schrödinger for the electron. Here, the quantum wave function (ϕ) is identical to the wave function of the Klein–Gordon equation.

Now, what happens if the electron experiences an external electric field? Suppose the electron is inside an atom, where there is a negative electrical potential $V = V(r)$. The electron has an electric charge q; its energy is shifted by an amount $-qV$, that is,

$$\text{Energy of electron} \to E - qV \to i\hbar\frac{\partial}{\partial t} - qV.$$

Equation (9.9) will then become

$$\left(i\hbar\frac{\partial}{\partial t} - qV\right)\phi = \left(mc^2 - \frac{\hbar^2\nabla^2}{2m}\right)\phi. \tag{9.10}$$

This equation can be further simplified by defining a new wave function,

$$\psi_s = e^{imc^2t/\hbar}\phi. \tag{9.11}$$

By substituting Eq. (9.11) into Eq. (9.10), the mc^2 term can be canceled out. Then, we have

$$i\hbar\frac{\partial\psi_s}{\partial t} = \left(-\frac{\hbar^2}{2m}\nabla^2 + qV\right)\psi_s. \tag{9.12}$$

This is the *Schrödinger equation* for an electron in the presence of an electric potential.

9.2 Physical Meaning of the Quantum Wave Function of the Schrödinger Equation

From the above derivation, one can easily see the physical meaning of the quantum wave function of the Schrödinger equation. Its wave function (ψ_s) is directly related to the wave function of the Klein–Gordon equation (ϕ); they differ only by a phase factor as shown in Eq. (9.11). Since ϕ represents only the longitudinal component of the matter wave, i.e., $\phi = \psi_{path}$, the quantum wave function of the Schrödinger equation (ψ_s) must also be associated with the longitudinal component of the matter wave (ψ_{path}), which describes the traveling wave of the electron.

From the above discussions, we see that all known quantum wave equations can be derived based on the *wave equation of the vacuum*. In other words, particles of different types (i.e., photons, scalar particles, and electrons) are all excitation waves of the same vacuum medium; all quantum wave functions are associated with the *longitudinal component* (i.e., path component) of the *vector potential* **Z** (see Table 9.2).

9.2.1 All Quantum Wave Equations Can Be Traced to the Wave Equation of the Vacuum

In this work, we showed that in both the mechanical medium and the vacuum medium, the wave equations always appear as a 4-D Laplace equation. We call this equation the "*Wave Equation of the Vacuum*" (WEV) where the wave medium is the vacuum. We think all free particles are excitation waves of the vacuum medium, and

Table 9.2 Physical meaning of the wave functions for different types of particles

Excitation mode (particle type)	**Wave equation**	**Wave function**	**Physical meaning of the wave function**
Excitation wave of the vacuum	Basic wave equation of the vacuum	$\boldsymbol{Z}(\boldsymbol{x}, t) = \boldsymbol{\epsilon_k}\, \psi_{\hat{\boldsymbol{k}}}(\boldsymbol{x}, t)$, $\psi_{\hat{\boldsymbol{k}}}(\boldsymbol{x}, t) = \psi_T\left(\hat{\boldsymbol{k}} \times \boldsymbol{x}\right)\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)$	Propagating wave of the *electric vector potential* (**Z**) in the vacuum
Photon	Wave equation of light	$\psi_{\hat{\boldsymbol{k}}}(\boldsymbol{x}, t) = \psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right) \propto e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}$	Plane wave solution of **Z**
Massive particle	Klein–Gordon equation	$\phi_{KG}(\boldsymbol{x}, t) = \psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)$	Longitudinal component of the wave of **Z** (in scalar form)
Electron (in scalar form)	Schrödinger equation	$\psi_{Schrodinger}(\boldsymbol{x}, t) = e^{imc^2t/\hbar}\psi_{path}\left(\hat{\boldsymbol{k}} \cdot \boldsymbol{x}, t\right)$	
Electron (in spinor form)	Dirac equation	$\psi_{Dirac}(\boldsymbol{x}, t) = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$	Longitudinal component of the wave of **Z** (in spinor form)

different particles are just different excitation modes. If the matter wave is described by the WEV, then different solutions of the WEV would represent different particles. These particles could have different mass or spin, but their wave functions should all satisfy the WEV. In the conventional theories, different particles are thought to satisfy different quantum wave equations. For example, the photon is supposed to satisfy the wave equation of light, the scalar particle is supposed to satisfy the Klein–Gordon equation, and the electrons are supposed to satisfy the Dirac equation (or the Schrödinger equation). In order to reconcile our matter wave model with the conventional quantum theories, one must demonstrate that all conventional quantum equations for different particles are derivable from the WEV. In Chaps. 7–9 of this book, we showed that this expectation can indeed be fulfilled. In other words, one could say that *the WEV is the mother of all quantum wave equations*. To demonstrate this point more clearly, we have summarized the relationship between the wave equations and wave functions of different types of free particles in Table 9.3.

Table 9.3 Physical properties of the wave function for different free particles

Particle type	Rest mass	Spin	Mathematical form of ψ	Nature of ψ	Wave equation
Photon	No	1	Vector	A special solution of WEV	Wave equation of light (by Maxwell)
Scalar particle	Yes	0	Scalar	A special solution of WEV	Klein–Gordon equation
Lepton/Electron (relativistic)	Yes	½	Spinor	A special solution of WEV	Dirac equation
Electron (non-relativistic)	Yes	½	Scalar	An approximate solution of WEV	Schrödinger equation

9.3 Transition from Classical Physics to Quantum Mechanics: The Mechanical View *Versus* the Wave View

It is generally believed among physicists that classical mechanics is a limiting case of quantum mechanics. So, one should be able to derive the equation of motion in quantum mechanics by extending what we know from classical physics. There can be two different conceptual approaches for deriving the quantum wave equation:

(1) ***The mechanical approach***. Here, the physical system is treated as a dynamic system composed of many mechanical particles, with their positions and momentums specified by q's and p's. The formalism of quantum mechanics was developed based on classical Newtonian mechanics, particularly the Lagrange's and Hamilton's form of classical analytical mechanics.
(2) ***The wave approach***. Here, the system is treated as excitation waves in the vacuum. The formalism is wave mechanics based on the Maxwell theory.

Historically, the *mechanical approach* was the mainstream view in quantum mechanics. However, this approach had great difficulties to explain certain quantum phenomena [11]. This is because the behavior of matter in the microscopic quantum system is different from the macroscopic mechanical world. For example, a particle's position coordinate q and its conjugating momentum p cannot be accurately defined at the same time. Also, the mathematical procedures in connecting classical mechanics to quantum mechanics are often complicated and must involve bold assumptions [9].

In this work, we proposed to use the *wave approach* to derive the quantum wave equations. Here, we showed that the transition from classical physics to quantum physics can be based on a conceptual unification between light wave and matter wave. From the Maxwell theory and Helmholtz decomposition, one can derive not only the wave equation of photons, but also quantum wave equations for massive particles (e.g., electrons). The only assumption here is that both the matter wave and the radiation wave are excitation waves of the quantum vacuum [12].

In this approach, the quantum conditions come from the understanding that, like the photon, massive particles also obey the Planck's relation and the de Broglie relation. For example, one can show $p \to -i\hbar\ \partial/\partial x$ when it operates on a traveling wave in the form of $e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}$. This *correspondence rule* simply reflects the de Broglie relation $\boldsymbol{p} = \hbar\boldsymbol{k}$. Similarly, the *correspondence rule* $E \rightarrow i\hbar\ \partial/\partial t$ is mainly a reflection of the Planck's relation.

There are further advantages in choosing the wave approach over the mechanical approach. For example, one can explain more easily the physical basis of particle creation/annihilation. If particles are excitation waves of the vacuum, the vacuum will be free of particles when it is at its resting state. Then, when the vacuum is excited with an energy stimulus, particles will be created due to the generation of new waves. Similarly, in the process of wave-wave interactions, some waves may be destroyed in order to create new waves. This may explain why particles can be converted into different types during interactions.

By comparison, it is far more difficult to explain particle creation/annihilation using the mechanical approach. In Dirac's theory of the electron, the creation of an electron was based on an assumption that the vacuum is filled with a sea of infinite number of negative-energy electrons [13]. When the system is stimulated with an energy input (such as the absorption of a γ ray), a negative-energy electron will gain energy and escape from the Dirac's sea to become a positive-energy electron, while the hole left behind becomes the anti-particle (positron). Although this hypothesis could explain pair creation of electron/positron, its assumption is difficult to be justified; there is no experimental evidence showing that the vacuum is filled with an infinite number of negative-energy particles at the resting state [14].

9.4 Chapter Summary

- Historically, the Schrödinger equation was derived by Erwin Schrödinger based on his conjecture in a classical mechanical view. It was very difficult to explain the origin of the Schrödinger equation based on first principle. In this chapter, we showed that the Schrödinger equation can be derived more naturally from the quantum wave model.
- In order to derive the Schrödinger equation, one needs to first introduce the "*correspondence rules*" (see Table 9.1). Historically, these rules were developed based on the Hamiltonian framework in classical mechanics. In this work, we proposed that both the photon and massive particles are quantized excitation waves of the vacuum; they obey the Planck's relation and the de Broglie relation. Then, one can show that the *correspondence rules* simply reflect the operation of the *de Broglie relation* and the *Planck's relation* on a traveling wave.
- Once one knows how to develop the *correspondence rules*, one can derive the Schrödinger equation by following the example of using the *correspondence rules* to derive the Klein–Gordon equation.
- The physical meaning of the quantum wave function of the Schrödinger equation is similar to the wave function of the Klein–Gordon equation; they differ only by a phase factor as shown in Eq. (9.11). In our model, all known quantum wave equations can be derived based on the *wave equation of the vacuum*. Particles

of different types (i.e., photons, scalar particles, and electrons) are all excitation waves of the same vacuum medium; all quantum wave functions are associated with the *longitudinal component* (i.e., path component) of the *electric vector potential* **Z** (see Table 9.2).

References

1. Cottingham, W. N., & Greenwood, D. A. (1998). *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press.
2. Schrödinger, E. (1926). Quantisierung als Eigenwertproblem (The first paper in this series). *Annalen der Physik, 384*, 361–376.
3. Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. *Annalen der Physik, 384*, 489–527.
4. Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. *Annalen der Physik, 385*, 437–490.
5. Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. *Annalen der Physik, 386*, 109–139.
6. Schrödinger, E. (1926). An undulatory theory of the mechanics of atoms and molecules. *Physical Review, 28*, 1049–1070.
7. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Chapter 16 The dependence of amplitudes on position. In *The Feynman Lectures on Physics*, (vol. 3). Addison-Wesley.
8. Messiah, A. (1965). *Quantum Mechanics* (Vol. 1). Wiley.
9. Dirac, P. A. M. (1981). *The Principles of Quantum Mechanics* (4th ed.). Clarendon Press.
10. Chang, D. C. (2017). *On the wave nature of matter: A transition from classical mechanics to quantum mechanics*. https://doi.org/10.48550/arXiv.physics/0505010.
11. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Chapter 37 Quantum behavior. In *The Feynman Lectures on Physics*, vol. 1. Addison-Wesley.
12. Chang, D. C. (2021). Review on the physical basis of wave–particle duality: Conceptual connection between quantum mechanics and the Maxwell theory. *Modern Physics Letters B, 35*, 2130004.
13. Dirac, P. A. M. (1930). A theory of electrons and protons. *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences, 126*, 360–365.
14. Dirac, P. A. M. (1934). Discussion of the infinite distribution of electrons in the theory of the positron. *Mathematical Proceedings of the Cambridge Philosophical Society, 30*, 150–163.

Chapter 10
A New Understanding on Wave-Particle Duality: Comparing the Quantum Wave Model with the Copenhagen Interpretation and Other Alternative Models

During the development of quantum mechanics, there was a dilemma in interpreting the physical meaning of the quantum wave function. In the mind of the traditional physicists, the electron is a physical *particle* (like a point mass); the quantum wave equation, on the other hand, can only describe the motion of a *wave* (matter wave). Thus, the physical identity of the two does not match! How to explain this dilemma becomes a big problem.

One major idea to resolve this problem was the *Copenhagen interpretation*. Niels Bohr and his colleagues (including Heisenberg and Pauli) thought that the connection between the concepts of "particle" and "matter wave" can be through the statistical interpretation. That is, the electron would remain behaving as a classical particle (like a point mass), while the matter wave would give the probability of finding the electron at a particular position in space–time.

Other leading scientists, including Einstein, Schrödinger, and de Broglie, however, did not agree with this statistical interpretation. They thought the matter wave should be a real physical wave. In fact, Einstein had several active debates with Bohr on this topic during the 1920s [1]. But neither man was able to convince the other.

Subsequently, many physicists tried to propose alternative models to replace the Copenhagen interpretation; these include the *pilot wave theory* [2, 3] and the *many-world theory* [4, 5]. These alternative theories, however, could not compete with the statistical interpretation proposed by Bohr. Today, the Copenhagen interpretation is still the leading theory for interpreting the quantum phenomenon of wave-particle duality.

Now, with the development of the *quantum wave model* discussed in this book, we will show that one can have a better understanding of the wave-particle duality in comparison to the Copenhagen interpretation and its alternative models.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_10

10.1 Bohr's Statistical Interpretation Can Be Explained by the Quantum Wave Model

Based on the understanding of our quantum wave model, one may say that both Einstein and Bohr could be partially right. From a philosophical point of view, Einstein's thinking was in the right direction; he knew intuitively that the matter wave must be a physical wave. However, from a technical point of view, Bohr's proposal was not unreasonable; his statistical interpretation can be quite useful for comparing quantum calculation with experimental results. In fact, it is possible to explain the statistical interpretation based on the quantum wave model presented in this book.

10.1.1 Why Can a Physical Wave Function Give the Probability of Detecting the Quantum Particle During Its Measurement? A Case Study Using the Photon as an Example

Let us first use a photon to demonstrate the above point. When a photon travels in the vacuum, it is in the form of a wave packet. As we pointed out earlier, the size of this wave packet is much larger than an atom (e.g., the size of the photon of visible light could be 10^4 times bigger than an atom).[1] When one conducts an experiment to detect a photon, the photon can simultaneously interact with many atoms within the detector. However, only a single atom among them can absorb the entire quantum energy of the photon. This is due to the *principle of all-or-none* during the photon-absorption process [6]. (More specifically, the photon is absorbed by one single electron within this "lucky" atom).

Therefore, there is a sudden change of energy distribution for the photon during the measurement process. Before the detection, the photon travels in the form of a wave packet; its energy is distributed over a relatively large volume (as large as the size of the wave packet). But as soon as the photon hits the detector, the entire quantum energy of the wave packet will suddenly be transferred into a single sub-atomic electron. In another word, the wave packet of the photon suddenly *collapses* (into a sub-atomic size) during the measurement process (see Fig. 10.1). The photon is now appearing as a small dot on the screen, and thus can be interpreted as a tiny particle.

To demonstrate this conceptual interchange of wave and particle during the measurement process, we can use a well-known magical story as an analogy. In a

[1] For example, a photon of visible light has a wavelength of about 0.5 μm. Since the photon is a wave packet, its width must be many times of its wavelength. Suppose the wave packet contains about 10 oscillating cycles, the wave packet will be about 0.005 mm long. Such a photon is about 10^4 times bigger than an atom.

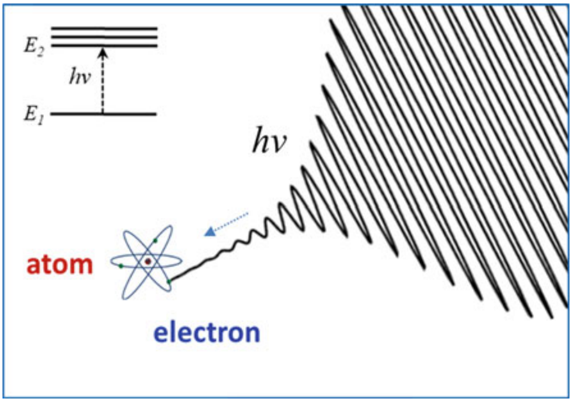


Fig. 10.1 Collapse of the quantum wave function during particle absorption. The free particle (photon) is an excitation wave of the vacuum; it travels as a wave packet. The size of this wave packet is far larger than an atom. During measurement, the entire photon is absorbed by a single electron inside the target atom. Thus, the energy of the wave packet is suddenly concentrated into a very small volume (of an electron) during the absorption process. In another word, the quantum wave function of the photon appears to collapse during a measurement, where the photon energy $h\nu$ is transferred to the orbital electron to allow its energy level to jump from E_1 to E_2. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Mod. Phys. Lett. B* **35**, 2130004 (2021)

famous novel called "*One Thousand and One Nights*", there is a story about Aladdin and the magic lamp. Aladdin accidentally got a magic lamp, in which lived a magical genie. When Aladdin wiped the lamp, the genie would emerge from the lamp and appear as a giant (see Fig. 10.2). The giant genie is omnipotent; no matter what Aladdin wishes; the genie can help him to accomplish it. After the genie completed the requested task, he would shrink back into the lamp at once. The size of the returning genie is now much smaller than the lamp.

In the physical world, the behavior of a photon is almost like the magic genie in the above story. When a photon is emitted from an atom, the wave packet of the photon is about ten thousand times larger than the atom. When the photon is absorbed by an atom, all its energy is instantly transferred to an electron inside the atom. This process is like that the entire wave packet of the photon would suddenly collapse into a small volume. This is similar to the situation of the genie in the Aladdin story when it shrinks back into the lamp (see Fig. 10.2).

Certainly, such magic performance of the photon is not due to the "magic power" as depicted in the Aladdin's lamp story. It is really due to the "*principle of all-or-none*" during energy transfer in the quantum world. That is, based on Planck's discovery, light emission or absorption is in "quanta"; either the entire package of energy of the photon is transferred (between the atom and the vacuum), or no energy is transferred at all.

By the way, there is one major difference between the photon emission/absorption and the magic lamp of Aladdin. In the Aladdin story, there was only one magic lamp;

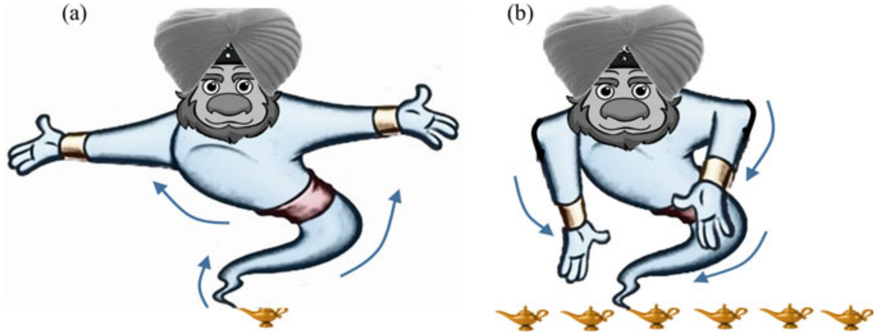


Fig. 10.2 The analogy of photons being absorbed by electrons and the magic lamp story. The process of photons emission/absorption can be illustrated using an analogy from the story of Aladdin's magic lamp. **a** When a photon is outside of an atom, its wave packet is hundreds of thousands of times larger than that of the atom. This situation is like the genie in the story of Aladdin's lamp. When this genie emerged from the lamp, he was a giant and his size was far bigger than the lamp. **b** But when the genie finished the task and returned to the lamp, he would suddenly shrink to a small size and become a part of the lamp. It is a process of collapse, which is analogous to the process of photon absorption by an atom. When the photon is absorbed by the atom, all its energy is instantly transferred to one of the electrons inside the atom. It's as if the photon's wave packet collapsed into a tiny volume in an instant

the genie could come out from this lamp and later return to the same lamp. But in the photon emission/absorption case, there are many atoms in the quantum system, each of which is capable of emitting a photon. The emitted photon does not need to return to the original atom; it can be absorbed by any atom along its pathway.

Because the photon wave packet is very large, many atoms inside the detector can interact with it. All of them will have a chance to absorb the photon. Which one can be the lucky atom? This involves a probability during the absorption process.

It is not difficult to see that the chance of an atom to absorb the photon is dependent on the relative position between the photon and the atom. For example, those atoms locating at the center of the photon wave packet would have a better chance of absorbing the photon. For those atoms located at the periphery of the wave packet, they would have less chance to absorb the photon. What is the factor that determines the probability of detecting the photon? It is most likely to be the energy carried by the excitation wave! In another word, those atoms that expose to the largest energy density of the wave packet would have a better chance to absorb the photon.

It is not difficult to calculate the energy distribution of a photon wave packet. From the Maxwell theory, we know the radiation energy is proportional to the square of the electric field. Earlier, we have shown that **E** is proportional to **Z**, the amplitude of which is represented by the wave function ψ. Thus, the probability of detecting the photon is proportional to the square of the wave function, i.e.,

Probability of detecting the photon at $(\boldsymbol{x}, t) \sim |\boldsymbol{Z}(\boldsymbol{x}, t)|^2 = |\psi(\boldsymbol{x}, t)|^2$

This relation is similar to the "*Born rule*" used in the Copenhagen interpretation, which proposed that the probability of detecting a quantum particle is proportional to the amplitude square of the wave function.

10.1.2 Similarly, the Probability of Detecting an Electron at a Particular Location Is also Related to the Amplitude of the Electron's Wave Function

From the above discussion, it is easy to see that a similar argument can be applied to an electron. According to our quantum wave model, the free electron is an excitation wave of the vacuum, just like the photon. When the free electron is traveling in the vacuum, it is in the form of a wave packet, the size of which is often bigger than an atom. For example, the electrons used in Davisson's diffraction experiment had an energy in the order of 100 electron-volts. The wavelength of such an electron is about 1.2×10^{-8} cm. Since the wave packet of the free electron contains many cycles of oscillations, the length of the wave packet should be many times of the wavelength. Thus, the size of the electron wave packet is likely to be larger than 10^{-7} cm. This would allow the free electron to interact simultaneously with many atoms in the detector.

Once the incoming electron is captured by an atom in the detector, the electron wave packet will suddenly collapse and transfer its entire energy to the target atom. Due to the principle of *all-or-none*, the entire electron can only interact with a single atom within the detector.

When one tries to measure the position of an incoming electron using a detector, the chance of detecting the electron by a particular atom is proportional to the energy density of the electron wave packet at the position of the target atom. According to our quantum wave model, the free electron is an excitation wave of the vacuum. The energy of the matter wave is proportional to the amplitude square of the electric vector potential, **Z**, which is represented by the wave function ψ. Thus, the probability of detecting the electron is proportional to the square of the wave function, i.e.,

$$\textbf{Probability of detecting the electron} \text{ at } (\boldsymbol{x}, t) \sim |\mathbf{Z}(\boldsymbol{x}, t)|^2 = |\psi(\boldsymbol{x}, t)|^2$$

This relation is in good agreement with the "*Born rule*" used in the Copenhagen interpretation (see Fig. 10.3).

Hence, from a technical point of view, the statistical interpretation proposed by Bohr could be partially justified. However, its classical view of treating the electron as a point-mass-like particle is not correct. As we showed in the earlier chapters, the electron is an excitation wave of the vacuum; it is a physical wave. The collapse of the wave function during measurement is really due to the fact that, the absorption process of an electron wave must obey the *principle of all-or-none* (i.e., an atom can absorb either one whole electron or no electron at all).

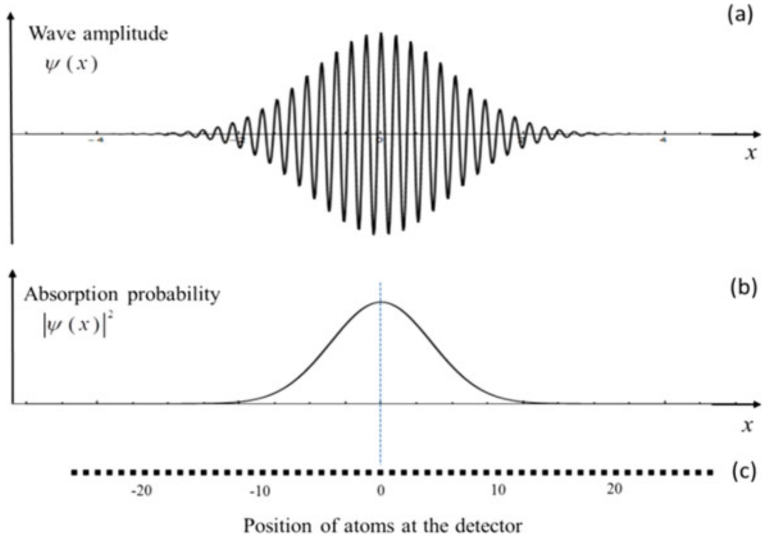


Fig. 10.3 The relationship between the wave function and the probability of detecting the particle. The wave packet of a free electron has a width much larger than an atom. So, before the electron is detected by the detector, it is widely distributed. Only when this electron is absorbed by an atom in the detector that the electron wave packet will collapse into a single atom. The probability for an atom at the x position to absorb the electron is proportional to the square of the wave function of the electron at the x position. **a** $\psi(x)$ is the wave function of the electron; **b** the absorption probability of the electron is proportional to $|\psi(x)|^2$; **c** the position of atoms within the detector; only one of these atoms can absorb the entire electron. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Mod. Phys. Lett. B* **35**, 2130004 (2021)

10.2 The Statistical Interpretation Does Not Work for the Electron Wave Function Inside an Atom

From the above discussion, one can see that, **for an electron traveling outside of the atom, the quantum wave function can be related to the probability of finding the electron** at a particular space and time. **The situation, however, is very different for an electron inside an atom**. According to the quantum wave model, the electron is a quantized wave packet; it is not a point mass. For an intra-atomic electron, the size of the electron wave packet is comparable to the size of the atom (see below). Thus, an intra-atomic electron could fill most of the space inside an atom. Under this situation, it is not possible to specify the exact location of an electron within the atom.

In the traditional literature, it is often stated that the "*electron cloud*" calculated based on quantum mechanics within an atom represents the *probability* of finding an electron at an infinitesimal element of space surrounding a given point. This view is a misunderstanding. The Copenhagen interpretation cannot be applied for an intra-atomic electron due to the following reasons:

(1) The electron is not a point-mass-like object; it is a wave packet, the size of which is comparable to the atom. Thus, the electron cannot be localized in "an infinitesimal element of space" within an atom.
(2) There is no physical way to detect the presence of "an electron located within an infinitesimal element of space within an atom". Such measurement is impossible.
(3) And because there is no measurement process to detect the location of an electron inside an atom, there is no "wave function collapse" (as observed when an external electron is absorbed by an atom).

Therefore, our interpretation based on the quantum wave model is different from the traditional view. We think **it is more reasonable to interpret the *electron cloud* as representing the charge distribution (of the electron) inside the atom** instead of a probability of finding a point-mass-like electron at a particular infinitesimal element of space.

How does one know that the size of an intra-atomic electron is comparable to the size of an atom? First, one can estimate the longitudinal wavelength of an electron (λ_L) from the Bohr's atomic model of hydrogen,

$$2\pi r = n\lambda_L, \tag{10.1}$$

where $n = 1, 2, 3, \ldots$ is a quantum number, r is the radius of the electron orbit within the atom. It can be shown from the Bohr's model that [7]

$$r = n^2 a, \tag{10.2}$$

where $$a = \frac{h^2(4\pi\varepsilon_0)}{4\pi^2 m e^2} = 0.53 \times 10^{-10}\text{m}.$$

Thus, even for the most inner orbital of the hydrogen electron, its radius is already comparable to the size of the hydrogen atom (its diameter is about 10^{-10} m).

From Eq. (10.1), one can see that the longitudinal wavelength of the orbital electron is also comparable to the atomic size. Since the width of the electron wave packet is usually many times of its wavelength, one can estimate that the length of the wave packet representing the orbital electron is far greater than the length of the atomic orbit.

Furthermore, the quantum wave function of the electron inside an atom (such as hydrogen) can be calculated directly using the Schrödinger equation. The results are well known [8–10]. The solution of the Schrödinger equation in the hydrogen atom (in polar coordinates) is given as

$$\psi(r, \theta, \varphi) = R_{nl}(r) Y_l^m(\theta, \varphi),$$

where R_{nl} is the radial distribution function, $Y_l^m(\theta, \varphi)$ is the spherical harmonics, which is made up of

$$Y_l^m(\theta,\varphi) = \left[\frac{2l+1}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}\right]^{1/2} P_l^{|m|}(\cos\theta)e^{im\varphi},$$

where $P_l^{|m|}(\cos\theta)$ is the *associated Legendre function.* The indices *n, l,* and *m* are quantum numbers that characterize the state of the electron; *n* is the *principal quantum number,* which must be a non-zero integer (1, 2, 3, ...); *l* is the *angular momentum quantum number*, which is any integer, 0, 1, 2, ..., $n - 1$; and *m* is the *magnetic quantum number*, which ranges from $-l$, $-l + 1$, $-l + 2$,..., $+ l$. Figure 10.4 is a plot of the various quantum wave functions of electrons in a hydrogen atom based on the solutions of the Schrödinger wave equation. These wave functions are generally referred to as the "*electron cloud*" of the atom.

From the above analysis, it is clear that the quantum wave function representing an intra-atomic electron is widely distributed. In fact, it can be shown from the solution of the Schrödinger equation that the size of the electron wave function is comparable to the size of an atom. According to the quantum wave model, the electron is a quantized wave packet; its wave function represents the dynamic variation of the vacuum medium as characterized by the electric vector potential **Z**. Thus, one should interpret the electron cloud as a physical distribution of charge density of the

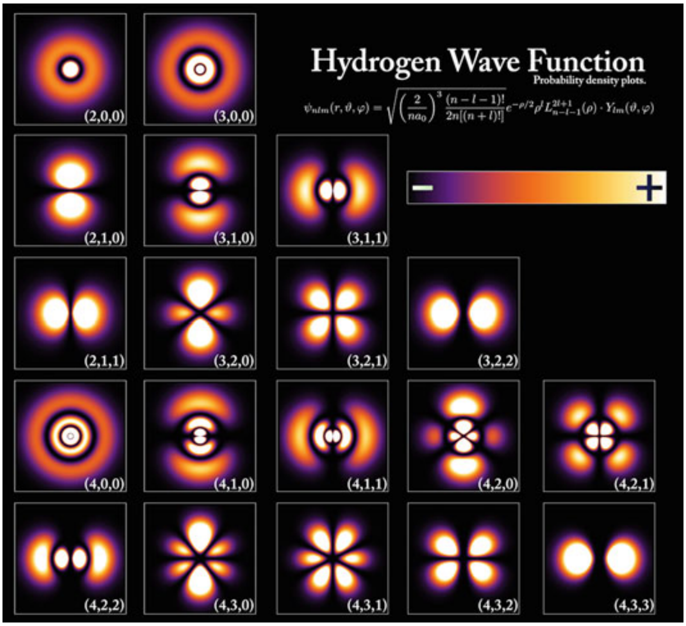


Fig. 10.4 Plots of some quantum wave functions of electrons in a hydrogen atom based on the solution of the Schrödinger equation. Each plot represents the "electron cloud" of a hydrogen atom. The three numbers in each parenthesis separated by commas are the quantum numbers *n*, *l*, and *m* respectively, which characterize the state of the electron. Photo Credit: PoorLeno, Wikimedia commons; Public domain

electron instead of representing the probability of finding a point-mass-like electron at a particular location.

In fact, this new interpretation of the electron cloud has several advantages. First, it can provide a physical basis to explain why two adjacent atoms cannot be pushed too close together (i.e., the Van der Waals force). Second, this can also explain the formation of molecular bonding between atoms. When two atoms form a bond between them, it is due to their sharing of the physical electronic wave functions. If their wave functions only represent probabilities, it would be difficult to explain the physical mechanism behind the bonding.

10.3 Controversy About the Different Interpretations of Quantum Mechanics

10.3.1 Skepticism About the Copenhagen Interpretation

The Copenhagen interpretation has long been regarded as the orthodox explanation of quantum mechanics. It contains at least 4 major assumptions:

(1) The quantum system is non-deterministic, instead, the quantum state of the system is only a superposition of multiple possible physical states.
(2) The calculation based on the quantum wave equation only gives the probability of finding the particle at a particular physical state.
(3) Before a measurement is made, the physical state of the system is unknown. When the measurement is made, the system will collapse into one of the possible physical states.
(4) The probability of finding the quantum particle in a particular physical state is governed by the *Born rule* (i.e., the probability of finding the particle in a particular state is proportional to the square of the amplitude of the wave function).

These assumptions need to be justified. In the above, we have pointed out that the points (2) and (4) of the Copenhagen interpretation can be easily explained using our quantum wave model. The points (1) and (3), however, are based not so much on the understanding of physics, but more on the philosophy of the physicists in the Copenhagen School. Some prominent physicists had expressed doubts about the probabilistic world implied by this Copenhagen interpretation. They believed that matter waves (the wave function in the Schrödinger equation) should be a real wave with physical properties, not just an elusive probability.

Einstein was one of the most well-known critics of the Copenhagen interpretation. He had always been skeptical of the concept of probability waves; he had a famous saying "*God does not throw dice*" to dramatize his view. At two well-known *Solvay conferences* (in 1927 and 1930), he debated this issue vigorously with Bohr and others. According to Pauli's recollection, Einstein once said: "*One cannot build a* *theory on many 'possibles'. Even if the theory is empirically and logically right, it is actually wrong*". Heisenberg recalled that every morning at breakfast, Einstein would come up with a thought experiment to challenge Bohr's quantum theory. By dinner time, Bohr and his collaborators would have presented counter-arguments and got Einstein to throw in the towel. The next day, Einstein would come up with new thought experiments to continue the challenge. In this cycle, the two sides debated for several days. But Einstein ultimately failed to convince Bohr [11].

Although Einstein could not win the debate with Bohr, he always believed that the statistical interpretation of quantum wave functions would not stand the test of time. Later at the Solvay Conference in 1930, Einstein engaged in a new round of debate with Bohr more openly at the conference. Although Einstein failed again, he was still not convinced [11].

Another well-known physicist, Schrödinger, was also not satisfied with the statistical probability explanation advocated by the Copenhagen School. As the inventor of the quantum wave equation for electrons, he has always been interested in what the physical nature of matter waves is. In the 1930s, he had several exchanges with Einstein on this topic. Inspired by Einstein, Schrödinger proposed a "thought experiment" to illustrate that the Copenhagen interpretation of the micro-world situation would contradict the macroscopic world observation [12]. Supposed that a cat is kept in a box containing a highly toxic gas; a control system in the box decides whether or not the gas is to be released. The switching of this control system depends on a quantum phenomenon (when a sample of isotopes decay in the box) (see Fig. 10.5). During this experiment, if the observer did not open the box, it was impossible for him to know whether the gas was being released, and whether the cat was alive or killed by the gas. According to the Copenhagen interpretation, the state of the cat during the experiment is the superposition of the probabilities of the two states. It was as if the cat died at the same time and was alive at the same time. This cannot be true. When Schrödinger told Einstein about the idea, Einstein liked it a lot. Einstein called this problem "**Schrödinger's cat**", and thought the example pointed out the paradox of using the Copenhagen interpretation to explain quantum mechanics.

Of course, other mainstream physicists who supported the Copenhagen interpretation could give a counter-argument. They said the outcome will only be known when the final measurement is made. Only then does the wave function collapse into one of several possible states. Therefore, it is meaningless to discuss the results before taking the measurements.

Quantum mechanics has been continuously developed and applied in the past century; its success is beyond doubt. But controversy over the interpretation of quantum wave functions has not stopped. Even today, many scientists are still not satisfied with the orthodox interpretation of quantum mechanics as proposed by the Copenhagen School. Particularly, many physicists are skeptical about Bohr's assumptions of "*superposition of states*" and "*the collapse of wave function upon measurement*". Thus, they tried to propose alternative interpretations of quantum mechanics. Among them, the most well-known ones are the *pilot wave theory* proposed by de Broglie and Bohm, and the *many-world interpretation* proposed by Everett.

Fig. 10.5 Schrödinger's cat. Suppose a cat is kept in a box containing poisonous gas stored in a glass bottle. There is a control mechanism in the box that decides whether to break the glass bottle or not. This control mechanism is triggered by a random event determined by the decay of a sample of radioactive isotope. Before the box is opened, it is impossible to know whether the gas was released or not, and thus, it is not certain whether the cat was alive or dead

10.3.2 The Many-World Interpretation of QM

A major dissatisfaction with Copenhagen interpretation is that before a measurement is made, the system is supposed to be a superposition of multiple quantum states, which are not reality. Now, when a measurement is made, the system suddenly collapses into one of the quantum states which now becomes real (i.e., observable). So there are two questions concerning this aspect of the Copenhagen interpretation: (1) Are the unobservable quantum states before measurement physically meaningful? (2) When the system collapses into one specific quantum state upon the measurement, what happen to the other quantum states? Where do they go?

In the 1950s, a student of John Wheeler at the Princeton University, Hugh Everett III, proposed an idea to solve the above two problems in his Ph.D. thesis, entitled "Relative State Formulation of Quantum Mechanics" [5]. He proposed that the entire universe is described by a gigantic wave function that contains within it an infinite number of increasingly divergent, non-communicating parallel *quantum worlds*. The universal wave function is objectively real, and all possible outcomes of quantum measurements are physically realized in some worlds. When an observer makes a measurement, he would only observe one of the possible physical results. The other possible quantum states are still there; they did not disappear upon measurement. They just go on with their own destiny in different worlds. Therefore, the wave function did not collapse upon measurement, it just made a choice of which world to go to.

The Everett's interpretation implies that there are an infinite number of worlds coexisting at the same time. It is one of many multiverse hypotheses in physics and philosophy. The theory proposed by Everett was later renamed the *Many-World Interpretation* (MWI) by Bryce DeWitt, who was responsible for making this interpretation of quantum mechanics more popular in the physics community.

MWI views time as a many-branched tree, wherein every possible quantum outcome can be realized. The observer, of course, can observe only one of the outcomes in his chosen world. This interpretation is aimed to resolve the measurement problem of the Copenhagen interpretation, such as the example of Schrödinger's cat. The MWI would argue that in some worlds, the cat was alive, while in other worlds, the cat was dead. When the observer opens the box, he would discover the cat is either alive or dead, depending on which world he is going to branch into.

When Everett presented his theory in 1956, he was totally rejected or ignored by the mainstream physicists [4, 13]. Subsequently, Everett left academia in 1956 and John Wheeler was not enthusiastic in promoting MWI at that time. However, in the 1970s, MWI suddenly became more popular owing to the effort of DeWitt. In addition, several influential physicists, including David Deutsch and Sean Carroll, also helped to actively promote the MWI. As a result, MWI became more known in later days. There was indication that MWI is now the second-most popular interpretation of quantum mechanics, just behind the Copenhagen interpretation [14].

10.3.3 The Pilot Wave Theory

Another alternative interpretation of quantum mechanics is the so-called "pilot wave theory", which was originally developed in the 1920s by de Broglie. He presented this pilot wave theory at the 1927 Solvay Conference but was criticized by Wolfgang Pauli. After that, de Broglie gave up this theory. Later, David Bohm, dissatisfied with the prevailing orthodoxy of the Copenhagen interpretation, rediscovered de Broglie's pilot wave theory in 1952 [2, 3].

The pilot wave theory was developed to avoid the "measurement problem" found in the Copenhagen interpretation. In the pilot wave theory, it was hypothesized that the quantum system involves two separate but related components, i.e., the propagation of the wave function and the motion of the particle along its trajectory. The wave function only plays the role of guiding the motion of the particle. The movement of all particles is governed by the common physical laws. The evolution of the wave function over time, on the other hand, is given by the Schrödinger equation.

The pilot wave theory is highly complicated and explicitly non-local: the velocity of any one particle depends on the value of the guiding equation, which depends on the configuration of all the particles under consideration. Theoretically, the motion of one particle depends on the positions of all other particles in the universe.

The pilot wave theory was rejected by most mainstream theorists, mainly because of its explicit non-locality. Many authors also expressed critical views of pilot wave theory by comparing it to Everett's many-worlds approach. Since many proponents of pilot wave theory interpreted the universal wave function as physically real, one could interpret the pilot wave theory as having the same many worlds as Everett's theory.

In the Everettian view, the Bohm particles are superfluous entities. Everett's comment on Bohm's 1952 paper was: "*Our main criticism of this view is on the grounds of simplicity—if one desires to hold the view that is a real field, then the associated particle is superfluous, since, as we have endeavoured to illustrate, the pure wave theory is itself satisfactory*" [4].

10.4 How Did These Different Theories Explain the Double-Slit Experiment for Electrons?

10.4.1 The Double-Slit Experiment

The various theories discussed in the above are all aimed to explain the observations of wave-particle duality in quantum physics. One of the most clear-cut examples in demonstrating wave-particle duality is the double-slit experiment. Thus, one can illustrate the different approaches of the above theories by examining how these different theories explain the results of the double-slit experiment.

In the double-slit experiment, a beam of electrons travel through a barrier that has two slits. If one puts a detector screen on the side beyond the barrier, one could observe an interference pattern of the detected particles. It shows interference fringes characteristic of waves arriving at the screen from two sources (the two slits). The interference pattern, however, is made up of individual dots corresponding to particles that had arrived on the detecting screen. If one modifies this experiment so that one slit is covered by placing another detector behind it, no interference pattern is observed at the screen. Thus, the state of both slits affects the final results. In this experiment, the system seems to exhibit the behavior of both waves (interference patterns) and particles (dots on the screen) (Fig. 10.6).

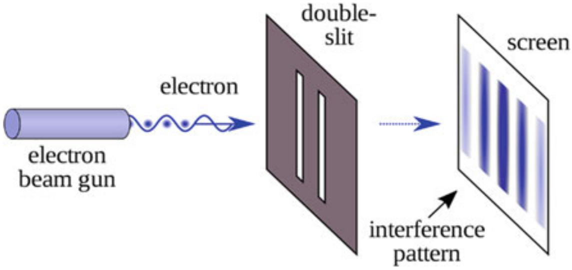


Fig. 10.6 A double-slit experiment with electrons yields a pattern of interference fringes. From the known wave-particle duality property of electrons, Feynman proposed that if a double-slit experiment was performed with electrons, the pattern of interference fringes would emerge. Image Credit Original: NekoJaNekoJa Vector: Johannes Kalliauer; Wikimedia commons; CC BY-SA 4.0

How did the different theories of quantum mechanics interpret the results of the double-slit experiment?

Explanation by the Copenhagen Interpretation (CI). The Copenhagen interpretation states that the particles are not localized in space until they are detected. Therefore, if no detector is placed at each slit, there is no information about which slit the particle has passed through. Since the particle detection is normally done at the detection screen behind both slits, its wave function is a superposition of two separated wave functions representing particles passing through either slit. So, the interference pattern is mainly caused by the superposition of wave functions; the dots on the detecting screen is caused by a collapse of the wave function during the measurement process.

If one slit has a detector on it, then the wave function collapses due to that detection. In that case, the particles reaching the detecting screen behind the double-slit barrier can only go through the slit not blocked by the detector.

Explanation by the Pilot Wave Theory (PW). In the pilot wave theory, the system is thought to involve two separate but related movements, i.e., the propagation of the wave function and the motion of the particle along its trajectory. The wave function plays the role of guiding the motion of the particle. The wave function is defined at both slits, but each particle has a well-defined trajectory that passes through exactly one of the slits. The final position of the particle on the detector screen and the slit through which the particle passes is determined by the initial position of the particle. Such initial position is not knowable or controllable by the experimenter, so there is an appearance of randomness in the pattern of detection.

In Bohm's 1952 papers, he used the wave function to construct a quantum potential that, when included in Newton's equations, gave the trajectories of the particles streaming through the two slits [2, 3]. In effect, the wave function interferes with itself and guides the particles by the quantum potential in such a way that the particles avoid the regions in which the interference is destructive and are attracted to the regions in which the interference is constructive, resulting in the interference pattern on the detector screen.

Explanation by the Many-World Interpretation (MWI). In the many-world interpretation, many quantum worlds can coexist at the same time; they are separated only when the measurement is taken. So, in the case of the double-slit experiment, there are particles that can pass through both slits. Because there is no measurement taken at the slit, it is not possible to know which slit the electron passes through. When the particle reaches the detecting screen, a measurement is taken place. The experimenter can observe only one of many possible results. That is, the measurement only gives the result of one single electron passing through a specific slit. This does not mean that the electrons with different trajectories are absent, they simply just cannot be detected in this quantum world; they can be detected in other quantum worlds.

Explanation by the Quantum Wave Model (QWM). In the quantum wave model discussed in this book, the explanation for the double-slit experiment is far simpler than the traditional quantum interpretations. It proposes that the quantum particle (such as an electron) is a quantized excitation wave. When the electron travels in the vacuum, its wave packet has a relatively large size; it is large enough to allow the electron wave to travel through both slits. Thus, a single electron can interfere with itself. When the electron hits the detecting screen, the entire wave packet of the electron will be absorbed by a single atom at the screen. There is a sudden collapse of the wave function of the electron during the measurement process. This collapse is really due to the *principle of all-or-none* in the electron absorption process.

10.5 Conclusion: Only the Quantum Wave Model Can Fully Explain the Quantum Phenomenon of Wave-Particle Duality

In order to fully explain wave-particle duality, a quantum theory must be capable to address the following **key questions** in a convincing manner:

(1) What is the physical meaning of the *matter wave*? Is it a physical wave or not? According to the Copenhagen interpretation, the matter wave is not a physical wave; it only has statistical significance, i.e., it gives the *probability* of finding the particle. In that case, it is very difficult to explain **why a "*probability wave*" can propagate in the vacuum like a traveling object**. What is the driving force for this "*probability wave*" anyway?
(2) How to explain the wave behavior of the electron in the double-slit experiment? If one regards the electron as a *point mass*, **how can a single electron pass through two separated slits** to generate an interference pattern?
(3) How to explain why an electron behaves like a particle at the detection screen in the double-slit experiment?
(4) When a measurement is made, why does the probability of detecting the particle is proportional to the amplitude square of the wave function?
(5) What is the physical basis of *deriving the quantum wave equation* for the *matter wave*? According to the Copenhagen interpretation, the matter wave is not a physical wave; it only represents the dynamic changes of the superposition of different quantum states. In that case, it is not possible to **explain the physical basis of deriving the quantum wave equation**. Can such a theory derive the Schrödinger equation based on the first principle?

From the discussion presented above in this chapter, one can see that only the quantum wave model can offer a logical basis to answer all five questions. (The comparison between different theories is summarized in the following Table 10.1). For example, in the case of **Copenhagen interpretation** (CI), although it can partially address Question #2 (by proposing a superposition of different quantum states), Question #3 (by proposing a collapse of the wave function during the measurement process), and Question #4 (by assuming the Born rule), it fails to answer Questions #1 and #5.

Table 10.1 Comparison of different quantum theories in their ability to address key questions about wave-particle duality

	CI	*MWI*	*PW*	*QWM*
Q1: Is matter wave a physical wave?	No	No	No	Yes
Q2: Why e^- behaves like wave in passing the slits?	Yes?	?	?	Yes
Q3: Why e^- looks like a particle at detection?	Yes	Yes	Yes	Yes
Q4: Why probability is related to $\|\psi(x)\|^2$?	Yes	?	?	Yes
Q5: What is the physical basis of the quantum wave equation?	No	No	No	Yes

In the case of the **Many-World interpretation** (MWI) and the **Pilot Wave theory** (PW), they have even more problems. Not only can't they answer Questions #1 and #5, it is not clear that they are capable of addressing Question #2. Furthermore, these theories involved many artificial assumptions; their arguments were also very complicated and appeared to be unnatural.

The **quantum wave model** (QWM), on the other hand, is very straightforward. It is based on the well-tested physical principles (i.e., the Maxwell theory, the quantum relations of Planck and de Broglie, and the Helmholtz decomposition theory). As we have shown in this chapter, the quantum wave model can directly address all key questions listed above. Furthermore, it has the advantage of simplicity and its assumptions and logic are far simpler than the other competing theories. Most importantly, the quantum wave model is the only model that can allow the quantum wave equation (the Schrödinger equation) to be derived based on the first principle! (see Chaps. 7–9).

10.6 Chapter Summary

- Explaining *wave-particle duality* is an outstanding problem in quantum mechanics. Niels Bohr and his colleagues (including Heisenberg and Pauli) proposed that the connection between the concepts of "particle" and "matter wave" can be through statistical interpretation. That is, the electron would remain behaving as a classical particle (like a point mass), while the matter wave would give the probability of finding the electron at a particular position in space–time. This statistical interpretation can be quite useful for comparing quantum calculation with experimental results.
- Such a statistical interpretation can be partially justified based on the *quantum wave model* presented in this book. When the free electron is traveling in the vacuum, it is in the form of a wave packet, the size of which is often bigger than an atom. Once the incoming electron is captured by an atom in the detector, the electron wave packet will suddenly collapse and transfer its entire energy to the target atom. Due to the principle of *all-or-none*, the entire electron can only interact with a single atom within the detector. The probability of detecting the

electron is proportional to the square of the wave function, in agreement with the "*Born rule*" used in the *Copenhagen interpretation*.

- The situation, however, is very different for an electron inside an atom. We showed that the *electron cloud* really represents the charge distribution (of the electron) inside the atom instead of a probability of finding a point-mass-like electron at a particular location.
- Today, many scientists are still not satisfied with the orthodox Copenhagen interpretation. People are particularly skeptical about Bohr's assumptions of "*superposition of states*" and "*collapse of wave function upon measurement*". Thus, they tried to propose alternative interpretations of quantum mechanics. Among them the most well-known ones are the *pilot wave theory* proposed by de Broglie and Bohm, and the *many-world interpretation* proposed by Everett.
- None of these theories, however, can offer a logical basis to fully explain the results of the double-slit experiment. Neither can they provide a physical basis to derive the quantum wave equation for the electron. At this time, only the quantum wave model can fully explain the quantum phenomenon of *wave-particle duality*. This model can also easily explain why the wave function suddenly collapses during a measurement.

References

1. Skibba, R. (2018). Einstein, Bohr and the war over quantum theory. *Nature, 555*, 582–584.
2. Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of hidden variables. I. *Physical Review, 85*, 166–179.
3. Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of hidden variables. II. *Physical Review, 85*, 180–193.
4. Everett, H. (1956). *The Theory of the Universal Wave Function*. Princeton University.
5. Everett, H. (1957). Relative state formulation of quantum mechanics. *Reviews of Modern Physics, 29*, 454–462.
6. Chang, D. C. (2017). Physical interpretation of the Planck's constant based on the Maxwell theory. *Chinese Physics B, 26*, 040301.
7. Slater, J. C. (1969). *Concepts and Development of Quantum Physics*. Dover Publications.
8. Schrödinger, E. (1926). An undulatory theory of the mechanics of atoms and molecules. *Physical Review, 28*, 1049–1070.
9. Houston, W. V. (1959). *Principles of Quantum Mechanics: Nonrelativistic Wave Mechanics with Illustrative Applications*. Dover Publications.
10. Messiah, A. (1965). *Quantum Mechanics* (Vol. 1). Wiley.
11. Isaacson, W. (2007). *Einstein: His Life and Universe*. Simon & Schuster.
12. Schrödinger, E. (1935). Die gegenwärtige situation in der Quantenmechanik. *Naturwissenschaften, 23*, 807–812.
13. Carroll, S. M. (2019). *Something Deeply Hidden: Quantum Worlds and the Emergence of Spacetime*. Dutton.
14. Tipler, F. J. (1994). *The Physics of Immortality: Modern Cosmology, God, and the Resurrection of the Dead* (1st ed.). Doubleday.

Part IV
The Physical Meaning of Mass and Energy From a Wave Perspective

Chapter 11
Why Can *Mass* and *Energy* Be Converted Between Each Other? *Energy*, *Momentum*, and *Mass* Have Geometrical Meanings in the Wave View

In classical physics, *energy* and *mass* are totally different physical concepts. It was a great surprise when physicists in the twentieth century discovered that *energy* and *mass* can be converted between each other. What is the physical basis of this discovery?

In popular science literature, it is often stated that the reason for mass-energy conversion is due to the special theory of relativity (STR). Such a claim, however, was a misunderstanding. Many recent reviews of the scientific literature had clearly shown that, the relation of mass-energy conversion had nothing to do with STR (see the session below and **Appendix E**). According to the quantum wave model discussed in this book, **the reason for mass-energy conversion is due to the fact that the particle is a quantized excitation wave**.

This new understanding is not surprising. In fact, it offers a useful hint to explain how matter could be created in our world. The logic of this thinking is very simple:

(1) Matter is made up of sub-atomic particles.
(2) Sub-atomic particles are excitation waves of the vacuum medium.
(3) Energy can be used to generate excitation waves.
(4) Thus, energy can be converted to matter.

In this chapter, we will explain in detail why the reason for mass-energy conversion is due to quantum physics instead of relativity. Furthermore, we will also try to answer the following questions:

- What is the physical meaning of ***mass***?
- Why should ***mass*** be treated on the same footing as ***energy*** and ***momentum***?
- How can a wave have ***mass***?
- What are the meanings of ***energy***, ***momentum***, and ***mass*** in the wave perspective?

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_11

11.1 The Discovery of Energy-Mass Equivalence Was Not Based on Special Relativity

The mass-energy equivalence relation $E = mc^2$ is one of the most well-known formulas in modern physics. In many textbooks, this relation is often thought to be arisen from the principle of relativity (PR) [1–5]. Such thinking, however, was not correct; it does not stand up if one conducted a careful examination of the literature [6–12]. Although Einstein was an active supporter of the mass-energy equivalence idea, he was not the inventor of this concept.

First, **the concept of mass-energy equivalence had been suggested by many other scientists before Einstein proposed STR** [13–17]. At the end of the nineteenth century, several European physicists had proposed that mass can be related to the energy-content of an object. For example, in 1881, J. J. Thomson showed that the magnetic field generated by a moving charged sphere could induce an effective mass on the sphere [13]. Oliver Heaviside in 1889 further suggested that the effective mass m should be proportional to E/c^2 [14]. Wien and Max Abraham called this the "electromagnetic mass" [18, 19]. The energy-mass relation $E = mc^2$ was first mentioned by Poincare in a paper published in 1900 [16]. In 1904, Fritz Hasenöhrl also proposed the concept relating energy with mass in a series of papers entitled "*On the theory of radiation in moving bodies*" [20, 21]. According to W. Fadner, there were many discussions on the topic of mass-energy equivalence before 1905 [8].

Second, **Einstein's arguments were not based on the principle of relativity**; instead, his arguments for $E = mc^2$ were mostly based on demonstrations in special hypothetical situations (he called them "*thought experiments*") [3, 8–12]. For example, the first publication of Einstein on this topic was a very short paper published in 1905, in which he considered an object sending out two identical pulses of light at opposite directions [22]. The energy of each of these light pulses is designated $\frac{1}{2}L$ as measured in the stationary frame S. After the emission of the light pulses, the total energy of the object was measured in the stationary frame S and moving frame S'. He argued that, the difference between the energy measured in the S and S' frames is equal to the kinetic energy K of the object. Using the Taylor expansion and ignoring higher-order terms, he showed that, *if a body gives off the energy L in the form of radiation, its mass diminishes by* L/c^2 [22]. (Details of this *thought experiment* and its critique are given in **Appendix E**).

His derivation, however, had several problems: (1) It was not a general derivation, instead, it was a result obtained from a special hypothetical situation. (2) The obtained result $E = mc^2$ was an approximation; it only applied when the moving speed of the object is much smaller than c. (3) His assumption that the total energy is a linear sum of the resting energy and the kinetic energy is incorrect for an object moving at high speed. (4) His hypothetical "*thought experiment*" was unrealistic; it cannot be performed in practice [9, 12]. (5) Finally, as pointed out by H. E. Ives, Einstein's 1905 paper on $E = mc^2$ was based on a circular logic [7].

Third, **Einstein was aware of the shortcomings of his derivation and thus wrote many more papers on this topic in the following years** [23–29]. However, most of Einstein's *thought experiments* had nothing to do with the principle of relativity; instead, **they were mainly based on his intuitive thinking that radiation and matter could behave similarly** [3, 12]. For example, Einstein proposed another *thought experiment* in 1906 based on a center-of-gravity argument [23]. (For details, see **Appendix E**). This thought experiment suggested that *the radiation wave can have an effective mass equal to its energy divided by* c^2.

In this *thought experiment,* Einstein's derivation of the mass-energy relation was based on the idea that *radiation wave can behave like a material object*. **His *thought experiment* actually violated the basic principle of STR**, since it assumed the entire box move simultaneously when a burst of radiation wave was emitted from its left end at $t = 0$. That would imply that information can be transmitted faster than the speed of light. This problem was recognized by Einstein himself in his later paper [30].

Finally, **a large number of historical reviews on the derivation of the mass-energy relation had been published in recent years**; **they concluded that Einstein failed to derive the mass-energy relation** [7–11]. For example, one review had explicitly pointed out that *Einstein was aware of the shortcomings of his derivation (in 1905 and 1906) and tried to write more papers to patch things up but arguably never succeeded* [11]. Recently, a series of elaborated reviews of Einstein's attempt to derive the mass-energy relation was published by E. Hecht [9, 31]. His conclusion was that: "*Einstein produced about 18 virtuoso derivations and demonstrations all aimed at establishing the mass-energy principle. … although each of them gave evidence for the applicability of* $E_0 = mc^2$ *to a particular set of circumstances, no one derivation, or collection of them taken together, succeeded in providing a definitive proof of its complete generality*" [9]. According to Hecht, "*The fact that Einstein continued to create demonstrations of the efficacy of* $E_0 = mc^2$ *up to 1946 tells us that he knew the definitive proof had not been accomplished*" [9].

A detailed summary of recent reviews on the derivation of the relation $E = mc^2$ is given in **Appendix E** at the end of this book. For readers who are interested in this issue, this summary may provide a useful reference.

11.2 Why Mass and Energy Are Convertible? It is a Quantum Wave Effect

One may then ask: If the relation of mass-energy equivalence was not derived from STR, what can be its physical basis? As we will show in the following, the concept of mass-energy equivalence is a consequence of quantum effect. More specifically, **the mass-energy equivalence is really due to the fact that matter is made of quantized excitation waves** [32].

11.2.1 The Relation of Mass-Energy Equivalence for Photon is Clearly a Quantum Effect

It can be easily shown that the relation of mass-energy equivalence for a photon is a consequence of its quantum property. We know photon and electron can behave as a particle as well as a wave. Such a wave property can make the particle behave differently from a point mass. This means that, a quantum object can have properties that are not found in classical mechanics. As shown in the following, if one treats the photon as a quantized wave packet, one can easily derive the mass-energy equivalence based on well-known quantum relations.

Let us first recall what the definition of "*mass*" is. From Newtonian mechanics, we know the *mass* (m) of an object is related to its momentum (p),

$$p = mv. \tag{11.1}$$

In the case of a photon, its speed is c. From the de Broglie relation, we know the momentum of a photon is $p = \hbar k$. (Here $\hbar$ is Planck's constant divided by 2π). So, the "mass" of a photon can be defined as

$$m = p/c = \hbar k/c. \tag{11.2}$$

Since $k = 2\pi/\lambda = \omega/c$, using Planck's relation, the above equation becomes

$$m = \hbar k/c = \hbar\omega/c^2 = E/c^2. \tag{11.3}$$

Hence, if one regards the quantized light wave as a particle, one can easily derive the mass-energy conversion relation of a photon, i.e.,

$$E = mc^2. \tag{11.4}$$

In the case of a photon, its mass is the "*effective mass*", not "*rest mass*". Since this effective mass is derived from the particle's momentum, we may call it the "*inertial mass*". Recently, we demonstrated that this *inertial mass* is also equal to the "*gravitational mass*" of the photon [33].

Thus, at least in the case of a photon, one can easily show that **the mass-energy equivalence is a consequence of the quantum relations of Planck and de Broglie**. For particles with non-zero rest mass, it is more complicated to derive the mass-energy equivalence relation. Nevertheless, as shown in the following, such a relation can indeed be derived based on the wave properties of the particle.

11.3 The Physical Meaning of *Mass*: Mass Should Be Treated on the Same Footing as Energy and Momentum

11.3.1 Where Does Mass Come From? The Physical Meaning of Mass According to Newton

What is the meaning of mass? This is an important question in the history of physics [10]. Newton is probably the first one to give a scientific concept of mass. In his famous work "*Mathematical Principles of Natural Philosophy*" [34], he thought "mass" is "the quantity of matter". He found that for any two objects, the ratio for their inertia and the ratio for their weight are the same. This implies that the inertia mass and the mass associated with weight are equal. Then, people could measure the mass of a body by determining its weight.

Furthermore, Newton proposed that the weight of an object is just a measure of the gravitational force for that object. He thus concluded that the inertia mass and the gravitational mass are the same thing.

In Newtonian mechanics, the mass is regarded as an intrinsic property of an object. Nowadays, the concept of mass is mainly defined by relating it to momentum, that is,

$$p = mv. \tag{11.1}$$

Thus, *mass* has a clear meaning, that is, ***m*** **is the proportional constant between the momentum** ***p*** **and the velocity** v. Although this definition was originated from Newton, it is widely accepted by most physicists, including Einstein.

11.4 How Can a Wave Have Mass?

11.4.1 The Meaning of Mass in the Wave View

As we stated earlier, one great discovery in modern physics is that an elementary particle can sometimes behave like a wave. In the case of photon, there is no doubt that it is a light wave. But even for particles with rest mass, such as electrons or neutrons, they can also behave like a wave. Then, one must ask: **How can a wave have mass?**

This question can be answered if one accepts the definition of mass as outlined in Eq. (11.1), i.e., "mass" is only the proportional constant between $\boldsymbol{p}$ and $\boldsymbol{v}$. We know that the quantum wave representing a particle has a momentum which is proportional to the wave number *k*:

$$p = \hbar k, \tag{11.5}$$

where $k = 2\pi/\lambda$ (λ is the wavelength). We also know that a wave packet has velocity, that is its *group velocity*, which is given as

$$v = \frac{d\omega}{dk}. \tag{11.6}$$

The energy of the quantum wave packet is proportional to its frequency, i.e.,

$$E = \hbar\omega. \tag{11.7}$$

We can calculate v easily once we know the dispersion relation of the quantum wave packet (i.e., $\omega = \omega\ (k)$). Substituting Eqs. (11.5) and (11.6) into Eq. (11.1), we have

$$m = \frac{p}{v} = \hbar k / \frac{d\omega}{dk}. \tag{11.8}$$

Thus, the mass can be calculated explicitly. This explains why a quantum wave packet can have mass.

From the above equation, it is easy to see that mass is not necessarily a constant. Since the dispersion relation $\omega = \omega\ (k)$ is not linear in general, m is a function of k according to Eq. (11.8). That means the *mass should vary with momentum change*. Thus, it is not a mystery that the mass of a particle would change with speed. Instead, it is a natural expectation if one accepts that the particle is a quantized excitation wave.

11.5 Origin of the *Energy–Momentum Relation* of a Quantum Particle

11.5.1 In the Teaching of Relativity, the Rest Mass is Simply an Integration Constant for Deriving the Energy–Momentum Relation

As we pointed out in the earlier parts of this chapter, although Einstein knew that E is related to mc^2, he was not able to derive the mass-energy equivalence relation from first principle. Furthermore, his 1905 STR paper gave the wrong prediction for the speed-dependence of mass [35]. He failed to derive the correct relation of $m = m_0\left(1 - v^2/c^2\right)^{-1/2}$. But today, many textbooks of relativity proclaim otherwise. They somehow maintain that both the mass-energy equivalence relation and the speed-dependence of mass are derived from STR.

How could they do that? The trick of these textbooks is to ignore Einstein's original argument in 1905 and invented alternative ways to obtain the correct relation of mass-energy equivalence (or speed-dependence of mass). Their derivations often have nothing to do with STR.

As a demonstration of this new argument, let us briefly review the derivation of these so-called *"relativistic relations"* in some current standard textbooks. A widely used textbook (which is part of *The MIT Introductory Physics Series*) is *Special Relativity* by A.P. French [3]. Its derivation of the relation $E = mc^2$ is given in its Chap. 1, before the principle of relativity is taught. In the following, let us summarize the essential steps of its derivation.

Its derivation starts by pointing out that, for a photon, its energy–momentum relation is

$$E = cp. \tag{11.9}$$

Then, one can get $$dE = c\,dp.$$

By multiplying both of these equations together, one has

$$E\,dE = c^2 p\,dp. \tag{11.10}$$

Although the above relation was derived from the energy–momentum relation of light, the author *argued that this same relation is also valid for matter* [3]. By integration of Eq. (11.10), one can obtain,

$$E^2 = c^2 p^2 + E_0^2, \tag{11.11}$$

where E_0^2 **is a constant of integration**.

Recall that the particle speed is $v = \dfrac{\partial E}{\partial p}$. One can differentiate both sides of Eq. (11.11) with respect to p and obtain

$$v = \frac{\partial E}{\partial p} = \frac{c^2 p}{E}. \tag{11.12}$$

At low speed ($v \ll c$), $E \to E_0$, the above equation becomes

$$p = \frac{E_0}{c^2} v. \tag{11.13}$$

Since we know at low speed, the momentum is equal to the rest mass times the velocity,

$$p = m_0 v, \tag{11.14}$$

from the above two equations, one can get

$$m_0 = \frac{E_0}{c^2}. \tag{11.15}$$

This suggests that E_0/c^2 is playing the role as a "*rest mass*". So, m_0 is just a measure of the resting energy E_0, which, is originally an integration constant of the *E versus p* relation.

From the above relation, one can get the mass-energy relation in the more familiar form,

$$E_0 = m_0\, c^2. \tag{11.16}$$

From this summary, one can see that, the classical derivation of the energy-mass relation is not really based on the principle of relativity. Instead, *it is based on a conceptual assumption that the variation of E vs p can be similar between matter and radiation.* It only demonstrated that the *rest mass* is related to an integration constant of the *E versus p* relation.

11.5.2 In the Quantum Wave Model, the Energy–Momentum Relation of a Particle Is Originated from the Dispersion Relation of the Quantum Wave Function

From the above, one can see that, in the derivation given in the relativity textbooks, the rest mass of an object is just an integrating constant in the energy–momentum relation; it has no clear physical meaning. This is not very satisfactory. In contrast, **the rest mass in the quantum wave model has a clear physical meaning**. As we showed in Chap. 7, the energy–momentum relation of a particle originated from the dispersion relation of the quantum wave packet. In that case, **the rest mass m_0 is related to an effective "wave number" in the direction perpendicular to the particle's trajectory**. This "wave number" is characterized by the parameter ℓ which describes the lateral oscillation motion of the quantum wave function. This finding is consistent with our new understanding that *mass, energy*, and *momentum* should be treated on the same footing. It is well known that *energy* and *momentum* are related to the frequency and wave vector, it is not surprising that mass could be related to a "wave number".

Recall that from Chap. 7, the quantum wave function describing a massive particle is given in Eq. (7.16), which contains several wave parameters, ω, $\boldsymbol{k}$, and ℓ. Here, ω and $\boldsymbol{k}$ are the frequency and wave vector of the traveling longitudinal wave, while ℓ is the transverse wave number characterizing the oscillation of the Bessel function, which describes the movement of the vacuum medium in the transverse direction. The dispersion relation of the quantum wave function is

$$\omega^2 = \left(k^2 + \ell^2\right)c^2. \tag{11.17}$$

What are the corresponding meanings of these wave parameters, ω, $\boldsymbol{k}$, and ℓ in a particle view? Apparently, based on the Planck's relation $E = \hbar\,\omega$ and the de Broglie relation $\boldsymbol{p} = \hbar\,\boldsymbol{k}$, ω and $\boldsymbol{k}$ can be identified with the energy (E) and momentum ($\boldsymbol{p}$) of the particle. Then, what is the physical meaning of ℓ?

If one multiplies the dispersion relation with $\hbar^2$ and using the Planck's relation and the de Broglie relation, one can obtain

$$E^2 = c^2 p^2 + c^2 \hbar^2 \ell^2. \tag{11.18}$$

By taking a square root for both sides of the above equation, it becomes

$$E = c\left(p^2 + \hbar^2\ell^2\right)^{1/2}. \tag{11.19}$$

Recall that the particle velocity (v) is equal to the group velocity of the wave packet, [36] $v = \dfrac{\partial \omega}{\partial k} = \dfrac{\partial E}{\partial p}$, the above equation gives

$$v = \frac{\partial E}{\partial p} = cp\left(p^2 + \hbar^2\ell^2\right)^{-1/2}. \tag{11.20}$$

Knowing the definition of momentum $p = mv$, the above equation becomes

$$v = \frac{cmv}{\sqrt{(mv)^2 + \hbar^2\ell^2}}, \tag{11.21}$$

one can re-arrange the above equation to have

$$m = \frac{\hbar\ell/c}{\left(1 - v^2/c^2\right)^{1/2}}. \tag{11.22}$$

When v approaches zero, m equals the rest mass, m_o. The above equation becomes

$$m_0 = \frac{\hbar\ell}{c}. \tag{11.23}$$

This indicates that the wave parameter ℓ is associated with the rest mass of the particle. Combining the above two relations, one has

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{11.24}$$

In contrast to the *rest mass* m_0, we may call m the *moving mass*. Substituting Eq. (11.23) into Eq. (11.18), the dispersion relation becomes

$$E^2 = p^2c^2 + m_0^2c^4, \tag{11.25}$$

which is exactly the so-called "*relativistic energy–momentum relation*" of a massive particle. Of course, our derivation was based on the quantum wave nature of the particle instead of relativity. So, this relation should be more correctly called "*quantum energy–momentum relation*", or "*energy–momentum relation*" of a quantum particle.

Furthermore, by combining the above two equations with $p = mv$, one can directly show that the total energy of a particle is proportional to its moving mass,

$$E = mc^2. \tag{11.26}$$

Thus, this demonstrated that the mass-energy equivalence relation really originated from the wave properties of the particle. In addition, from the energy–momentum relation as expressed in Eq. (11.25), it is obvious that, when the particle is at rest ($p = 0$), the rest-energy of a particle is proportional to its rest mass, i.e.,

$$E_0 = m_0c^2. \tag{11.27}$$

Hence, the mass-energy equivalence relation is not only applicable for a photon, as we showed earlier, it also holds for a particle with non-zero rest mass.

11.6 Energy, Momentum, and Mass Are All Related to the Curvature of Bending the Vacuum Medium

From the above discussion, it is apparent that the resting energy E_0 is contributed solely from the rest massm_0, which in turn is related to the parameter ℓ in the wave function. This provides a very useful hint about the physical nature of the rest mass. This in fact suggests that the rest mass could be related to a geometrical property of the vacuum.

To understand this point, let us first review the physical nature of energy and momentum according to the view of wave-particle duality. From the Planck's relation and the de Broglie relation, it is easy to see that *the energy and momentum of a free particle are related to the periodicity of oscillation of the vacuum medium*. More specifically, from $E = \hbar\omega$, we know **energy is related to the periodicity of oscillation in the time dimension**, and from $p = \hbar k$, **momentum is related to the periodicity of oscillation in the spatial dimension along the direction of the trajectory**. Our finding that $m_0 = \hbar\ell/c$ suggests that the rest mass can be associated with some sort of oscillation periodicity too. Indeed, from Eq. (7.16), we know that the transverse component of the free particle wave function is described by a Bessel function, the asymptotic form of which is

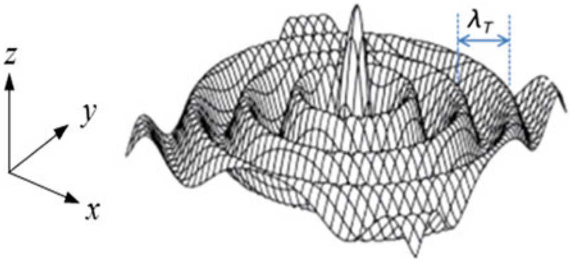


Fig. 11.1 **A plot of the transverse component of the wave function ψ_T for n = 0**. The wavelength is denoted $\boldsymbol{\lambda_T}$. The particle is traveling along the axis *z*. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, arXiv preprint physics/0404044v2 (2016)

$$J_n(\ell r) \to \left(\frac{2}{\pi \ell r}\right)^{1/2} \cos\left(\ell r - \frac{2n+1}{4}\pi\right). \tag{11.28}$$

(The variation of Bessel function for $n = 0$ is shown in Fig. 11.1.) Thus, ℓ can be regarded as the "*transverse wave number*" of the free particle, i.e., it is the inverse of the wavelength in the transverse oscillation, $\ell = 2\pi/\lambda_T$ (see Fig. 11.1). Thus, Eq. (11.23) means that **the rest mass of a particle is associated with the oscillation periodicity of the wave function in the transverse plane**.

This result appears to make very good sense. In essence, our analysis suggests that **energy, momentum, and mass are all related to the oscillations of the quantum wave function, which characterize the curving of the vacuum medium in different dimensions**. More specifically, if one uses the natural unit ($c = 1$), one can see:

$$\begin{cases} E = \hbar\omega \sim \dfrac{1}{T} (T \text{ is the period of the wave oscillation}) \\[2ex] p = \hbar k \sim \dfrac{1}{\lambda_L} \left(\lambda_L \text{ is the wavelength along the longitudinal direction}\right) \\[2ex] m_0 = \hbar\ell \sim \dfrac{1}{\lambda_T} \left(\lambda_T \text{ is the wavelength along the transverse direction}\right) \end{cases}$$

These imply that, *the energy (E) is inversely proportional to the wavelength in the time dimension*; *the momentum (p) is inversely proportional to the wavelength in the longitudinal spatial dimension*; and *the rest mass* (m_0) *is inversely proportional to the wavelength in the transverse spatial dimension.* Since "*the inverse of wavelength*" is a measure of the ***curvature*** of bending the wave medium, the above results suggest that, the particle properties including ***energy, momentum,*** **and** ***rest mass*****, are all related to the** ***curvature of bending the vacuum medium during the propagation of the excitation wave***. Such bending curvatures are just taking place in different dimensions.

This finding is highly interesting. We know both energy and mass must be created by "work". When there is no excitation wave, there is no stress in the vacuum medium; its curvature is zero. When an excitation wave emerges, it creates stress at a local region of the vacuum medium. Since it takes work to bend the wave medium, the larger the bending curvature, more work is required. This is true for bending the medium in all dimensions (spatial and temporal). Thus, a shorter wavelength of a propagating wave should always associate with a higher "energy state", which may be reflected in an increase in energy, momentum, or "mass" of the excitation wave.

11.6.1 The Resting Energy and the Kinetic Energy of a Single Particle Appear to Form a Two-Dimensional Hilbert Space

The energy–momentum relation as given in Eq. (11.25) provides an important insight about the physical nature of energy of a free particle. Namely, it suggests a special geometrical linkage between the particle energy and its momentum and rest mass. For a moving particle, there are two types of energies associated with it: (1) The "kinetic energy" E_K, which associates with the momentum, i.e., $E_K = cp$, and (2) The "resting energy" $E_0 = m_0c^2$, which is the intrinsic energy of a particle that does not depend on its momentum. However, the total energy of the particle (E) is not a linear combination (algebraic sum) of E_K and E_0, instead, they form a triangular relationship following the Pythagoras law, i.e.,

$$E^2 = \boldsymbol{p}^2c^2 + E_0^2 = E_K^2 + E_0^2. \tag{11.29}$$

This suggests that *the kinetic energy E_K and the resting energy E_0 form a two-dimensional Hilbert space. The particle energy (E) is a vector sum of E_K and E_0* (see Fig. 11.2).

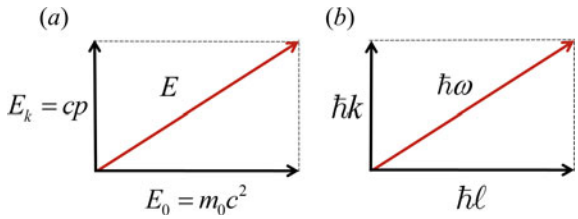


Fig. 11.2 The resting energy and kinetic energy of a single particle form a two-dimensional Hilbert space. a Geometrical relationship between the corpuscular properties *E, p*, and *m*. **b** Geometrical relationship between the wave parameters ω, k, and ℓ. (Here we use the natural unit, $c = 1$)

This relationship can be understood easily from a wave perspective. Recall that Eq. (11.25) was derived from the dispersion relation of a free particle, i.e., Eq. (11.17). Using the natural unit ($c = 1$), the dispersion relation can be rewritten as

$$(\hbar\omega)^2 = (\hbar k)^2 + (\hbar\ell)^2. \tag{11.30}$$

It suggests that $\hbar\omega$ is a "vector sum" of two perpendicular vectors with amplitudes equal to $\hbar k$ and $\hbar\ell$ (see Fig. 11.2b). We know $\boldsymbol{k}$ is the wave vector parallel to the trajectory of the particle, while ℓ is a "wave number" that characterizes the oscillation of the wave function in a plane transverse to $\boldsymbol{k}$. Thus, the directions of oscillations in the vacuum characterized by $\boldsymbol{k}$ and ℓ apparently are associated with two perpendicular axes. We have demonstrated that $\hbar\ell$ is associated with the rest mass and thus the resting energy, and, from the de Broglie relation, we know that $\hbar k$ is associated with momentum and thus the moving energy. Hence, it is not surprising that the resting energy E_0 and moving energy E_K may form a two-dimensional Hilbert space. Naturally, their vector sum becomes the total energy of the particle.

This provides a simple explanation to the physical nature of the mechanical properties of a particle [37]. This geometrical interpretation suggests that the physical natures of energy and mass are very similar; they are all related to the curvatures of bending the wave medium. There is no wonder why energy and mass can be converted between each other.

11.7 How Can an Excitation Wave Behave Like a Particle?

Based on the discussion in this chapter, we believe that at the quantum level, all particles are excitation waves of the vacuum medium. Thus, not only photons, but all sub-atomic particles, including electrons, protons, neutrons, etc., are all quantized wave packets. Because their creation and annihilation must follow the "*principle of all-or-none*", they look like individual particles.

11.7.1 The "Quantum" Phenomenon is Just a Manifestation of the "Principle of All-or-None"

Let's first review what a "quantum phenomenon" is. As we mentioned in Chap. 3 of this book, the discovery of quantum phenomena began with Planck's work on blackbody radiation. Planck discovered that the radiant energy of light is not distributed continuously, but consists of many undividable small units. The smallest unit of this radiated energy is called a "quantum" of light. From the results of these works, it is believed that light transfers energy from one object to another in the form of "photons". That is, it seems that light is transmitted in the form of "particles".

A few years later, when Einstein studied the photoelectric effect, he discovered that when an electron in an atom absorbs the energy of light, it also has a smallest unit; this smallest unit is equal to the quantum of light discovered by Planck. So, when an atom is illuminated with light, the electrons inside it either absorb the entire quantum energy of a photon, or they do not absorb any light energy at all.

Conversely, when an electron inside an atom jumps from a high-energy level to a low-energy one, it produces a photon. And the energy of the photon is exactly equal to the potential energy difference between the two levels of the electron. That is to say, all the energy released by the electron goes to the newly generated photon.

Thus, in nature, light behaves in a quantum form. In this process, the light energy involved has a smallest unit (that is, a “photon”) and cannot be sub-divided. We can say that the energy contained within the photon shares the same destiny: either to live or to die together. Thus, either the entire photon is absorbed by a quantum particle (such as an electron), or it is not absorbed at all. In another word, **a photon is like a digital object; it can only exist in two states**, either “**1**” or “**0**”.

Thus, the quantized light wave can be regarded as a kind of “particle”. In fact, this is not only true for photons, but also for electrons. The same goes for the electron’s anti-particle, the positron. We can even generalize this conclusion to all particles in the universe.

Furthermore, not only the creation and annihilation of the quantized excitation wave is “particle-like”, the wave packet can also have particle properties such as *energy, momentum*, *velocity,* and *mass* (see Fig. 11.3). It is well known that the *velocity* of the quantized excitation wave is equal to the group velocity of the wave packet. We have also shown earlier in this chapter that the wave packet has an *effective mass* which is equal to the ratio between its momentum and speed. In the macroscopic view, this effective mass behaves just like the *inertial mass* of a mechanical object. In fact, we showed recently that such an *effective mass* can generate the same gravitational effect as the *gravitational mass* of an object [33].

Since the excitation wave of the vacuum can have well-defined *particle properties*, including *energy*, *momentum*, *velocity,* and *mass*, it will look like a *particle* in the macroscopic view (i.e., in the classical limit) (see Fig. 11.3).

11.7.2 There is a One-to-One Correspondence Between the Particle Properties and the Wave Properties

One nice thing about the wave model is that it makes things very simple. In the quantum wave model, a particle is a quantized excitation wave of the vacuum. One can look at a quantum particle from two different perspectives, i.e., either from a particle view or from a wave view. For the particle view, one can talk about the energy of the particle, the momentum of the particle, and the rest mass of the particle. But for a wave view, one can talk about the frequency, the wave vector, and the transverse wave number. (The “transverse wave number” is the oscillation in the transverse plane.)

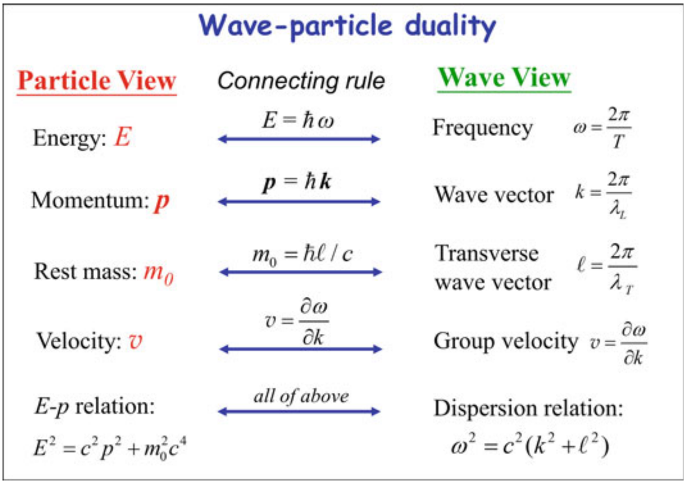


Fig. 11.3 **Wave-particle duality**. There is a one-to-one correspondence between particle properties and wave properties in the quantum view

From the discussions above, one can easily see that, there is a one-to-one correspondence between the particle properties and the wave property (see Fig. 11.3 above). For example, *energy* is related to the *frequency,* and the rest mass m_0 is related to the transverse oscillation of the wave packet. Therefore, *energy*, *momentum* and *mass* are basically the same thing. All of them are a measure of the curvature of bending the vacuum medium. **The physical properties of a particle now have their own geometrical meanings!** Naturally, a sharper bending of the medium is associated with more work, and thus requires more energy. *These geometrical relations explain why mass and energy are equivalent*.

11.8 Chapter Summary

- The discovery of *energy-mass equivalence* was not based on special relativity. According to recent literature reviews, Einstein failed to derive the mass-energy relation based on STR (see **Appendix E**).
- Why mass and energy are convertible? It is a quantum wave effect. The relation of *mass-energy equivalence* is really due to the fact that matter is made of quantized excitation waves. In the case of a photon, one can easily show that the *mass-energy equivalence* relation is a consequence of the quantum relations of Planck and de Broglie. For particles with non-zero rest mass, we can also show that the *mass-energy equivalence* relation is derived from the wave properties of the quantum particle.

- *Mass* should be treated at the same footing as *energy* and *momentum*. In Newtonian mechanics, *mass* is regarded as an intrinsic property of an object. The concept of *mass* is mainly defined by relating it to *momentum*, that is, m is the proportional constant between the momentum and the velocity.
- What is the physical meaning of *mass*? In the teaching of relativity, the *rest mass* is simply an integration constant for deriving the *energy–momentum relation*. In the quantum wave model, the *energy–momentum relation* of a particle is originated from the *dispersion relation* of the quantum wave function. *Mass* is related to a wave property.
- Our analysis suggests that *energy*, *momentum*, and *mass* are all related to the oscillations of the quantum wave function, which characterize the curving of the vacuum medium in different dimensions. More specifically, the *energy* (E) is inversely proportional to the wavelength in the *time* dimension; the *momentum* (p) is inversely proportional to the wavelength in the longitudinal *spatial* dimension; and the *rest mass* (m_0) is inversely proportional to the wavelength in the transverse *spatial* dimension. Since "*the inverse of wavelength*" is a measure of the *curvature* of bending the wave medium, the above results suggest that particle properties including *energy*, *momentum*, and *rest mass*, are all related to the *curvature* of bending the vacuum medium during the propagation of the excitation wave.
- How can an *excitation wave* behave like a *particle*? The "*quantum*" phenomenon is just a manifestation of the "*principle of all-or-none*". Since the *excitation wave* of the vacuum can have well-defined *particle properties*, including *energy*, *momentum*, *velocity,* and *mass*, it will look like a *particle* in the macroscopic view.

References

1. Coughlan, G. D., & Dodd, J. E. (1991). *The ideas of particle physics: an introduction for scientists* (2nd ed.). Cambridge University Press.
2. Longair, M. S. (1984). *Theoretical concepts in physics: an alternative view of theoretical reasoning in physics*. Cambridge University Press.
3. French, A. P. (1968). *Special relativity*. W. W. Norton & Company.
4. Okun, L. B. (1989). The concept of mass. *Physics Today, 42*, 31.
5. Hartle, J. B. (2003). *Gravity: an introduction to einstein's general relativity*. Addison-Wesley.
6. Whittaker, E. (1951). *A history of the theories of aether and electricity*. Thomas Nelson and Sons Ltd.
7. Ives, H. E. (1952). Derivation of the mass-energy relation. *Journal of the Optical Society of America, 42*, 540–543.
8. Fadner, W. L. (1988). Did Einstein really discover $E = mc^2$? *American Journal of Physics, 56*, 114–122.
9. Hecht, E. (2011). How Einstein confirmed $E_0 = mc^2$. *American Journal of Physics, 79*, 591–600.
10. Jammer, M. (2000). *Concepts of mass in contemporary physics and philosophy*. Princeton University Press.
11. Rothman, T. Was Einstein the first to invent $E = mc^2$?. https://www.scientificamerican.com/article/was-einstein-the-first-to-invent-e-mc2/

12. Chang, D. C. (2018). A new interpretation on the non-Newtonian properties of particle mass. *Journal of Modern Physics, 9*, 215–240.
13. Thomson, J. J. (1881). On the electric and magnetic effects produced by the motion of electrified bodies. *Philosophical Magazine, 11*, 229–249.
14. Heaviside, O. (1889). XXXIX. On the electromagnetic effects due to the motion of electrification through a dielectric. *Philosophical Magazine, 27*, 324–339.
15. Lorentz, H. A. (1899). Simplified theory of electrical and optical phenomena in moving systems. *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings, 1*, 427–442.
16. Poincaré, H. (1900). La théorie de Lorentz et le principe de réaction. *Archives Néerlandica Science Exactes Natural, 2*, 252–278.
17. Lorentz, H. A. (1904). Electromagnetic phenomena in a system moving with any velocity smaller than that of light. *Proceedings R. Netherland Academy Arts Science, 6*, 809–831.
18. Wien, W. (1901). Ueber die Möglichkeit einer elektromagnetischen Begründung der Mechanik. *Annalen der Physik, 310*, 501–513.
19. Abraham, M. (1902). Prinzipien der Dynamik des Elektrons. *Physikalishce Zeitschrift, 4*, 57–63.
20. Hasenöhrl, F. (1904). Zur Theorie der Strahlung in bewegten Körpern. *Annalen der Physik, 15*, 344–370.
21. Hasenöhrl, F. (1905). Zur Theorie der Strahlung in bewegten Körpern. Berichtigung. *Annalen der Physik, 16*, 589–592.
22. Einstein, A. (1952). Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig? English translation: The principle of relativity: A collection of original memoirs on the special and general theory of relativity, (pp. 69–71). Dover Publications; *Annalen Physics*, *323*, 639–641 (1905).
23. Einstein, A. (1906). Das Prinzip von der Erhaltung der Schwerpunktsbewegung und die Trägheit der Energie. *Annalen der Physik, 325*, 627–633.
24. Einstein, A. (1909). On the development of our views concerning the nature and constitution of radiation. *Physikalishce Zeitschrift, 10*, 817–826.
25. Einstein, A. (1990). Manuscript on the special theory of relativity. Section 14. The inertia of energy (1912). In M. J. Klein, A. J. Kox, J. Renn, & R. Schulmann, (Eds.) *The collected papers of Albert Einstein, The Swiss years: writings, 1912–1914*, (vol. 4, pp. 50–55). (Transl. by Anna Beck). Princeton University Press.
26. Einstein, A. (1990). Manuscript on the special theory of relativity (1912–1914). In M. J. Klein, A. J. Kox, J. Renn, & R. Schulmann, (Eds.) *The collected papers of Albert Einstein, The Swiss years: writings, 1912–1914*, (vol. 4, pp. 3–88). (Transl. by Anna Beck). Princeton University Press.
27. Einstein, A. (1922). *The meaning of relativity*. Princeton Univ.
28. Einstein, A. (1935). Elementary derivation of the equivalence of mass and energy. *Bulletin of the American Mathematical Society, 41*, 223–230.
29. Einstein, A. (1946). An elementary derivation of the equivalence of mass and energy. *Technion Yearbook, 5*, 16.
30. Einstein, A. (1990). On the inertia of energy required by the relativity principle (1907). In J. Stachel, D. C. Cassidy, J. Renn, & R. Schulmann, (Eds.) *The collected papers of Albert Einstein, The Swiss years: writings, 1900–1909*, (vol. 2, pp. 238–250), (Transl. by Anna Beck). Princeton University Press.
31. Hecht, E. (2009). Einstein on mass and energy. *American Journal of Physics, 77*, 799–806.
32. Chang, D. C. (2020). A quantum interpretation of the physical basis of mass–energy equivalence. *Modern Physics Letters B, 34*, 2030002.
33. Chang, D. C. (2018). A quantum mechanical interpretation of gravitational redshift of electromagnetic wave. *Optik, 174*, 636–641.
34. Newton, I. (1729). *Mathematical principles of natural philosophy*, (vol. 1). Benjamin Motte.
35. Einstein, A. (1905). Zur elektrodynamik bewegter körper. *Annalen der Physik, 322*, 891–921.

36. Messiah, A. (1965). *Quantum mechanics*, (vol. 1). Wiley.
37. Chang, D. C. (2016). *What is the physical meaning of mass in view of wave-particle duality?*. arXiv preprint physics/0404044v2.

Part V
How to Resolve the Conflict Between Quantum Physics and Relativity?

Chapter 12
The Quantum Origin of the So-Called "Relativistic Relations"

As shown in many textbooks, the foundation of modern physics is built on two theories, i.e., quantum mechanics (QM) and the special theory of relativity (STR). These textbooks, however, often fail to mention that **there is a major conflict between QM and STR; their assumptions about the vacuum are totally different**. According to STR, the vacuum must be empty; otherwise, it will provide a reference frame to determine which inertial frame is stationary and which one is not. This would violate the principle of relativity. Quantum mechanics, on the other hand, requires the vacuum to be not empty. The vacuum is regarded as the ground state of the quantum system; it has very rich physical properties. Thus, the vacuum cannot be empty as assumed in the STR.

Now, we are facing with a dilemma: Do we chose to believe QM or STR? Or, can we find some way to reconcile between the two? In this chapter, we will show that many of the so-called "relativistic effects" can indeed be explained based on quantum physics.

In classical physics, our physical world is described using Newtonian mechanics. However, it was discovered in the early twentieth century that the movement of sub-atomic particles (e.g., photon and electron) appears to follow different rules. These observed deviations from Newtonian mechanics were often attributed to be due to relativity. It was claimed that the physical concepts used in the Newtonian mechanics are only low-speed approximations; they are no longer valid when the speed of the object is increased to approach the speed of light. Thus, many physicists called these non-Newtonian relations "relativistic relations".

This conclusion, however, is a misinterpretation. It fails to recognize that the sub-atomic particles behave differently from Newtonian mechanics is because the quantum physics in the microscopic world is different from the classical physics in the macroscopic world. For the sub-atomic particles, **the observed deviation from Newtonian mechanics was not due to relativity. Instead, it is because the quantum particles making up matters are excitation waves.**

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_12

In this chapter, we will examine in detail the physical basis of the various so-called “relativistic relations”. We will show that these relations have a clear root in quantum physics.

12.1 The Quantum Basis of the So-Called “Relativity Relations”

In the last chapters, we had shown that the *relation of mass-energy equivalence* was not arisen from relativity; instead, it is a direct consequence of the fact that the quantum particle is an excitation wave of the vacuum. Furthermore, the so-called “*relativistic energy–momentum relation*” is really arisen from the *dispersion relation* of the quantum wave function.

In the following, we will show in detail that the other so-called “*relativistic relations*” are also due to the wave nature of the quantum particles. At present, there are many important questions about fundamental physics, including: (1) *Why is the speed of light constant?* (2) *Why can no particle travel faster than light?* (3) *Why is the mass of a particle not constant? Why is mass speed-dependent?* These questions are often explained using STR. In this chapter, we will show that the above questions can be explained more easily based on the quantum wave model.

12.2 Why is the Speed of Light Constant?

One advantage of the quantum wave model is that it can provide a clear physical basis to explain why light must travel at a constant speed of c and no particle can travel faster than c. In classical mechanics, the vacuum is just an empty space. In principle, a particle (including photon) should be able to travel at any speed. Also, the speed of a particle should appear to be different in different moving frames. The results of many light speed measurement experiments, however, indicated that the speed of light is constant [1, 2]. Furthermore, results of Michelson-Morley experiment suggested that the speed of light appears to be the same in all inertial frames [3]. These observations are not consistent with the general expectation of a classical mechanical system.

In order to make his relativity theory agree with the experimental results, Einstein simply assumed that light must travel at a constant speed c in all inertial frames. He did not know the physical basis of this assumption, so he just called it a “postulate” [4].

In our quantum wave model, there is no need to assume that the speed of light is a constant, since this is a prediction of the theory. Here, the photon is just a quantized excitation wave. Like all waves, its traveling speed is determined by the physical properties of the wave medium. The wave equation of light was derived from the Maxwell’s equations. The velocity of the excitation waves in the vacuum is $c = 1/\sqrt{\mu_0 \varepsilon_0}$. Hence, *c is entirely determined by the physical properties of the vacuum medium*. Because the Maxwell equation of light transmission is Lorentz invariant, it remains the same form in different inertial frame. The value of c does not change with the transformation. Thus, the speed of light should be the same in all inertial frames. Therefore, the postulate of STR about the constant speed of light is not a truly required "postulate"; it can be predicted based on Maxwell's electromagnetic theory of light.

12.3 Why Can No Particle Travel Faster Than Light?

In the STR, it requires that no particle can travel faster than the speed of light, otherwise the theory would breakdown. However, one needs to give a physical reason to justify why a particle cannot travel faster than light. In STR, Einstein simply regarded that as a conjecture; he just hypothesized that Nature somehow does not allow any object to travel faster than the speed of light.

The quantum wave model, on the other hand, clearly predicts that no particle can travel faster than the speed of light. The reason is because all particles are quantized excitation waves of the vacuum medium. Their speed of transmission is entirely determined by the physical properties of the wave medium. Since all particles are excitation waves of the same medium (i.e., the vacuum), they must have the same speed limit.

This point can be easily demonstrated mathematically using the results we obtained in Chap. 7. Because the particle is a wave packet, we know its traveling speed is its "group velocity" instead of its "phase velocity". The group velocity is determined by $\partial\omega/\partial k$ and is generally slower than the phase velocity (ω/k).

For photon, its dispersion relation is given by Eq. (7.9),

$$\omega = ck.$$

Its group velocity happens to be the same as its phase velocity, $c = 1/\sqrt{\mu_0\varepsilon_0}$. Thus, the photon always travels at the speed c.

For particles with non-zero rest mass, their *energy–momentum relation* is based on the dispersion relation of the excitation wave, which is given in Eq. (7.14),

$$\omega^2 = \left(k^2 + \ell^2\right)c^2, \tag{12.1}$$

where ℓ is proportional to the rest mass. From this dispersion relation, one can directly calculate the traveling speed of the particle, which is equal to the group velocity of the wave packet, i.e., $v = \dfrac{\partial\omega}{\partial k}$. From Eq. (12.1), its group velocity is

$$v = \frac{\partial\omega}{\partial k} = \frac{ck}{\sqrt{k^2 + \ell^2}} < c. \tag{12.2}$$

For a massive particle, $m_0 \neq 0$; according to Eq. (11.23), $\ell \neq 0$. This explains *why the speed of a massive particle is not constant* and *its maximum speed cannot exceed the speed of light*, *c*.

Thus, particles in general must travel at a speed less than the speed of light (c); and c is the ultimate speed for all particles. Unlike STR, this conclusion is **not** a postulate; it is a consequence of the fact that all particles are excitation waves of the quantum vacuum.

Finally, there is an important point that one should not overlook. The fact that *the maximum speeds for all particles (with or without rest mass) are the same* strongly suggests that, the wave medium carrying the excitation waves of all particles (including massive particles and photons) must be the same (i.e., the same vacuum medium). In another word, **all particles are different excitation waves of the same vacuum medium**!

12.4 Why is Mass Speed-Dependent?

In classical mechanics, the mass of an object is considered to be an intrinsic mechanical property of that object, and it is a constant. At the beginning of the twentieth century, a number of experiments had been conducted to measure the mass of an object at different speed [5–7]. It was found that the mass appears to increase with its speed. Later, many particle accelerators were built to study particle physics. It was fully verified that the mass of a particle is not a constant; instead, it increases with the speed in the following manner:

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{12.3}$$

This phenomenon can be explained very easily using the quantum wave model. In fact, we have explicitly derived the above relation in Chap. 7 (see Eq. (7.21)). Using this wave model, it is not difficult to explain **why mass should increase with the speed of the particle**.

Recall that in our model, all sub-atomic particles are quantized excitation waves of the quantum vacuum. Hence, their limiting speed of propagation is determined by the physical properties of the vacuum medium, which is the speed of light c. Thus, no particle can travel faster than c. When the particle speed v is much smaller than c, one can easily apply force to accelerate the particle. The energy received by the particle is used mainly to increase the particle speed. But this situation will change when the particle speed approaches the speed of light, because it will become very difficult to further increase the speed of the particle. That means, the particle cannot be accelerated as before using the same amount of force. It thus appears that the inertial mass of the particle increases dramatically at high speed. In another word, the input energy is no longer used fully to increase the speed of the particle, instead, it seems that part of the input energy is used to increase the particle mass.

Therefore, if one understands that particle is just a wave, one will expect that the particle cannot be accelerated in a linear manner. For a wave packet, the general relationship between p and v is not a straight line. In Chap. 7, we showed that the energy–momentum relation of a quantum particle is

$$E^2 = p^2c^2 + m_0^2c^4. \tag{12.4}$$

Recall that the velocity of the particle is the group velocity of the wave packet,

$$v = \frac{\partial\omega}{\partial k} = \frac{\partial E}{\partial p}.$$

Using these two relations, one can easily show that

$$p = m_0 v\ (1 - v^2/c^2)^{-1/2}. \tag{12.5}$$

Figure 12.1 is a plot of the momentum p versus the speed v for the Newtonian model and the wave model.

In physics, the momentum p is defined as the product of mass m and speed v, i.e.,

$$p = mv. \tag{12.6}$$

Thus, the mass m is defined as the *slope* of the line connecting the origin with the *p versus v* curve. From Fig. 12.1, it is apparent that this slope can vary with v. In the quantum wave model, the p *vs* v curve appears as a straight line only when v

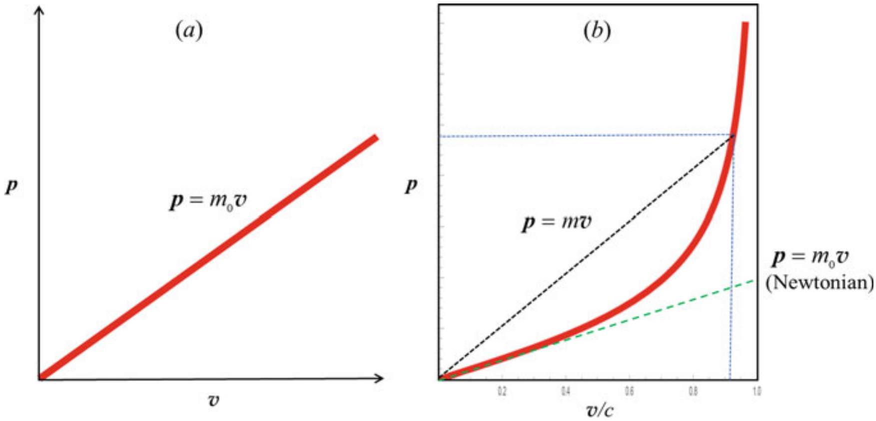


Fig. 12.1 The relationship between momentum and velocity according to different models. a The red line represents the linear relationship between momentum and velocity according to the Newtonian mechanics, which is $p = m_0v$. **b** The red line is a plot of momentum (p) vs speed (v) according to the wave model: $p = m_0v\ (1 - v^2/c^2)^{-1/2}$. Credit: Fig. 12.1(b) is reproduced from an earlier publication of the author: D. C. Chang, arXiv preprint physics/0404044v2 (2016)

is much smaller than c. When the speed of the particle is increased to approach the speed of light, the p *vs* v curve quickly bends upward; this explains why the mass of a quantum particle increases rapidly when the speed of the particle increases to near c (see Fig. 12.1). In fact, from the above two Eqs. (12.5) and (12.6), one can explicitly obtain

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{12.7}$$

Therefore, the mass of a particle under acceleration is not constant. This relation agrees exactly with the results of experiments (i.e., Eq. (12.3)).

From the above discussion, it is clear that the particle view used in classical mechanics is not an accurate description of the dynamic process in the microscopic world. This is because at the sub-atomic level, a quantum particle is a wave packet instead of a point mass. Thus, its energy, momentum, and mass relationships are different from that of Newtonian mechanics. In the old days, people thought that *the mass is an intrinsic physical entity* of an object, thus, m should be a constant. But such an impression is just an illusion at low speed.

12.5 The Physical Basis for the Speed-Dependence of Mass Was Not from STR

In many physics textbooks, it is often claimed that the discovery of mass being speed-dependent is attributed to Einstein [8–10]. For example, in the textbook *Feynman's Lectures on Physics* (Vol 1, Chap. 15), it was explicitly stated: "*For over 200 years the equations of motion enunciated by Newton were believed to describe nature correctly, and the first time that an error in these laws was discovered, the way to correct it was also discovered. Both the error and its correction were discovered by Einstein in 1905.*

Newton's Second Law, ... was stated with the tacit assumption that m is a constant, but we now know that this is not true, and that the mass of a body increases with velocity. In Einstein's corrected formula m has the value $m = \frac{m_0}{\sqrt{1-v^2/c^2}}$, *where the "rest mass"* m_0 *represents the mass of a body that is not moving...*" [9]

Such teaching is aimed to promote Einstein as a genius; this practice is quite common nowadays. However, it is not clear whether such textbook statements were based on facts. A careful examination of the literature record would indicate otherwise. First, it is not true that Einstein was the first one to discover that mass is not a constant. The discovery of *mass varying with speed* had been known from theoretical and experimental studies many years before Einstein published his first relativity paper in 1905 (see below). Second, Einstein's derivation of the speed-dependent mass relations did not give the right results [4]. The correct relation of speed-dependent mass was not discovered by Einstein. In the followings, let us examine the historical records in detail.

At the end of the nineteenth century, many physicists were interested in studying the electrodynamics and kinematics of an electron in the electromagnetic field. They found that the electromagnetic contents of a charged particle could be related with its "inertial mass". In 1899, Lorentz applied his modified aether theory to show that, the mass of a charged particle is not constant. Instead, the mass appears to be dependent on the speed of the particle, and, such speed-dependence is different between the longitudinal direction and the transverse direction [11]. Later in 1904, Lorentz explicitly showed that, the mass of an electron parallel to the direction of motion is $m_L = \gamma^3 m_0$ and the mass perpendicular to the direction of motion is $m_T = \gamma m_0$, where $\gamma = 1/\sqrt{1 - v^2/c^2}$ is called the "Lorentz factor" [12].

In 1902, Max Abraham also published a paper entitled "*Principles of the Dynamics of the Electron*", in which he showed that the "transverse electromagnetic mass" of an electron is also dependent on the speed of the particle [13]. Although his formula was slightly different from Lorentz's results, it also showed that the ratio of m_T/m_0 is a function of v^2/c^2.

At about the same time, several experimental physicists tried to test the theories of Lorentz and Abraham by measuring the variation of mass as a function of speed. In 1901 and 1902, Kaufmann published two papers on this topic [5, 6]. He demonstrated in experiment that the mass of a particle is indeed speed-dependent. In Kaufmann's experiment, the mass to velocity relation was obtained by measuring the deflection of high-speed electrons emitted from a radioactive isotope (radium) in an electric and magnetic field. The momentum of the electron was determined from the magnetic deflection, and its kinetic energy was determined by the electric deflection. Thus, both the velocity v and the *e/m* ratio can be determined for the electron in his experiment. (His experimental apparatus is shown in Fig. 12.2).

The results of Kaufmann's experiments clearly demonstrated that when the velocity of the electron increases toward the speed of light, its mass increases rapidly (see Fig. 12.3). Several years later, more experiments on the speed-dependence of mass were conducted by Bucherer, as well as Guye and Lavanchy [7, 14]. Their results are also plotted on Fig. 12.3. Taking these results together, it appears that all of their data could fit with the so-called "Lorentz relation":

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

which is identical to the mass versus speed equation derived in the quantum wave model, i.e., Eq. (12.7).

Following the works of Lorentz and Kaufmann, Einstein was aware that the mass of an object could vary with speed. He tried to derive such a relation in his 1905 paper on relativity [4]. Like Lorentz, Einstein thought that the mass of an object could be sub-classified into "longitudinal mass" and "transverse mass"; and they are speed-dependent. From his calculation, Einstein showed that the mass of an "electron" can change with its traveling speed, such that

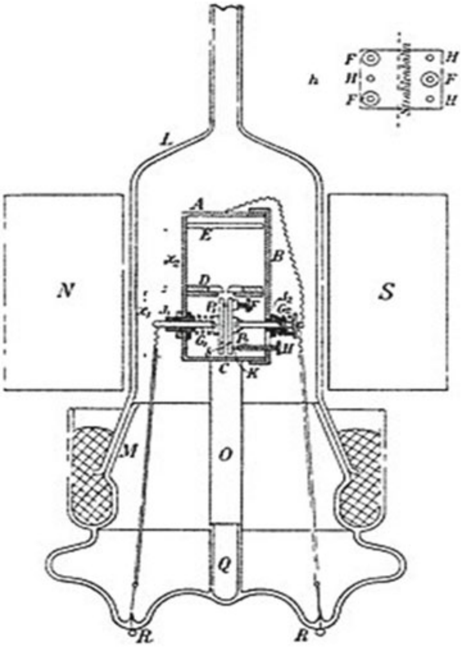


Fig. 12.2 The experimental apparatus used by Kaufmann in 1901 for measuring the variation of the electron mass with different speeds. Image Credit: Fig. 1 of Kaufmann 1901 paper: W. Kaufmann, “Die magnetische und electrische Ablenkbarkeit der Bequerelstrahlen und die scheinbare Masse der Elektronen”, *Göttinger Nachrichten* **2**, 143–168 (1901). Public domain

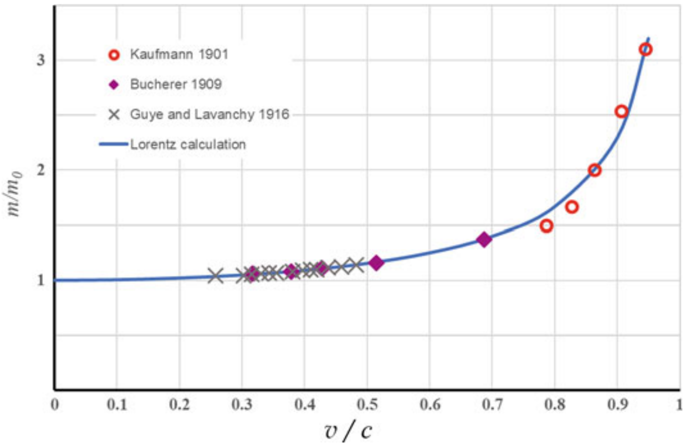


Fig. 12.3 **The mass of an electron varies with its speed**. The red circles, purple diamonds, and gray crosses are the experimental data from Kaufmann 1901, Bucherer 1909, and Guye and Lavanchy 1916, respectively. The solid blue line is the calculated result based on the Lorentz relation

$$\begin{cases} \text{Longitudinal mass} = \dfrac{m_0}{\left(1 - v^2/c^2\right)^{3/2}} & (12.8) \\ \\ \text{Transverse mass} = \dfrac{m_0}{1 - v^2/c^2}. & (12.9) \end{cases}$$

These predictions were later shown to be incorrect; they did not agree with the experimental data shown in Fig. 12.3. Many experiments had been conducted since the beginning of the twentieth century to measure particle mass at different speeds [5–7, 14]. These experimental results generally supported the prediction of Eq. (12.7), but not the results of Einstein's 1905 paper (see Fig. 12.3).

In conclusion, **the experimental findings are entirely consistent with the quantum wave model but not with STR**. From the wave perspective of our model, it is not difficult to understand the physical foundation behind the speed-dependence of mass. This is because the quantum particle is a quantized excitation wave; its traveling speed cannot exceed the phase velocity of the vacuum medium, which is c. When the speed of the particle approaches the speed of light, it will become harder and harder to be accelerated; thus, the mass will appear to be heavier and heavier.

12.6 Newtonian Mechanics is a Limiting Case of Wave Mechanics for a Quantum Particle

In the traditional physics teaching, it is often said that the deviation from Newtonian mechanics at high speed is due to relativity. Now we have a new understanding. According to the quantum wave model, the reason for the particle behaving differently from Newtonian mechanics is because the quantum object is an excitation wave in nature; it looks like a particle in the macroscopic view. This wave packet will behave like a Newtonian particle only at certain physical limit, namely, when the longitudinal wave vector is much smaller than the transverse wave vector, i.e., $k << \ell$.

12.6.1 The Speed of the Quantum Particle Depends on the Ratio of k and ℓ

From the dispersion relation of the quantum wave function representing a massive particle, one can directly calculate the particle's traveling speed, which is the group velocity of the wave packet,

$$v = \frac{\partial \omega}{\partial k} = \frac{ck}{\sqrt{k^2 + \ell^2}}. \tag{12.10}$$

When $k << \ell$,

$$v = \frac{ck}{\sqrt{k^2 + \ell^2}} \approx c\frac{k}{\ell} << c.$$

Thus, the speed of the particle is far smaller than the speed of light when the longitudinal wave vector is much smaller than the transverse wave vector.

On the other hand, in the opposite case that $k >> \ell$, the speed of the particle can approach (but not exceed) the speed of light,

$$v = \frac{ck}{\sqrt{k^2 + \ell^2}} \leq c.$$

Recall that k is related to the particle's momentum p and ℓ is related to the particle's rest mass m_o, the above relation implies that: (1) The speed of a particle with no rest mass ($\ell = 0$) must always travel at the speed of light. (2) The particle's speed will approach the speed of light when the longitudinal wave vector is much larger than the transverse wave vector, (i.e., when the kinetic energy contributed by the momentum is much larger than the resting energy contributed from the rest mass). One may notice that this conclusion is independent of the absolute value of ℓ (i.e., the rest mass of the particle). This explains why *all particles must have the same speed limit c regardless of their rest mass*. It is a consequence of the fact that *all particles are excitation waves of the same vacuum medium*, and the physical properties of the vacuum determine the ultimate speed of wave transmission.

12.6.2 The Energy–Momentum Relation in the Newtonian Mechanics Can Be Derived from the Dispersion Relation of the Quantum Particle When the Kinetic Energy Is Much Smaller Than the Resting Energy

According to the quantum wave model, the wave packet representing a particle has a dispersion relation of

$$\omega^2 = \left(k^2 + \ell^2\right)c^2, \tag{12.1}$$

where ω is the oscillation frequency of the wave packet. When $k << \ell$, this dispersion relation of the wave packet can be approximated using Taylor expansion

$$\omega = c\ell\left(1 + k^2/\ell^2\right)^{1/2} \approx c\ell\left(1 + \frac{1}{2}\frac{k^2}{\ell^2}\right).$$

Multiply $\hbar$ to each side of the above equation, one can obtain

$$\hbar\omega = \hbar\ell c\left(1 + \frac{1}{2}\frac{\hbar^2k^2}{\hbar^2\ell^2}\right). \tag{12.11}$$

Recall that ω, k, and ℓ are related to the energy, momentum, and rest mass of a particle, and $m_0 = \hbar\ell/c$, the above equation becomes

$$E = m_0c^2\left(1 + \frac{1}{2}\frac{p^2}{(m_0c)^2}\right) = m_0c^2 + \frac{1}{2}\frac{p^2}{m_0}. \tag{12.12}$$

At low speed, $p = m_0v$, the above equation becomes

$$E = m_0c^2 + \frac{1}{2}m_0v^2. \tag{12.13}$$

The first term on the right-hand side is identified as the "resting energy", which is a constant for a given particle. The second term on the right-hand side is the kinetic energy of the particle as given in Newtonian mechanics. Thus, from Eq. (12.12), one can clearly see that, the dispersion relation of the quantum particle can be reduced into the energy–momentum relation of Newtonian mechanics when $k << \ell$. Furthermore, it also gives the kinetic energy formula as used in the Newtonian mechanical theory (see Eq. (12.13)).

One may wonder what is the physical meaning of the condition $k << \ell$. As we showed earlier, k is related to the momentum p of the quantum particle. This is just the de Broglie relation. In Chap. 11, we also showed that ℓ is related to the rest mass of the particle. From an energy perspective, the condition $k << \ell$ would mean that the kinetic energy is much smaller than the resting energy of the particle.

12.7 Modification of Newton's Gravitation Law Based on Our New Understanding of Mass in the Wave View

As we know, the gravitational theory in classical mechanics was developed by Newton (see Fig. 12.4), a giant figure in physics. The original formula proposed by the Newtonian gravitational theory had an ambiguity; it is not clear whether the "mass" defined in that equation represents the rest mass or the moving mass of an object. In the day of Newton, people did not know that mass can change with speed. So they thought there is only one type of mass, and which is a constant. With our knowledge today, we know the understanding of Newton's time is not correct.

Now, we must decide whether the mass term in the Newtonian gravitational law represents the moving mass or the rest mass. The answer is not that difficult to find. In the Newtonian mechanics, the mass is defined from the momentum. So the mass there is called "inertial mass", which is equivalent to the "moving mass" we understand today. **In the Newtonian theory, the gravitational mass should be identical to the inertial mass**. Thus, *the gravitational mass included in Newton's gravitational law*

Fig. 12.4 Isaac Newton. Isaac Newton (1642–1726) was an English mathematician and physicists. He was like the father of classical mechanics. The studies of mechanics and astronomy are based on Newton's gravitational theory and Newton's laws. Furthermore, he had contributed to a wide range of topics including optics and calculus. Photo credit: Portrait of Isaac Newton (1642–1727) by Godfrey Kneller (1646–1723); Wikimedia commons, Public domain

must be equal to the moving mass instead of the rest mass of that object. This means that in the Newtonian gravitational law, we should identify the two masses involved (M and M') as the *moving masses*. In another word, we can generalize Newton's gravitational law by identifying

$$M \equiv m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{12.7a}$$

Then, the gravitational force will depend not only on the rest masses of the interacting objects and the distance between them, it will also depend on the speeds of movement of the two interacting objects. In another word, the generalized Newton's gravitational law now becomes

$$\begin{aligned} F = G\frac{MM'}{r^2} &= G\frac{m_0 m_0'}{r^2}\left\{\left[1 - (v/c)^2\right]\left[1 - \left(v'/c\right)^2\right]\right\}^{-1/2} \\ &= G\frac{m_0 m_0'}{r^2}\left[1 + \frac{1}{2}\left(v^2/c^2 + v'^2/c^2\right) + \boldsymbol{O}\right], \end{aligned} \tag{12.14}$$

where v and v' are the speed of the interacting objects as measured from the vacuum system, $\boldsymbol{O}$ is the higher-order terms (v^4/c^4,v'^4/c^4,$v^2 v'^2/c^4$ and higher), which could be ignored in most astronomical studies. This revised Newton's law can be tested in future experiments. One can carefully analyze the movement of objects around a star or movements of stars in a galaxy. One may find that the movement could deviate slightly from the calculation based on the original Newton's law. Such deviations can be tested against the predictions of the revised Newton's law as described in Eq. (12.14).

12.8 Gravity Can Be Understood as a Consequence of Energy-Attracting-Energy

Now, since the moving mass is not a constant of the object, it does not enjoy a special role as an inherent property of that object. In Newton's gravitational law, the gravitational force between two objects is supposed to be generated by their masses. If mass is no longer regarded as an intrinsic property of the particle in the wave view, could one find a deeper reason for the generation of the gravitational force?

A possible answer could be that: *Newton's gravitational law can be rewritten based on the energy of the object rather than the mass of the object*. Since the mass m is proportional to the energy of a particle E by the relation $E = mc^2$, we can express Newton's gravitational law as

$$F = G\frac{m_1 m_2}{r^2} = \frac{G}{c^4}\frac{E_1 E_2}{r^2} = G'\frac{E_1 E_2}{r^2}, \tag{12.15}$$

where $G' = G/c^4$. In another word, we can interpret the source of the gravitational force as *energy-attracting-energy* instead of *mass-attracting-mass*!

This energy here is the total energy of the object, which includes both the resting energy E_0 and the kinetic energy cp. Thus, for particles that have no resting energy, they can still be attracted in a gravitation field. This explains why light can be bent when it passes through the vicinity of a massive object. It has been observed in astronomical studies that when light passes through a galaxy, it can generate a "lensing effect" [15–17]. This is a convincing demonstration that the massless photon can interact with a gravitational field.

12.8.1 Speed of Gravitational Wave Equal c Implies that Gravitational Force Is Transmitted Through the Vacuum Medium

Another major question about gravity is: what is the medium that transmits the gravitational force? In our model, the vacuum is filled with only one type of medium, that is, the quantum vacuum. Therefore, the simplest physical assumption is that, *like the electromagnetic force, the gravitational force is also transmitted through the same quantum vacuum*. In fact, there is an important piece of evidence for supporting this assumption. According to the current understanding, the gravitational force is transmitted at a speed equal of c (the speed of light) [18]. Therefore, it is highly possible that the gravitational force is transmitted through the same medium as the transmission of light. This may further suggest that the mechanism of the gravitational transmission could also be based on a similar mechanism related to the excitation of the quantum vacuum.

12.9 Chapter Summary

- There is a serious conflict between quantum mechanics and relativity. In relativity, the vacuum must be empty; otherwise, it will provide a reference frame to determine which inertial frame is stationary and which one is not. Quantum mechanics, on the other hand, regards the vacuum as the ground state of the quantum system, which cannot be empty.
- Why is the speed of light constant? This is because the photon is a quantized excitation wave of the vacuum, its traveling speed is determined by the physical properties of the wave medium. From the Maxwell's theory, the velocity of light in the vacuum is $c = 1 \big/ \sqrt{\mu_0 \varepsilon_0}$, which is a constant regardless of the chosen coordinates.
- Why can no particle travel faster than light? This is because all particles are excitation waves of the vacuum; their traveling speed v is determined by the *group velocity* of its wave packet. According to the *dispersion relation* obtained in the quantum wave model, the group velocity of a massless particle (photon) is c, while the group velocity of a massive particle cannot exceed c.
- Why is the mass of a particle speed-dependent? This is because the quantum particle is a wave, its traveling speed cannot exceed the speed of light c. Suppose the particle is accelerated using a fixed amount of force. When the particle speed v is much smaller than c, one can easily accelerate the particle. This situation changes when the particle speed v approaches the speed of light, it will be very difficult to further increase v. It thus appears that the inertial mass of the particle increases dramatically at high speed.
- Since the gravitational mass is equal to the inertial mass, which is speed-dependent, the Newton's gravitational law needs to be revised to accommodate this new finding. We propose that the gravitational force will depend not only on the rest masses of the interacting objects and the distance between them, but it will also depend on the speeds of movement of the two interacting objects.
- Conceptually, *gravity* can now be understood as a force of *energy-attracting-energy*, instead of *mass-attracting-mass*. Finally, the recently reported finding that the speed of gravitational wave equals to c suggests that the gravitational force, like the electromagnetic force, is transmitted through the vacuum medium (i.e., the *quantum vacuum*).

References

1. This Month In Physics History. (2010). July 1849: Fizeau publishes results of speed of light experiment. *APS News*, *19*.
2. Michelson, A. A. (1879). Experimental determination of the velocity of light. *American Journal of Science*, *s3–18*, 390–393.
3. Michelson, A. A., & Morley, E. W. (1887). On the relative motion of the Earth and the luminiferous ether. *American Journal of Science, 34*, 333–345.

4. Einstein, A. (1905). Zur elektrodynamik bewegter körper. *Annalen der Physik, 322*, 891–921.
5. Kaufmann, W. (1901). Die magnetische und electrische Ablenkbarkeit der Bequerelstrahlen und die scheinbare Masse der Elektronen. *Göttinger Nachrichten, 2*, 143–168.
6. Kaufmann, W. (1902). Die elektromagnetische masse des elektrons. *Physikalishce Zeitschrift, 4*, 54–56.
7. Bucherer, A. H. (1909). Die experimentelle bestätigung des relativitätsprinzips. *Annalen der Physik, 28*, 513–536.
8. French, A. P. (1968). *Special relativity*. W. W. Norton & Company.
9. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Chapter 15. The special theory of relativity. In *The Feynman lectures on physics*, (Vol. 1). Addison-Wesley.
10. Resnick, R. (1968). *Introduction to special relativity*. Wiley.
11. Lorentz, H. A. (1899). Simplified theory of electrical and optical phenomena in moving systems. *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings, 1*, 427–442.
12. Lorentz, H. A. (1904). Electromagnetic phenomena in a system moving with any velocity smaller than that of light. *Proceedings R. Netherlands Academia Arts Science, 6*, 809–831.
13. Abraham, M. (1902). Prinzipien der Dynamik des Elektrons. *Physikalishce Zeitschrift, 4*, 57–63.
14. Guye, C.-E., & Lavanchy, C. (1916). Vérification expérimentale de la formule de Lorentz-Einstein par les Rayons cathodiques de grande vitesse [suite et fin]. *Archives des Sciences Physiques et Naturelles*, *4e période*, 441–448.
15. Refsdal, S., & Surdej, J. (1994). Gravitational lenses. *Reports on Progress in Physics, 57*, 117.
16. Wong, K. C., Tran, K.-V.H., Suyu, S. H., Momcheva, I. G., Brammer, G. B., Brodwin, M., Gonzalez, A. H., Halkola, A., Kacprzak, G. G., Koekemoer, A. M., Papovich, C. J., & Rudnick, G. H. (2014). Discovery of a strong lensing galaxy embedded in a cluster at $z = 1.62$. *Astrophysics Journal Letters, 789*, L31.
17. Refregier, A. (2016). Weak gravitational lensing by large-scale structure. *Annual Review of Astronomy and Astrophysics, 41*, 645–668.
18. Berti, E. (2016). The first sounds of merging black holes. *Physics, 9*, 17.

Chapter 13
The Physical Basis of Lorentz Transformation and Minkowski's Four-Dimensional Space–Time

In the special theory of relativity (STR), the coordinates in a stationary inertial frame and the coordinates in a moving frame can be converted between each other using the Lorentz transformation. Thus, many people thought the applicability of Lorentz transformation is a verification of STR. Such thinking is not true. The Lorentz transformation was first proposed by Hendrik Lorentz at the end of the nineteenth century based on his aether theory; it was developed to explain the finding of the Michelson-Morley experiment that the propagation of light is unchanged in different moving frames [1–3]. Later, Einstein showed in 1905 that one could greatly simplify the derivation of the Lorentz transformation using STR [4]. However, his derivation was not unique. In many models, Lorentz transformation can be derived based on the requirement that the equation of motion of light does not change between two different inertial frames.

Therefore, the derivation of the Lorentz transformation does not require the assumption of an empty vacuum. There is a clear difference between Lorentz invariance and the principle of relativity. In the case of Lorentz invariance, it only requires that a transformation of coordinate from (t, x, y, z) into (t', x', y', z') would not change the mathematical form of the equation of motion. It is purely a mathematical concept; it does not prove whether the universe has a fixed resting frame or not. The principle of relativity is far more stringent; it assumes that there is no fixed resting frame in our universe. Thus, the vacuum must be an empty space. In such a case, any inertial frame can be chosen as a stationary frame.

In the quantum wave model discussed in this book, the vacuum of course is not empty. In fact, we hypothesize that the quantum vacuum is a dielectric medium. In this chapter, we will show that, with this non-empty vacuum, one can easily derive the Lorentz transformation.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_13

13.1 The Measurements of *Space* and *Time* in a Wave Propagating System

13.1.1 How to Define Space and Time? The Particle View vs the Wave View

Our physical world can be described as four-dimensional, i.e., three dimensions in *space* and one dimension in *time*. For an inertial system, how can one define the length in *time* or *space*? Apparently, one needs some sort of measuring ruler (i.e., a reference system) to do that. Then, what sort of reference system can one use?

In the classical mechanical system, the basic building blocks of the material world are particles, which can move freely in an empty space following certain mechanical laws, such as Newton's Laws. There is no natural reference system for measuring the length of *space* and *time*. The situation is different in a wave system. For example, in the view of the *quantum wave model* discussed in this book, our Universe is filled with a vacuum medium; the "particles" are excitation waves of the vacuum. In this case, there is a natural reference system to measure the time and space. For instance, one can use the propagation of light (photon) to define the length of time and space. More specifically, *time* can be determined from the *frequency* of a specific type of light, while the *space* can be measured using the *wavelength* of this light.

In this case, one can easily see that *space* and *time* could appear as relative in a wave propagating system. We know from the *Doppler effect* that the frequency of a wave can shift in a moving frame in comparison to a stationary frame. Hence, if one uses a wave system as the reference for an inertial system, it is not surprising that one will find the *time* and *space* are not absolute; they can vary depending on the motional state (speed) of the inertial system.

13.1.2 Graphical Analysis of the Changes of Space and Time Between a Stationary Frame and a Moving Frame. The Physical Basis of Time Dilation and Lorentz Contraction from the Wave View

Why can time and space vary in different inertial frames? It is basically like an "*optical Doppler effect*". Let us use a concrete example to demonstrate this point. Suppose S is an inertial frame which is stationary relative to the vacuum; S' is an inertial frame that moves at a speed v relative to the stationary frame S (see Fig. 13.1).

In the stationary frame S, light wave is propagating in the way shown in the Fig. 13.2. For simplicity, we assume light wave is a plane wave. This light wave has regular frequency and wavelength. Suppose there is a clock attached to the inertial frame S, how can one measure the time in the S frame using the wave system? One

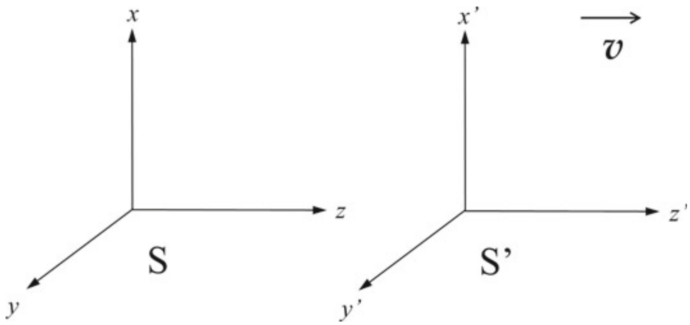


Fig. 13.1 A stationary frame *S* and a moving frame *S*' with a speed *v*

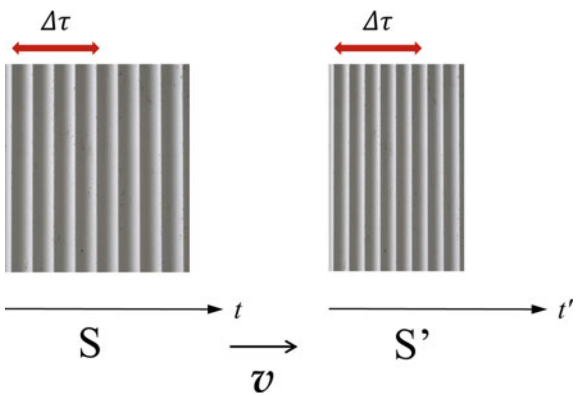
can count the number of oscillations within a fixed time period ($\Delta\tau$); suppose this number is *n*. Then, one can define one unit of time (Δt) in the *S* frame as the time period that contains *n* number of oscillations of the wave, i.e.,

$$\Delta t = \text{a time period for observing } n \text{ oscillations.}$$

Now, for an observer located in the moving frame *S*', he will also observe that the light wave is propagating as a series of plane waves. But because *S*' is moving, these plane waves become slightly compressed (See Fig. 13.2). That means the observed frequency of light wave in the *S*' frame is higher than that of the *S* frame. Suppose a clock identical to the clock in *S* frame is attached to the *S*' frame, one can measure the time in the *S*' frame by counting the number of oscillations of the light wave within the fixed time period $\Delta\tau$. Suppose this number is *n*'. Then, the unit of time in the *S*' frame ($\Delta t'$) is determined to be

$$\Delta t' = \text{a time period for observing } n' \text{ oscillations.}$$

Fig. 13.2. The time dilation in the moving frame *S*' compared to the stationary frame *S*. Suppose an identical plane wave light is propagating in the *S* frame and *S*' frame. The frequency of light waves observed in the *S* frame and *S*' frame would appear to be different. This could be interpreted as that the clocks in the *S* and *S*' frames may run at different rates



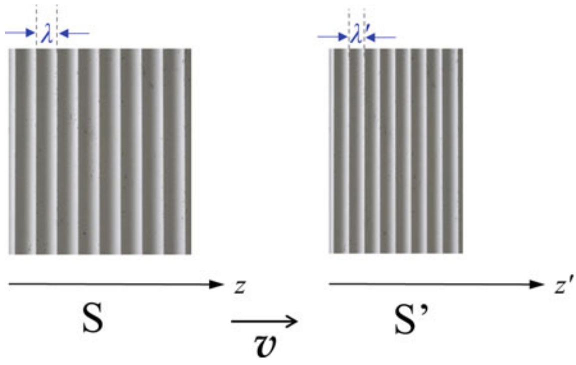


Fig. 13.3 The length contraction in the moving frame *S*' in comparing to the stationary frame *S*. The wavelength is denoted λ in *S* frame and λ' in *S*' frame. As indicated by the blue arrows in the figure, one can see that λ > λ'. This could explain the Lorentz contraction for the moving frame

Since the observed oscillation frequency in the *S*' frame is higher than that in the *S* frame, the number of oscillations within the fixed time period ($\Delta\tau$) observed in the *S*' frame (*n*') is larger than the number of oscillations observed in the *S* frame (*n*), i.e., $n' > n$. Thus, based on the above analysis, it would appear that $\Delta t' > \Delta t$. This explains why one may observe a "time dilation" in the moving frame.

Similarly, if one compares the wave oscillation in the spatial domain between the *S* frame and the *S*' frame, one should also find that the wave oscillation in the moving *S*' frame is compressed in comparison to the wave oscillation observed in the stationary *S* frame (see Fig. 13.3). Suppose the light wave is propagating along the *z*-direction, along which the *S*' frame is moving away from the *S* frame. The length of space (in the *z*-direction) in each inertial frame can be determined from the wavelength of the oscillation wave (λ). Due to the compressing of the oscillation wave in the moving frame, one will find that the wavelength in the *S*' frame is shorter than the wavelength in the *S* frame. This explains why the length of space appears to be contracted in a moving inertial frame; this observation is called "Lorentz contraction".

13.2 Implication of the Michelson-Morley Experiment: Light Propagation is Independent of the Inertial Frame

13.2.1 The Wave Equation of the Vacuum is Unchanged Under the Lorentz Transformation

Based on the results of the Michelson-Morley experiment, we know light propagation is independent of the inertial frame [5]. That means the wave equation of light in the vacuum should be unchanged between the stationary frame and the moving frame. From the Maxwell's theory, we know the wave equation for light is

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = 0, \tag{13.1}$$

or,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = 0,$$

where $\psi = \psi(\boldsymbol{x}, t)$. Results of the Michelson-Morley experiment demanded that this equation should remain the same in different inertial frames. Suppose we designate two arbituary inertial frames as S_1 and S_2, where S_2 frame is moving at a speed v relative to S_1 frame along the z-axis. The wave equations of light in the S_1 and S_2 frame will be written as

$$\begin{cases} \nabla_1^2\psi_1 - \dfrac{1}{c^2}\dfrac{\partial^2\psi_1}{\partial t_1^2} = 0 \\ \nabla_2^2\psi_2 - \dfrac{1}{c^2}\dfrac{\partial^2\psi_2}{\partial t_2^2} = 0. \end{cases} \tag{13.2}$$

This requirement that the wave equation must appear the same in the S_1 frame as in the S_2 frame is called "*Lorentz invariant*". (Note: In the literature, this is sometimes called "*Lorentz covariant*".) The transformation of $(\boldsymbol{x}_1, t_1) \to (\boldsymbol{x}_2, t_2)$ is called the "*Lorentz transformation*".

Without saying, since the transformation of the wave equation from S_1 frame to S_2 frame is Lorentz invariant, the solution of this wave equation is also Lorentz invariant. That means when the wave function is transformed from the S_1 frame to S_2 frame, the wave function would look exactly the same in form.

13.2.2 Solutions of the Wave Equation of Light in Different Inertial Frames

As shown in Chap. 7, the wave function of the photon is the plane wave solution of the above wave equation, i.e.,

$$\psi(\boldsymbol{x}, t) \sim e^{i(k\boldsymbol{x} - \omega t)}. \tag{13.3}$$

For simplicity, let us assume that the light wave is propagating in the z-axis, i.e., $\hat{k} \| \hat{z}$. The wave function in the stationary S_1 frame then is in the form of

$$\psi_1(\boldsymbol{x}_1, t_1) \sim e^{i(k_1 z_1 - \omega_1 t_1)}, \tag{13.4}$$

where $\omega_1 = ck_1$. This is the well-known dispersion relation of light. The wave function can be simplified as

$$\psi_1 \sim e^{i(z_1 - ct_1)}. \tag{13.5}$$

Since the wave equation of light is Lorentz invariant, the wave function of light in the inertial frame S_2 should also appear as

$$\psi_2(\boldsymbol{x}_2, t_2) \sim e^{i(k_2 z_2 - \omega_2 t_2)}, \tag{13.6}$$

where $\omega_2 = ck_2$. Thus, the wave function of light in the S_2 frame can be written as

$$\psi_2 \sim e^{i(z_2 - ct_2)}. \tag{13.7}$$

From the above analysis, one can easily identify the equation of motion for the trajectory of the photon. We know light is a propagating wave. In order to follow the track of the light wave, one can follow the movement of the peak of the wave. It is easy to see that, in the S_1 frame, the peaks of $\psi_1 \sim e^{i(z_1 - ct_1)}$ is moving in a way that satisfies the condition

$$z_1 - ct_1 = 0. \tag{13.8}$$

Similarly, the trajectory of the photon in the S_2 frame is given by

$$z_2 - ct_2 = 0. \tag{13.9}$$

13.3 Derivation of the Lorentz Transformation in the Quantum Wave Model

In order to meet the requirement of invariance of light propagation in different inertial frames, one must transform both the spatial coordinate z and the time dimension t during the "*Lorentz transformation*". Since S_2 frame is moving with velocity v in the z-axis relative to S_1 frame, their spatial coordinates in the transverse direction (i.e., x and y axes) do not need to change, i.e.,

$$x_1 = x_2$$
$$y_1 = y_2$$

Then, one only needs to transform $(z_1, t_1) \rightarrow (z_2, t_2)$. Such a *two-dimensional* coordinate transformation can be taken by assuming

$$\begin{cases} z_2 = a_1 z_1 + b_1 t_1 \\ z_1 = a_2 z_2 + b_2 t_2 \end{cases} \tag{13.10}$$

where a_1, b_1, a_2, b_2, are converting factors to be determined later. From the previous analysis, we know that, at $z_2 = 0$,

$$z_1 = vt_1,$$

which implies $b_1/a_1 = -v$. Also, at $z_1 = 0$,

$$z_2 = -vt_2,$$

which implies $b_2/a_2 = v$. Therefore, Eq. (13.10) becomes

$$\begin{cases} z_2 = a_1(z_1 - vt_1) \\ z_1 = a_2(z_2 + vt_2) \end{cases}. \tag{13.11}$$

Because S_1 and S_2 frames can be arbitrarily chosen, the factor a must be frame-independent, i.e., $a_1 = a_2 = a$. Thus, the above relations become

$$\begin{cases} z_2 = a(z_1 - vt_1) \\ z_1 = a(z_2 + vt_2) \end{cases}. \tag{13.12}$$

From Eqs. (13.8) and (13.9), we know $z_1 = ct_1$ and $z_2 = ct_2$ along the trajectory of the photon. Thus, from Eq. (13.12), one can have

$$\begin{cases} ct_2 = a(ct_1 - vt_1) \\ ct_1 = a(ct_2 + vt_2) \end{cases}. \tag{13.13}$$

Combining the above two relations, one can get

$$c^2 = a^2(c^2 - v^2).$$

Then, one can identify the factor a to be

$$a = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} = \gamma.$$

where γ is the famous "*Lorentz factor*". Thus, Eq. (13.12) becomes

$$\begin{cases} z_2 = \gamma(z_1 - vt_1) \\ z_1 = \gamma(z_2 + vt_2) \end{cases}. \tag{13.14}$$

Table 13.1 The Lorentz transformation

$x_2 = x_1$	$x_1 = x_2$
$y_2 = y_1$	$y_1 = y_2$
$z_2 = \gamma(z_1 - vt_1)$	$z_1 = \gamma(z_2 + vt_2)$
$t_2 = \gamma(t_1 - vz_1/c^2)$	$t_1 = \gamma(t_2 + vz_2/c^2)$
where v is the velocity of S_2 frame as measured in the S_1 frame, and $\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1/2}$	

This is the Lorentz transformation for the spatial coordinate z between the inertial frame S_1 and S_2. Using the above equations, one can show that the Lorentz transformation for the time coordinate t between the inertial frame S_1 and S_2 is

$$\begin{cases} t_2 = \gamma(t_1 - vz_1/c^2) \\ t_1 = \gamma(t_2 + vz_2/c^2) \end{cases}. \tag{13.15}$$

In summary, the Lorentz transformation between the coordinates in the S_1 frame and S_2 frame is given in the Table 13.1.

From the above discussions, one can see that the Lorentz transformation is really based on the requirement of invariance of the propagation of light in different inertial frames. Such a requirement is reflected in the results of the optical interferometer experiments and the Maxwell theory of light wave propagation. The derivation of the Lorentz transformation does not require the vacuum to be empty.

13.3.1 How Does the Quantum Wave Model Differ from STR on Lorentz Transformation?

From the above analysis, one can see that the Lorentz transformation can be easily derived based on the quantum wave model. There is a major difference between the results shown above and the results of STR. In the case of STR, it excludes the aether model and regards the vacuum as an empty space. Hence, any frame could be regarded as a stationary frame.

In the quantum wave model, the argument is different. We showed that, in order to explain the optical Doppler effect, a two-dimensional coordinate transformation is required. From the requirement of invariance for light propagation in different frames, one can derive the Lorentz transformation. Thus, the coordinate transformation does not imply the absence of a wave medium. In our derivation of Lorentz transformation, it never requires the light wave to propagate in an empty space. (In fact, if the vacuum is empty, it will be impossible to support the propagation of the light wave.) The Lorentz transformation is between any two arbitrary inertial frames; it does not require one of these frames to be the stationary frame.

In Einstein's original paper published in 1905, he stated that his theory of relativity is based on two postulates [4]:

(1) *The same laws of electrodynamics and optics will be valid for all frames of reference.*
(2) *Light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body.*

Later, as the theory of relativity became more popular, the first postulate was greatly expanded to cover all physical laws. Thus, in the modern-day version of the STR, it is claimed to be built on two postulates [6, 7]:

(1) *All inertial frames are equivalent with respect to the laws of physics.*
(2) *The speed of light in empty space always has the same value c.*

As we had shown in the above, it is not necessary to have these postulates for deriving the Lorentz transformation.

First, **Postulate 1 is an unsubstantiated over-claim.** What one needs to show is that the propagation of light appears to be the same in different inertial frames. This can be accomplished based on the results of the Michelson-Morley experiment. Certainly, the optical interferometer results only implied that *light propagation* is independent of the inertial frame, they did not imply that *all laws of physics* are the same in all inertial frame.

Second, **Postulate 1 cannot be true** if the vacuum is a wave medium. In this case, one can determine which frame is stationary.

Finally, **Postulate 2 is unnecessary**. It does not need to have the status of a "*postulate*", since it is a direct result of the Maxwell theory of light and the Michelson-Morley experiment. We know $c = \sqrt{1/\varepsilon_0\mu_0}$, which is determined by the physical properties of the vacuum. Thus, it is obvious that the speed of light should not change in different inertial frames.

13.4 The Four-Dimensional Space–Time as Proposed by Minkowski

One attractive feature of STR is the concept of four-dimensional space–time. This concept is made famous by Minkowski's proposal of treating space–time in a four-vector framework. His proposal of using the *matrix tensor* to connect the *covariant four vector* with the *contravariant four vector* is now a standard treatment in particle physics.

The beautiful mathematical framework proposed by Minkowski has helped tremendously for the acceptance of STR in general public. But in reality, is our world a truly four-dimensional world? Actually, there is room for debate. As we will show in **Appendix F**, if one uses the *Hilbert space* (which appears more natural in comparing to human perception) to describe a multi-dimensional world, time is not exactly a real dimension, like what we experience in the axis of a rectangular system.

It may be more appropriate to call *time* as a "*pseudo-dimension*". If one chooses to use the mathematical framework of Minkowski, then, the wave equation of light propagation may look like a four-dimensional world of space–time. (For details, see **Appendix F**.)

The Lorentz transformation is the basis for the establishment of Minkowski's four-dimensional space–time framework. But in reality, the proposal of regarding time as the 4^{th} dimension was first made by Poincare in around 1906 [8]. A short history on the development of the four-dimensional space–time concept is summarized in **Appendix F**.

13.4.1 Mathematical Physics and Reality

In the development of a physical theory, it usually involves two parts: (1) the physical concept and (2) the mathematical framework. A good mathematical framework can help to make the physical theory look simple or beautiful. Usually, in the eye of the readers, people tend to accept more easily a theory if it can be expressed in a very compact form in mathematics, or, if the mathematical presentation looks beautiful, e.g., the notation or equation is highly symmetrical. That is why sometimes a certain mathematical physics work can become very popular although its physical concept might be highly questionable.

One way to make a physical concept look simple is to use a matrix to describe the physical entities in different dimensions of space–time. In this regard, Minkowski (and later Einstein) had done a very successful job.

Of course, physics is for the understanding of nature, it is more than a beautiful theory of mathematical physics. To evaluate the validity of an important physical theory, one must carefully test its postulates in experiment. There is no exception for STR. So, in order to test the validity of STR, one must design some critical experiments. Recently, we have proposed such an experiment based on a precise measurement of the moving mass of a particle [9]. This experiment will be summarized in the next chapter.

13.5 Chapter Summary

- In the *quantum wave model* discussed in this book, our Universe is filled with a vacuum medium; the *particles* are excitation waves of the vacuum. In this case, there is a natural reference system to measure *time* and *space*. For instance, one can use the propagation of light (photon) to define the length of *time* and *space*. More specifically, *time* can be determined from the *frequency* of a specific type of light, while *space* can be measured using the *wavelength* of this light.

- Then, both *space* and *time* could appear as being *relative*. We know from the *Doppler effect* that the *frequency* of a wave can shift in a moving frame in comparison to a stationary frame. Hence, if one uses a wave system as the reference for an inertial system, it is not surprising that one will find the *time* and *space* are not absolute; they can vary depending on the motional state (speed) of the inertial system.
- Based on the results of the Michelson-Morley experiment, we know light propagation is independent of the inertial frame. That means the wave equation of light in the vacuum should be unchanged between two inertial frames moving at different speeds. This requirement that the wave equation must appear the same in the S_1 frame as in the S_2 frame is called "*Lorentz invariant*".
- The Lorentz transformation can be easily derived based on the quantum wave model. Unlike the derivation in STR, which regards the vacuum as an empty space, and hence any frame could be regarded as a stationary frame, the quantum wave model uses a different argument. We showed that, in order to explain the optical Doppler effect, a two-dimensional coordinate transformation is required. From the requirement of invariance for light propagation in different frames, one can easily derive the Lorentz transformation. Thus, the coordinate transformation does not imply the absence of a wave medium.
- The Lorentz transformation is the basis for the establishment of Minkowski's four-dimensional space–time framework. But in reality, the proposal of regarding *time* as the 4^{th} dimension was first made by Poincare in around 1906. A short history on the development of the four-dimensional space–time concept is summarized in **Appendix F**.
- Is our world truly a four-dimensional world? There is room for debate. If one uses the *Hilbert space* to describe a multi-dimensional world, *time* is not exactly a real dimension. It may be more appropriate to call *time* a "*pseudo-dimension*". If one chooses to use the mathematical framework of Minkowski, then, the wave equation of light propagation may look like a four-dimensional world of space–time. (For details, see **Appendix F**.)

References

1. Lorentz, H. A. (1895). *Versuch Einer Theorie Der Electrischen Und Optischen Erscheinungen in Bewegten Körpern.* E. J. Brill.
2. Lorentz, H. A. (1899). Simplified theory of electrical and optical phenomena in moving systems. *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings, 1*, 427–442.
3. Lorentz, H. A. (1904). Electromagnetic phenomena in a system moving with any velocity smaller than that of light. *Proceeding of Royal Netherland Academic Arts Science, 6*, 809–831.
4. Einstein, A. (1905). Zur elektrodynamik bewegter körper. *Annalen der Physik, 322*, 891–921.
5. Michelson, A. A., & Morley, E. W. (1887). On the relative motion of the Earth and the luminiferous ether. *American Journal of Science, 34*, 333–345.
6. French, A. P. (1968). *Special relativity.* W. W. Norton & Company.

7. Einstein, A. (2005). *Relativity : The special and general theory.* Pi Press.
8. Poincaré, H. (1906). Sur la dynamique de l'électron. *Rendiconti del Circolo matematico di Palermo, 21*, 129–176.
9. Chang, D. C. (2017). Is there a resting frame in the universe? A proposed experimental test based on a precise measurement of particle mass. *European Physical Journal Plus, 132*, 140.

Chapter 14
Experimental Tests on the Principle of Relativity and the Twin Paradox

Both quantum mechanics and relativity are important theories in modern physics. However, there is a serious conflict between them, namely, their views on the vacuum are totally different. In quantum mechanics, the vacuum is not regarded as an empty space; the vacuum is just the ground state of the quantum system, which is thought to have very rich physical properties. The special theory of relativity (STR), on the other hand, requires that the vacuum must be an empty space, otherwise, the vacuum would become an absolute resting frame in our universe. Using such a reference frame, one can easily determine which inertial frame is at rest and which one is moving. This would defeat the *Principle of Relativity*.

Physicists today are thus facing with a dilemma: which view of the vacuum should one believe? In science, all important questions must be settled in experiment. Therefore, there is a strong need for us to design new experiments to test whether there is an absolute resting frame in our universe or not.

In this chapter, we will discuss two new experiments for testing the *Principle of Relativity*. The first experiment is to test whether the speed-dependence of moving mass is the same in different inertial frame. The second experiment is to test the famous *Twin Paradox* using extra-terrestrial clocks.

14.1 Experimental Test #1: Is There a Universal Resting Frame in Our World? Measuring the Particle's Mass in Different Moving Frames

A simple experiment to test whether there is a resting frame in our universe or not would be to measure the moving mass of a particle in an inertial frame oriented in different direction in reference to the solar system. Detailed design of this proposed experiment had been reported in our earlier paper [1].

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_14

We know the mass of a particle is not constant; it is dependent on its traveling speed, such that,

$$M = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \tag{14.1}$$

where M is the moving mass and m_0 is the rest mass. When the speed of the particle increases, the moving mass of the particle will also increase. Using such a principle, it is possible to determine if a laboratory frame is in motion relative to the resting frame of our universe.

The basic experimental setup is shown in Fig. 14.1. For simplicity, the charged particles used for this measurement could be electrons. Using an accelerator, the electrons are accelerated to a speed u. The speed of the outgoing electrons can be measured using a Time-of-Flight device. Using a switching magnet, the electron beams can be directed either to a spectrometer at the right or a spectrometer at the left. The masses determined in these two spectrometers (M_R and M_L, respectively) are then compared.

This experiment will be repeated at different time of the day, and in different days of the year. We will examine if any non-zero reading for $\Delta M = M_R - M_L$ can be detected, and whether the measured ΔM will vary with the orientation of the laboratory frame.

If the STR is correct, all inertial frames will be equivalent. Then, the laboratory frame can be regarded as the stationary frame. The moving mass should be the same regardless of the particle's moving direction. But if the vacuum is a dielectric medium as proposed in the quantum wave model, there is a resting frame in our universe. One will observe a difference in the moving mass when the particle changes its traveling direction. This is because the speed of the particle is a vector sum of the particle's

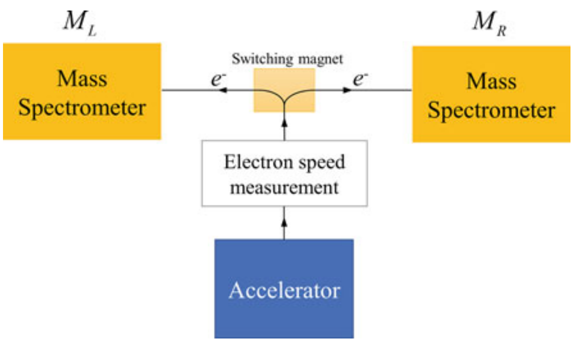


Fig. 14.1 Conceptual diagram of the experimental setup. Electrons accelerated by an accelerator are analyzed by two identical mass spectrometers located at the left side and the right side. If the STR is correct, the laboratory frame can be regarded as the stationary frame; there is no speed difference between particles going to the left or going to the right. The measured moving mass of the two particles will be the same. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Eur. Phys. J. Plus* **132**, 140 (2017)

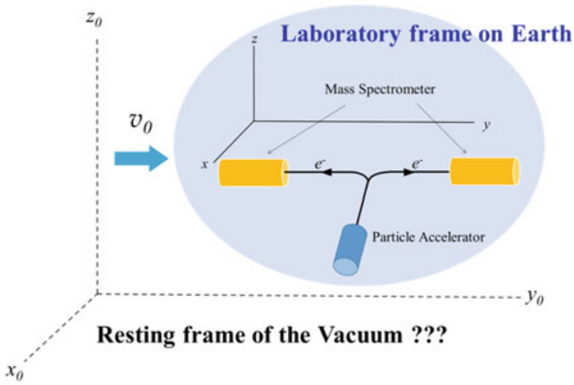


Fig. 14.2 The basic idea of our proposed experiment. If the vacuum is a universal resting frame, the particle traveling to the right should have a combined speed faster than the particle traveling to the left. The measured moving mass of the two particles will be different

velocity relative to the laboratory frame and the velocity of the laboratory frame relative to the vacuum (see Fig. 14.2).

We proposed to conduct this experiment in a way similar to the Michelson-Morley experiment. Since the experimental apparatus is fixed at the surface of the Earth, its orientation depends on the Earth's rotation. Thus, the direction of the particle movement will change at different times of the day or in different seasons of the year (see Fig. 14.3). This means that the observed difference in the moving mass will change with the hours of the day.

The speed of the Earth can be estimated from the Earth's orbital speed, the speed of the Sun in our galaxy, and the speed of the Milky Way *vs* the rest of the universe. We estimated that the mass variation due to the change of apparatus orientation can be as high as $\frac{\Delta m}{m_0} \approx 2 \times 10^{-4}$ [1]. Such a change of particle mass should be measurable using existing technology.

For details of this proposed experiment, please see the original article: (D. C. Chang, "Is there a resting frame in the universe? A proposed experimental test based on a precise measurement of particle mass", *Eur. Phys. J. Plus* **132**, 140 (2017)).

14.2 Experimental Test #2: Can Any Inertial Frame Be Regarded as the Stationary Frame?

A second set of experiments to test the Principle of Relativity can be based on the "*Twin Paradox*". In the special theory of relativity (STR), it is hypothesized that all inertial frames are equal (the so-called "principle of relativity"); thus, any inertial frame can be regarded as the stationary frame [2, 3]. There was a famous challenge to this hypothesis; it was called the "*Twin Paradox*" (or "*Clock Paradox*") [3–7].

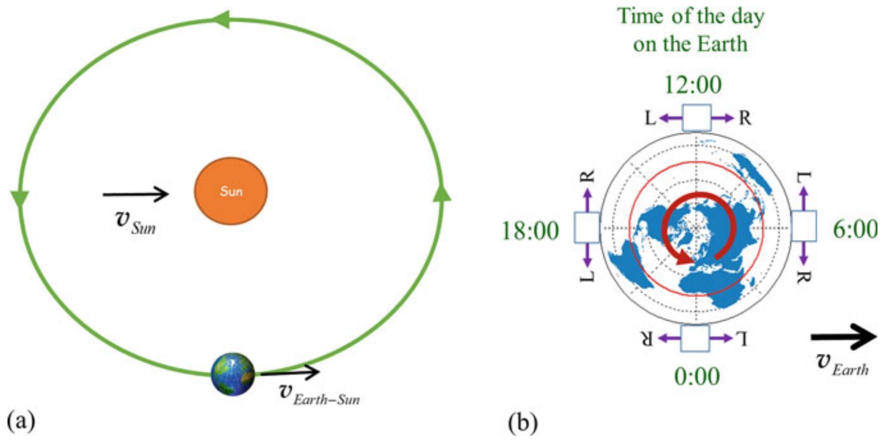


Fig. 14.3 A simplified diagram showing the essence of the experimental design. If there is a resting frame in our universe, it is expected that, for two electrons traveling in opposite directions (right and left), there will be a difference in their moving mass. This mass difference will be seasonal-dependent and change with the time of the day. **a** A top view of the movement of the Earth around the Sun. **b** The R and L arms of the apparatus are pointing to the East–West direction. Because of the movement of the Earth, the orientation of the apparatus is different in different times of the day. Thus, the electrons moving toward Right and Left will have different velocities relative to the resting frame of our universe. This means that the difference in their moving mass will also change with the hours of the day. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Eur. Phys. J. Plus* **132**, 140 (2017)

14.2.1 The Twin Paradox: A Challenge to the Principle of Relativity

One of the most peculiar results of STR is its prediction of *time dilation*. More specifically, based on the Lorentz transformation, it was predicted that time would become slower in a moving frame in comparison to the stationary frame. The elapse times between two events measured in the stationary frame and the moving frame are related by

$$\Delta t' = \gamma \Delta t = \frac{\Delta t}{\sqrt{1 - v^2/c^2}}, \tag{14.2}$$

where v is the speed of the moving frame and c is the speed of light, the factor γ is called "the Lorentz factor" which is always larger than 1. This relation implies that the time interval measured in a given clock is *less* in the rest frame of the clock than in any other frame. Thus, it would appear that "*the moving clocks run slow*" [3, 7].

This prediction of STR raised a serious challenge [7]. It was argued that, if two identical twin brothers (A and B) travel in different ways, they will have conflicting age expectations. Suppose twin brother A stays on the Earth while twin B travels in a rocket moving away in space. After sometime, B returns to the Earth. A thinks B is the moving twin. Due to the time dilation predicted by STR, B should be younger than A (see Fig. 14.4). But according to the *principle of relativity*, B could think he is the stationary one and regard A as the moving twin, so A should be younger than B. These two conclusions are contradicting. Thus, there is a paradox. (This is generally referred to as the "*Twin Paradox*" or "*Clock Paradox*").

Many physicists had dismissed this paradox by pointing out that, it is easy to determine who the moving brother is, because the twin in the rocket would experience acceleration and deceleration, but the twin staying on Earth would not [3, 7]. Thus, there is no paradox.

The above argument, however, cannot entirely dismiss the paradox. What happens if we modify the arrangement of traveling for the two twin brothers? Suppose both the twin brother A and twin brother B ride on rockets to travel in opposite directions. After a certain time period, both of them return to Earth (see Fig. 14.5). In this case, both of them would experience the same acceleration and deceleration. Their experiences are symmetrical, so each one can claim that the other brother is moving relative to himself.

Alternatively, what happens if the twin brothers are placed on different planets of the solar system? In this case, they are clearly staying in different inertial frames, and none of them would experience acceleration and deceleration. Then, which one of the twin brothers will age slower?

In order to clearly test the principle of relativity, we propose to conduct an experiment to examine the Twin Paradox using extra-terrestrial clocks. For example, we can compare the time readings between two sets of atomic clocks placed separately in Mars and Earth. The detailed experimental design will be discussed in the later part of this chapter.

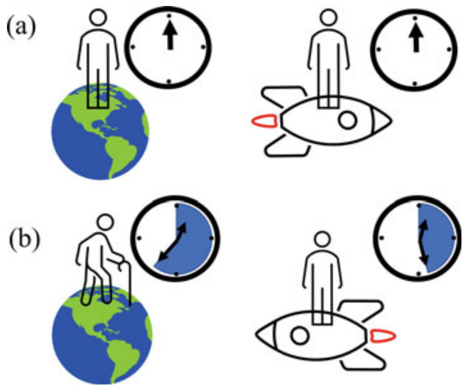


Fig. 14.4 An illustration showing the basic arrangement of the twin paradox. Suppose twin brother A stays on the Earth while twin brother B travels in a rocket moving away in space. After sometime, B returns to the Earth. A thinks B is the moving twin. Due to the time dilation predicted by STR, twin B should be younger than twin A. However, according to the principle of relativity, B can regard A as the moving twin. Then, twin A should be younger than twin B. There is a paradox

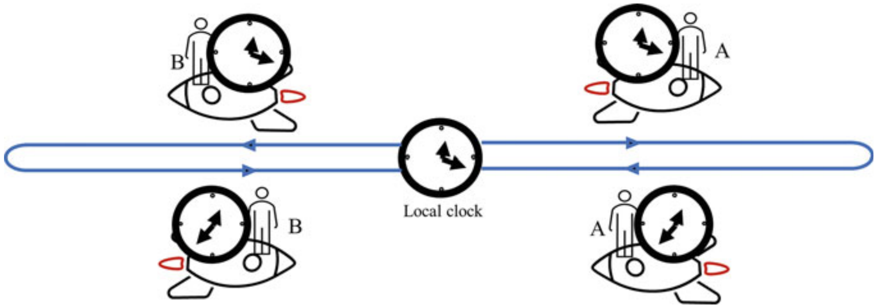


Fig. 14.5 An illustration showing a modified arrangement in the twin paradox. Suppose both the twin brother A and twin brother B ride on rockets to travel in opposite directions. After a certain time period, both of them return to Earth

14.2.2 Re-Examining the Previous Experiments on the Twin Paradox: Did They Ask the Key Questions?

Previously, a number of experiments had claimed to test the Twin Paradox [5, 8]. However, most of them were designed to test time dilation instead of the principle of relativity. **To fully test the Twin Paradox, the experiment needs to answer two independent questions:**

(1) **Will the moving twin be younger than the stationary twin?** This is to test whether time is different in different reference frames and this difference agrees with the Lorentz transformation.
(2) **Can either twin be regarded as the stationary twin?** This is to test the principle of relativity: Could any inertial frame be taken as the stationary frame? Will it create a paradox?

Most previous experiments were designed to test Question #1 but not Question #2. For example, a very well-known experiment claiming to test the *Twin Paradox* was an experiment conducted by Hafele and Keating [8, 9], in which a set of clocks was carried by commercial airlines around the Earth; then, times were compared between these traveling clocks and clocks located in the laboratory (see Fig. 14.6). More specifically, four atomic clocks were flown around the world on commercial jet flights during October 1971, once eastward and once westward. They observed directionally dependent time differences. Relative to the atomic clocks staying on ground, the flying clocks lost 59 ns during the eastward trip and gained 273 ns during the westward trip. Their observed results appeared to be in good agreement with predictions of conventional relativity theory. Thus, they claimed that their results “provide an unambiguous empirical resolution of the famous clock ‘paradox’…” [8].

A careful examination of the Hafele-Keating experiment, however, raises a question on whether they had truly tested the Twin Paradox. Their experiment only demonstrated that the times between the laboratory clocks and the moving clocks were

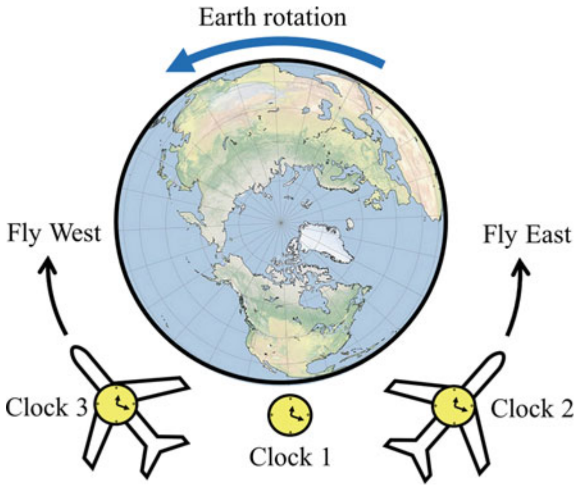


Fig. 14.6. The design of the Hafele-Keating experiment. A set of clocks was carried by commercial airlines around the Earth; then, times were compared between these traveling clocks and clocks located in the laboratory. More specifically, four atomic clocks were flown around the world on commercial jet flights during October 1971, once eastward and once westward. They observed directionally dependent time differences

different and were in agreement with the prediction of STR. Their experiment did not test whether any inertial frame can be regarded as the stationary frame. In fact, in their experiment, they could only use the virtual clock in the center of the Earth as the stationary frame to fit their data. If they chose either the clock flying Eastward or the clock flying Westward as the stationary clock, the result would not fit with the prediction of STR.

This point could be easily overlooked because there was a technical bias in their experimental design; it is obvious that the clocks carried in airplanes could not be regarded as a stationary frame. Since the movement of an airplane involved acceleration, taking off and landing, the airplane is not an inertial frame. So, one can naturally choose the virtual clock at the center of the Earth as the only inertial frame. The design of their experiment had excluded the possibility of testing the second question in the Twin Paradox, i.e., can either twin brothers be regarded as stationary?

Similar limitations also existed in other experiments designed to test time dilation. For example, Bailey et al. measured the relativistic time dilation for positive and negative muons in a circular orbit within the CERN Muon Storage Ring [10]. Their results were found to be in accordance with the prediction of special relativity. In this experiment, the laboratory was taken as the stationary frame, while the stored muons were taken as the moving frame. So, such experiment could not test whether any inertial frame can be chosen as the stationary frame.

14.3 Testing the Twin Paradox Using Extra-Terrestrial Clocks

Therefore, in order to truly test the Twin Paradox, we need to design a new experiment to address Question #2 specifically. That is, the clocks must be placed in two different inertial frames, where there is no acceleration and deceleration. The symmetrical arrangement would allow one to assume either one of these inertial frames as the stationary frame. One can then test whether there is a time dilation between these clocks as predicted by the STR.

Since both the Earth and Mars are rotating around the Sun at almost constant speeds, each of them can be regarded as an independent inertial system in their own right. Furthermore, the moving speeds of the Earth and Mars are high enough and thus can easily generate a large time dilation effect which is not difficult to measure. This is a very ideal situation to test the Twin Paradox (see Fig. 14.7).

With the current technology, it is not difficult to put an experimental clock on Mars and compare its time with a clock stationed on Earth. In the past two decades, several space agencies, including NASA, ESA, and CNSA, had already launched many exploratory probes to Mars. These probes include land rovers and orbiters. Many of them are known to carry highly precise clocks. Some of these clocks could be used as the Mars clock in this experiment.

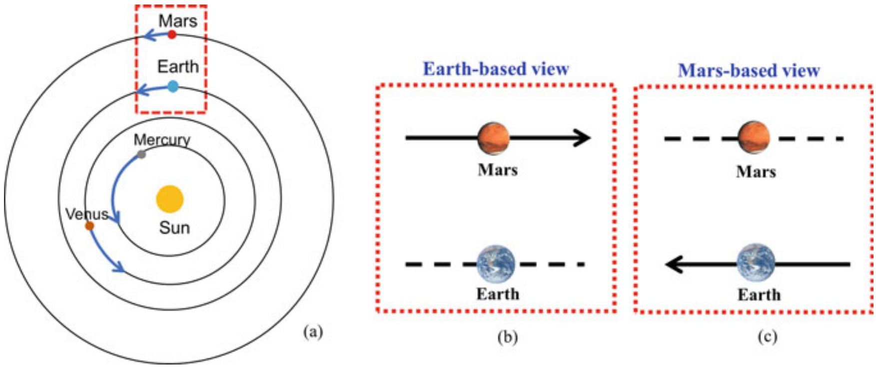


Fig. 14.7 Different views on the inertial frames of planet Earth and planet Mars. **a** In the solar system, the Earth moves at a higher speed than Mars. **b** If one regards the Earth as the Stationary frame, the Mars is a moving frame with a speed of $(v_{Mars} - v_{Earth})$. **c** Alternatively, if one regards the Mars as the Stationary frame, the Earth is a moving frame with a speed of $(v_{Earth} - v_{Mars})$

14.3.1 Key Models to be Tested

The experimental results will be used to compare with the predictions of three different theoretical models:

(1) **The Galilean model,** which represents the view of classical mechanics. *Time* is supposed to be the same in every inertial frame.
(2) **The special theory of relativity** (STR), in which the Earth is assumed to be the stationary frame.
(3) **The special theory of relativity** (STR), in which Mars is assumed to be the stationary frame.

We propose to conduct this experiment by putting one clock on Mars and comparing its time with a laboratory clock staying on Earth. We can then use the measured time dilation data to test the three models discussed above.

Based on the Lorentz factors, one can calculate the theoretical values of time dilation between the Mars clock and the Earth clock according to different theoretical models. We know the moving speed of Earth around the Sun (v_E) is 29.78 km/s; the moving speed of Mars (v_M) around the Sun is 24.07 km/s; the rotation speed at the surface of Earth (v_{RE}) is 465 m/s; the rotation speed of the Mars surface (v_{RM}) is 240 m/s. Using these known values, we can calculate the time difference between different clocks using the relevant Lorentz factors. The predicted values of the effective time dilation factors in the above three models are summarized in Table 14.1.

Since the *effective time dilation factor* $\overline{\gamma}$ values are very close to 1.0, it may be more convenient to see their differences using another representation. Thus, we introduce a new term called "*Time Dilation Ratio*" (TDR), which is defined as

$$\frac{\Delta t_B - \Delta t_A}{\Delta t_A} = \overline{\gamma} - 1. \tag{14.3}$$

Table 14.1 Calculated time dilation between Clock B (on Mars) and Clock A (on Earth) based on the Lorentz factor

Model	Stationary frame	Predicted result	Effective time dilation factor ($\overline{\gamma}$)	Time dilation ratio ($\overline{\gamma} - 1$)
Galilean	Not applied	No time difference between Clock A and Clock B	$\overline{\gamma} = 1$	0
Special relativity (Earth-based view)	**Earth**	Mars clock is slower	1.000000000180517	1.805×10^{-10}
Special relativity (Mars-based view)	**Mars**	Earth clock is slower	0.999999999817717	-1.823×10^{-10}

The predicted values of the *time dilation ratios* for the above three models are also included in Table 14.1.

14.3.2 Calculation of the Total Time Dilation by Including the Gravitational Redshift Effect

In the current literature, it has been proposed that time dilation can have multiple physical origins. That is, not only *the relative motion between two inertial frames* can give rise to time dilation (due to the "Lorentz factor" as discussed in the above); time dilation can also be resulted due to *the gravitational potential difference* of the two clocks. This understanding is based on the *gravitational redshift effect*. It has been demonstrated that, when an electromagnetic wave travels from a lower gravitational potential position to a higher gravitational potential position, its frequency becomes redshifted, such that,

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\Phi}{c^2} \tag{14.4}$$

where $\Delta\nu = \nu_B - \nu_A$ is the difference in the light frequency, and $\Delta\Phi = \Phi_B - \Phi_A$ is the difference of the gravitational potential between the two points. This phenomenon is called "*gravitational redshift of light*" [11–16].

In general relativity, the gravitational redshift effect is interpreted as a consequence of time dilation due to gravitational potential difference [11, 12]. It was thought that the clock at a higher gravitational potential runs faster than an identical clock located at a lower gravitational potential [11, 12]. Their clock rate difference is determined by the factor $\Delta\Phi/c^2$. In the quantum wave model discussed in this book, we have a different explanation for the gravitational redshift effect [17]. Recently, we showed that the gravitational redshift effect can also be explained based on quantum physics [17]. The key point is that, the gravitational mass of a photon is not its rest mass; instead, it is its quantum mass which can be determined from its momentum as described by the de Broglie relation [18]. Thus, the gravitational mass of a photon is not zero. Then, the condition of energy conservation will require a frequency shift when the photon travels between two points with different gravitational potentials. Therefore, the gravitational redshift of light is essentially a quantum effect [17]. (For details, please see Chap. 15).

In the condition that the speed of the moving object is far less than the speed of light and the gravitational energy is far less than the resting mass energy, the time dilation caused by the gravitational redshift effect can be combined with the time dilation caused by the Lorentz factor [16], that is,

$$\Delta t_B = \left(\overline{\gamma} - \frac{\Delta\Phi}{c^2}\right)\Delta t_A. \tag{14.5}$$

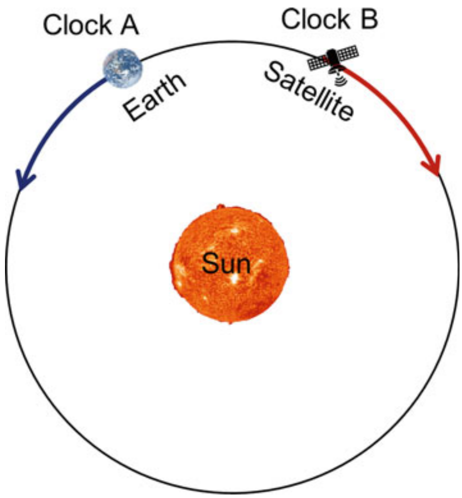


Fig. 14.8 A simpler experimental design for the clock paradox. Clock A is placed in a laboratory on Earth; Clock B is placed on a satellite that orbits the Sun in the opposite direction from Earth. One can compare the time dilation between Clock B (on the satellite) and Clock A (on Earth)

Thus, if one includes the gravitational redshift effect into the calculation of time dilation between the Mars clock and the Earth clock, the result would be somewhat different from those presented in Table 14.1. Nevertheless, the STR prediction for the Earth-based view will still be different from that of the Mars-based view. So, the measured time dilation can clearly differentiate these two views.

Technically, it is highly feasible to do this proposed experiment. Based on our calculation, the predicted time dilation between the Mars clock and the Earth clock is quite large. There is no problem for most atomic clocks to detect such time difference.

Furthermore, if one wants to make the data analysis simpler, one can remove the gravitational redshift effect by launching a satellite flying in the opposite direction of the Earth orbit. One can then compare the clocks on the satellite with those in the Earth laboratory (see Fig. 14.8).

This experiment in fact is simpler to conduct and involves less technical challenges. The cost of doing this experiment is also less than that of comparing clocks between Mars and Earth. So, the proposed experiment as outlined in Fig. 14.8 is highly feasible.

14.4 Chapter Summary

- In this chapter, we discussed two proposed experiments for testing the *Principle of Relativity*. The first experiment is to test whether the speed-dependence of moving mass is the same in different inertial frame. The second experiment is to test the famous *Twin Paradox* using extra-terrestrial clocks.

- In the first experiment, we propose to measure the moving mass of a particle traveling in different directions in reference to the solar system. We know the mass of a particle is not constant; it is dependent on its traveling speed. If the STR is correct, the moving mass should be the same regardless of the particle's moving direction. But if the vacuum is a dielectric medium as proposed in the *quantum wave model*, there is a resting frame in our universe. One will observe a difference in the moving mass when the particle travels in different directions.
- We predicted that the observed difference in the moving mass will change with the hours of the day if the vacuum is a universal resting frame.
- The second proposed experiment is to test the *Principle of Relativity* based on the "*Twin Paradox*". In the special theory of relativity, any inertial frame can be regarded as the stationary frame. There was a famous challenge to this hypothesis; it was called the "*Twin Paradox*".
- Previously, a very well-known experiment claiming to test the *Twin Paradox* was the experiment conducted by Hafele and Keating (in 1972), in which a set of clocks was carried by commercial airlines around the Earth; then, times were compared between these traveling clocks and clocks located in the laboratory. This experiment had certain limitations. In order to truly test the *Twin Paradox*, one needs to design a new experiment. That is, the clocks must be placed in two different inertial frames, where there is no acceleration and deceleration.
- Since both the Earth and Mars are rotating around the Sun at almost constant speeds, each of them can be regarded as an independent inertial system in their own right. Furthermore, the moving speeds of the Earth and Mars are high enough and thus can easily generate a large time dilation effect which is not difficult to measure. This is a very ideal situation to test the *Twin Paradox*.

References

1. Chang, D. C. (2017). Is there a resting frame in the universe? A proposed experimental test based on a precise measurement of particle mass. *European Physical Journal Plus, 132*, 140.
2. Einstein, A. (1905). Zur elektrodynamik bewegter körper. *Annalen der Physik, 322*, 891–921.
3. French, A. P. (1968). *Special relativity.* W. W. Norton & Company.
4. Bondi, H. (1957). The space traveller's youth. *Discovery, 18*, 505–510.
5. Crawford, F. S. (1957). Experimental verification of the 'Clock-Paradox' of relativity. *Nature, 179*, 35–36.
6. McMillan, E. M. (1957). The "Clock Paradox" and space travel. *Science, 126*, 381–384.
7. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Chapter 16. Relativistic energy and momentum. In *The Feynman lectures on physics*, Vol. 1. Addison-Wesley.
8. Hafele, J. C., & Keating, R. E. (1972). Around-the-world atomic clocks: Observed relativistic time gains. *Science, 177*, 168–170.
9. Hafele, J. C., & Keating, R. E. (1972). Around-the-world atomic clocks: Predicted relativistic time gains. *Science, 177*, 166–168.
10. Bailey, J., Borer, K., Combley, F., Drumm, H., Krienen, F., Lange, F., Picasso, E., von Ruden, W., Farley, F. J. M., Field, J. H., Flegel, W., & Hattersley, P. M. (1977). Measurements of relativistic time dilatation for positive and negative muons in a circular orbit. *Nature, 268*, 301–305.

11. Einstein, A. (1911). Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes. *Annalen der Physik, 340*, 898–908.
12. Longair, M. S. (1984) *Theoretical concepts in physics: An alternative view of theoretical reasoning in physics.* Cambridge University Press.
13. Pound, R. V., & Rebka, G. A. (1959). Gravitational red-shift in nuclear resonance. *Physical Review Letters, 3*, 439–441.
14. Pound, R. V., & Snider, J. L. (1965). Effect of gravity on gamma radiation. *Physical Review, 140*, B788–B803.
15. Vessot, R. F. C., Levine, M. W., Mattison, E. M., Blomberg, E. L., Hoffman, T. E., Nystrom, G. U., Farrel, B. F., Decher, R., Eby, P. B., Baugher, C. R., Watts, J. W., Teuber, D. L., & Wills, F. D. (1980). Test of relativistic gravitation with a space-borne hydrogen maser. *Physical Review Letters, 45*, 2081–2084.
16. Ashby, N. (2002). Relativity and the global positioning system. *Physics Today, 55*, 41.
17. Chang, D. C. (2018). A quantum mechanical interpretation of gravitational redshift of electromagnetic wave. *Optik, 174*, 636–641.
18. Chang, D. C. (2023). A quantum view of photon gravity: The gravitational mass of photon and its implications on previous experimental tests of general relativity. *Modern Physical Letter B* 2250179.

Chapter 15
A Quantum View of Photon Gravity: Implications of the Quantum Wave Model on General Relativity

Today, it is well known that the theory of general relativity (GR) is incompatible with quantum mechanics (QM) [1–6]. First, QM is supposed to be a theory for explaining physical events in the microscopic world; it is applicable at the atomic or sub-atomic level. GR is a physical theory for explaining the macroscopic world; it is applicable at the cosmological level. Thus, the objects to be described by these two theories are very different.

Second, all attempts to develop a quantum theory of gravity had failed. That is because the particle supposed to carry the gravitational force in the GR (graviton) has a spin equals to 2; it could not be renormalized. Thus, no one could develop a quantum gravitational theory based on GR.

Finally, the concepts of vacuum in these two theories are very different. General relativity, as a classical theory, treats the vacuum as an empty space. In fact, it must assume that there is no resting frame in our Universe. Otherwise, one can use this universal resting frame (i.e., the vacuum) to determine whether an object is in motion or not. This would allow one to differentiate between an object under acceleration and an object resting under gravity. This would violate the *principle of equivalence* (PE).

On the other hand, quantum mechanics treats the vacuum as the ground state of a complex physical system [7]. As shown in the *quantum wave model* discussed in this book, the vacuum has very rich physical properties. Furthermore, in the Standard Model of particle physics today, it is thought that virtual particle pairs can be created or annihilated instantly in the vacuum. When energy is provided, the virtual particles can become real particles. If the vacuum is just an empty space, it is not possible to explain where these real particles come from.

Thus, there is a conceptual conflict between the empty vacuum assumption in GR and the non-empty vacuum assumption in QM. Now, we are facing a big problem. Which theory should we believe? In the past one hundred years, we know QM is supported by countless experimental evidence. There is no doubt about its validity. But GR also appears to be well supported too. In the literature, there had been many

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_15

experimental tests on the *principle of equivalence* of GR. All of them claimed that their results support the prediction of GR. Then, we must carefully review these experimental tests and to see if their original interpretations are indisputable or not. For example, could some of these experimental results be explained by more than one theory? Particularly, could these experimental observations be explained based on quantum physics instead of GR?[1]

15.1 Does a Quantum Particle with no Rest Mass Interact with Gravity?

One of the most convincing evidence for supporting the theory of general relativity (GR) is the discovery that light does not travel in a straight line in a gravitational field. This finding is often cited as a definitive proof that space–time is curved by the presence of mass, as suggested by GR [8, 9]. In fact, many experiments aiming to test the GR were based on determining the gravitational effects of light, including (a) the observation of light bending near a star [10, 11], (b) lensing effect of a galaxy [12–15], (c) gravitational redshift of electromagnetic wave [16–19], and (d) discovery of black holes [20, 21]. All of these experiments involved measuring the behavior of light in a gravitational field. So far, results of these experiments all claimed to be supportive of the GR [10–23]. However, there is still a question on whether such interpretation is unequivocal. Are there possible alternative interpretations? Can these observed gravitational effects of photon also be explained based on other physical principles?

15.1.1 What is the Gravitational Mass of a Photon?

Thus, we have undertaken a careful investigation on whether the reported gravitational effects of photon can be explained based on quantum physics. Most of the previous experimental tests in support of GR were based on the assumption that light has no gravitational mass and thus should not interact with gravity. Their methods were to examine the pathway of light near a gravitational field. If light is observed to be bended near a gravitational source, it would indicate that the space–time is curved by the gravitational field, as suggested by GR [10, 11, 15, 24]. For example, a very well-known experiment claiming to support GR was the experiment done by Eddington during a solar eclipse [10]. He reported that star light passing the edge of the Sun was bended by an angle consistent with Einstein's calculation based on GR (see Fig. 15.1a). Similar experiments conducted later using radio-frequency wave also confirmed that electromagnetic wave is bended near a star [11].

[1] The discussions presented in this chapter is based on a recent paper of the author: D. C. Chang, *Mod. Phys. Lett. B,* 2250179 (2023).

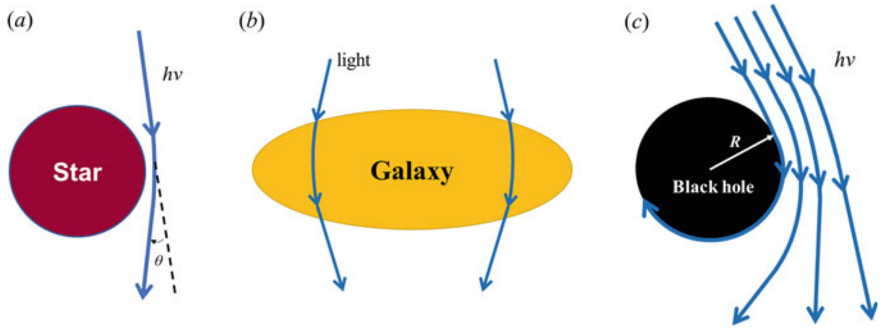


Fig. 15.1 **Bending of light by gravity**. With the understanding that the gravitational mass of a photon is not zero, one can predict that the passage of light will be bent in a gravitational field. **a** Light deflection near a star; **b** lensing effect of a galaxy; **c** a strong gravitational field can prevent light to escape from the black hole. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Mod. Phys. Lett. B* **36**, 2250179 (2023)

Another class of light-bending experiments is the observation of the "*lensing effect*" of certain galaxies [12–14, 25]. Many people had used such lensing effect to support GR [9, 18]. This lensing effect demonstrated that light passing a gravitational field is not in a straight line. Thus, such findings are consistent with the prediction of general relativity which proclaims that space–time can be curved by the presence of mass.

The design of these experiments, however, **involved a questionable hidden assumption**. That is, they assumed that, because photon has no rest mass, it should not interact with the gravitational field; thus, light should travel in a straight line near a star according to Newton's theory.

But, is this assumption really correct? The key question here is: ***What is the gravitational mass of a particle?*** Is it the *rest mass* or the *moving mass*?

According to the Newtonian theory of mechanics, the *gravitational mass* of an object is identical to the *inertial mass* of the same object. This understanding was also supported by Einstein [26–28]. Since the inertial mass of a particle is its moving mass, there should be no doubt that **the gravitational mass of a quantum particle should be its moving mass**.

In Newton's day, people thought the mass of an object is a constant. This understanding has changed now [29–32]. As we showed in Chap. 12, the inertial mass (m) of a particle increases with speed such that

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \tag{15.1}$$

where m_0 is the rest mass of the particle and v is the velocity of the particle as measured against a stationary frame [33]. For a massive particle, which is usually modeled as a point mass in the classical view, it is easy to calculate its inertial mass using the above equation. For a quantum particle such as the photon, the situation is slightly more complicated. This is because the photon has no rest mass (i.e., $m_0 = 0$). One may think that $m_0 = 0$ implies $m = 0$ according to Eq. (15.1). But this is not correct, because the speed of light is equal to c, so both the denominator and numerator of Eq. (15.1) are zero. The zero rest mass does not mean the moving mass (m) of a photon is zero. Then, we need to find another way to calculate the inertial mass of a photon.

Let us review first on what is the meaning of *mass* in physics. In Newtonian mechanics,

$$p = mv, \tag{15.2}$$

mass (m) is simply the proportional coefficient between the particle's momentum (p) and its velocity (v). In quantum mechanics, we know the momentum of a photon is given by the de Broglie relation,

$$p = \hbar k, \tag{15.3}$$

where k is the wave vector. We also know the speed of light is always c. Then, from Eqs. (15.2) and (15.3), we have

$$mc = \hbar k \tag{15.4}$$

Thus, the *inertial mass* of a photon is

$$m = \frac{\hbar k}{c}. \tag{15.5}$$

So, it is clear that, for a quantum particle with no rest mass (such as *photon*), it can still have a non-zero *inertial mass*; we may call it the "*effective mass*" or the "*quantum mass*".

Then, the gravitational mass of a photon is simply its quantum mass as given in Eq. (15.5).

15.1.2 The Quantum Interpretations of Light Deflection and Lensing Effects

If one realized that **the gravitational mass of a photon is not zero**, one would expect that photon should naturally interact with a gravitational field; that means the trajectory of light should not be a straight line near a star (see Fig. 15.1a). Thus, *the design of the light-deflection experiments was on a faulty basis. Their experimental results could be explained either by the non-zero gravitational mass of a photon or by GR. It was ambiguous.* (Of course, in the day of Eddington, people knew very little about quantum mechanics; Eddington probably did not know that the quantum mass of a photon is not zero.)

The realization of photon having a non-zero gravitational mass can also easily explain the observations of the "*lensing effect*" caused by certain galaxies [12–14, 25] (see Fig. 15.2). Many studies had used such lensing effect as evidence for supporting the prediction of GR that gravity is caused by space–time bending. Their conclusions, however, were also based on a faulty assumption similar to the light-deflection experiment; namely, they assumed that the photon has no gravitational mass. But as we pointed out in the above, their assumption was wrong. The lensing effect of a galaxy can be easily explained based on the fact that photon has a non-zero quantum mass and so it can be deflected in the gravitational field of a galaxy (see Fig. 15.1b).

Once one recognizes that the photon has a non-zero gravitational mass, one can immediately predict that a galaxy or a cluster of galaxies can produce a lensing effect to light rays emitted from distant stars. In most galaxies, a large amount of matters (including ordinary matters as well as dark matters) are concentrated at the center; their distribution is almost disk-like. This produces a gravitational field gradient. Since a photon has gravitational mass, its pathway will be bent by the galaxy's gravitational force when a light ray passes through the galaxy. As a result, the galaxy would appear to act as a lens (see Figs. 15.1b and 15.2).

Such lensing effect in fact is quite common in modern astrophysical studies. It was first observed in the double-imaged quasar in 1979 [12]. Later, hundreds of gravitational lensing effects were reported by different groups [13]. For example, a strong lensing galaxy in the cluster IRC 0218 was identified using the Hubble Space Telescope. Due to its lensing effect, the image of the distant galaxy behind it was distorted to produce a counter image [24].

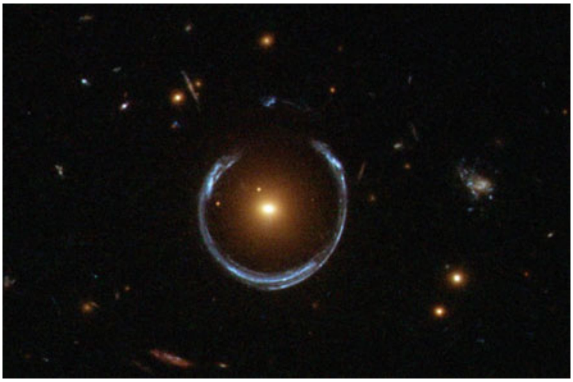


Fig. 15.2 "Lensing effect" of a galaxy. The gravity of a luminous red galaxy (LRG) was found to distort the light from a more distant blue galaxy. The image of LRG 3-757 was first observed in 2007 from the Sloan Digital Sky Survey (SDSS); this image is a follow-up observation taken with the Hubble Space Telescope. Photo Credit: ESA/Hubble & NASA

15.2 Origin of Black Holes

In the literature, another type of evidence for supporting GR is the discovery of black holes. The black hole is an object that generates a very strong gravitational force such that even light cannot escape from it. In the last 30 years, many black holes have been observed in different places of our universe [20, 34]. Some of the black holes at the center of galaxies were found to be very massive, about many million times of the mass of our Sun [20].

The existence of black hole was reported to be one of the predictions given in GR [18, 34–36]. Thus, the observation of black holes is considered to be a strong evidence for supporting GR. However, with the realization that the photon has a non-zero gravitational mass (m), one can also predict the existence of black hole based on the Newtonian gravitation theory.

The behavior of photon in a gravitation field is no different than an ordinary particle with mass. Suppose a massive object has a mass M_b, a particle (with moving mass m) flying nearby this object will experience a gravitational force. Suppose the velocity of this moving particle is u and the perpendicular distance between the particle and the massive object is R (see Fig. 15.1c). In order for the particle not to fall into that object, it needs to have a centrifugal force at least equal to that of the gravitational force, i.e.,

$$\frac{mu^2}{R} = G\frac{mM_b}{R^2}, \tag{15.6}$$

or

$$u = \sqrt{\frac{GM_b}{R}}. \tag{15.7}$$

This is the velocity required for the moving particle to counter act the attractive gravitational force of the massive object; it is called the "escape velocity". Now, if this escape velocity is equal to (or larger than) the speed of light (i.e., $u \geq c$), no particle (including photon) can escape the gravitational field of this massive object. Since the maximum velocity for any particle is c, all flying-by particles (including photons) will be captured by the massive object, which now becomes a "black hole" (see Fig. 15.3).

From Eq. (15.7), one can calculate the size of a black hole; its radius is

$$R = \frac{GM_b}{c^2}. \tag{15.8}$$

This radius is equivalent to the "event horizon" commonly used in describing a black hole. Thus, the existence of black holes not only can be explained based on GR, it can also be explained based on the non-zero quantum mass of a photon (and the Newtonian theory).

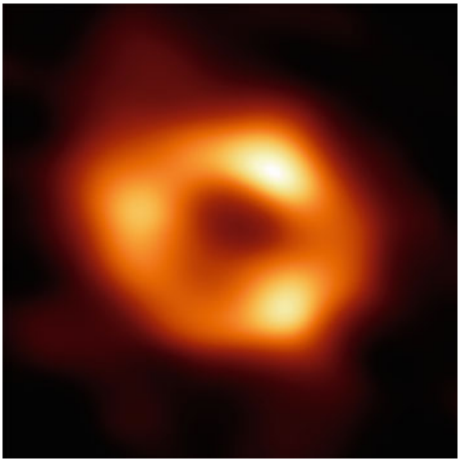


Fig. 15.3. Black hole. The first black hole image taken by The Event Horizon Telescope in 2019. It is the image of a supermassive black hole called Sagittarius A* at the center of our galaxy. Photo Credit: EHT Collaboration; Creative Commons Attribution 4.0

15.3 The Quantum Interpretation of Gravitational Redshift in Electromagnetic Waves

15.3.1 Gravitational Redshift of EM Wave According to GR

In the literature, the most strong evidence cited for supporting the principle of equivalence (PE) of GR was based on the measurements of the gravitational redshift of electromagnetic waves [16–19]. According to GR, time can be affected by gravity; and thus, it predicts that there should be a gravitational redshift of light [37]. Suppose a beam of laser light is transmitted from a ground station (point A) to a receiver in a satellite (point B) orbiting above the Earth (Fig. 15.4). The theory of GR predicts that the light ray will experience a gravitational redshift [38]

$$\nu' = \nu \exp\left(-\frac{\Delta\phi}{c^2}\right), \tag{15.9}$$

where *v'* is the frequency of light at the satellite, *v* is the initial light frequency at the ground station, $\Delta\phi$ is the difference of gravitational potential between point *A* and point *B*. Since the photon energy is much larger than the change of gravitational potential energy (i.e., $c^2 >> \Delta\phi$), one can apply Taylor's expansion to the above equation and obtain

$$\nu' = \nu\left(1 - \frac{\Delta\phi}{c^2}\right). \tag{15.10}$$

Now, denoting the change of frequency as $\Delta\nu = \nu' - \nu$, the above relation gives

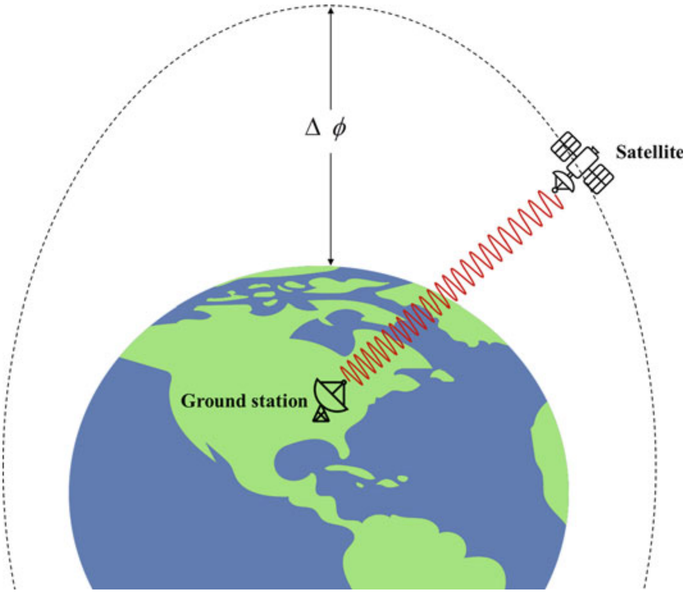


Fig. 15.4 **Gravitational redshift of photon.** When a beam of electromagnetic wave is transmitted from a ground station at the Earth surface to a satellite, its frequency undergoes a redshift. $\Delta\phi$ is the gravitational potential difference between the transmitter and the receiver. (In this illustration, the satellite is assumed to move in a circular orbit). Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Mod. Phys. Lett. B* **36**, 2250179 (2023)

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\phi}{c^2}. \tag{15.11}$$

This is the well-known relation of *gravitational redshift*; it had been used in many experiments to test the validity of GR [16–18]. Such a relation is actually used in satellite communication today. For example, many receiver systems designed for satellite navigation are now incorporating the *gravitational redshift effect* as given in the above [39].

15.3.2 Explanation of the Gravitational Redshift Effect Based on Quantum Physics

So far, many experiment tests on the principle of equivalence (PE) were based on measuring the gravitational redshift effect of photon. Their data were all consistent with the prediction given in Eq. (15.11). Thus, it was claimed that the PE is well verified [18, 19, 23]. However, we found that the gravitational redshift effect can be explained in more than one way; besides the GR explanation, *the gravitational* *redshift effect can also be explained by the fact that the gravitational mass of a photon is not zero* [40].

With the understanding that photon has a non-zero gravitational mass, one can easily explain why the frequency of light will change in a gravitational field. Since a photon has moving mass, it can interact with the gravitational field. This implies that,

$$\begin{aligned}\text{Total energy of a photon} &= \text{Its quantum energy} \\ &\quad + \text{Its gravitational potential energy}.\end{aligned}$$

The quantum energy of a photon is given by Planck's relation; its gravitational potential energy can be calculated based on the Newton's Law. Thus, the total energy of a photon within a gravitational field is

$$E_{\text{total}} = h\nu + m\phi, \tag{15.12}$$

where m is the quantum mass of the photon as given in Eq. (15.5), ϕ is its gravitational potential at a particular position.

According to the principle of *conservation of energy*, the total energy of a photon moving freely in space should be conserved. When a photon moves from point A to point B and there is a gravitational potential difference between these two points, the photon will change its frequency from ν to ν' in order to satisfy the requirement of conservation of energy. That is,

$$\Delta E_{\text{total}} = h\Delta\nu + m\Delta\phi = 0, \tag{15.13}$$

where $\Delta\nu = \nu' - \nu$, and $\Delta\phi = \phi_B - \phi_A$ is the difference of the gravitational potential between point A and point B. By substituting Eq. (15.5) into Eq. (15.13), we get

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\phi}{c^2}. \tag{15.14}$$

Thus, one can predict that the photon must be redshifted when it moves from the Earth surface to a satellite above the Earth (see Fig. 15.4). Hence, **the gravitational redshift of electromagnetic wave is a consequence of the fact that the gravitational mass of a photon is not zero**.

Very interestingly, our theoretical result obtained based on the quantum mass of a photon (i.e., Eq. 15.14) is identical to the theoretical result based on GR (i.e., Eq. 15.11). That means **the gravitational redshift effect can be explained either based on GR or quantum physics**. Thus, the observation of gravitational redshift cannot be regarded as an unequivocal evidence for supporting GR.

15.4 Testing the Principle of Equivalence by Measuring the Mass Variation Due to Speed Changes

From the above discussions, one can see that most of the supporting evidence for GR are not unequivocal; they can also be explained by quantum physics. Thus, in order to critically evaluate the validity of GR, one needs to propose new experiments to directly test its basic hypothesis, the *principle of equivalence* (PE).

Recently, we suggested that one possible way to test the *principle of equivalence* is to measure the moving mass of an object over time. We know the inertial mass of a particle is its moving mass, which varies with the particle's velocity,

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = \gamma m_0, \tag{15.15}$$

where γ is called the "Lorentz factor". One can use the above relation to determine if a particle is at rest or under acceleration. The design of such an experiment is relatively simple. Suppose an object (with rest mass m_0) is placed on top of a sophisticate "*electronic balance*" inside a rocket. The weight of this object is determined by the gravitational force and/or by the acceleration of the rocket. Let us consider two different motional states for the rocket (see Fig. 15.5):

(A) *The rocket is resting in a gravitational field (gravitational acceleration* $= g$*).*
(B) *The rocket is accelerating in space (acceleration* $= a$*) where there is no gravitational force.*

If the acceleration $a = g$, the weight of the object will appear the same in both Case A and Case B. This is the basis for PE. However, if one can continue to measure

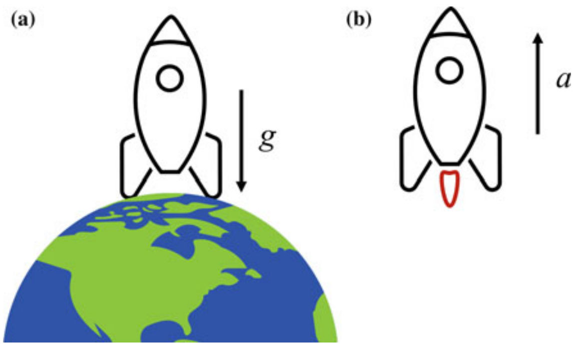


Fig. 15.5 Graphical presentation of the Principle of Equivalence. Two rockets in different situations: **a** The rocket is resting in a gravitational field, and **b** the rocket is under acceleration without gravity; its acceleration rate a is the same as the gravitational rate g in **a**. According to GR, all physical experiments conducted within these two rockets should give identical results. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Mod. Phys. Lett. B* **36**, 2250179 (2023)

the weight of the object over a long period of time, it will be possible to determine whether the rocket is under acceleration or not.

In this proposed experiment, the weight of the object will be measured as a function of time and see if there is any change in this weight. At the beginning of the experiment, one cannot differentiate if the rocket is in motion or not. One could think the rocket is resting in a gravitational field (i.e., Case A) or under acceleration (with $a = g$) (i.e., Case B). But as time goes by, the speed of the rocket in case B should increase with time, while the speed of the rocket in Case A will remain unchanged. That means the moving mass of the object in Case B should increase with time, while the moving mass of the object in Case A should remain constant (see Fig. 15.6). Based on such a measurement, one can determine whether the rocket is resting under gravity or accelerating without gravity.

One may worry that whether this proposed experiment is technologically feasible. The experimental design for measuring the moving mass of an object could be relatively simple. For example, one can use a magnetic force (or electric force) to determine if the mass of an object inside the rocket has changed or not. The design of such an experimental apparatus is shown in our recent publication [41]. There could be other experimental designs for detecting whether the mass of an object inside the rocket can change due to acceleration. For example, one could use the Coulombic force between two charged objects to balance the gravitational force (or acceleration). A more sophisticated experiment can be done by measuring the *mass-to-charge ratio* of a charged particle using a mass spectrometer or a Penning trap [42, 43]. Using the

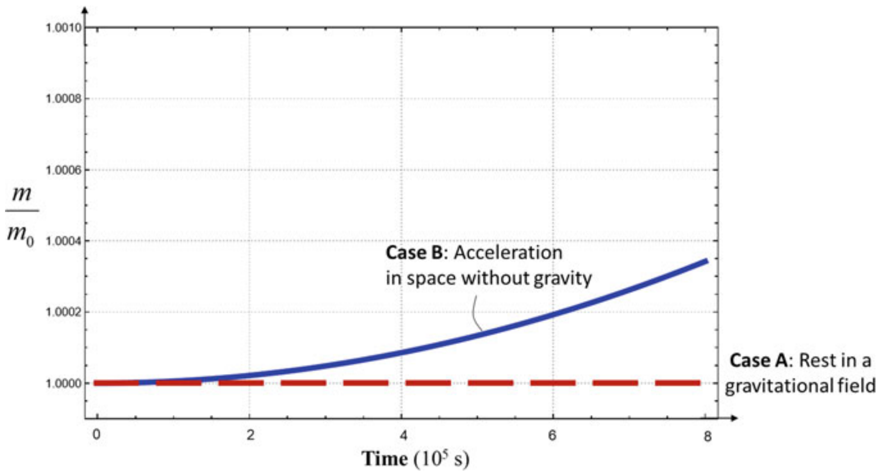


Fig. 15.6 Differentiation between an accelerating rocket and a rocket sitting in a gravitational field. By measuring the mass variation over time, one can determine whether the rocket is under acceleration or resting in gravity. The red dashed line represents the Case A scenario that the rocket stays at rest in a gravitational field. The blue solid line represents the Case B scenario that the rocket accelerates in space without gravity. Credit: This figure is reproduced from an earlier publication of the author: D. C. Chang, *Mod. Phys. Lett. B* **36**, 2250179 (2023)

advanced technology available at present, one can determine the *mass-to-charge ratio* of an electron to a precision of 10^{-8} [44]. This is more than sufficient to determine whether the Lorentz factor γ would change due to rocket acceleration.

In summary, there are multiple ways to determine the moving mass (inertial mass) of an object when this object is placed inside a rocket in motion. Since the moving mass is a function of speed as given in Eq. (15.15), it is not difficult to differentiate whether an object is resting in gravity or accelerating without gravity. Under this situation, one can experimentally determine whether the *principle of equivalence* is true or not.

15.5 Re-examination of Eddington's Eclipse Expedition Experiment: Did It Prove that GR is the Correct Gravitational Theory While the Newtonian Gravitational Theory is not?

In the above discussion, we followed the modern practice that the test of general relativity is based on whether light can pass through a gravitational field in a straight line or not. This is a standard argument today. For example, Stephen Hawking wrote in his famous book "*A Brief History of Time*":

> the fact that space is curved (according to general relativity) means that light no longer appears to travel in straight lines in space. So general relativity predicts that light should be bent by gravitational fields [6].

Some people may take issue with this statement, because according to the famous Eddington experiment conducted during the 1919 eclipse expedition, it was claimed that the predicted difference between Newton's Law and general relativity is not whether light will be deflected in a gravitational field; the difference is the degree of deflection [10]. And this was what Eddington's experiment tried to determine. It was widely reported that their experimental results clearly supported the gravitational theory of Einstein over the gravitational theory of Newton [10].

This report had tremendous influence on people's subsequent acceptance of the GR. So, let us re-examine it carefully in detail.

15.5.1 What Did Eddington's Experiment Really Test?

The first serious test of Einstein's theory of general relativity was the light deflection experiment conducted by Arthur Eddington's team in 1919 [10]. In the Newtonian theory of gravity, light should travel in straight lines in the gravitational field. Einstein's general relativity, however, predicted that light should be bent near a gravitational source, such as the Sun. This prediction can be tested by measuring the position of stars near the Sun during an eclipse. This was the goal of the experiment conducted by the British astronomer, Arthur Eddington at 1919 [10].

More specifically, the Eddington experiment was designed to test three possibilities: (1) the classical light theory, (2) the general relativity proposed in Einstein's 1911 paper, and (3) the general relativity proposed by Einstein's 1916 paper (see Table 15.1) [10].

In Einstein's 1911 paper entitled "*On the influence of gravity on the propagation of light*" (Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes, *Annalen der Physik*, **35**, 1911) [37], he predicted that based on the principle of equivalence, light could be deflected near a gravitational source (such as the Sun). His prediction of light deflection near the gravitational field was not based on the realization that the photon has non-zero gravitational mass. Instead, Einstein thought that the speed of light would change in the gravitational field due to the principle of equivalence. Thus, "*by means of Huyghens's principle, that light-rays propagated across a gravitational field undergo deflexion*" [37, 45]. This is somewhat similar to the observation of light deflection at the air–water boundary. Einstein acknowledged that this thinking of light speed change could be in conflict with the 2nd postulate in his 1905 special relativity paper [46].

This prediction is different from the conventional Newtonian theory, which predicts that light should not be bent. So, one can test it. After the publication of the 1911 paper [37], Einstein had actively searched for astronomers to conduct the experimental test of his calculation result. A young astronomer from Berlin named Erwin Freundlich was interested in the project and took the job. They planned to conduct their measurement for the eclipse to be taken place on August 21, 1914, in the Crimea. Unfortunately, just twenty days before the eclipse, the World War I started and Germany declared war on Russia. Freundlich and his German colleagues were captured by Russian soldiers. This experimental test was abolished [47].

Later, when Einstein developed a more completed theory of general relativity in 1916 [26], he found the predicted deflection angle was twice as large as what he predicted in the 1911 paper [26, 37].

At 1918, a team of British astronomers, including Arthur Eddington and Sir Frank Dyson, decided to conduct an experiment to test Einstein's theory of general relativity. So, they organized an expedition to measure the deflection of star light by the

Table 15.1 Three possibilities tested in the 1919 eclipse expedition experiment

	Model to be tested	**Assumption**	**Deflection angle predicted**	**Remarks**
1	**Classical light theory**	Light does not interact with gravity	0	Regarding light as EM wave
2	**General Relativity** Einstein (1911)	Based on the principle of equivalence	0.83 s arc	Regarding light as EM wave
3	**General Relativity** Einstein (1916)	Based on the principle of equivalence and space–time curving	1.7 s arc	

Sun during the May 29, 1919, eclipse. They sent out two teams to do the measurement. One team went to Sobral in Brazil while the other team went to Principe near the equator in Africa [10].

When the expedition experiment was done, they found the result was closer to the prediction of Einstein's second paper. When they publish their report in 1920, they somehow changed the emphasis of their experiment. Particularly, they changed the model #2 of Table 15.1 (i.e., *the prediction of Einstein's general relativity in 1911*) to ***the prediction of Newtonian gravitation theory***. In their words, **Possibility #2** became: "*The energy or mass of light is subject to gravitation in the same way as ordinary matter. If the law of gravitation is strictly the Newtonian law, this leads to an apparent displacement of a star close to the sun's limb amounting to 0''·87 outwards*" [10].

This description is clearly different from what Einstein discussed in his 1911 paper [37]. First, in Einstein's 1911 paper, light was treated as an electromagnetic wave; it was not treated as "*ordinary matter"*. Second, the calculation of deflection angle in Einstein's 1911 paper was based on the *principle of equivalence*, not based on "*strictly the Newtonian law*".[2] As to why Eddington mis-represented the essence of Einstein's 1911 paper, we are highly curious. One possible reason is for generating sensational publicity. Indeed, the expedition experiment was widely reported in many major press with very sensational headlines, such as "***Revolution in science***" and "*New Theory of the Universe:* ***Newtonian Ideas Overthrown***" [48].

Another problem with Eddington's report on the 1919 solar eclipse experiment is that the quality of their data was really not conclusive. As pointed out by Stephen Hawking, "*It is ironic, therefore, that later examination of the photographs taken on that expedition showed the errors were as great as the effect they were trying to measure. Their measurement had been sheer luck, or a case of knowing the result they wanted to get, not an uncommon occurrence in science*" [6].

In conclusion, if their data were accurate, it may conclude that the 1911 version of Einstein's GR is wrong. The measurement made by Eddington in the 1919 expedition experiment only showed that *Einstein's 1916 GR paper gave a better prediction than Einstein's 1911 GR paper*.

15.6 Outstanding Questions on GR that Need to Be Investigated

GR is a very important theory today, particularly in cosmology. Thus, in the future it will be highly worthwhile to design new experiments to critically test it. Currently, there are several key questions about GR that need to be investigated:

[2] In the calculation, Einstein also made use of the common relation $\Phi = -km/r$, which is from the Newtonian theory.

(1) ***How can mass curve space–time? What is the physical mechanism?*** A major proposal by Einstein in his 1916 paper is that the space–time of our Universe can be curved by the presence of mass. This is a very bold assumption. Is there any experimental evidence that the space–time of our Universe is indeed curved? Are we living in a flat four-dimensional space–time or a curved space–time? Furthermore, why and how can the presence of mass or energy change the curvature of our space–time? This is a conjecture proposed by Einstein. But to prove a conjecture, one needs solid physical evidence. What is the physical basis for the curving of space–time by mass or energy? Einstein did not explain that in his original paper. After one hundred years, we still have not found its physical basis. So, whoever wants to use general relativity as a law of nature will have the burden to explain what is the physical basis to make the presence of the mass to curve the space–time.

(2) ***Is PE correct?*** Einstein started his general relativity based on the *principle of equivalence*. As we pointed out above, up to this time, we have not yet found unequivocal evidence to verify the validity of PE. Of course, many experiments, including light bending experiment, lensing experiment, etc., had claimed to verify the principle of equivalence. But, as we show in this chapter, these experimental results can also be explained by the fact that the gravitational mass of a photon is not zero. Another class of experiments which claimed to support the principle of equivalence is *gravitational redshift*. However, as shown in the above sections, that can also be explained based on quantum physics. Thus, the experiments claiming to support PE were unequivocal. Therefore, in order to fully verify the principle of equivalence, future scientists must come up with some critical experiments to test that.

(3) ***Can PE be applied to other forces?*** In the original paper of Einstein, the principle of equivalence was proposed to connect between the *acceleration* of an inertial system and the presence of *gravitational force*. Since all forces involve acceleration, not only gravitational force, other forces such as electromagnetic forces also involve acceleration. Would the principle of equivalence also apply to these other forces? In fact, the gravitational force is the weakest force among the known four forces in nature. The electric force is 10^{38} times stronger than the gravitational force. That means the acceleration involved for electric force is very strong. Is there any experimental evidence indicating that the *principle of equivalence* may be applied to *electric force*?

(4) ***Would electric or magnetic charge also curve the space–time?*** If the *principle of Equivalence* is a law of nature, it would suggest that other forces can also be equivalent to acceleration. In that case, if the gravitational force is explained by the presence of the mass curving the space–time, should the presence of electric charge also curve the space–time? We know the electric force is much stronger than the gravitational force; that means the ability of the electric charge to curve the space–time should also be much stronger than the gravitational mass. Since one can measure the results of a much stronger effect more easily than a much weaker effect, should we have already observed many experimental evidence

that the space–time around the electric charge or a magnetic charge is highly curved? Do we have that evidence?

15.7 Chapter Summary

- Most of the previous experimental tests in support of GR were based on the assumption that light has no *gravitational mass* and thus should not interact with gravity. This assumption was not correct. The *gravitational mass* of a photon is its *inertial mass*, which is not zero.
- The momentum of a photon is given by the de Broglie relation and the speed of light is always c. So, the *photon* has a non-zero *inertial mass*, which is equal to $m = \hbar k/c$.
- If one realizes that the gravitational mass of a photon is not zero, one expects that photon should naturally interact with a gravitational field; that means the trajectory of light is not a straight line near a star. It should be bent.
- The realization of photon having a non-zero gravitational mass can also explain the "*lensing effect*" observed in some galaxies. Furthermore, with the knowledge that light has non-zero gravitational mass, one can predict the existence of *black holes* based on the Newtonian gravitation theory.
- In the literature, the strongest evidence for supporting the GR was based on the observations of *gravitational redshift* in electromagnetic waves. This effect, however, can also be explained by quantum physics. It can be shown that the *gravitational redshift* of electromagnetic wave is a consequence of the fact that the gravitational mass of a photon is not zero and the requirement of energy conservation when photon travels between two points with different gravitational potentials. Therefore, the *gravitational redshift* effect can be explained either based on GR or quantum physics.
- The general theory of relativity (GR) is based on the *principle of equivalence* (PE). We proposed a new experiment to directly test the validity of PE. The key is to measure the mass of a moving object over time. There are multiple ways to determine the moving mass of an object when this object is placed inside a rocket. Since the moving mass is a function of speed, it is not difficult to differentiate whether an object is resting in gravity or accelerating without gravity. Under this situation, one can experimentally determine whether the *principle of equivalence* is true or not.
- We re-examine the famous eclipse expedition experiment conducted by Eddington's team in 1919. It had been widely reported that this experiment clearly proved that GR is the correct gravitational theory while the Newtonian gravitational theory is not. Our re-examination showed that this claim could not be substantiated.
- Finally, we listed a number of outstanding questions on GR that need to be investigated.

References

1. Penrose, R. (2005). *The road to reality: A complete guide to the laws of the universe*. Knopf.
2. Greene, B. (2004). *The fabric of the cosmos: Space, time, and the texture of reality*. A.A. Knopf.
3. Carroll, S. M. (2004). *Spacetime and geometry: An introduction to general relativity.* Addison Wesley.
4. Anderson, E. (2017). *The problem of time: Quantum mechanics versus general relativity.* Springer.
5. Powell, C. S. (2015). Relativity versus quantum mechanics: The battle for the universe. The Guardian.
6. Hawking, S. W. (1988) *A brief history of time: From the big bang to black holes.* Bantam Books
7. Ryder, L. H. (1996). *Quantum field theory.* Cambridge University Press.
8. Rindler, W. (1977). *Essential relativity: Special, general, and cosmological* (2nd ed). Springer.
9. Hartle, J. B. (2003). *Gravity: An introduction to Einstein's general relativity*. Addison-Wesley.
10. Dyson, F. W., Eddington, A. S., & Davidson, C. (1920). A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A, 220*, 291.
11. Counselman, C. C., Kent, S. M., Knight, C. A., Shapiro, I. I., Clark, T. A., Hinteregger, H. F., Rogers, A. E. E., & Whitney, A. R. (1974). Solar gravitational deflection of radio waves measured by very-long-baseline interferometry. *Physical Review Letters, 33*, 1621–1623.
12. Walsh, D., Carswell, R. F., & Weymann, R. J. (1979). 0957 + 561 A, B: Twin quasistellar objects or gravitational lens? *Nature, 279*, 381–384.
13. Wambsganss, J. (1998). Gravitational lensing in astronomy. *Living Reviews in Relativity, 1*, 12.
14. Inada, N., Oguri, M., Pindor, B., Hennawi, J. F., Chiu, K., Zheng, W., Ichikawa, S.-I., Gregg, M. D., Becker, R. H., Suto, Y., Strauss, M. A., Turner, E. L., Keeton, C. R., Annis, J., Castander, F. J., Eisenstein, D. J., Frieman, J. A., Fukugita, M., Gunn, J. E., … York, D. G. (2003). A gravitationally lensed quasar with quadruple images separated by 14.62 arcseconds. *Nature, 426*, 810–812.
15. Refregier, A. (2016). Weak gravitational lensing by large-scale structure. *Annual Review of Astronomy and Astrophysics, 41*, 645–668.
16. Pound, R. V., & Snider, J. L. (1965). Effect of gravity on gamma radiation. *Physical Review, 140*, B788–B803.
17. Vessot, R. F. C., Levine, M. W., Mattison, E. M., Blomberg, E. L., Hoffman, T. E., Nystrom, G. U., Farrel, B. F., Decher, R., Eby, P. B., Baugher, C. R., Watts, J. W., Teuber, D. L., & Wills, F. D. (1980). Test of relativistic gravitation with a space-borne hydrogen maser. *Physical Review Letters, 45*, 2081–2084.
18. Will, C. M. (2006). The confrontation between general relativity and experiment. *Living Reviews in Relativity, 9*, 3.
19. Müller, H., Peters, A., & Chu, S. (2010). A precision measurement of the gravitational redshift by the interference of matter waves. *Nature, 463*, 926–929.
20. Begelman, M. C. (2003). Evidence for black holes. *Science, 300*, 1898.
21. Gibbons, G. W., & Hawking, S. W. (1971). Evidence for black holes in binary star systems. *Nature, 232*, 465.
22. Dodwell, G. F., & Davidson, C. R. (1924). Determination of the deflection of light by the Sun's gravitational field from observations made at Cardillo Downs, South Australia, during the total eclipse of 1922 September 21. *Monthly Notices of the Royal Astronomical Society, 84*, 150–162.
23. Wojtak, R., Hansen, S. H., & Hjorth, J. (2011). Gravitational redshift of galaxies in clusters as predicted by general relativity. *Nature, 477*, 567.
24. Wong, K. C., Tran, K.-V.H., Suyu, S. H., Momcheva, I. G., Brammer, G. B., Brodwin, M., Gonzalez, A. H., Halkola, A., Kacprzak, G. G., Koekemoer, A. M., Papovich, C. J., & Rudnick, G. H. (2014). Discovery of a strong lensing galaxy embedded in a cluster at $z = 1.62$. *Astrophysical Journal Letter, 789*, L31.

25. Refsdal, S., & Surdej, J. (1994). Gravitational lenses. *Reports on Progress in Physics, 57*, 117.
26. Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik, 354*, 769–822.
27. Einstein, A. (1907). Relativitätsprinzip und die aus demselben gezogenen Folgerungen. *Jahrbuch der Radioaktivität, 4*, 411–462.
28. Roll, P. G., Krotkov, R., & Dicke, R. H. (1964). The equivalence of inertial and passive gravitational mass. *Annals of Physics, 26*, 442–517.
29. Poincaré, H. (1900). La théorie de Lorentz et le principe de réaction. *Archives Neerlandaises des Sciences Exactes et Naturelles, 2*, 252–278.
30. Lorentz, H. A. (1904). Electromagnetic phenomena in a system moving with any velocity smaller than that of light. *Proceeding of Royal Netherland Acadamy of Arts Science, 6*, 809–831.
31. Kaufmann, W. (1901). Die magnetische und electrische Ablenkbarkeit der Bequerelstrahlen und die scheinbare Masse der Elektronen. *Göttinger Nachrichten, 2*, 143–168.
32. Abraham, M. (1902). Prinzipien der Dynamik des Elektrons. *Physikalishce Zeitschrift, 4*, 57–63.
33. Chang, D. C. (2020). A quantum interpretation of the physical basis of mass–energy equivalence. *Modern Physics Letters B, 34*, 2030002.
34. Narayan, R. (2005). Black holes in astrophysics. *New Journal of Physics, 7*, 199.
35. Einstein, A. (1939). On a stationary system with spherical symmetry consisting of many gravitating masses. *The Annals of Mathematics, 40*, 922.
36. Hawking, S. W. (1972). Black holes in general relativity. *Communications in Mathematical Physics, 25*, 152–166.
37. Einstein, A. (1911). Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes. *Annalen der Physik, 340*, 898–908.
38. Longair, M. S. (1984). *Theoretical concepts in physics: An alternative view of theoretical reasoning in physics.* Cambridge University Press.
39. Ashby, N. (2002). Relativity and the global positioning system. *Physics Today, 55*, 41.
40. Chang, D. C. (2018). A quantum mechanical interpretation of gravitational redshift of electromagnetic wave. *Optik, 174*, 636–641.
41. Chang, D. C. (2023). A quantum view of photon gravity: The gravitational mass of photon and its implications on previous experimental tests of general relativity. *Modern Physics Letters B, 36*, 2250179.
42. Blaum, K., Novikov, Y. N., & Werth, G. (2010). Penning traps as a versatile tool for precise experiments in fundamental physics. *Contemporary Physics, 51*, 149–175.
43. Sturm, S., Köhler, F., Zatorski, J., Wagner, A., Harman, Z., Werth, G., Quint, W., Keitel, C. H., & Blaum, K. (2014). High-precision measurement of the atomic mass of the electron. *Nature, 506*, 467–470.
44. Marshall, A. G., Hendrickson, C. L., & Jackson, G. S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews, 17*, 1–35.
45. Einstein, A. (1952) *The principle of relativity: A collection of original memoirs on the special and general theory of relativity.* Dover Publications.
46. Einstein, A. (1905). Zur elektrodynamik bewegter körper. *Annalen der Physik, 322*, 891–921.
47. Isaacson, W. (2007). *Einstein: His life and universe.* Simon & Schuster.
48. Revolution in science. New theory of the universe: Newtonian ideas overthrown. The Times (November 7, 1919).

Part VI
Outstanding Questions and Remaining Challenges

Chapter 16
Further Thoughts on Quantum Physics

The *quantum wave model* discussed in this book is intended to provide a fundamental understanding of quantum physics at the sub-atomic level. For things on smaller scales, at the subnuclear level, today's physics community relies on quantum field theories, commonly referred to as the "*Standard Model of Particle Physics*". This Standard Model has been shown to be very successful in agreeing with experimental observations. However, today's Standard Model still cannot fully explain many fundamental questions. For examples:

- What is the origin of *rest mass* in the elementary particle?
- According to the Standard Model, the mass of an elementary particle is obtained from a coupling between its quantum field and the Higgs field. Is this Higgs mechanism consistent with general relativity? How can a coupling constant with the Higgs field curve the space–time, as proposed in the GR?
- What is the physical meaning of *anti-mass*? What is its origin?
- Why are particle and anti-particle created in pair? Does it imply that the vacuum is filled with a "*Dirac sea*" of an infinite number of negative-energy particles?
- What is the origin of electric charge? Why can some particles have electric charge? Where does the negative charge or positive charge come from?

In this chapter, we would like to explore if the *quantum wave model* can offer some useful ideas to help to resolve the above questions.

16.1 Essence of the Standard Model of Particle Physics

The Standard Model of particle physics is the theory describing our current understanding of elementary particles. It was developed in the second half of the twentieth century through the work of many scientists worldwide. The current formulation was completed around the mid-1970s.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_16

The development of the Standard Model was driven by both theoretical and experimental particle physicists. The Standard Model is a paradigm of a quantum field theory. More specifically, it is a gauge quantum field theory based on the internal symmetries of the unitary product group SU(3) $\times$ SU(2) $\times$ U(1). In 1954, two young theoretical physicists, C.N. Yang and R. Mills, proposed to use nonabelian groups for the gauge theory to explain strong interactions [1]. Their work inspired others to use the nonabelian groups to investigate the weak force. In 1961, S. Glashow combined the electromagnetic and weak interactions [2]. In 1967, S. Weinberg [3] and A. Salam [4] incorporated the Higgs mechanism [5–7] into Glashow's electroweak interaction, giving it its modern form.

The Standard Model describes the strong, weak, and electromagnetic fundamental interactions using mediating *gauge bosons* (see Fig. 16.1). The species of *gauge bosons* are photons, $W-$, $W+$ and Z bosons, and eight gluons. The Standard Model also has 24 fundamental fermions (12 particles and their associated anti-particles), which are the constituents of all matter. The quarks can combine to form composite particles, accounting for the hundreds of species of particles that were observed in experiment so far. The Higgs mechanism is believed to give rise to the masses of all elementary particles in the Standard Model. This includes the masses of the W^{+-} and Z bosons, and the masses of fermions (i.e., the quarks and leptons).

The Standard Model has been shown to agree with almost all experimental tests conducted up to now. However, many particle physicists believe that it is an incomplete description of nature, and a more fundamental theory will be discovered in the future. Although the Standard Model is believed to be theoretically self-consistent, it leaves a lot of physical phenomena unexplained. There are still many questions remaining to be answered. It is well aware in the physics community that the Standard Model (in the present form) has certain limitations. For example:

- It cannot explain the observed predominance of matter over anti-matter.
- It cannot explain gravitation.
- The model does not contain any viable particles that can constitute dark matter.
- It does not explain the recent findings of neutrino oscillations and their non-zero masses.
- Many scientists consider the Standard Model to be *ad hoc* in nature, requiring 19 numerical parameters whose values are arbitrary and can only be determined by experiments [8].
- It is believed that explaining neutrino mass will require an additional 7 or 8 fitting parameters for the Standard Model [9].

Thus, there is still a long way to go for physicists to develop a comprehensive theory that can explain our physical world from the subnuclear scale to the cosmos scale. There is no wonder why many physicists in recent years are exploring new theories beyond the Standard Model (e.g., string theory, brane theory, super-symmetry, etc.). These new theories, however, are not supported by experiments in spite of many years of trying. There were doubts on whether these theories could work [10–12].

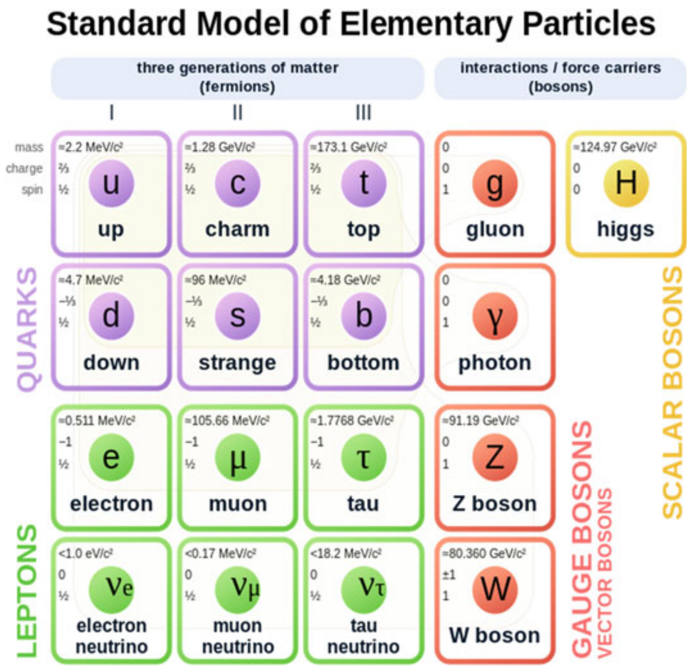


Fig. 16.1 Current Standard Model of particle physics. The above diagram summarizes the current view of elementary particles according to the Standard Model of particle physics. The three columns on the left list the fermions that make up matter: the upper two rows are three groups of quarks, and the lower two rows are three groups of leptons. The forth column from the left lists the force-transmitting bosons (also known as "gauge bosons"). The rightmost column shows the hypothetical Higgs particle. Image Credit: Wikimedia Commons, Public domain

16.2 Conceptual Difference Between the Standard Model and the Quantum Wave Model: *Particle Versus Waveticle*

Today, there are still many unanswered questions about fundamental physics. They cannot be answered by the current version of the Standard Model. As we have shown in the earlier chapters of this book, some of these questions can be answered using new ideas suggested in the *quantum wave model*. (For details, see the following discussions.)

The Standard Model and the *quantum wave model* use very different starting points. In the *Standard Model*, particle is conceptually regarded as a "*point mass*". In the quantum field theory, the particle is an excitation of its own field. For example, photons are excitations of the electromagnetic field, and electrons are quantum excitations of the electron field. Since there are many different types of sub-atomic particles, one must assume the existence of many different quantum fields. Then, people must wonder: **Why does nature need so many different *quantum fields*?**

Where do these *quantum fields* come from? What is the physical nature of all these *quantum fields*? So far, these questions have not been satisfactorily answered.

In the Standard Model, the properties of the particle, including rest mass, electric charge, and spin, are pre-existing inherent properties of the particle. They cannot be derived from the model.

In the *quantum wave model*, it was hypothesized that *all free particles are excitation waves of the vacuum; they just represent different excitation modes.* In this way, one does not need to deal with many different *fields*; one can only deal with one common wave medium, i.e., the *vacuum medium*. Here, all particles are quantized excitation waves in the form of solitons. Not only photons, but all sub-atomic particles, including electrons and neutrinos, are quantized wave packets of the *vacuum medium*.

In this case, the inherent properties of the so-called "particle" could be derived from the model. For example, as we have shown in Chaps. 7 and 11, the mass of the particle is found to be associated with the wavelength in the transverse oscillation of the traveling wave packet. This gives the *mass* a physical meaning at the same footing as *energy* and *momentum.*

We think it is misleading to call the quantum object a "*particle*". The word "*particle*" implies that the sub-atomic particle behaves as a "point mass", somewhat like a tiny billiard ball. This is clearly not true [13–15]. In the case of photon, we know it is a quantized electromagnetic wave. In the case of electron, the Bragg diffraction experiment and the double-slit experiment clearly indicated that it does not behave like a *point mass* in the classical sense [16, 17].

A second reason for rejecting the name "*particle*" is that an electron cannot be described as a point object in the classical way. In classical mechanics, one can describe an object by both its *position* and *momentum.* But in quantum mechanics, the *position* and *momentum* of an electron cannot be determined simultaneously [18]. The only explanation is that the quantum object is a wave packet rather than a point mass [13] (see Fig. 16.2).

Therefore, it is misleading to call the electron or photon a "*particle*". We need to find a better term to describe such quantum objects. One may call them "*wave/particles*". However, this name is a little bit long; thus, we suggest simplifying it by calling them "*waveticles*". With this new name, the quantum nature of *wave-particle duality* for electrons and photons becomes more clear (see Fig. 16.2).

In the following discussion, we will use the word "*waveticle*" to describe the quantum particle according to the *quantum wave model*.

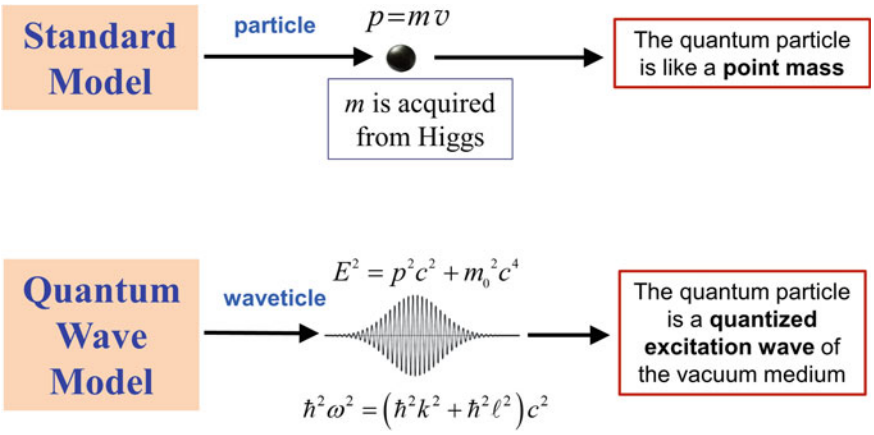


Fig. 16.2 Conceptual view of a particle in the Standard Model of particle physics versus that in the quantum wave model. In the Standard Model of particle physics, the particle is like a point mass. The mass of the particle is acquired from the interaction between the particle field and the hypothetical Higgs field. In the quantum wave model, the quantum particle is a quantized excitation wave of the vacuum medium. It is more appropriate to call it a "waveticle". The mass of the waveticle is determined from the dispersion relation of the excitation wave

16.3 What is the Origin of the Rest Mass? What is the Physical Meaning of Anti-Mass?

We know the universe is made up of particles; most of which have a rest mass. Where does this rest mass come from? Do particles obtain their mass from interaction with the Higgs field, as suggested in the Standard Model?

In the quantum field theory, a particle is regarded as an excitation of a field. Then, how can a particle acquire mass? According to the Standard Model, the mass of the particle is acquired through their interaction with the Higgs field, the excitation of which is a scalar particle called "the Higgs boson" [19]. In the Standard Model, the gauge field is introduced solely to allow the Lagrangian density to be invariant under a local gauge transformation [20]. The excitations of this field are the gauge bosons. The gauge boson originally has no mass. The gauge boson acquires mass due to its interaction with a hypothetical Higgs field with a broken symmetry. This mechanism was proposed by Higgs and others in the mid-1960s [5–7]. According to the Standard Model, not only the gauge boson particles (W^+, W^- and Z) acquire their rest mass due to their interaction with the Higgs field, leptons and quarks also acquire their mass in a similar way [3, 4, 21].

The Higgs particle had been theorized for a long time but was not verified in experiment until ten years ago. Experiments to hunt for the Higgs boson started at CERN around 2008 using the newly built Large Hadron Collider (LHC). On July 2012, two experimental groups at LHC reported independently that they found a new particle with a mass of about 125 GeV, which could be the Higgs boson [22, 23].

There was a lot of excitement in the physics community about their findings. Their results were considered to be a major triumph of the Standard Model!

16.3.1 What is the Problem with the Higgs Model of Mass Acquisition?

In the Higgs model, the rest mass of a particle is just a parameter associated with the strength of coupling between the particle field and the Higgs field. This thinking raises several conceptual problems. First, the physical meaning of the rest mass m defined in this way would be intrinsically different from energy E or momentum p. This does not seem to be very satisfactory in view of our traditional understanding of the physics concept.

Second, the Higgs mechanism does not seem to be consistent with the hypothesis of GR. How can a coupling constant between the particle field and the Higgs field cause a curving in space–time?

Third, how can the mass generated from Higgs convert to energy, or vice versa?

Fourth, the origin of the Higgs field is entirely based on conjecture. It is very strange. What is the physical origin of the Higgs field? What is the physical meaning of the Higgs particle? It is neither a kind of matter, nor is it a kind of force-carrier.

Finally, there seems to be a conceptual difficulty in the Higgs model. It is very difficult to use common-sense ideas to explain to a layman how particle may acquire mass through the Higgs mechanism. In the last 40 years, various analogies have been invented to describe the particle's interaction with the Higgs field, including analogies with a number of symmetry-breaking effects, such as the formation of rainbow from sunlight, separation of color using a prism, and resistance affecting some objects moving through syrup or snow [24]. However, none of these analogies appears to be satisfactory. In fact, analogies based on simple resistance to motion are misleading, since the Higgs field does not work by resisting motion of the particle.

Furthermore, **what is the physical meaning of *anti-mass*** according to the Higgs model? In the Standard Model, the particle itself does not have rest mass. The rest mass comes from the interaction between the particle and the hypothetic Higgs field. Then, how can one explain the origin of *anti-mass*. We know nature has both particles and anti-particles. The anti-particles do not carry "mass"; they carry "*anti-mass*". What does *anti-mass* mean according to the Higgs mechanism? We know mass is related to energy. If one believes the particle behaves like a point mass, does *anti-mass* mean *negative energy*? Can the Higgs mechanism within the Standard Model explain that?

16.3.2 The Explanation of Mass and Anti-Mass in the Quantum Wave Model is Far Simpler

In the *quantum wave model*, this question can be answered in a more natural way. We have shown in Chaps. 7 and 11 that the *rest mass* of a particle is actually associated with ℓ, the "*transverse wave number*" of the free particle. Recall that a quantum wave packet is undergoing both the translational motion and the transverse oscillation following the Bessel function of the first kind. The transverse oscillation is characterized by a "transverse wave length" (λ_T). The rest mass of the particle is shown to be proportional to the "transverse wave number", which is defined as $\ell = \dfrac{2\pi}{\lambda_T}$. This interpretation actually makes good sense, because now the *energy*, *momentum,* and *mass* can all mean similar things: *the energy (E) is inversely proportional to the wavelength in the time dimension*; *the momentum (p) is inversely proportional to the wavelength in the longitudinal spatial dimension*; and *the rest mass (m_0) is inversely proportional to the wavelength in the transverse spatial dimension.* Since "*the inverse of wavelength*" is a measure of the *curvature* of bending the wave medium, the particle properties, including *energy, momentum,* and *rest mass*, are all related to the *curvature of bending the vacuum medium during the propagation of the excitation wave*. Thus, the meaning of *mass* will be at the same footing as *energy and momentum.*

Now, what about *anti-mass*? Can one explain *anti-mass* according to the *quantum wave model*? The answer is yes. Since the rest mass of the *waveticle* is related to the transverse wave number ℓ, it would suggest that the *anti-mass* is associated with the negative transverse wave number $-\ell$. One may recall that in the *quantum wave model*, the wave packet representing a massive particle (i.e., the *waveticle*) looks like a vortex wave. In the direction perpendicular to the trajectory of the particle, the wave packet oscillates following the Bessel function of the first kind. This transverse wave number is ℓ. So, for the particle (*waveticle*) with anti-mass, it simply means that it has a negative transverse wave number $-\ell$. That would mean the oscillation of the excitation wave in the transverse direction is in an opposite phase (see Fig. 16.3).

Since the *waveticle* with positive and negative ℓ oscillates following the same Bessel function with identical wavelength, the magnitudes of the mass and anti-mass are the same. But because they oscillate at opposite phase, their $\pm$ signs become opposite.

Thus, the *quantum wave model* not only explains the origin of *mass* for the quantum particles, but also explains the physical basis of *anti-mass*.

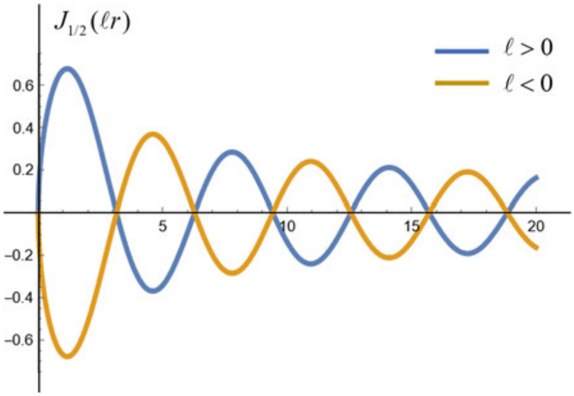


Fig. 16.3 Physical meaning of anti-mass for a waveticle. In the quantum wave model, mass means a positive ℓ while anti-mass means a negative ℓ. The anti-particle with anti-mass represents a wave packet with a transverse oscillation in an opposite phase in comparing to a particle. They both follow the same Bessel function

16.4 What is the Origin of Electric Charge in a Quantum Particle?

A quantum particle (such as an electron) not only has mass and spin, it also has electric charge. What is the origin of charge?

In the conventional quantum field theory, the origin of charge is totally unknown. It regards "electric charge" as an inherent property of the particle. Thus, the origin of *charge* cannot be explained in the Standard Model.

This question can be answered in the *quantum wave model* discussed in this book. Here, we can find a way to explain the origin of electric charge in a quantum particle. The key is that the charge comes from the vacuum medium.

According to the *quantum wave model*, the vacuum behaves like a dielectric medium. That means, the quantum vacuum is composed of a mixture of very refined primordial positive charges and negative charges. Thus, the quantum vacuum can be regarded as a superposition of two charged *mediums*, i.e., the negatively charged "*n-type medium*" and the positively charged "*p-type medium*". (For details, see Chap. 6.) Since the vacuum contains exactly the same amount of *n-type medium* and *p-type medium*, it has zero net charge at the resting state. The vacuum is entirely electrically neutral.

In the *quantum wave model*, the elementary particles are simply quantized excitation waves of the vacuum medium. (The "particle" is now called a "*waveticle*".). When there is an excitation wave generated in the vacuum, the primordial charges will be separated at the point of excitation. There can be three different possibilities: (1) The excitation wave could contain an equal amount of *n-type medium* and *p-type medium,* and thus is electrically neutral. This is like the case of a photon. (2) The excitation wave may contain mainly the *n-type medium*, and it will have a negative charge; the *waveticle* will become a particle with negative charge (such as an electron). (3) The excitation wave may contain mainly the *p-type medium,* and the *waveticle* will become a particle with positive charge (such as a positron).

Thus, the charge of the *waveticle* depends on whether it is a wave packet primarily generated from the *n-type medium* or the *p-type medium.*

In the *quantum wave model* discussed in this book, one can explicitly calculate the charge contained within the quantized wave packet in theory. As we showed in Chap. 7, the *wave function* of the quantum wave equation ψ essentially represents the *electric vector potential* $\mathbf{Z}$, which is related to the *charge displacement* $\mathbf{D}$ by the relation, $\mathbf{D} = \nabla \times \mathbf{Z}$. Now, using the Gauss Law $\nabla \cdot \mathbf{D} = \rho_e$, one can associate the *quantum wave function* ψ with the *charge density* ρ_e. Thus, it is possible to calculate the charge contained within a *waveticle*.

16.5 What is the Origin of Anti-Particle? What is the Physical Basis of Pair-Creation?

In nature, every type of particle has its own anti-particle. Why does nature need to have anti-particles? Furthermore, particle and anti-particle are often created in pair, why?

Historically, the origin of *anti-particle* was first proposed by Dirac. In Dirac's model, he assumed an infinite number of negative-energy electrons are pre-existing in the vacuum. When this "*Dirac sea*" of negative-energy electrons are excited by a sudden energy input, such as an incoming gamma ray, one of the negative-energy electrons can be kicked up to become a positive-energy electron, with a hole left behind. This hole will behave like an *anti-particle* of electron. Since an electron carries negative charge, the hole will have the opposite electric charge; it now becomes a positively charged *positron*. This is the explanation of the origin of anti-particles.

In the *quantum wave model*, there is no need to assume the pre-existence of an infinite "*Dirac sea*" of negative-energy electrons. It is easy to explain the origin of *anti-particles* from the wave view. According to the *quantum wave model*, the vacuum is a dielectric medium which is composed of a superposition of the negatively charged *n-type medium* and the positively charged *p-type medium.* At the resting state, the positive charges and the negatives charges are exactly balanced so that the vacuum medium is electrically neutral. When the vacuum is excited to generate a negatively charged *waveticle* representing an electron, the unbalanced positive charges left behind will form another *waveticle* with opposite charges, which becomes the anti-particle *positron* (see Fig. 16.4).

So, the creation of *anti-particle* is mainly for satisfying the requirement of *conservation of charge in the vacuum*. In fact, the pair-creation of particle and anti-particle also can satisfy the requirements of *conservation of momentum* and *conservation of spin*. When a negatively charged excitation wave is created in the vacuum, another excitation wave with positive charges must also be created. These two wave packets have opposite momentum. They travel at the same speed but in opposite directions, when one *waveticle* goes to the left, the other *waveticle* goes to the right. They also carry opposite spins.

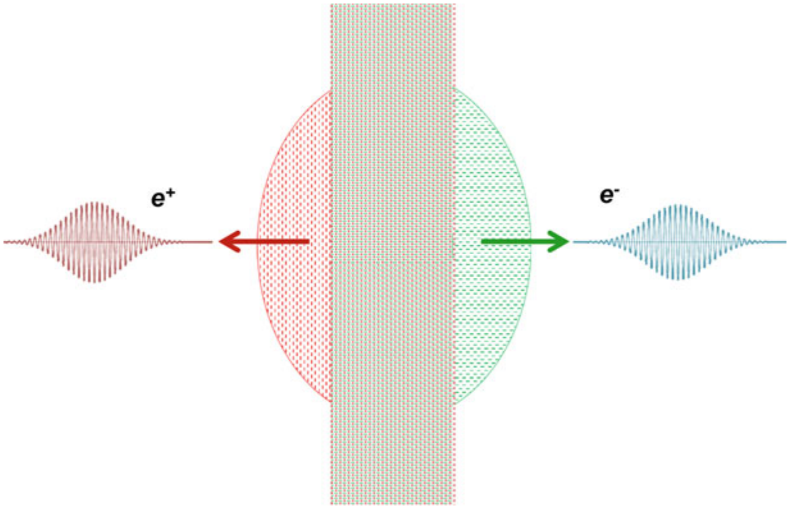


Fig. 16.4 Pair-creation of particle and anti-particle. Pair-creation of particle and anti-particle is mainly due to the requirements of conservation of charge, momentum, and spin. This is a conceptual illustration of the pair-creation of electron–positron pair from the vacuum medium. The excitation of the vacuum medium in the middle generates a wave packet of a positron to the left (red) and a wave packet of electron to the right (green)

From the foregoing argument, one can easily see why in nature, *particle and anti-particle must be created in pairs*. If a particle is created without generating an accompanying anti-particle, it will be impossible to satisfy the condition of *conservation of charge* in the vacuum. The amount of net charge in the non-exciting vacuum must always be kept at zero due to the exact balance between the negatively charged "*n-type medium*" and the positively charged "*p-type medium*".

16.6 Chapter Summary

- In this chapter, we discuss the similarity and difference between the basic ideas of the *quantum wave model* and the *Standard Model of particle physics*.
- At present, there are still many important questions that the Standard Model cannot fully explain. The quantum wave model may offer some new ideas for inspiration.
- Unlike the Standard Model which assumes that nature has many different *quantum fields*, the quantum wave model hypothesizes that *all free particles are excitation waves of one single wave medium, i.e., the vacuum*.
- We point out that a quantum particle such as an electron is not like a *point mass*. It is a *quantized excitation wave* of the vacuum medium. Thus, it is more appropriate to call the quantum object (such as an electron) a "*waveticle*" instead of a "*particle*".

- Unlike the Standard Model, which assumes that particles obtain their mass from interaction with the Higgs field, the quantum wave model proposes that the *rest mass* of a particle is originated from its "*transverse wave number*". This proposal can also explain the origin of "*anti-mass*".
- What is the origin of *electric charge* in a quantum particle? According to the quantum wave model, the charge comes from the vacuum medium.
- This also explains why *particle* and *anti-particle* are created in pair. The pair-creation of *particle* and *anti-particle* is to satisfy the requirements of conservation of *electric charge*, *momentum*, and *spin* in the vacuum.

References

1. Yang, C. N., & Mills, R. L. (1954). Conservation of isotopic Spin and isotopic gauge invariance. *Physical Review, 96*, 191–195.
2. Glashow, S. L. (1961). Partial-symmetries of weak interactions. *Nuclear Physics, 22*, 579–588.
3. Weinberg, S. (1967). A model of leptons. *Physical Review Letters, 19*, 1264–1266.
4. Salam, A. (1968). Elementary particle physics: Relativistic groups and analyticity. In N. Svartholm (Ed.), *Eighth nobel symposium.* Almquvist and Wiksell.
5. Englert, F., & Brout, R. (1964). Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters, 13*, 321–323.
6. Higgs, P. W. (1964). Broken symmetries and the masses of gauge bosons. *Physical Review Letters, 13*, 508–509.
7. Guralnik, G. S., Hagen, C. R., & Kibble, T. W. B. (1964). Global conservation laws and massless particles. *Physical Review Letters, 13*, 585–587.
8. Chang, D. C. (2013). A classical approach to the modeling of quantum mass. *Journal of Modern Physics, 4*, 21–30.
9. Strumia, A., & Vissani, F. (2006). Neutrino masses and mixings and.... arXiv: hep-ph/0606054
10. Woit, P. (2006). *Not even wrong: The failure of string theory and the search for unity in physical law.* Basic Books.
11. Smolin, L. (2006). *The trouble with physics: The rise of string theory, the fall of a science, and what comes next.* Houghton Mifflin.
12. Hossenfelder, S. (2017). Science needs reason to be trusted. *Nature Physics, 13*, 316–317.
13. Chang, D. C. (2021). Review on the physical basis of wave–particle duality: Conceptual connection between quantum mechanics and the Maxwell theory. *Modern Physics Letters B, 35*, 2130004.
14. Chang, D. C. (2017). Physical interpretation of the Planck's constant based on the Maxwell theory. *Chinese Physics B, 26*, 040301.
15. Hey, T., & Walters, P. (2003). *The new quantum universe.* Cambridge University Press.
16. Davisson, C. J., & Germer, L. H. (1927). The scattering of electrons by a single crystal of nickel. *Nature, 119*, 558–560.
17. Bach, R., Pope, D., Liou, S.-H., & Batelaan, H. (2013). Controlled double-slit electron diffraction. *New Journal of Physics, 15*, 033018.
18. Messiah, A. (1965). *Quantum mechanics,* Vol. 1. John Wiley & Sons.
19. Liu, G.-Z., & Cheng, G. (2002). Extension of the Anderson-Higgs mechanism. *Physical Review B, 65*, 132513.
20. Oerter, R. (2006). *The theory of almost everything: The standard model, the unsung triumph of modern physics.* Pi Press.
21. Cottingham, W. N., & Greenwood, D. A. (1998). *An introduction to the standard model of particle physics.* Cambridge University Press.

22. Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Khalek, S. A., Abdelalim, A. A., Abdinov, O., Aben, R., Abi, B., & Abolins, M. (2012). Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B, 716*, 1–29.
23. Chatrchyan, S., Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Aguilo, E., Bergauer, T., Dragicevic, M., Erö, J., & Fabjan, C. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B, 716*, 30–61.
24. Ellis, J. What is the Higgs Boson? https://home.cern/resources/video/physics/what-higgs-boson-john-ellis-theoretical-physicist

Chapter 17
Investigating the Quantum Properties of Nucleons from a Wave View

In this book, we proposed that free particles are excitation waves of the vacuum. These free particles (we call them "*waveticles*") include photons, electrons, and other leptons. They are stand-alone particles with no internal sub-component. On the other hand, we know other sub-atomic particles (such as protons and neutrons) are composed of sub-particles (quarks). These composite particles are called "*hadrons*". Are these hadrons excitation waves of the vacuum too?

We think the answer is "yes". However, there are basic differences between the elementary particles discussed so far and the hadrons. The size of the photons and electrons is apparently relatively large (in comparison with an atomic nucleus). Their wave packet could be very long (say, in the order of nanometer or longer). The size of a hadron is very small; the size of a proton is estimated to be less than 10^{-15} m. Apparently, the size of quarks must be even smaller. (Note: the current estimated size of the quark is in the order of 10^{-18} m.)

In nature, there are *long-range* and *short-range* forces. We think the *photon/lepton* and the *hadron/quark* belong to two different types of excitation waves of the vacuum. Photons and leptons are excitation waves driven by the *long-range* force, which is the electromagnetic force. The hadrons and quarks are quantized excitation waves of the vacuum driven by the *short-range* forces, which include the strong and weak nuclear forces.

At present, we have very good knowledge about the physical properties of the *long-range* force. They are basically described by the Maxwell equations. Thus, we can derive the wave equations for the free particles associated with the excitation waves of the *long-range* force. For the *short-range* forces, their physical properties are not yet clearly known. At this time, it is not possible to write down the quantum wave equations for hadrons or quarks. Thus, there are many outstanding questions about the structure and properties of the nucleons (for details, see below).

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_17

17.1 The Building Blocks of Nature: What is an Elementary Particle?

17.1.1 Elementary Particles versus Composite Particles

In our universe, most matters are in the form of atoms, which are composed of electrons and nucleons (protons and neutrons). In the old days, all sub-atomic particles are called "*elementary particles*". Now, it is no longer true. In the study of particle physics using colliders, it was found that new particles can be created during the collision, and some particles can decay to other particles. These observations suggested that some of the sub-atomic particles can be composed of other particles. They should be called "*composite particles*". Protons and neutrons are found to belong to this class. On the other hand, leptons, including electron, muon, tau, and neutrinos, are not composite particles. They remain to be classified as "elementary particles".

With the development of the Standard Model in the later part of the twentieth century, it became widely accepted that nucleons are composed of more elementary particles called "*quarks*". Scientists now believe that there are six different quarks which can be grouped into 3 generations (*up/down, strange/charm, top/bottom*). Furthermore, each quark can have three different "*colors*" (*red*, *green*, and *blue*). This so-called "*color*" here is not a real color, but a hidden degree of freedom for the quarks. Each quark also has its own anti-particle. All together, we have 36 quarks.

Besides protons and neutrons, many new particles were found to be composed of quarks. Particles formed by 3 quarks are called "*baryons*", while particles formed by 2 quarks are called "*mesons*". Together they are called "*hadrons*".

From now on, we will call all *sub-atomic particles* found in nature as "*quantum particles*". These quantum particles are classified into two categories: (1) *Elementary particles*, which include *leptons* and *gauge bosons* (such as photon), as well as *quarks*. (2) *Composite particles*, which include mainly *hadrons* (see Fig. 17.1).

In the Standard Model of particle physics today, "elementary particles" can be subdivided into "fermions" and "bosons". *Fermions* are the building blocks of matters, which can be sub-divided into *leptons* and *hadrons*, while *bosons* are thought to be mediators of forces. For example, photons are the mediator of electromagnetic force,

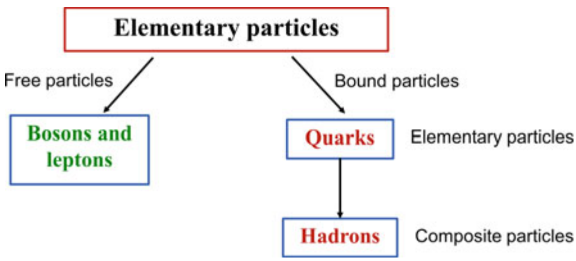


Fig. 17.1 Classification of elementary particles. Elementary particles can be classified into free particles and bound particles. Gauge bosons, leptons, and quarks are elementary particles; hadrons are composite particles made up of quarks

Z and $W^{\pm}$ are mediators of the weak nuclear force, while gluons are the mediators of the strong nuclear force (see Fig. 16.1 in the last chapter).

17.1.2 Free Particles versus Bound Particles

In our physical world, not all elementary particles are free particles, and not all free particles are elementary particles. For particles that we have discussed so far, like electrons and photons, they are elementary particles, and they can travel in the space freely. Other elementary particles, such as quarks, cannot travel freely; they can only be confined with other quarks to form a complex structure. They are not free particles. Some sub-atomic particles, such as protons and neutrons, on the other hand, can travel freely by themselves. But they are not elementary particles. They are "composite particles".

So, the elementary particles can be sub-divided into "*free elementary particles*" and "*bound elementary particles*". In the *quantum wave model* discussed in this book, it is proposed that all *free elementary particles* (such as photons and electrons) are quantized excitation waves of the vacuum medium, which are driven by the long-range force. We will further hypothesize that the *bound elementary particles* (quarks) and their composite particles (hadrons) are also excitation waves of the same vacuum medium, but they are driven by the short-range forces. Using our terminology, they are all "*waveticles*"!

Why do we think hadrons are wave? This hypothesis is based on the following reasons. First, if the hadron is not a wave, one cannot explain why it can be created in the vacuum or annihilated into nowhere. Second, we know most hadrons are not stable; they can decay to give leptons and photons. This decaying process suggests that hadrons are also excitation waves of the vacuum.

17.2 Long-Range Force *versus* Short-Range Forces in the Vacuum Medium

In our model, there is one major difference between the leptons and the hadrons. We know there are four different types of forces in nature, the electromagnetic force, the weak nuclear force, the strong nuclear force, and gravity. The electromagnetic force is long-range, while the strong and weak nuclear forces are short-range (in the order of 10^{-15} m or less). In nature, a force must be transmitted through a medium. So, what are the media for transmitting the long-range and short-range forces? We know the vacuum is a mediator for the electromagnetic force. It is very likely that the same vacuum is the mediators also for the strong and weak forces. Otherwise the space of our Universe must be filled with multiple media, which would make the **nature** far more complicated. In the quantum wave model, we proposed that **the**

free elementary particles (such as photons and electrons) **are quantized excitation wave of the vacuum medium driven by the long-range force**. That is why the quantum wave equation of electron can be derived from the Maxwell theory. For **the *bound elementary particles*** (quarks), they **are excitation waves driven by the short-range nuclear forces** (see Table 17.1). At this time, it is not yet known how to model the action of short-range forces on the vacuum medium. That requires a set of equations different than Maxwell equations. We know very little about it. This will be the work of the future.

In summary, there could be two different types of excitation waves transmitted through the vacuum: **The first type is driven by the long-range force**; photons, scalar particles, electrons, and other leptons all belong to this type. They appear as *free elementary particles*. **The second type is mainly driven by the short-range forces**. It includes all quarks; they appear as *bound elementary particles*. Interestingly, since the *free elementary particles* are driven by the long-range force, their wavelength is usually very long (say, from Å to meters). Conversely, since the *bound elementary particles* (quarks) are driven by the short-range forces, their wavelength usually is very short (estimated to be far less than 10^{-15} m, the size of a nucleus). For the *composite particles* such as the hadrons, they can be in a *bound* state (inside a nucleus) or in a *free* state (outside of a nucleus). They may be driven by both the short-range and long-range forces within the vacuum.

Table 17.1 Classification of various quantum particles

	Particle species	Boson *versus* fermion	Matter *versus* force carrier	Long-range force *versus* short-range force	Remarks
Elementary particle (free)	Photon	Boson	Force carrier	Long	Radiation wave
	Leptons	Fermion	Matter	Long	Electron is the major constituent of an atom
	Other bosons	Boson	Force carrier	Long	They include gauge bosons, Higgs, and other yet to be discovered bosons
Elementary particle (bound)	Quarks	Fermion	Matter	Short + long	It cannot be ruled out that quarks are formed by more elementary objects
Composite particle	Hadrons (baryon)	Fermion	Matter	Short + long	The atomic nucleus is composed of protons and neutrons
	Hadrons (meson)	Boson	Matter??		Could mesons form matter?

17.3 Internal Structure of Nucleons and the Atomic Nucleus

17.3.1 Current Understanding Based on the Particle View

Hadrons are believed to be composed of quarks. For example, a proton is thought to be composed of two *up* quarks (each with electric charge 2/3 e) and one *down* quark (with $-1/3$ e) (see Fig. 17.2). Each of these quarks is supposed to have a different "color", so that the combined "color" of the hadron is zero (i.e., "white").

The quarks within the proton are supposed to be held together by the short-range nuclear force called "strong interaction". In the Standard Model, this strong interaction is mediated by a massless gauge boson called "gluon". Gluon is considered to be a quantum particle associated with the "color" force between quarks [1]. It is thought that gluons couple to the color charges of the quarks according to the theory of quantum chromodynamics (QCD).

Quarks only account for a small amount of mass. Most of the hadron mass is contributed by the binding energy through gluons. For example, a proton is thought to be made up of two *up* quarks and one *down* quark. The resting energies of the quarks only account for a small fraction of the rest mass of the proton (see Table 17.2). The majority of the proton rest mass is thought to be contributed by the binding energies.

Starting from the early days of nuclear physics, it has been believed that the atomic nucleus is made up of individual protons and neutrons. These nucleons are thought to be held together by the strong nuclear force. Since gluon is the only known mediator of the strong force, it would imply that protons and neutrons are held together in the nucleus by gluons. The detailed mechanism, however, is not yet clear.

Hadrons in general are not stable. With the exception of proton, all "free" hadrons (not bound within an atomic nucleus) are unstable. They will eventually decay into other particles. Proton is the only hadron that is stable either in a free state or bound within an atomic nucleus. Neutron, by comparison, is stable only inside the atomic

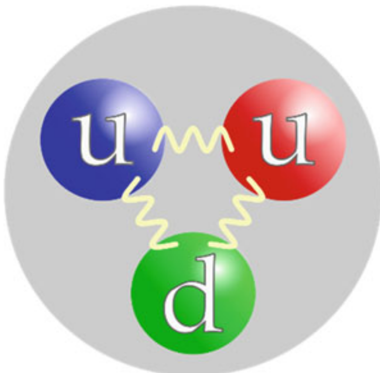


Fig. 17.2 Structure of a proton. A proton is thought to be composed of two up quarks and one down quark, holding together by gluons. Image Credit: Arpad Horvath, Wikimedia Commons, CC BY-SA 2.5

Table 17.2 Mass of two hadrons and their containing quarks

Particles	**Mass** (MeV)	**Reference**
Proton (uud)	938.27	Ref. [2, 3] https://pdg.lbl.gov/
Neutron (udd)	939.57	Ref. [3]
Up quark	1.8–3.0	Ref. [2]
Down quark	4.5–5.3	Ref. [2]

Note: The mass listed here is actually the equivalent energy, i.e., $E_0 = m_0c^2$

nucleus. When it is outside of the nucleus, it becomes unstable and will decay in about 879 s.

17.3.2 Speculation from the Wave View

According to the Standard Model, hadrons are composite particles composed of *quarks*. In the wave view, quarks are also excitation wave of the vacuum. But unlike the free elementary particle (such as the electron), the quark is a bound particle which does not travel freely in space. It is supposed to be trapped inside a small space within a hadron. In our model, we proposed that the vacuum can have two different types of waves. The first type is a propagating wave. The second type is a localized wave, like a vortex. The free particle such as an electron is a waveticle like the first type. The quark is a waveticle more like the second type. As an analogy, we can think of the second type of wave (localized wave) as a vortex-like structure (e.g., water whirlpool, tornado or hurricane) (see Fig. 17.3). In fact, the idea of *vortex* had been used to model a wide variety of physical objects, from the tiny *Majorana fermion* to the huge structure of a galaxy [4, 5]. In our view, the quantum *wave function* of the quarks should represent a displacement of the vacuum medium, just like the other elementary particles (*waveticles*).

Since the nucleon is made up of multiple quarks, a nucleon at rest would look like a complex wave structure composed of multiple vortex-like quarks (see Fig. 17.4). But unlike an individual quark, the composite particle (such as a proton or neutron) can travel freely in space with high speed. It is a free *waveticle*.

How can the quarks be held together inside a hadron? In the Standard Model, they are held together by the exchange of gluons. In our wave model, it is not clear whether gluons really exist or not. We know the strong force is responsible for holding the quarks together. But how this works is still unclear. In fact, there is no direct evidence showing the existence of gluon, since the gluon cannot be isolated from the hadron. The confirmation of the gluon concept can only rely on indirect evidence.

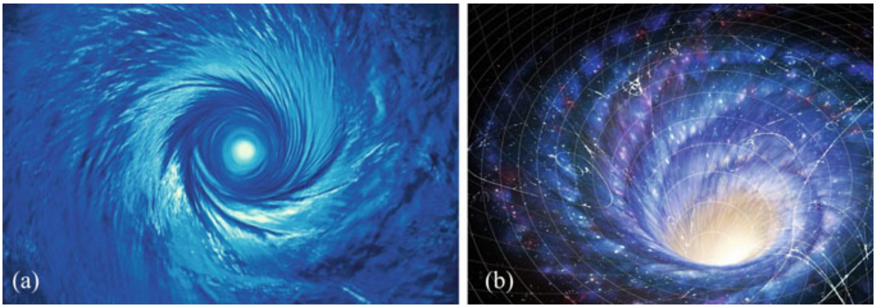
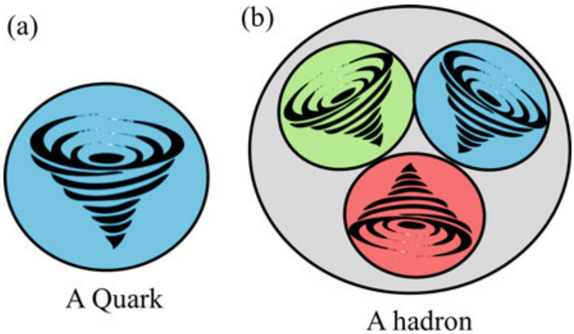


Fig. 17.3 Bound elementary particle (quark) can be regarded as a localized vortex of the vacuum. Vortex models are commonly used to model a wide range of physical systems, from **a** very microscopic Majorana fermion to **b** the very macroscopic galaxy formation. In this work, we propose that the bound elementary particle (quark) can also be modeled as a vortex. Image Credit: **a** Pacchioni, G. Into the vortex. Nat Rev Mater **4**, 79 (2019); **b** University of Warwick/Mark Garlick

Fig. 17.4 A conceptual wave model for a nucleon. **a** A quark can be modeled as a localized vortex. **b** Then, a nucleon such as a proton (a hadron) can be considered as a combined structure made of three different such vortexes



17.3.3 Could the Atomic Nucleus be an Aggregation of Quarks Instead of an Aggregation of Nucleons?

In the traditional view, the atomic nucleus is thought to be an aggregation of nucleons (protons and neutrons) (see Fig. 17.5a). But this may not be the true picture. Since nucleons are supposed to be held together by the force of strong interaction, and so are the quarks within an individual nucleon, there may not be a clear boundary between neighboring nucleons. Quarks belonging to different nucleons could interact just like quarks within the same nucleon. If this is the case, the entire atomic nucleus is just an aggregation of interacting quarks, instead of a cluster of individual protons and neutrons.

Take the helium nucleus as an example. It is not necessary to be a composite of two individual protons and two neutrons. It can be a "cage" containing 12 quarks (6*u* and 6*d*); these quarks are moving rapidly inside it (see Fig. 17.5b). In other words, the atomic nucleus is just an aggregation of entangled quarks; there is no individual

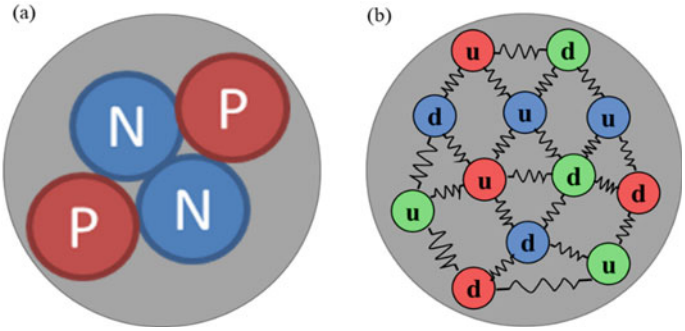


Fig. 17.5 Two views of the helium nucleus structure. **a** The traditional view of the helium nucleus. **b** One may regard the helium nucleus as an aggregation of entangled quarks

proton or neutron inside it. The proton and neutron may appear as individual particles only when they are ejected outside of the nucleus.

From such a view, there is no conceptual difference between the nucleus of helium atom and the nucleus of hydrogen atom (i.e., a proton). They are all composite particles of quarks. This can also be true for nuclei of higher atomic numbers. This means that **a nucleus can be generally regarded as a composite particle**. Indeed, for radioactive isotopes emitting alpha particles, the radiation particle is identical to the nucleus of the helium atom.

Thus, we can expect that a nucleus under motion will have similar wave properties as a hadron. This expectation is indeed confirmed in experiments. Diffraction experiments using particle beams had indicated that, like protons and neutrons, the nuclei of helium were also found to behave as waves and follow the de Broglie relation [6, 7].

17.4 Outstanding Questions in Particle Physics

17.4.1 What Hold Nucleons Together in a Nucleus?

Most nuclear physicists today agree that nucleons inside an atomic nucleus are held together by the strong force. So how to explain the strong force in the nucleus? In the 1930s, a Japanese physicist named Hideki Yukawa (see Fig. 17.6) proposed a model to describe the strong force. He hypothesized that the strong force between protons and neutrons is expressed through the exchange of an intermediary particle. This idea could be inspired by the implication of the quantum field theory of electron at that time. After the establishment of Dirac's electron theory, the quantum field theory (called "quantum electrodynamics") was soon developed. In this theory, the interaction between electrons can be explained by the creation and annihilation of photons. This was interpreted as the start of a new idea that "the interaction between electrons is through the exchange of photons".

In Yukawa's model, he proposed that the intermediary particle responsible for the nucleon-nucleon interaction is a kind of "*meson*" (mediating particle) [8]. According to the distance that the strong force can exist, he estimated that the mass of this meson is about hundreds of times that of an electron, and much smaller than that of a proton. At that time, no such particles were found. In 1936, a new particle called "muon" was discovered from the observation of cosmic rays [9]. The mass of the muon (~100 meV) is about 200 times the mass of the electron. Since the muon mass is between the mass of an electron and a proton, it fits Yukawa's prediction. People at that time thought that this muon was the meson predicted in the Yukawa's model.

Later experiments, however, found that the muon is a lepton and does not participate in the strong interaction; it cannot be the "meson" predicted in Yukawa's model [10].

After World War II, the energy of accelerators became larger and larger, and more and more new particles were discovered. A new particle called "pion" was discovered in 1947 by C F Powell and his colleagues [11]. Its mass is larger than that of muons (~140 meV), and its properties are more in line with the predictions of Yukawa's model. So, people thought that the pion is the "meson" predicted in Yukawa's model [10]. Yukawa was awarded the Nobel Prize in 1949; Powell also won the Nobel Prize in 1950.

Ironically, people later discovered that the pion is actually a hadron composed of two quarks, it is not a particle for mediating force. Therefore, the earlier identification of the pion being the transmitter of the strong force between nucleons was just another misunderstanding.

Fig. 17.6 Hideki Yukawa. Hideki Yukawa (1907–1981) was a Japanese theoretical physicist. In the 1930s, Yukawa proposed a model to explain the strong force; namely, the nucleons are held together by exchanging a force mediating particle called "meson". The idea that force is transmitted by particles had a great impact on the development of particle physics. His model won him a Nobel Prize in Physics in 1949. Photo Credit: General Atomic Division of General Dynamics Corporation courtesy AIP Emilio Segrè Visual Archives, Physics Today Collection

Later, Murray Gell-Mann and others developed the quark model, which proposed that the strong force is transmitted through a gauge boson called "gluon" [12]. This quark theory is now generally accepted. However, although Yukawa's strong force model was proven to be incorrect, the view that "***forces are mediated through the exchange of intermediary particles***" has become a mainstream idea in particle physics.

17.4.2 Could Quarks be Composed by Even More Elementary Objects? Could the Concept of Particle Go Away at Higher Energy?

In the last century, physicists have made tremendous progress in understanding the way how nature works. The Standard Model of particle physics is no doubt one of the triumphal achievements in the study of physics. However, the subject of understanding nature in the most microscopic scale is an exceedingly difficult challenge. The Standard Model was developed mainly half a century ago. In comparing the human history of studying nature, half a century is a very short time. So, it is not surprising that we are still far from getting the last answer of how nature works. No matter how good or how reasonable the model we are building today, it still requires many experimental tests and theoretical refinements.

Take the Standard Model of elementary particle for example. How good is it now? According to one of the leaders in this field, Steven Weinberg (see Fig. 17.7), we are still very far from getting the final theory. For example, in one of Weinberg's paper entitled "*What is An Elementary Particle?*", he said: "*I would have to admit that no one really knows*" [13]. According to Weinberg:

> From this point of view, we are entitled only to say that the quarks and gluons are more elementary than nucleons and pions, because their fields appear in a theory, the Standard Model, that applies over a much wider range of energies than the effective field theory that describes nucleons and pions at low energy. We cannot reach any final conclusion about the elementarity of the quarks and gluons themselves. The Standard Model itself is probably only an effective quantum field theory, which serves as an approximation to some more fundamental theory whose details would be revealed at energies much higher than those available in modern accelerators, and which may not involve quark, lepton, or gauge fields at all [13].

About whether quark could be composite of more elementary particles, Weinberg thought that is highly likely: *"The fact that we see no structure in the quarks and leptons only tells us that the energies involved in their binding must be quite large—larger than several trillion electron volts. But so far no one has worked out a convincing theory of this sort*" [13].

Many physicists are now wondering whether the Standard Model of particle physics today is going to be replaced by a more complete theory in the future. At this

Fig. 17.7 Steven Weinberg. Steven Weinberg was a major contributor to the Standard Model of particle physics. He was awarded the Nobel Prize in Physics in 1979 for the contributions to the theory of the unified weak and electromagnetic interaction. Photo Credit: AIP Emilio Segre Visual Archives

point, the so-called "elementary particles" may not be elementary at all. According to Weinberg:

> We will not be able to give a final answer to the question of which particles are elementary until we have a final theory of force and matter. When we have such a theory, we may find that the elementary structures of physics are not particles at all. Many theorists think that the fundamental theory is something like a superstring theory, in which quarks, leptons, etc. are just different modes of vibration of the strings [13].

17.4.3 Could the Problem be Solved by the String Theory?

In order to go beyond the Standard Model of particle physics and to solve the well-known problem that the quantum theory is incompatible with general relativity, some physicists started to develop a new theory called "*string theory*" about half a century ago. The string theory is a theoretical framework in which the point-like particles in the Standard Model are replaced by one-dimensional objects called *strings.* Instead of treating sub-atomic particles as the fundamental building blocks of matter, the string theory proposed that everything is made of extremely tiny strings, whose vibrations produce effects that we interpret as atoms, electrons, and quarks.

At the early stage, the String theory was promoted in a very high-profile manner. It was called the "*theory of everything*" [14]. That is, it is a single mathematical framework capable of describing the entirety of the known universe.

In order for that to work, string theory has to make several radical assumptions [14]. For example, it hypothesized that we must be living in a world with 10–11 dimensions. Some versions of the theory would require even higher number of dimensions of space–time. Furthermore, it assumes that these extra dimensions are curled up so tightly so that we will not notice them in the ordinary life.

The string theory was highly popular among some of the particle physicists and cosmologists during the past 30 years. Its popularity has subsided more recently due to two reasons. First, so far, it has failed to make any substantial predictions that could supply a breakthrough on the current theory. Second, it is exceedingly difficult or even impossible to design experiment to test the predictions of this theory.

To construct models of particle physics based on string theory, physicists typically begin by specifying a shape for the extra dimensions of space–time. Each of these different shapes corresponds to a different possible universe, or "vacuum state", with a different collection of particles and forces. The current version of string theory has an enormous number of vacuum states, estimated to be around 10^{500} [15]. Although this could be an advantage for using the string theory to model a multiverse theory, it is sufficiently diverse to accommodate almost any phenomenon. Many physicists regard this profligacy as string theory's fatal flaw. If a theory makes so many different, contradictory predictions, then almost any set of observations could be found to confirm it. In other words, that makes it almost impossible to falsify the theory.

Many critics of string theory have expressed concerns about the large number of possible universes described by the string theory [15–17]. For example, in a recent book entitled "*Not Even Wrong*", it was argued that the large number of different physical scenarios will make the string theory useless as a framework for constructing models of particle physics [15].

17.5 The Mass of a Composite Particle

In Chap. 11, we proposed that the origin of mass in a free particle (such as an electron) is associated with a transverse wave number. For the mass of a composite particle, it is much more complicated. This is because the internal structure of a composite particle is not clearly known.

At present, it is not possible to write down a simple quantum wave equation for hadrons. But this may not prevent us to get some ideas on their basic properties. Based on experimental observations and from general analogies between free particles (simple and composite), we can still obtain some basic information about the relationship between energy, momentum, and mass for the composite particles. The most important thing is that, like the free elementary particles, *composite particles also behave like waves*. As we showed earlier, electrons can be diffracted from a crystal following the Bragg diffraction law [18]. It was found later that neutrons and protons can also be diffracted in a similar manner [6, 7]. This implies that composite particles also have wave properties like the electron and they follow the de Broglie relation too.

This similarly allows us to speculate that *the energy–momentum relation for a composite particle is similar to that of a free elementary particle*. In Chap. 11, we already showed that, for a free particle such as an electron, the resting energy and moving energy of a single particle appear to form a two-dimensional Hilbert space. The total energy of the particle is the vector sum of these two energies. We propose that the same is true for a composite particle. That is, Eq. (11.29) of Chap. 11 is also applicable for the composite particles,

$$E^2 = c^2 p^2 + E_0^2. \tag{17.1}$$

Here, E_0 is the zero-point energy of the composite particle, which is the amount of energy possessed by the particle when there is no translational motion. What is the source of the resting energy? A natural expectation is that it must be contributed by the intrinsic energy of individual quarks and their binding energies.

Once we know the relation between energy and momentum, we can derive the relationship between energy and mass. This can be done by differentiating Eq. (17.1) versus p,

$$2E\frac{\mathrm{d}E}{\mathrm{d}p} = 2c^2 p$$

Since the particle is a wave packet, $\mathrm{d}E/\mathrm{d}p = \mathrm{d}\omega/\mathrm{d}k = v$. Recall that $p = mv$, the above equation becomes

$$E = mc^2. \tag{17.2}$$

At $v = 0$, $p = 0$, then

$$E_0 = m_0 c^2. \tag{17.3}$$

Thus, no matter whether a free particle is a simple particle or a composite particle, one can always define a rest mass which is directly related to the resting energy. And, by substituting Eq. (17.3) into Eq. (17.1), we can obtain the "relativistic" energy–momentum relation for a composite particle,

$$E^2 = c^2 p^2 + m_0^2 c^4. \tag{17.4}$$

Using Eq. (17.2) and $p = mv$, this can further give

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{17.5}$$

By comparing Eqs. (17.2)–(17.5) with Eqs. (11.24)–(11.27) in Chap. 11, it is clear that the relations between E, p, and m are the same for both free elementary particles and composite particles.

17.5.1 Origin of Rest Mass for a Composite Particle

The origin of the rest mass is different between a free elementary particle and a composite particle. As we had discussed in the earlier chapters, for the free particle (such as an electron), the rest mass is associated with a transverse wave number ℓ. For the composite particles (such as a proton or neutron), the rest mass is apparently associated with the internal energy of the constituents which make up the composite particle. For example, a proton is thought to be made up of two *up* quarks and one *down* quark. Each of these quarks is an elementary particle; it has its own resting energy (rest mass). But in addition to that, there are also binding energies between quarks. These energies are associated with the strong force needed to hold the quarks together. In fact, the resting energies of the quarks only account for a small fraction of the rest mass of the proton (see Table 17.2). The majority of the proton rest mass appears to be contributed by the binding energies.

Finally, one may ask: What is the relation between *E, p,* and *m* for the atomic nucleus? We know the nucleus of the hydrogen atom is made up of a single proton. This suggests that there can be some sort of similarity between a nucleus and a hadron. Indeed, as we discussed in the earlier part of this chapter, the atomic nucleus can be regarded as an aggregation of interacting quarks instead of a cluster of protons and neutrons (see Fig. 17.5b). In another word, the nucleus can be thought of as a soup of quarks.

Thus, we can expect that a free nucleus could have similar wave properties as a hadron. This suggests that Eqs. (17.1)–(17.5) are applicable also to the nucleus of an atom. **The relationships between energy, momentum, and mass are no different between an atomic nucleus and a hadron particle**. Then, we can conclude that, for free particles of different kinds, including elementary particles like photons and electrons, and composite particles like protons and atomic nuclei, they all follow the same energy–momentum relation, i.e.,

$$E^2 = c^2 p^2 + E_0^2. \tag{17.1}$$

In our model, all types of free particles are waveticles; they have similar energy–momentum relations, which appear as parallel lines in the E^2 *versus* p^2 plot (see Fig. 17.8). The only difference between particles is that their intersects with the vertical axis are not the same. That means, the resting energy E_0 for different particles is different. Since the resting energy is directly related to the rest mass by $E_0 = m_0 c^2$, this reflects that the rest mass is different for different kinds of free particle.

The predictions shown in Fig. 17.8 are fully consistent with observations in experiments conducted so far. The fact that the slope of the energy–momentum relation is the same for all particles (slope $= c^2$) suggests that both the "simple particles" (i.e., free elementary particles) and the composite particles are excitation waves of the same medium. We know the propagation speed of a wave is determined by the physical properties of the wave medium. It can be shown easily that the ultimate speed of a waveticle is determined by the slope of the E^2 *versus* p^2 plot. Since the slope of photons is the same as that of electrons and hadrons, there is no basic difference between the radiation wave and the matter wave. They are just different excitation modes of the vacuum medium! Thus, the results plotted in Fig. 17.8 are consistent with the basic assumption of the quantum wave model that all free particles in our universe are excitation waves of the same *quantum vacuum*.

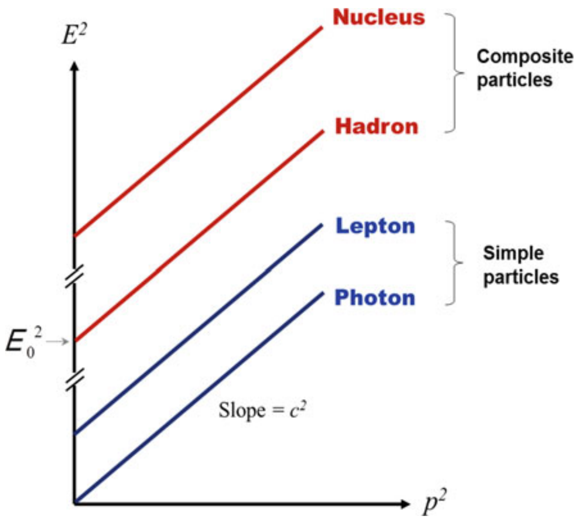


Fig. 17.8 All types of free particles (waveticles) have similar energy–momentum relations. The slope of E^2 versus p^2 plot is the same for all particles (slope $= c^2$), but the intercept E_0^2 is different for different particles. Here, we called the free elementary particles collectively as "simple particles". Credit: This figure is partially reproduced from an earlier publication of the author: D. C. Chang, arXiv preprint physics/0404044v2 (2016)

17.6 Chapter Summary

- According to the quantum wave model, free particles are excitation waves of the quantum vacuum (here we call them "*waveticles*"). There are two different types of waveticles: The first type is driven by the *long-range force*; photons, scalar particles, electrons, and other leptons all belong to this type. The second type is driven by *short-range forces*. It includes all quarks and hadrons.
- The electromagnetic force is the long-range force. The short-range forces include the strong interaction and the weak interaction.
- The first type of quantum particles, such as photons, electrons, and other leptons, are stand-alone *elementary particles* with no internal sub-component. They can travel freely in space.
- Quarks are *elementary particles* of the second type. They are *bound particles* and cannot travel freely in space.
- Hadrons, including protons and neutrons, are not *elementary particles*. Instead, they are *composite particles* made up of quarks. Except for proton, these composite particles are unstable outside of the nucleus.
- Is the Standard Model the final theory? Many leading physicists suspect that it is the approximation of a more fundamental theory.
- One could think of the atomic nucleus as a large *composite particle*; it is just an aggregation of interacting quarks, instead of a cluster of individual protons and neutrons.
- The energy–momentum relation for a *composite particle* is similar to that of a *free elementary particle*.

- All free particles in our universe are excitation waves of the same vacuum medium. There is no basic difference between the radiation wave and the matter wave; they are just different excitation modes of the *quantum vacuum*!

References

1. Hey, T., & Walters, P. (2003). *The new quantum universe.* Cambridge University Press.
2. Beringer, J., & Particle Data Group et al. Review of particle physics. *Physical Review D, 86*, 010001.
3. National Institute of Standard and Technology. CODATA Internationally Recommended 2018 Values of the Fundamental Physical Constant. https://physics.nist.gov/cuu/Constants/index.html
4. Pacchioni, G. (2019). Into the vortex. *Nature Reviews Materials, 4*, 79–79.
5. University of Warwick, News and Events. Galaxy sized twist in time pulls violating particles back into line. https://warwick.ac.uk/newsandevents/pressreleases/galaxy_sized_twist/
6. Halban, H. V. J., & Preiswerk, P. (1936). Preuve expérimentale de la diffraction des neutrons. *CR Acadimic Science Paris, 203*, 73–75.
7. Estermann, I., & Stern, O. (1930). Beugung von molekularstrahlen. *Zeitschrift für Physik, 61*, 95–125.
8. Yukawa, H. (1935). On the interaction of elementary particles. I. *Proceeding Physical Mathematical Social Japan, 17*, 48.
9. Neddermeyer, S. H., & Anderson, C. D. (1937). Note on the nature of cosmic-ray particles. *Physical Review, 51*, 884–886.
10. Yukawa, H. (1949). Meson theory in its developments," presented at Nobel Lecture.
11. Lattes, C. M. G., Muirhead, H., Occhialini, G. P. S., & Powell, C. F. (1947). Processes involving charged mesons. *Nature, 159*, 694–697.
12. Ryder, L. H. (1996). *Quantum field theory.* Cambridge University Press.
13. Weinberg, S. (1997). What is an elementary particle? *Beam Line, 27*, 17.
14. Greene, B. (2004). *The fabric of the cosmos: Space, time, and the texture of reality*. A.A. Knopf.
15. Woit, P. (2006). *Not even wrong: The failure of string theory and the search for unity in physical law.* Basic Books.
16. Penrose, R. (2005). *The road to reality: A complete guide to the laws of the universe.* Knopf.
17. Smolin, L. (2006). *The trouble with physics: The rise of string theory, the fall of a science, and what comes next.* Houghton Mifflin.
18. Davisson, C. J., & Germer, L. H. (1927). The scattering of electrons by a single crystal of nickel. *Nature, 119*, 558–560.

Chapter 18
Exploration in Cosmology from the Quantum Wave View: Is There a Beginning or an End to Our Universe?

The quantum wave model discussed in this book is not a *theory of everything*! It is mainly used for explaining the physical phenomena in the quantum world. However, it can also give some useful hints to explain what we observed in the cosmos, for example, the origin of *dark matter* (see below).

Cosmology is undoubtedly a very important field of study. From the early days in human history, people have already wondered what our world is made of. Modern cosmology is a very young field; many powerful observational tools, such as high-resolution telescopes and satellites, only emerged in the last century. As a result, people have very limited time to build the fundamental theory. Many current models are based on conjectures from research in particle physics or general relativity.

Naturally, modern cosmological theories are dominated by the views of particle physicists. As we pointed out in earlier chapters of this book, the particle view has certain limitations. For future cosmological research, it may be helpful to explore new ideas based on alternative approaches, such as the idea of the quantum wave model. This chapter is an exploration of such an approach.

The Standard Model of cosmology today is far from the final theory. There are still many challenges that need to be addressed. Therefore, future cosmology researchers should have plenty of opportunities. The most mysterious things in cosmology at present are *dark matter* and *dark energy*. No one knows what they are. So, it might be interesting to see if the quantum wave model can shed light on this problem.

At the end of this chapter, we will also discuss briefly a very important question: Is there a beginning or an end to our universe?

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_18

18.1 What is Dark Matter Made of? The View of the Quantum Wave Model

In recent study of cosmology, it was discovered that our universe is not only composed of visible matters, but also *dark matters* and *dark energy* [1, 2]. The name "*dark matter*" implies that it is invisible. This is because *dark matter* is transparent to electromagnetic radiation and thus cannot be detected using current imaging technologies. Its existence is indirectly inferred mainly from its gravitational effects on the universe‘s large-scale structure, such as the rotational motions of stars around galaxies and gravitational lensing [3, 4].

The primary evidence for dark matter was from calculations showing that many galaxies would behave quite differently if they did not contain a large amount of unseen matter. The first proposal of "dark matter" is often attributed to Fritz Zwicky's work in the 1930s [5, 6]. Based on the study of the *Coma Cluster*, Zwicky obtained evidence of unseen mass ("dark matter"). Zwicky estimated its mass based on the motions of galaxies near its edge and compared that to an estimate based on its brightness and number of galaxies. He estimated the cluster had far more mass than was visually observable.

In the 1960s and 1970s, further evidence was obtained using galaxy rotation curves based on measurements of the velocity curve of spiral galaxies; Rubin and Ford showed most galaxies must contain about six times dark matter in additional to visible mass [7].

Later, radio astronomers also made use of new radio telescopes to map the 21 cm line of atomic hydrogen in nearby galaxies. The radial distribution of interstellar atomic hydrogen often extends to much larger galactic radii than those accessible by optical studies, extending the sampling of rotation curves [8]. Results of these measurements also suggested very large values of *mass-to-light* ratio in the outer parts of the galaxy (see Fig. 18.1).

How to explain such strange phenomena became an important topic in cosmology in the 1980s. Cosmologists called this invisible matter "*dark matter*". The Standard Model of cosmology at present indicates that the total mass/energy of the universe contains about 5% ordinary matter, 26% dark matter, and 69% dark energy [1, 2]. So, the amount of dark matter is about five times of that of visible matter. At this point, very little is known about the nature of dark matter. There are two major questions waiting to be answered: (1) What is dark matter composed of? Is it composed of particles? What could these particles be? (2) Why are there more dark matters than visible matters?

The early proposed dark matter candidates include neutrinos and exotic particles in high-energy particle physics [9]. Other suggested candidates include *axions* and supersymmetric particles such as *photinos* (the supersymmetric partner of photon) [10], and massive compact halo objects such as black holes [11].

At present, the leading hypothesis about dark matter is that it is composed of *weakly interacting massive particles* (WIMPs) which interact only through gravity and the weak nuclear force [12]. These WIMPs are supposed to be new particles in the

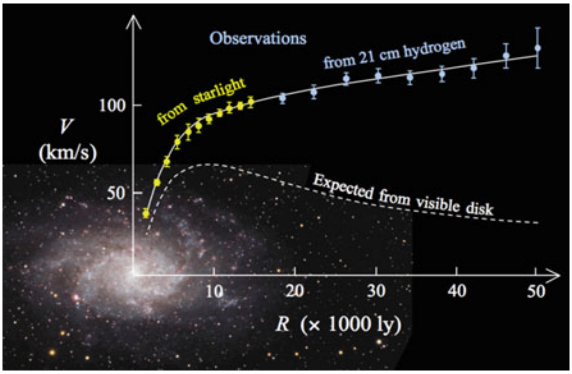


Fig. 18.1 **Rotation curve of spiral galaxy Messier 33**. The observed rotation curve of spiral galaxy Messier 33 is represented by the yellow and blue points, and the predicted rotation curve from distribution of the visible matter is shown as the dashed line. The discrepancy between the two curves can be accounted for by adding a dark matter halo surrounding the galaxy. (For details, see E. Corbelli and P. Salucci, arXiv preprint astro-ph/9909252 (1999).) Image Credit: Stefania.deluca; Wikimedia Commons, Public domain

100 GeV mass range. So far, none of the experiments designed to detect WIMPs has produced any evidence for their existence [11, 13–15]. Currently, there are several ongoing projects attempting to detect the WIMPs either directly or indirectly [16, 17].

We think the quantum wave model discussed in this book may provide a basis for searching new sources of dark matter. According to our model, **all particles in nature are excitation waves of the vacuum** (here we call them "*waveticles*"). Some of these excitation waves could behave as dark matter. These waveticles do not interact easily with other particles because they have certain specific properties, such as:

- **Having no electric charge**. The dark matter (DM) particles have no electric charge. Thus, they do not interact with other charged particles through electromagnetic interactions. That means these DM particles will not interact with the electrons or nucleus of any atom. Also, because they have no electric charge, the DM particles will not interact with electromagnetic radiation, including light. Thus, they will look dark.
- **Having no color charge, i.e., not being composite particles**. These DM particles are unlikely to be hadrons. Since hadrons are made of quarks, any hadron can interact strongly with the atomic nuclei when it passes through an ordinary object. For example, neutron is a hadron without electric charge. It does not behave like dark matter because it can interact easily with the atomic nuclei of visible matters.
- **With small cross section**. For the dark matter not to interact with visible matter, its interaction cross section with particles of ordinary matters must be very small. That means, the DM particle may behave like neutrino, which has very small interaction cross section.

- **Possessing very short wavelength**. Why the dark matter is dark? That is because we cannot see it. But what is *seeing*? It involves either absorption or scattering of the particle by the atoms. Take the photons for example. We see light because EM wave can interact with the orbital electrons of the atoms in our retina (or in a light sensor). For light to be absorbed by an orbital electron, it requires a matching of the particle energy ($E = \hbar\omega$) with the electron transition energy (the energy difference between the initial state and the final state of the electron, $\Delta E_{\text{electron}}$). In another word, absorption or emission of light requires $\hbar\omega \approx \Delta E_{\text{electron}}$. For most condensed matters, the energy level of an orbital electron (or electron bands) is in the order of several electron volts. If the energy of the incoming particle is much larger than the electron energy, the chance of absorption is extremely small. The particle will just pass by without interacting with the atom. This is why *X*-ray is more transparent to matter than visible light. If the DM particles have a very short wavelength, it will not interact easily with the orbital electrons of an atom or the band electrons of a solid. It thus would appear as "dark".

The properties listed above are not unusual. It is reasonable to expect that **many excitation waves in the vacuum can satisfy the above requirements**. Thus, the number of particles meeting these criteria could be very large. We did not notice them before because we cannot observe them. Also, since these DM particles have no charge, they cannot form atoms. And thus, they cannot aggregate together to form bulb objects.

Our model can also shed some light on the second question: *How can there be more dark matters than visible matters?* According to the quantum wave model, all particles are excitation waves of the vacuum medium. Each particle represents a different excitation mode. Since there can be many different excitation modes, the number of possible particle types in nature can be very large. Some of them may satisfy the required properties listed above. So, the number of candidates for dark matter could be quite large according to the wave view.

Furthermore, since particles are excitation waves of the vacuum, they normally do not interact with each other. Take the sound waves for example. Each sound can be transmitted independently without interference by other sounds. Imagine that when we talk with a friend in a noisy marketplace, there are many different sounds. Yet, we can still hear the voice of our friend. Hence, according to the wave model, it is the norm that particles without electric charge or color charge do not interact with each other. They would appear as "dark".

Therefore, it is possible to use the *quantum wave model* to explain the origin of *dark matter*. A more difficult challenge is "*dark energy*". The Standard Model of cosmology at present assumes that our Universe is filled with *dark energy*, a totally unknown object. Why do scientists need to hypothesize its existence? As explained in the following, it is a concept originated from the development of the *Big Bang* model, which plays a major role in the modern studies of cosmology.

18.2 Is Our Universe Expanding? The Big Bang Model

The Big Bang model is the prevailing cosmological theory that explains the origin and evolution of the universe. The theory is based on the observation made by Edwin Hubble in the 1920s that the universe is expanding. The Big Bang model suggests that the universe began as a single point of infinite density and temperature (called a "singularity"). After that, the Universe has been expanding ever since.

18.2.1 Hubble's Observation in 1929

In 1929, Edwin Hubble reported the first evidence for the expansion of our universe based on his study of 24 extra-galactic nebulae [18]. Using the largest telescope at the time (at Mount Wilson Observatory), Hubble noted that light from faraway galaxies appeared to be stretched to longer wavelengths, a phenomenon called redshift. He discovered that the more distant a nebula is from us, the faster it appears to move away (see Fig. 18.2). This implies that the universe is expanding uniformly in all directions.

Hubble plotted the velocity-distance relation of these nebulae and calculated the slope (now called the "*Hubble's constant K*"). He found that the value of the constant K from the 24 individual nebulae was about 500 km/sec per million parsecs. (The Hubble's constant is often denoted as H_0 in modern cosmology). Hubble's discovery suggested that our Universe is not static, as scientists had always believed; instead, our Universe is expanding.

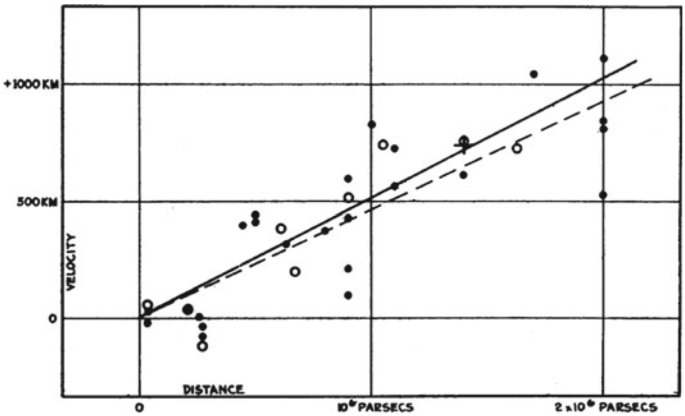


Fig. 18.2 Plot of the velocity-distance relation in the 1929 Hubble's paper. The black dots and the solid line represent 24 nebulae data, while the circles and dashed lines represent the combined nebulae data of 9 groups. Photo Credit: E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae", Proc. Natl. Acad. Sci. U.S.A. 15, 168–173 (1929)

18.2.2 The Story of Einstein's Cosmological Constant

When Albert Einstein developed his theory of general relativity, he thought that the universe was neither expanding nor contracting, in another word, the universe was static. In order to account for this, Einstein modified his theory of general relativity in 1917 by introducing a *cosmological constant* term into his equations. This *cosmological constant* represented a repulsive force that counteracted the gravitational attraction between matters, preventing the universe from collapsing under its own weight.

When Edwin Hubble published his observation that distant galaxies were moving away from us at a speed proportional to their distance, scientists started to realize that the universe was expanding, and the idea of a static universe was abandoned. Physicists at that time were very excited about Hubble's observation. Einstein even went to Mount Wilson to visit Hubble in 1931 and happily played with the telescope which led to Hubble's discovery of the expanding universe.

Hubble's discovery made a big impression on Einstein. If our universe is not static, the *cosmological constant* Einstein added earlier to the general relativity equation to make the universe static is no longer needed. Einstein thus renounced the *cosmological constant,* saying it is "*a term which was not required by the theory as such, nor did it seem natural from a theoretical point of view*" [19]. It was said that Einstein later wrote: "*Since I introduced this term, I had always a bad conscience....* ***I am unable to believe that such an ugly thing is actually realized in nature***" [20].

George Gamow recalled that Einstein regarded the introduction of the cosmological constant to be the "***biggest blunder***" in his life [21]. It was a big surprise that Einstein's "biggest blunder" would revive after more than half a century later.

18.2.3 The Rise of the Big Bang Model

With the knowledge of the expansion of the universe, scientists proposed that, by extrapolating back in time, there should be a beginning of the universe. The early universe should be much denser and hotter. The **Big Bang model** suggests that the universe began as a singularity, which expanded rapidly, and the temperature dropped from almost an infinite value to a few billion degrees. As the universe cooled, subatomic particles began to form, and eventually, atoms formed. The formation of atoms allowed light to travel freely, and the universe became transparent [22].

The early proponents of the Big Bang model include Georges Lemaître and George Gamow. In 1931, Georges Lemaître published a short paper in Nature entitled "The Beginning of the World from the Point of View of Quantum Theory". He wrote: "*If we go back in the course of time we must find fewer and fewer quanta, until we find all the energy of the universe packed in a few or even in a unique quantum*" [23].

In the 1940s, Gamow, Alpher, and Herman proposed the existence of the relics of radiations from early hot universe. They used the Big Bang idea to explain how different elements formed [24]. They also calculated that the present relics of the early radiation should be around 5 K [25]. This was called the *cosmic background radiation* [26].

At that time, the competing model of the Big Bang is called "Steady State Model". It believed that matters in the universe were created under rather constant rate (in a steady state), and there was no singularity (or a "Big Bang") in the early universe. This model was championed by astronomer Fred Hoyle [27].

The debates between these two models lasted for several years. The Big Bang model was finally becoming the preferred choice of the cosmology community. This was mainly due to the observation of *cosmic microwave background* (CMB) radiation in 1964. The CMB was interpreted as relics of radiation from the early hot universe. It was regarded as strong evidence for the Big Bang model.

18.2.4 Cosmic Microwave Background: Evidence of the Big Bang Model

In the past century, the most important experimental observation in cosmology is probably the measurement of the **cosmic microwave background** radiation (CMB), which was thought to be the leftover radiation from the Big Bang; it was emitted when the universe was only 380,000 years old.

In 1964, Arno Penzias and Robert Wilson at Bell Labs built a large horn antenna to detect radio waves from the sky. However, they found some background noise at about 3 K which cannot be eliminated [28] (see Fig. 18.3). At the same time, a group of cosmologists led by Robert Dicke in Princeton University were actively trying to detect the cosmic background radiation but without success.

The Princeton team visited Penzias and Wilson to see their measurement and the large horn equipment. They were very excited and thought that the background noise Penzias and Wilson found was the cosmic background radiation leftover from the Big Bang of the Universe. But Penzias and Wilson were more cautious about the interpretation of their measurement. So, the two teams decided to publish their papers side by side in the *Astrophysical Journal* in the form of *letters to the editor*. Penzias and Wilson's paper entitled "*A Measurement of Excess Antenna Temperature at 4080 Megacycles per Second*" simply reported the measurement of the background noise radiation [29]. In the Dicke's group paper entitled "*Cosmic Black-body Radiation*", they argued that the Big Bang theory implied the existence of cosmic microwave radiation following the black-body spectrum [22]. They suggested that the cosmic microwave background radiation detected by Penzias and Wilson was an important experimental support to the Big Bang theory [22].

Since the Big Bang theory had dominating influence in the cosmology community, Penzias and Wilson were awarded the Nobel Prize in Physics in 1978 for detecting

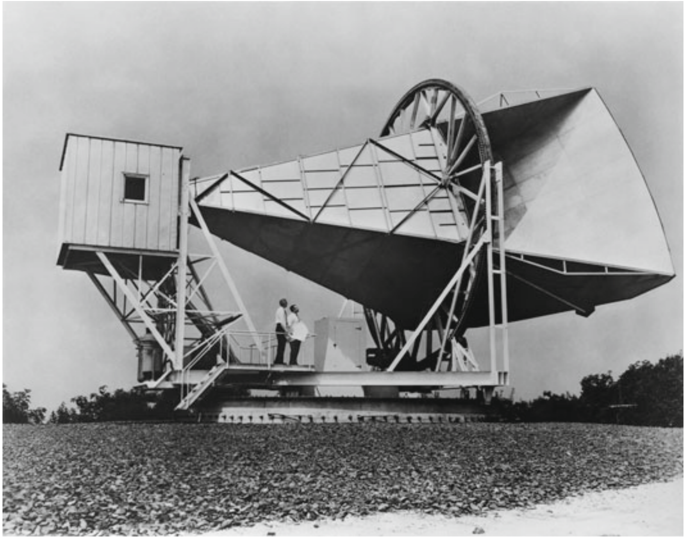


Fig. 18.3 Antenna used to detect the Microwave Background Radiation. The 15 m horn antenna at Bell Telephone Laboratories in Holmdel, New Jersey, was built in 1959 for pioneering work in communication satellites for the NASA ECHO I. The horn was later modified to work with the Telstar Communication Satellite frequencies as a receiver for broadcast signals from the satellite. In 1964, Robert Wilson and Arno Penzias discovered the cosmic microwave background radiation with it, for which they were awarded the 1978 Nobel Prize in physics. Credit: NASA

the microwave background noise from their antenna. [28]. Since then, the CMB has been extensively studied using satellites. The *Cosmic Background Explorer* (COBE) satellite was launched in 1989 to study the CMB in detail. COBE made precise measurements of the CMB, which confirmed its uniformity and showed that the CMB does follow strictly the rule of black-body radiation (see Fig. 18.4). Their data also provided the first evidence for tiny variations in its temperature. These variations are thought to be the seeds of the large-scale structure of the universe, such as galaxies and clusters of galaxies.

Later, other satellites such as the *Wilkinson Microwave Anisotropy Probe* (WMAP) and the *Planck* satellite have made more precise measurements of the CMB (see Fig. 18.5). These measurements generally confirmed the predictions of the Big Bang model and allowed scientists to obtain detailed information about the age, composition, and structure of the universe. [1, 2]

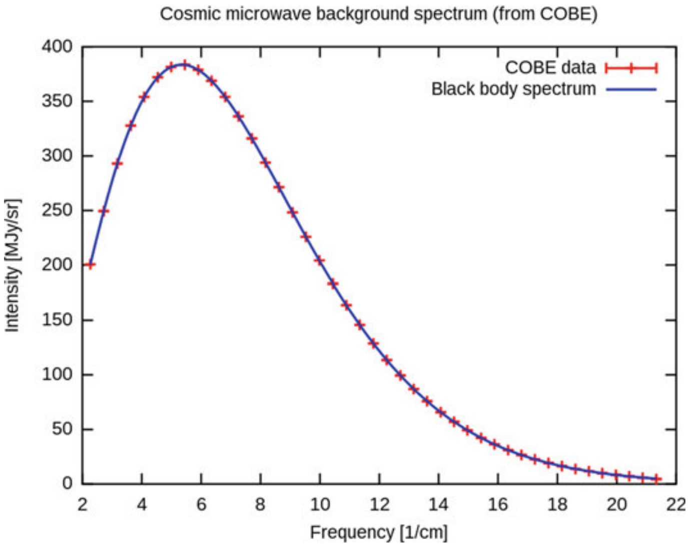


Fig. 18.4 CMB fitting the black-body radiation spectrum. Graph of cosmic microwave background spectrum measured by the FIRAS instrument on the COBE satellite. The CMB fits perfectly with the black-body spectrum. Image Credit: Quantum Doughnut, PD

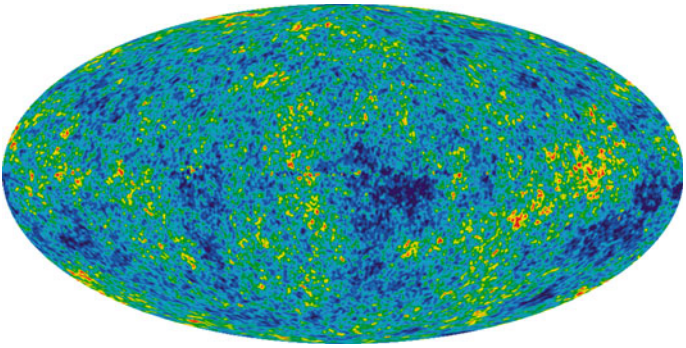


Fig. 18.5 CMB map as observed by WMAP. Heat map of temperature fluctuations in the cosmic microwave background based on Wilkinson Microwave Anisotropy Probe measurement. The image reveals 13.77 billion-year-old temperature fluctuations (shown as color differences) thought to be the seeds that grew to become galaxies. The signal from our Galaxy was subtracted using the multi-frequency data. This image shows a temperature range of ± 200 microKelvin. Credit: NASA/ WMAP Science Team

18.3 Is Our Universe Expanding in an Accelerating Rate? The Idea of Dark Energy

18.3.1 Observations on Type Ia Supernova Suggested that the Universe is Expanding in an Accelerating Rate

The most recent supporting evidence for the Big Bang theory is the discovery of accelerating expansion of the universe based on studies of Type Ia supernovae [30–32]. This type of supernova occurs in a binary star system where one star is a white dwarf. The white dwarf could accumulate matter from its companion star. When it eventually reaches a critical mass, it undergoes a catastrophic explosion, releasing an enormous amount of energy (see Fig. 18.6).

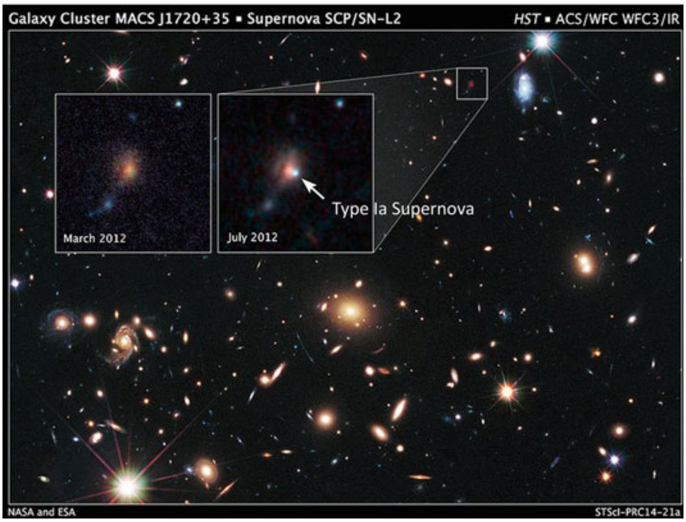


Fig. 18.6 Type Ia supernovae. The heart of a vast cluster of galaxies called MACS J1720 + 35 is shown in this image, taken in visible and near-infrared light by NASA's Hubble Space Telescope. The small white box at upper right marks the location of an exploding star called a supernova, located behind the cluster. An enlarged view of the supernova, catalogued as SCP/SN-L2, is shown in the inset image at top right, taken during July 2012. An arrow marks the location of the supernova, which resides near the bright core of the host galaxy. The supernova is seen as it appeared 7.7 billion years ago. The inset image at top left, taken in March 2012, shows the same region before the supernova blast. This image underscores the transient nature of exploding stars. The supernova is a member of a special class of exploding star called Type Ia, prized by astronomers because it provides a consistent level of peak brightness that is useful for making distance estimates. Credit: NASA, ESA, S. Perlmutter (UC Berkeley, LBNL), A. Koekemoer (STScI), M. Postman (STScI), A. Riess (STScI/JHU), J. Nordin (LBNL, UC Berkeley), D. Rubin (Florida State University), and C. McCully (Rutgers University)

Since the Type Ia Supernova explosions are extremely bright and visible billions of light-years away, astronomers realized that they could be used as a "standard candle" for measuring distance. In the 1980s, the Supernova Cosmology Project, led by Saul Perlmutter at Lawrence Berkeley National Laboratory, began searching for Type Ia supernovas. In the 1990s, the High-Z Supernova Search team, led by Brian Schmidt of the Australian National University and Adam Riess of the Space Telescope Science Institute, also joined the search. They expected to observe the deceleration of supernovas caused by the gravitational attraction of mass, according to Einstein's gravitational theory.

However, both teams soon found that the Type Ia supernovas they observed were fainter than expected from Hubble's Law, suggesting that the universe was not decelerating but accelerating. Both teams quickly announced their findings. In January 1998, the Supernova Cosmology Project announced at a press conference that they had analyzed 40 supernovas and found that the universe's expansion would continue forever, and that the data could be explained by a cosmological constant. In February, the High-Z team presented their supernova data at a conference, also showing that the expansion of the universe is accelerating and interpreted that the acceleration was driven by the energy of the vacuum which they thought was Einstein's "*cosmological constant*".

Later that year, cosmologist Michael Turner coined the term "**dark energy**" to describe this mysterious repulsive force, in analogy with the invisible dark matter that makes up most of the matter in the universe. The finding of an accelerating expanding universe was surprising, yet it was soon accepted. In December 1998, Science magazine selected the discovery of accelerating expansion of the universe as the Breakthrough of the Year.

18.4 The Standard Model of Cosmology Today and Some of Its Current Debates

Today, the Standard Model of cosmology is known as the **Lambda-Cold Dark Matter** (ΛCDM) model [11]. This model is based on Einstein's theory of general relativity (GR). The term "**Lambda** (Λ)" stands for Einstein's "*cosmological constant*", a term which Einstein once considered the "*biggest blunder*" in his life.

The Lambda-CDM model is a combination of the Big Bang model and the theory of *cosmic inflation*. It suggests that the universe began at a tiny point called "*singularity*"; the universe was in a state of infinite density and extremely high temperature. At this point, the laws of physics as we know them do not hold.

During the first fraction of a second after the Big Bang, the universe underwent a period of rapid expansion known as *inflation*. During this period, the size of the universe increased by a factor of 10^{26} in 10^{-35} s [33]. After inflation, the universe continued to expand and cool, and energies were converted into particles. These early particles included quarks, which combined to form protons and neutrons. These

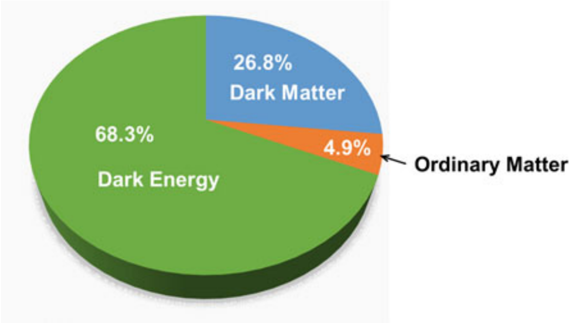


Fig. 18.7 Composition of matter and energy in our Universe according to the Lambda-CDM model. Estimate of the composition of our universe according to the Lambda-CDM model based on Planck's high-precision cosmic microwave background map: Normal matter that makes up stars and galaxies contributes just 4.9% of the Universe's mass/energy inventory; dark matter, which is detected indirectly by its gravitational influence on nearby matter, occupies 26.8%, while dark energy, a mysterious force thought to be responsible for accelerating the expansion of the Universe, accounts for 68.3%

particles then combined with electrons to form atoms, specifically hydrogen and helium. The formation of atoms marks the end of the "*cosmic dark ages*" and the beginning of the era of *recombination*, which occurred about 380,000 years after the Big Bang.

From then on, the universe continued to expand and cool, and the first stars and galaxies formed. The first stars were likely massive and short-lived, exploding as supernovae and releasing heavy elements into the universe. These heavy elements provided the building blocks for the formation of subsequent generations of stars and planets.

Using a six-parameter fitting for the CMB data, the Lambda-CDM model suggests that the universe is currently composed of approximately 68% dark energy, 27% dark matter, and 5% normal matter [2] (see Fig. 18.7). Dark energy is believed to be responsible for the accelerating expansion of the universe, while dark matter is thought to play a crucial role in the formation and evolution of galaxies.

18.4.1 Some Current Debates on the Standard Model

Despite the remarkable success of the Big Bang theory and the Lambda-CDM model in explaining astronomical observations, there are still many unanswered questions. For example, the nature of *dark matter* and *dark energy* remains a mystery. These phenomena require an extension of our current understanding of particle physics or suggest a breakdown of general relativity on cosmological scales [34].

Proposing the breakdown of physical laws at the early universe is also not a very attractive hypothesis. Thus, many leading investigators in cosmology had expressed concerns on whether the current theory is examining nature from the right direction. For example, Michael Turner, who first coined the term "dark energy", suggested that the current questions about *dark matter*, *dark energy*, and *inflation* may not be the right ones to ask. It is possible that *dark energy* is a mirage, and the real explanation for cosmic acceleration could lie in a replacement for general relativity [35]. *Inflation*, in its current form, is another major challenge. Its loose ends, such as the prediction of *multiverse* (multiple universes), and its lack of a clear driving mechanism, may also make it out of touch with reality.

One early proponent of the *inflation* model, Paul Steinhardt, criticized the Standard Model of cosmology in a 2005 paper [36]. He argued that many key questions remain unanswered, particularly, the *anthropic principle* advocated by some cosmologists is unreliable. Paul Steinhardt published another paper entitled "*The Inflation Debate*" in 2011, in which he argued that the original *inflation* theory was incorrect, and the current theory is difficult to test experimentally. Steinhardt notes that *bad inflation* is more likely than *good inflation*, and obtaining a flat universe without *inflation* is more likely than with *inflation*. He concludes that the inflation theory is problematic, and there is a need for alternative theories to explain the origins of the universe [37].

Another problem with the Standard Model is the discrepancy between the observed vacuum energy density and the energy required for driving the accelerated expansion of the Universe. As we discussed earlier, the studies of Type Ia supernovae led to the discovery of the accelerating expansion of the universe. In the Standard Model, this expansion was thought to be driven by dark energy. *Vacuum energy* is the leading candidate for dark energy, and it is mathematically equivalent to the *cosmological constant* "*lambda*" that Einstein introduced in 1917 [30, 34, 38]. According to some of the existing quantum theories, vacuum fluctuations, or virtual particles created and annihilated in the vacuum, provide energy to the vacuum. However, theoretical calculations suggest that the energy density associated with these fluctuations is at least 10^{55} times greater than the energy density measured by cosmologists [34, 39]. If the vacuum energy density were truly that high, all matter in the universe would instantly fly apart, and galaxies would never have formed. This discrepancy has been referred to as "the worst embarrassment in all of theoretical physics" [34].

18.5 New Ideas for Cosmological Research from the Quantum Wave View

From the above discussion, one can see that there are many challenging questions for the current version of the Standard Model of cosmology. For example,

- The Big Bang model is based on extrapolating what we observed in recent time to the very early time of the Universe. How reliable is this extrapolation?
- At present, there are too many big mysteries in the Standard Model of cosmology. For example, we know nothing about the nature of ***dark energy*** or the mechanism of ***inflation***.

- The Lambda-CDM model is based on the theory of general relativity (GR). However, there is a conceptual conflict between them. The GR assumes the vacuum is an empty space. But the Lambda-CDM model assumes that the vacuum is the ground state of a quantum system, which contains a tremendous amount of energy. How to reconcile between the two?

Thus, there is a strong need to explore new ideas to resolve these mysteries. The quantum wave model discussed in this book could offer some help. This model examines nature in a different philosophical point of view in comparison with the traditional particle view. It can offer a new approach to understand our Universe. In the following, we will explore some new ideas for the future studies of cosmology.

18.5.1 Is There a Need to Re-Examine the Idea of Big Bang?

We think there is a need to re-examine if our Universe is truly expanding or not. So far, the proposal of an expanding Universe was based mainly on observations of nearby galaxies moving away from us [18, 20]. From the quantum wave point of view, these observations do not directly imply that the Universe is expanding.

First, according to the quantum wave model, matter is made of excitation waves of the vacuum. Thus, the movements of material objects, including hydrogen atoms, stars, nebula, or galaxies, are all movements of excitation waves. The movement of excitation waves does not mean a movement of the wave medium. Take the sound wave as an example, a sound wave propagating in the sea does not mean the sea water is moving in the same direction or speed as the sound wave. So, the observation of the material objects (galaxies) moving away from us does not imply that the vacuum is expanding. In the quantum wave model, the vacuum is a medium that fills the entire Universe. So, it is the vacuum that is directly connected to the Universe, not the excitation waves of the vacuum. Then, **the observed movement of the material objects could not imply whether the Universe itself is expanding or not**.

Second, observations of distant galaxies moving away were made in nearby astronomical regions. This means they are very local phenomena. We know the largest structure of our Universe is the **cosmic web**, in which billions and billions of galaxies are distributed in a non-uniform way (see Fig. 18.8). The structure of the cosmic web is expected to be dynamic; there are local movements and fluctuation within it. Thus, there will be local stretching and contraction. **It is not surprising for an observer within a local area to observe nearby objects moving away or moving inward**. Since the observations of Hubble and the measurements of Type Ia supernovae both represent local events observed within the cosmic web, it is questionable whether these local observations can be extrapolated to the entire Universe.

Finally, **there are technical issues** on the reliability of the reported observations of galaxies moving away from us. For example, in the original study reported by Hubble, he did not have reliable "standard candles" to measure the distance of extra-galactic nebulas. So, he just used some brightest stars in the nebula and assumed

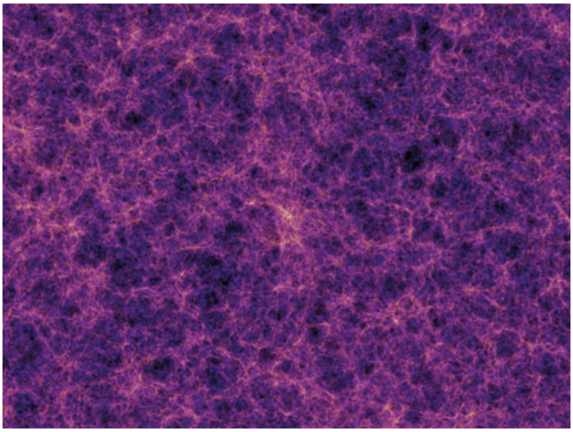


Fig. 18.8 Cosmic web. In shaping the Universe, gravity builds a vast cobweb-like structure of filaments tying galaxies and clusters of galaxies together along invisible bridges hundreds of millions of light-years long. This is known as the cosmic web. This plot is a simulation using a super computer. Credit: Volker Springel (Max Planck Institute for Astrophysics) et al. CC BY-SA 4.0

their apparent luminosities had a maximum. He used them as a measure of distance. In fact, according to his original data, Hubble estimated that the age of the universe is only a few billion years, which is younger than our solar system. It does not make sense. Hubble also recognized that. He said that apparently one had to be careful when using the redshift to measure the age of the universe [18].

Then, what about the observations suggesting that our Universe is expanding in an accelerating rate? This finding was based on observations of the Type Ia supernova. In the circle of cosmology, Type Ia supernova was commonly used as a "standard candle" to calculate its distance from Earth. But these "candles" were found recently not to be so standard [40, 41]. For example, a recent study found evidence that supernovae can arise by two different processes, which would give different brightness. This raised suspicion that standard candles are not so standard after all [41]. Another study examined the local environment of 28 Type Ia supernovae and found those with more elements heavier than helium (so-called "metals") were dimmer than their less metallic counterparts [40]. Therefore, Type Ia supernova may not be a very reliable "standard candle". Apparently, more extensive studies are needed to substantiate the earlier findings of accelerated expansion.

18.5.2 Is Extrapolation in Time Justified?

Extrapolation in time is a dangerous game, it is not scientific. There is no assurance that one observed in the world for a short period will be applied for all times. Take the history of mankind for example. One cannot judge the trend of the stock market by extrapolation. One also cannot use extrapolation to judge the historical trend; it is well known that empires could rise and fall. In the case of cosmology, we do not yet know the mechanism behind the moving away of nearby galaxies at the present time. Is it a perpetual effect or a temporary trend? **Big Bang** is certainly a bold idea; it will need very strong evidence to support it.

So, unless we have overwhelming evidence that our Universe was created by a Big Bang, there is no need to hypothesize the existence of **dark energy**. Then, what about the recent claim that dark energy accounts for most of the matters/energy in our world? This was based on the analysis of CMB using the Lambda-CDM model. It is a model-dependent estimate. In science, one needs to carefully differentiate an estimate from a fact.

18.5.3 Should Time Be Continuous?

Time should be continuous; there is no evidence that it has a beginning or an end. Hence, the idea of time having a beginning and an end is an artificial assumption, whoever suggested that will have the burden of proof. The current version of the Standard Model of cosmology has not yet produced overwhelming evidence for proving that time has a beginning or an end.

According to the quantum wave model discussed in this book, all matters are quantized excitation waves of the vacuum. How long has the vacuum been there? Naturally, the vacuum should exist with our world from whatever time our world exists. Since the material objects in our world are excitation waves of the vacuum medium, they should have plenty of time to interact with each other. This would suggest that the material objects in our world should have been at some sort of thermal equilibrium a long time ago.

This could explain why the distribution of matter in our present world is so homogeneous (in the large scale) and why the temperature of CMB appears to be so uniform. This problem is commonly referred to as the "*horizon problem*". One of the major reasons for the proposal of the *inflation* theory is to solve the *horizon problem.* With the realization that time is continuous and the age of the vacuum in our Universe can be far longer than the model predicted by the Big Bang theory, the *horizon problem* is not a problem. Then, there is no need to propose the *inflation* theory.

18.5.4 What Could Be the Origin of CMB?

From the above discussion, one can see that the Big Bang model is far from convincingly proven. There are still multiple challenges for it. One may raise the question of CMB. Had the studies of CMB already proven beyond doubt that the Big Bang model is correct?

The presence of CMB was important evidence for supporting the Big Bang Model. However, we do not think this evidence is overwhelming. The major problem is that no one knows exactly what the origin of CMB is. Can one prove that the CMB is indeed the relic of radiation from the Big Bang? Could one exclude the other alternative sources for the cosmic background radiation?

For example, can one exclude the contribution of local thermal radiation from the observed CMB? We know in any thermal system, there should be background radiation within it. The space environment we live in, such as the solar system, can be regarded as a "thermal cavity" (see Fig. 18.9). Furthermore, our Sun is located at a spiral arm of the Milky Way; the aggregation of nearby stars in this spiral arm could also form a bigger "thermal cavity". Thus, one can expect that there will be background radiation in the thermal cavity we live in. As we had discussed in Chaps. 2 and 3, the electromagnetic radiation within a thermal cavity can be described by the formula of black-body radiation as worked out by Planck. In fact, based on the COBE satellite measurements, the CMB observed outside of the Earth fits exactly with the prediction of black-body radiation. This may suggest that the cosmic background radiation observed by mankind could be originated from the thermal radiation in the local space environment near Earth.

In order to test whether CMB originated from local thermal radiation or from the remnants of the Big Bang, we need to measure the CMB outside of the solar system, and preferably, outside of the Milky Way. If the CMB is indeed representing the remnants of the Big Bang, its distribution should appear the same regardless of

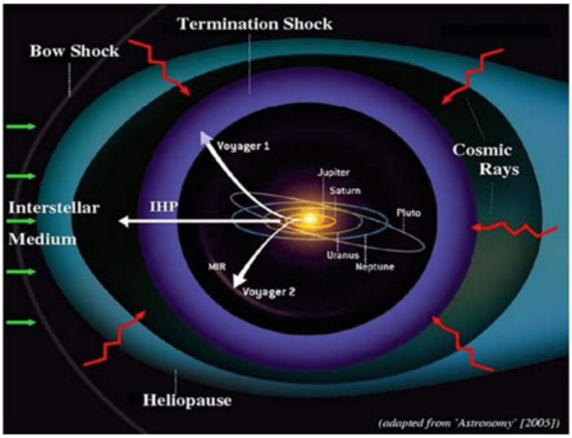


Fig. 18.9 Cavity of the solar system Heliosphere. The Sun's stellar wind bubble, known as the heliosphere, is a region of space dominated by the Sun. Its boundaries are located at the termination shock, approximately 80–100 AU from the Sun upwind of the interstellar medium and approximately 200 AU from the Sun downwind of the interstellar medium. Here, the solar wind collides with the interstellar medium, forming a huge elliptical structure called the heliosheath. The outer boundary of the heliosphere, the heliopause, is the point where the solar wind finally ends and interstellar space begins. Voyager 1 and Voyager 2 passed the termination shock and entered the heliosheath at 94 and 84 AU from the Sun, respectively. *Image Credit* Wimmer-Schweingruber, R.F., McNutt, R., Schwadron, N.A. et al. Interstellar heliospheric probe/heliospheric boundary explorer mission—a mission to the outermost boundaries of the solar system. Exp Astron **24**, 9–46 (2009)

whether the measurement is made inside the solar system, outside the solar system, or outside the Milky Way.

18.6 Is There a Beginning or an End to Our Universe? The Model of an Endless Recycling Universe

The study of the origins and the ultimate fate of the universe is a topic of great interest to scientists. At present, there is no clear evidence on whether there is a beginning or an end of our universe. The mainstream model of cosmology at present assumes that there is a beginning, i.e., the Big Bang, but it has no clue on whether there is an end. Depending on the different estimates of dark energy, many exotic end models of the universe had been suggested, including “Big Freeze” and “Big Crunch” [42, 43]. In the “Big Freeze” scenario, the universe will continue to expand forever, eventually becoming so spread out that all matter is too far apart to interact with each other, resulting in a cold, dark, and lifeless universe. In the “Big Crunch” scenario, the universe will stop expanding and then start to contract, eventually collapsing in on itself in a massive explosion.

Another possibility is the “Big Bounce” model, which suggests that the universe goes through an infinite number of cycles, with each cycle beginning with a Big Bang and ending with a Big Crunch. After the Big Crunch, a new universe is born from the remnants of the old one, leading to an endless cycle of creation and destruction [44].

We, however, think that a more probable scenario is that the universe has no beginning and no end; it simply goes on continuously without dramatic events such as the Big Bang or Big crunch. Our thinking is that, at the macroscopic level, the universe is roughly at a steady state. But of course, in different local regions of the universe, various astronomical objects could undergo a variety of recycling processes. For example,

- ***Recycling of Energy and Matter***

According to the quantum wave model, matter and radiation are just different forms of energy; they are all excitation waves of the vacuum. Hence, matter and energy will convert between each other almost constantly. It is well known that intense radiation can be generated when matter is destroyed in various forms of explosions, and massive particles can be created in the vacuum when the vacuum is irradiated with energetic radiation. Also, exotic particles with high energy can be created during particle–particle collisions. Such energetic particles will eventually decay into stable particles like electrons and protons, which can combine to form hydrogen atoms.

- ***Recycling of Atoms***

Atoms are the building blocks of matter, and they are constantly recycled in the universe. When a star explodes, it releases not only radiation energy but also matter in the form of atoms. These atoms can be used to create new stars, planets, and everything on top of them. For example, the iron atoms in our blood were created by nuclear fusion in the cores of massive stars, which later exploded as supernovae and spread their contents into space.

Furthermore, heavier elements can be created within the neutron stars or during supernova explosions; they can also be broken down to become light elements through nuclear fission. So, different kinds of atoms can be recycled in the universe.

- ***Recycling of Stars and Planets***

Stars and planets are also recycled in the universe. When a star dies, it releases its energy and matter, which can be used to create new stars and planets. In fact, our Sun and its planets were formed from the remnants of previous generations of stars that had exhausted their fuel and ended their lives. Also, the debris from a planetary collision can be used to form new planets or moons.

When a star is at the end of its life, it can release its energy and materials back to the universe. During the active stage of a star's life, it fuses hydrogen atoms into helium in its core. Once the hydrogen supply is depleted, the star enters the next stage, becoming a red giant. During this phase, the star expands and cools, and helium fusion begins in the core. Eventually, the nuclear fusion stops, and the remaining core becomes a white dwarf.

For more massive stars, the end of the red giant phase leads to a supernova explosion. During this event, the star releases an enormous amount of energy and heavy elements, including iron and nickel. The core of the star collapses, either forming a neutron star or a black hole, depending on the mass of the star.

In addition to supernova explosions, stars also release energy through stellar winds. These winds are composed of charged particles that are ejected from the star's surface, carrying away mass and energy. The winds can last for millions of years. The release of energy and materials back into the universe through supernova explosions and stellar winds is essential for the creation of new stars.

Then, what about the white dwarfs and neutron stars, can they release their energy and materials back to the universe? White dwarfs can release energy in the form of residual heat, which is slowly radiated away over billions of years [45]. They can also be merged with neutron stars or black holes. Furthermore, a white dwarf can merge with another star (or another white dwarf) to ignite a Type Ia supernova [46]. Since such supernova leaves no compact remnant, the whole mass of the former white dwarf could dissipate into the space.

Neutron stars can recycle their material through merging. When two neutron stars merge, they produce a *kilonova*, which releases an enormous amount of energy, including gamma rays, X-rays, and visible light [47]. The neutron star merging can also cause the release of heavy elements, such as gold and platinum [48]. These elements are dispersed into the surrounding interstellar medium, contributing to the chemical evolution of the universe (see Fig. 18.10).

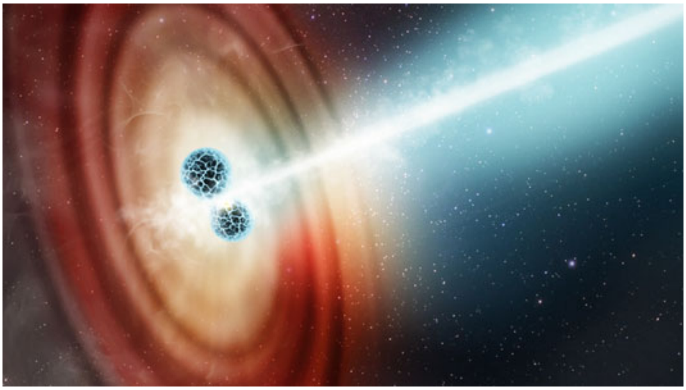


Fig. 18.10 Energy released during neutron star merging. An artist's concept showing two neutron stars collided. In the aftermath, a blowtorch jet of radiation was ejected at nearly the speed of light. Astronomers used Hubble telescope to measure the motion of a blob of material the jet slammed into. Credits: NASA's Goddard Space Flight Center; Lead Producer: Paul Morris

- ***Recycling of Galaxies and Black Holes***

Finally, galaxies and black holes are also part of the recycling process in the universe. When galaxies collide, they can merge to create new galaxies. The Milky Way is currently merging with another nearby galaxy, the Andromeda Galaxy, which will eventually result in the formation of a new, larger galaxy. The Hubble Space Telescope has imaged numerous galaxy mergers in various stages of evolution (see Fig. 18.11).

Black holes, which are formed when a massive star collapses, can also merge to create larger black holes or be ejected from their galaxies. Stars within galaxy can be swallowed by the supermassive black hole. Due to this extreme gravitational pull, matter and radiation in the vicinity of the black hole are accelerated to high speeds

Fig. 18.11 Merging of a pair of galaxies. Two galaxies, dubbed IC 694 and NGC 3690, made a close pass some 700 million years ago. As a result of this interaction, the system underwent a fierce burst of star formation. Credit: NASA, ESA, the Hubble Heritage Team (STScI/AURA)-ESA/Hubble Collaboration and A. Evans (University of Virginia, Charlottesville/NRAO/Stony Brook University)

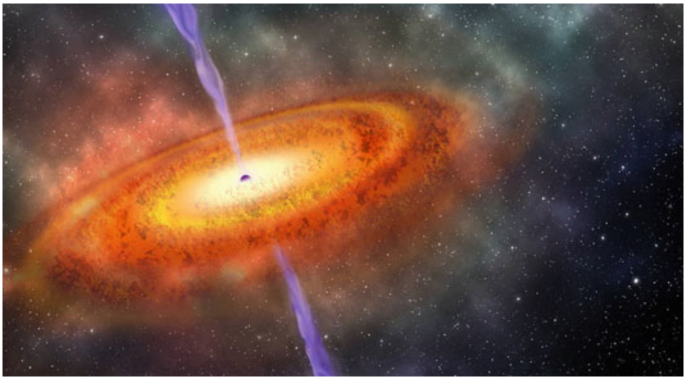


Fig. 18.12 Black hole releasing energy in the form of jets. An artist's concept showing a distant supermassive black hole. Credits: Robin Dienel/Carnegie Institution for Science

and emit intense radiation, including X-rays and gamma rays. This radiation comes from the accretion disk [49]. Black holes can also release energy in the form of jets, which are narrow streams of particles that are ejected from the black hole at close to the speed of light [49–51] (see Fig. 18.12). These jets can extend for millions of light-years. Jets from black holes play an important role in the evolution of galaxies, as they can influence the formation of stars and regulate the growth of supermassive black holes in the centers of galaxies.

One of the most intriguing aspects of black holes is their ability to merge with one another. The merger of black holes releases an enormous amount of energy, which could be detected as gravitational waves. At present, the mechanism through which black hole releases its trapped energy/materials is not yet fully understood. Scientists anticipate that more novel mechanisms will be discovered in the future. So, it is very likely that the energy trapped inside the black hole will eventually be released back into the surrounding environment.

In conclusion, the recycling of matter and energy in different forms is a constant process within the universe. While the origin and ultimate fate of the universe remain a mystery, we propose that the "endless recycling model" could be the most probable course for the universe's evolution. In other words, our universe is most likely going for ever, there is no foreseeable end to our universe.

18.7 Does the Vacuum Have an End? Is There a Boundary in Our Universe?

Finally, before ending our discussion on cosmology, we would like to ask a philosophical question: **Does the vacuum have an end?** Since the vacuum is supposed to fill the entire space of the Universe, would our Universe have an end?

According to the quantum wave model discussed in this book, matter is composed of quantized excitation waves of the vacuum medium. Hence, we can think of matter as a kind of energy. The entire contents of the material world, including atoms, stars, and galaxies, are all composed of excitation waves of the vacuum. So, what we observe in our Universe, the so-called "material world", is only the manifestation of the Universe; the body of the Universe is actually the vacuum itself.

Now, does the vacuum have an end? If the vacuum has an end, what is the space outside of it? What happens to the world at the end of the vacuum?

In scientific investigation, we usually go with a common presumption, that is, **the simplest answer is most likely to be the correct answer**. In that spirit, our simplest assumption is that the vacuum must fill the entire space of the Universe, and there is no end for the vacuum.

Then, we need to investigate whether this assumption is realistic. Throughout this book, we regard the vacuum as a medium occupying a three-dimensional space. **Is it possible that this three-dimensional vacuum has no end?**

From the study of geometry, we know some of the objects can have no end. For example, for a one-dimensional string, it can have no end if the string is connected end to end as a circle (see Fig. 18.13a). Also, for a two-dimensional plane, it can have no end if this plane is a part of the surface area of a sphere (see Fig. 18.13b).

Then, perhaps one could generalize the above concepts to a mathematical conjecture: **For any N-dimensional object, it can be continuous with no end, if and only if this object is the boundary of a $N + 1$ dimensional object.**

Previously, we have already demonstrated that this conjecture can hold when $N = 1$ or 2. For example, the string is a one-dimensional object, and a flat disk is a two-dimensional object. The boundary of the two-dimensional disk is a circle. When the string is wrapped around this circle, it becomes endless. (Note: A line segment of the circle can become a straight line when the radius of the circle becomes infinitively large.)

Similarly, a plane is a two-dimensional object. It can become endless if the plane becomes a part of the surface of a three-dimensional sphere. In this case, the spherical **surface** is the **boundary** of the sphere itself. (Note: A segment of the spherical surface plane can become flat when the radius of the sphere becomes infinitively large.)

So, our conjecture is true for one-dimensional and two-dimensional objects. It is reasonable to expect that it may also be true for a three-dimensional object.

Based on this thinking, it is highly possible that the vacuum medium filling the three-dimensional space of our world can have no end. The only requirement is that *the Universe must have an extra hidden spatial dimension*, which is not observable in our material world. Testing this speculation could be an interesting topic of future study.

So, we can hypothesize that the vacuum is endless. This hypothesis in fact makes good sense because it can easily explain what we have observed in nature. Because the vacuum medium is endless, its excitation waves thus can travel forever in any direction. This is equivalent to say, any stable particle (such as a photon, electron, and proton) can travel in space forever until it is captured or collides with other objects, just as we observe in a mechanical world. In such a world, **there is no center and**

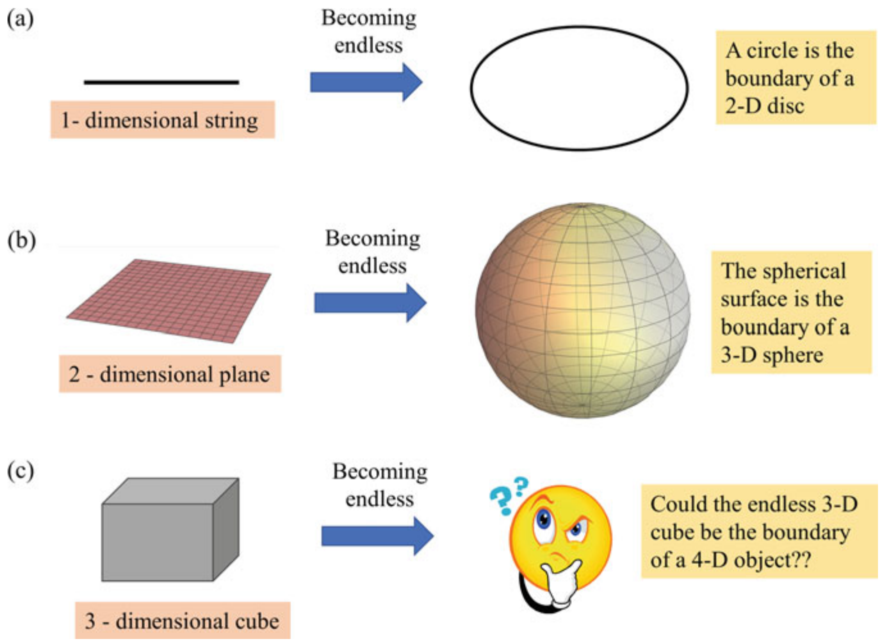


Fig. 18.13 Examples showing that a lower dimensional object can be endless if it is the boundary of a higher dimensional object. **a** For a one-dimensional string, it can have no end if the string is connected end to end as a circle. **b** For a two-dimensional plane, it can have no end if this plane is a part of the surface area of a sphere. **c** For a three-dimensional cube, can it be without an end?

there is no edge. **Everywhere in our Universe is the same**! Also, because space has no end, the world would look similar in any direction and is truly isotropic. There is no *horizon problem*!

18.8 Great Opportunities in Experimental Observation: Moon-Based Astronomy

As noted above, the current understanding of the evolution of the universe is not the final theory; there are many important questions that remain to be answered. This will provide many opportunities for future generations of cosmologists. In particular, we now live in a world of technological revolution. Now is a great time for new experimental discoveries. Our understanding of nature has improved enormously over the past century. This is really due to the development of powerful telescopes. For example, the space-based Hubble Telescope has greatly increased our knowledge of cosmology. A new generation of space-based telescopes, the James Webb Telescope, is expected to perform even better. New advances could be achieved by building telescope arrays on the back side of the Moon. These may include optical (including infrared and ultraviolet) telescopes, radio telescopes, and X-ray and gamma-ray detectors. In addition, we can also set up cosmic ray detectors on the Moon to study high-energy particles.

These moon-based telescopes have many advantages. In addition to eliminating interference from Earth and providing clearer images, they can also have longer baselines, which can be as large as the diameter of the Moon's orbit. (In fact, the baseline can be further extended to the diameter of Earth's orbit around the Sun.) They can use triangulation to determine the distance of distant objects.

These telescope arrays can allow us to peer deeper into the universe, and thus enabling us to observe events happening at an earlier time.

Therefore, these future tools may help us answer/clarify some of the major questions raised in this chapter, such as the *Big Bang* problem, the meaning of Type Ia supernova measurements, the *inflation* hypothesis, and the recycling of energy from black holes.

18.9 Chapter Summary

- The most mysterious things in cosmology at present are *dark matter* and *dark energy*. No one knows what they are. The quantum wave model may provide some useful hints. According to our model, all particles in nature are excitation waves of the vacuum, and some of these excitation waves can appear as *dark matter* because they have certain specific properties.
- The *Big Bang* model is a popular cosmological theory that explains the origin and evolution of the universe. The theory is based on Edwin Hubble's observations in the 1920s that the universe was expanding. Observations of cosmic microwave background (CMB) radiation in 1964 further supported this idea. CMBs are interpreted as relics of early thermal cosmic radiation after the Big Bang.
- The Standard Model of cosmology today is known as the Lambda-Cold Dark Matter (ΛCDM) model, which is primarily based on general relativity.
- Today's Standard Model suggests that during the initial stages of the Big Bang, the universe went through a period of extremely rapid expansion (called "cosmic *inflation*"). After *inflation*, the universe continued to expand and cool, converting energy into particles.
- There are still many unanswered questions in the Lambda-CDM model. For example, very little is known about the nature of *dark energy* or the mechanism of *inflation*. Furthermore, the Lambda-CDM model and general relativity have conflicting views on the physical nature of the vacuum.
- Should *time* be continuous? The idea of time having a beginning and an end is a bold assumption, and current cosmological research has not yet produced overwhelming evidence that time had a beginning. With the realization that time

is continuous, and the age of the Universe could be far longer than what was predicted by the Big Bang theory, the proposal of *inflation* is perhaps unnecessary.

- Many exotic models for the end of the universe have been proposed, including the "*Big Freeze*" and the "*Big Crunch*". We think the universe probably had no beginning and no end; it just went on continuously, without dramatic events like the *Big Bang* or the *Big Crunch.* On a macroscopic level, the universe is almost in a stable state. However, the recycling of matter and energy in different forms is an ongoing process in the universe.
- Is there an end to the vacuum? Does our universe have boundaries? We think the vacuum filling the three-dimensional space of our world can have no end. The only requirement is that the universe must have an extra hidden dimension of space that is unobservable in our physical world. Testing this speculation may be an interesting topic for future research.
- We see great opportunities in experimental observation. In the near future, it is technically feasible to develop Moon-based astronomy. These future studies will help us to answer some of the key questions raised in this chapter.

References

1. Bennett, C. L., et al. (2013). Nine-year Wilkinson microwave anisotropy probe (WMAP) observations: final maps and results. *Astrophysical Journal Supplement Series, 208*, 20B.
2. Planck Collaboration, Ade, P. A. R., et al. (2016). Planck 2015 results—XIII. Cosmological parameters. *Astronomy Astrophysics*, *594*, A13.
3. Sofue, Y., & Rubin, V. (2016). Rotation curves of spiral galaxies. *Annual Review of Astronomy and Astrophysics, 39*, 137–174.
4. Refregier, A. (2016). Weak gravitational lensing by large-scale structure. *Annual Review of Astronomy and Astrophysics, 41*, 645–668.
5. De Swart, J. G., Bertone, G., & Van Dongen, J. (2017). How dark matter came to matter. *Nature Astronomical, 1*, 0059.
6. Spergel, D. N. (2014). Cosmology today. *Daedalus, 143*, 125–133.
7. This Month in Physics History. (2023). June 1980: Vera Rubin Publishes paper hinting at dark matter. *APS News, 32*.
8. Corbelli, E., & Salucci, P. (1999). The extended rotation curve and the dark matter halo of M33. arXiv preprint astro-ph/9909252.
9. Krauss, L. M. (1986). Dark matter in the universe. *Scientific American, 255*, 58–71.
10. Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. (1984). Formation of galaxies and large-scale structure with cold dark matter. *Nature, 311*, 517–525.
11. Spergel, D. N. (2015). The dark side of cosmology: Dark matter and dark energy. *Science, 347*, 1100–1102.
12. Bertone, G. (2010). The moment of truth for WIMP dark matter. *Nature, 468*, 389–393.
13. LUX Collaboration, Akerib, D. S., et al. (2014). First results from the LUX dark matter experiment at the Sanford underground research facility. *Physical Review Letter, 112*, 091303.
14. SuperCDMS Collaboration, Agnese, R., et al. (2014). Search for low-mass weakly interacting massive particles with SuperCDMS. *Physical Review Letter, 112*, 241302.
15. XENON100 Collaboration, Aprile, E., et al. (2013). Limits on spin-dependent WIMP-nucleon cross sections from 225 live days of XENON100 data. *Physical Review Letter, 111*, 021301.

16. The PanadaX Collaboration. PandaX dark matter experiment homepage. http://pandax.physics.sjtu.edu.cn/
17. Reich, E. S. (2013). Dark-matter hunt gets deep. *Nature, 494*, 291–292.
18. Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National academy of Sciences of the United States of America, 15*, 168–173.
19. Einstein, A. (1931). *Relativity: The Special and General Theory/Appendix IV*.
20. Kirshner, R. P. (2004). Hubble's diagram and cosmic expansion. *Proceedings of the National academy of Sciences of the United States of America, 101*, 8–13.
21. Gamow, G. (1970). *My world line: An informal autobiography.* Viking Press.
22. Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. (1965). Cosmic black-body radiation. *Astrophysical Journal, 142*, 414–419.
23. Lemaître, G. (1931). The beginning of the world from the point of view of quantum theory. *Nature, 127*, 706–706.
24. Gamow, G. (1946). Expanding universe and the origin of elements. *Physical Review, 70*, 572–573.
25. Alpher, R. A., & Herman, R. C. (1949). Remarks on the evolution of the expanding universe. *Physical Review, 75*, 1089–1095.
26. Gamow, G. (1948). The evolution of the universe. *Nature, 162*, 680–682.
27. Hoyle, F. (1948). A new model for the expanding universe. *Monthly Notices of the Royal Astronomical Society, 108*, 372–382.
28. Wilson, R. W. (1979). The cosmic microwave background radiation. *Science, 205*, 866–874.
29. Penzias, A. A., & Wilson, R. W. (1965). A measurement of excess antenna temperature at 4080 Mc/s. *ApJ, 142*, 419.
30. Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., & Tonry, J. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal, 116*, 1009–1038.
31. Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., Project, T. S. C. (1999). Measurements of Ω and Λ from 42 high-redshift supernovae. *ApJ, 517*, 565–586.
32. Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., Spyromilio, J., Riess, A. G., Filippenko, A. V., Hamuy, M., Smith, R. C., Hogan, C., Stubbs, C., Diercks, A., Reiss, D., Gilliland, R., … Ciardullo, R. (1998). The high-Z supernova search: Measuring cosmic deceleration and global curvature of the universe using type ia supernovae. *ApJ, 507*, 46–63.
33. Guth, A. H., & Kaiser, D. I. (2005). Inflationary cosmology: Exploring the universe from the smallest to the largest scales. *Science, 307*, 884–890.
34. Riess, A. G., & Turner, M. S. (2004). From slowdown to speedup. *Scientific American, 290*, 62–67.
35. Turner, M. S. (2022). The road to precision cosmology. *Annual Review of Nuclear and Particle Science, 72*, 1–35.
36. Steinhardt, P. J., & Turok, N. (2005). The cyclic model simplified. *New Astronomy Reviews, 49*, 43–57.
37. Steinhardt, P. J. (2011). The inflation debate. *Scientific American, 304*, 36–43.
38. Carroll, S. M., Press, W. H., & Turner, E. L. (1992). The cosmological constant. *Annual Review of Astronomy and Astrophysics, 30*, 499–542.
39. This Month in Physics History. (2009). January 1998: The accelerating expansion of the universe. *APS News, 18.*

40. Moreno-Raya, M. E., Mollá, M., López-Sánchez, Á. R., Galbany, L., Vílchez, J. M., Rosell, A. C., & Domínguez, I. (2016). On the dependence of type Ia SNe luminosities on the metallicity of their host galaxies. *ApJ, 818*, L19.
41. Hall, S. (2017). Supernova's messy birth casts doubt on reliability of astronomical yardstick. *Nature, 548*, 379–380.
42. Skibba, R. (2020). Crunch, rip, freeze or decay—how will the Universe end? *Nature, 584*, 187–187.
43. Khoury, J., Ovrut, B. A., Seiberg, N., Steinhardt, P. J., & Turok, N. (2002). From big crunch to big bang. *Physical Review D, 65*, 086007.
44. Steinhardt, P. J., & Turok, N. (2002). Cosmic evolution in a cyclic universe. *Physical Review D, 65*, 126003.
45. NASA/Goddard Space Flight Center, "White Dwarfs". https://imagine.gsfc.nasa.gov/science/objects/dwarfs1.html
46. Mazzali, P. A., Röpke, F. K., Benetti, S., & Hillebrandt, W. (2007). A common explosion mechanism for type ia supernovae. *Science, 315*, 825.
47. Rastinejad, J. C., Gompertz, B. P., Levan, A. J., Fong, W., Nicholl, M., Lamb, G. P., Malesani, D. B., Nugent, A. E., Oates, S. R., Tanvir, N. R., De Ugarte Postigo, A., Kilpatrick, C. D., Moore, C. J., Metzger, B. D., Ravasio, M. E., Rossi, A., Schroeder, G., Jencson, J., Sand, D. J., Smith, N., Fernández, J. F. A., Berger, E., Blanchard, P. K., Chornock, R., Cobb, B. E., De Pasquale, M., Fynbo, J. P. U., Izzo, L., Kann, D. A., Laskar, T., Marini, E., Paterson, K., Escorial, A. R., Sears, H. M., & Thöne, C. C. (2022). A kilonova following a long-duration gamma-ray burst at 350 Mpc. *Nature, 612*, 223–227.
48. Chen, H.-Y., Vitale, S., & Foucart, F. (2021). The relative contribution to heavy metals production from binary neutron star mergers and neutron star-black hole mergers. *ApJL, 920*, L3.
49. NASA/Goddard Space Flight Center. Black hole accretion disc energies. https://svs.gsfc.nasa.gov/10545
50. NASA. Famous black hole has jet pushing cosmic speed limit. https://www.nasa.gov/mission_pages/chandra/images/famous-black-hole-has-jet-pushing-cosmic-speed-limit.html
51. Snios, B., Nulsen, P. E. J., Kraft, R. P., Cheung, C. C., Meyer, E. T., Forman, W. R., Jones, C., & Murray, S. S. (2019). Detection of superluminal motion in the X-ray Jet of M87. *ApJ, 879*, 8.

Part VII
Conclusion and Summary

Chapter 19
Conclusion: Matter is Composed of Waves

This book presents a revolutionary idea: *matter is made of waves*! In traditional physics, matter is thought to be made up of massive particles, each of which is like a "*point mass*". In this book, we propose that sub-atomic particles are quantized excitation waves of the vacuum; each particle is just a *wave packet*. Furthermore, the vacuum is not an empty space; instead, it behaves like a dielectric medium. We show that the wave equation in quantum mechanics can be directly derived from the wave excitation mechanism of the vacuum. From this derivation, one can easily see the physical meaning of the quantum wave function, that is, the wave function is a measure of the physical movement of the vacuum medium during wave excitation.

19.1 The Objective of Writing This Book

This book is a serious attempt to resolve a long-standing mystery in quantum physics. That is, how to explain the phenomenon of *wave-particle duality*? This problem has long been a conceptual challenge in quantum mechanics. So far, no consensus has been found to explain this phenomenon. The quantum wave model proposed in this book can easily explain this quantum mystery. (For details, see below.)

In fact, the quantum wave model can explain a lot more than this. As we showed in Chaps. 6–11, this model can provide the physical basis for explaining many discoveries in modern physics.

The work discussed in this book is based on a very simple idea: ***The unification of radiation wave and matter***. More specifically, it proposes that *both light and matter are composed of quantized excitation waves of the vacuum medium.* It is well known that light is a quantized radiation wave. This work proposes that matter is also made of waves. Thus, one can unify both the physical basis of light and matter. This proposal is based on experimental observations that massive particles have wave properties like photons. We just take these findings one step further. Namely, we propose that

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4_19

a massive **particle not only can behave like a wave**, but that **it is actually made of waves!** It is more appropriate to call the quantum object a "*waveticle*" instead of a "*particle*".

A key assumption of the quantum wave model is that the vacuum is a wave medium. In physics, there are two major schools of thought about the property of the vacuum. In classical mechanics, the vacuum is regarded as an empty space. But in the study of electromagnetism and optics, the vacuum has long been regarded as a medium. We know **all physical waves require a wave medium to propagate**. For example, sound waves are known to propagate in physical media such as air, water, or elastic solids. The same must be true for the propagation of light or electromagnetic waves. That is why in the nineteenth century, many physicists believed in the hypothesis that the vacuum is filled with a medium called "aether" [1]. This aether hypothesis, however, was disfavored in the twentieth century.

It can be said that, this quantum wave model is inspired by the aether hypothesis in the nineteenth century and the discovery of the quantum property of wave-particle duality in the twentieth century. Its contribution is really on building a simple and coherent theory that can accommodate both the physical foundation of light (Maxwell's theory) and the physical foundation of matter wave (quantum mechanics).

19.2 What Problems Can the Quantum Wave Model Solve?

The reason for proposing this quantum wave model is to provide simple explanations to problems that are difficult to explain with current quantum theory. Such problems include:

(1) ***The Physical Basis of Wave-Particle Duality***. As demonstrated in diffraction experiments and double-slit experiments, we know the electron behaves just like a light wave. So, from an experimental point of view, there is no doubt that the physical behaviors of electrons and photons are very similar. However, the current quantum theory has great difficulty to explain it. This is because traditional physicists tend to view our world using a classical perspective. In classical physics, matter is composed of particles which behave like point mass. So, an electron should not behave like a photon!

The quantum wave model discussed in this book can easily solve this problem. If electrons and photons are both quantized excitation waves of the vacuum medium, it is natural that an electron should behave like a photon. Then, there is no mystery why a single electron can pass two slits to generate an interference pattern (in the double-slit experiment) or be diffracted from a crystal surface following the Bragg's law (see Chaps. 4 and 10).

(2) ***Why Can Particles Be Created or Annihilated?*** In the particle view of quantum physics, it is very difficult to explain why point mass particles can be created or annihilated in a vacuum. These observations can be easily explained in the quantum wave model. Since particles are excitation waves rather than rigid

objects, when the vacuum medium is excited by an energetic stimulation, new waves can be generated, and this new wave will appear in the form of new particles. Moreover, when such waves/particles collide with other waves/particles, they can also create new types of waves. Therefore, waves can be created, annihilated, or converted into other types of waves. From a macroscopic perspective, these excitation waves appear as "particles". This explains why we can observe the creation and annihilation of various particles in collision experiments.

(3) ***The Physical Basis for the Planck's Relation, de Broglie Relation, and Heisenberg's Uncertainty Principle***. As shown in Chap. 3, the quantum wave model can directly explain the physical origin of the well-known quantum relations, including the Planck's relation and the de Broglie relation, based on the Maxwell theory. Also, it clearly showed that the Heisenberg's Uncertainty Principle is really based on the fact that the quantum particle is a wave packet, in which the frequency spread and the width of the wave packet are related by the Fourier condition $\Delta\omega \cdot \Delta t \approx 1$.

(4) ***The Physical Basis for the Derivation of the Quantum Wave Equations***. In the current particle-based quantum theory, there is no clear physical basis for the derivation of the quantum wave equation. For example, it was pointed out by Feynman: "*Where did we get that (Schrödinger equation) from? Nowhere. It's not possible to derive it from anything you know. It came out of the mind of Schrödinger*" [2]. Similarly, the Dirac equation was derived based on a conjecture that the quantum Poisson bracket has certain commutation properties [3]. The quantum wave model can now fill in this gap; it shows that the quantum wave equations of electrons can be directly derived based on the wave excitation mechanism of the vacuum, which can be traced to the Maxwell theory (see Fig. 19.1). This suggests that there is a natural transition between classical physics and quantum physics (see Chaps. 6–9).

(5) ***The Physical Meaning of the Quantum Wave Function***. The wave function in the Schrödinger equation is often called "matter wave". However, it was not clear what the physical meaning of matter waves is. There has been a lot

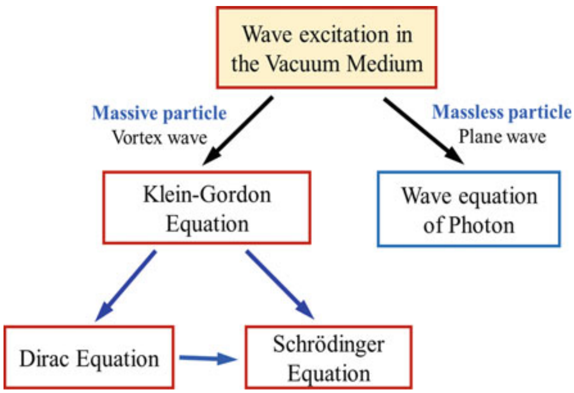


Fig. 19.1 Derivation of quantum wave equations in the quantum wave model. Since a quantum particle is a quantized excitation wave of the vacuum, the wave equation should be derivable from the wave excitation mechanism of the vacuum medium. This is shown to be the case for all known quantum wave equations for electrons and photon

of debates about this. In fact, there was a famous debate between Bohr and Einstein on this particular issue [4]. According to the particle view, the electron is like a point mass; it cannot be a wave. Therefore, one can only interpret the wave function as a statistical parameter which gives the probability of finding an electron. This is known as the "Copenhagen interpretation". Many physicists disagreed, including Schrödinger and Einstein. This problem can now be solved using the quantum wave model. It is shown that the quantum wave function is related to the newly defined *electric vector potential* $\boldsymbol{Z}$, which is a measure of the displacement of the vacuum medium [5]. Based on this idea, the physical meaning of "matter wave" becomes very clear (see Chaps. 6–9).

19.3 Implications of the Quantum Wave Model

This quantum wave model not only can resolve some of the outstanding mysteries in quantum physics, but it also has strong implications on our understanding of some fundamental issues in physics. At present, the study of modern physics is built on a combination of quantum mechanics and relativity. These two theories, however, have conflicting assumptions on the physical property of the vacuum. Particularly, the special theory of relativity assumes that the vacuum is an empty space; such an assumption is not consistent with the understanding in quantum mechanics. To resolve this conflict, one needs to find a new basis in quantum physics to explain the effects currently attributed to relativity.

The quantum wave model can indeed meet this need. In this book, it is shown that the so-called "relativistic effects" are actually consequence of the fact that quantum particles are quantized excitation waves of the vacuum medium. The following is a summary of this new understanding:

(1) ***What is the Meaning of Mass? Why Are Mass and Energy Convertible?***

In the literature today, the basis of mass-energy conversion is often attributed to relativity. Recent literature reviews, however, indicated that this is not true [6, 7]. This book shows that the mass-energy conversion actually arises from the wave nature of quantum particles (see Fig. 19.2). We know energy and momentum are related and both of them have their counter-physical meanings in the wave view. What about *mass*? Historically, mass is regarded as a particle property only. In this work, we show that mass could also be a wave property. In fact, it is found that the physical meanings of mass, energy, and momentum are very similar; they are all related to the *curvature of bending the vacuum medium* during wave excitation. With this new understanding, one can easily show that mass and energy can indeed be converted between each other (for details, see Chap. 11).

(2) ***Why Can No Particle Travel Faster Than c? Why Do All Particles Have the Same Traveling Speed Limit?***

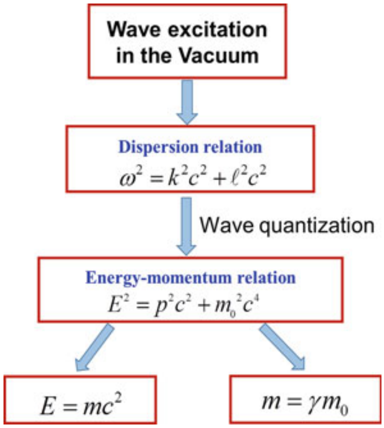


Fig. 19.2 Origin of the mass-energy conversion relation in the wave view. According to the quantum wave model, the basis for mass-energy conversion is due to the fact that the quantum particle is an excitation wave of the vacuum medium. Based on the same argument, one can also predict that the mass of a particle is speed-dependent

It is well known that no particle can travel faster than the speed of light. But why is it? Previously, there was no explanation for this experimental finding. It was just regarded as a postulate in STR. With the quantum wave model, it is very easy to explain it. We know the speed of wave propagation is determined by the physical properties of the transmitting medium. Since all particles are quantized excitation waves of the vacuum medium, they should have the same speed limit (see Chap. 12).

(3) ***Why Is the Mass of a Particle Speed-Dependent?***

As shown in the above, all quantum particles cannot travel faster than the speed of light. Suppose a particle is accelerated in an accelerator, the energy and momentum of the particle will increase continuously. Since the particle's speed v has an upper limit c, the particle cannot increase its speed further when its speed approaches the speed of light. Its momentum p can only be increased through the increase of mass. Hence, the energy absorbed by the particle is not used to increase its speed v, but is mainly used to increase its mass m.

In other words, the faster a particle travels, the harder it is to accelerate. This implies that the larger the velocity of the particle, the greater its inertial mass. Thus, one can predict that the inertial mass of this particle will increase with its speed. And all this is because the particle is a quantized excitation wave; its traveling speed cannot exceed the phase velocity of the vacuum medium, which is the speed of light (see Chap. 12).

In fact, the quantum wave model has many more interesting implications. For example, it can provide helpful hints for solving the following questions:

- What is the physical nature of anti-mass? (see Chap. 16).
- Why are particles limited to Fermions and Bosons? (see Chap. 7).
- What is Gravity? Is it mass-attracting-mass, or energy-attracting-energy? (see Chap. 12).

For readers interested in these questions, we strongly urge them to read the relevant chapters.

In summary, we believe that the quantum wave model presented in this book is a very attractive model. Its assumptions are simple, its arguments are based on well-documented experimental observations, and its results are highly interesting. This model has important implications for our future understanding of quantum physics. Furthermore, it has a clear advantage over current particle theory. The quantum wave model requires only a single vacuum medium, whose properties are known from Maxwell's theory. In contrast, current quantum field theory requires each particle to be the excitation of its own *field*. Therefore, the number of *fields* in the physical world is unlimited. Between "*particle*" and "*field*", it is not clear which one is more fundamental.

19.4 Unification of Concepts in the Understanding of Nature

The quest to comprehend nature has been a long-standing pursuit of humankind. Over time, our understanding of the natural world has greatly expanded; we have witnessed the unification of concepts that were once thought to be unrelated. This has led to major milestones in our understanding of nature. For example,

(1) ***Newton's theory***, which unified the concept of matter and gravity. This is a groundbreaking theory, which revealed that matter is the source of gravity, and that gravity determines the weight of matter.
(2) ***Maxwell's theory***, with contributions from Faraday, Ampere, and others. This theory not only unified electricity and magnetism, but it also further unified light with electromagnetic radiation. It is a truly conceptual breakthrough.
(3) ***The quantum wave model***, as discussed in this book. This model unifies the concepts of matter and radiation. It suggests that both are excitation waves of the vacuum. Previously, it is well known that radiation is a kind of wave. Here, we propose that matter is also made up of physical waves. So, the physical nature of matter and radiation are similar.

Because of these breakthroughs, concepts that were once believed to be entirely separate are now understood as interconnected and unified, providing us with a more comprehensive understanding of nature.

19.5 To Go Beyond *Fashion*, *Faith*, and *Fantasy*

Recently, Roger Penrose[1] published a widely noticed book entitled "*Fashion, faith, and fantasy in the new physics of the universe.*", in which he pointed out that theoretical physics today is a field that is often plagued by concepts that are based on *fashion*, *faith*, and *fantasy* [8]. In his book, Penrose highlights the importance of going beyond these concepts. He strongly believed that "*Theoretical physics should be based on empirical evidence and objective analysis, rather than on fashion, faith, and fantasy*". He emphasizes the importance of using rigorous mathematical methods to develop theories that are testable through experiments. To him, theories that cannot be tested through experiments are not truly scientific and should not be considered as such.

Penrose was not alone in seeing the danger of the current trend of theoretical physics development. For example, a theoretical physicist, Sabine Hossenfelder, recently published an article entitled "*Science needs reason to be trusted*", in which she explicitly pointed outed that, "*the particle physics community has always been subject to fads and fashions*". "*And it's not only theoretical high-energy physics. You also see this in cosmology, where models for inflation abound. Theorists introduce one or several new fields and potentials that drive the Universe's dynamics before decaying into normal matter. Current observational data can't distinguish the different models. And even if new data comes in, there will still be infinitely many models left to write papers about*" [9].

In Penrose's view, "*The ultimate test of a theory is its ability to make predictions that can be tested through experiments*". Experimental tests are crucial for verifying the validity of a theory and for identifying any flaws or limitations in it. Also, experimental tests can sometimes lead to surprising results that can challenge established beliefs and theories.

Another way to go beyond *fashion*, *faith*, and *fantasy* is to encourage open-mindedness and critical thinking. Scientists should be willing to challenge established beliefs and theories, and to consider alternative explanations for phenomena. By maintaining a spirit of inquiry, we can avoid falling into the traps of dogma and ideology, and instead remain open to new ideas and possibilities.

Hence, there is a strong need today to go beyond *fashion*, *faith*, and *fantasy* in fundamental physics. As we continue to push the boundaries of our knowledge in quantum physics and cosmology, it is crucial that we should remain grounded in empirical evidence and objective analysis.

We fully agree with this spirit. In this book, therefore, we try our best not to judge a theory or model by the influence of *fashion*, *faith*, and *fantasy*. Instead, we would like to evaluate the validity of a theory primarily by its logical consistency and ability to withstand experimental tests. In fact, as shown in Chaps. 14, 15, and 18, we even suggest new experiments to be conducted in the future for testing some of the most basic hypotheses.

---

[1] Penrose is a well-known mathematical physicist. He is the recipient of the 2020 Nobel Prize in physics.

Besides experimental evidence, we also believe that the truth of Nature should be conceptually simple. **Simplicity is a kind of beauty!**

This is why we think the *quantum wave model* is a step in the right direction.

References

1. Whittaker, E. (1951). *A history of the theories of aether and electricity.* Thomas Nelson and Sons Ltd.
2. Feynman, R. P., Leighton, R. B., & Sands, M. (1964). Chapter 16 the dependence of amplitudes on position. In *The Feynman lectures on physics* (Vol. 3). Addison-Wesley.
3. Dirac, P. A. M. (1981). *The principles of quantum mechanics* (4th ed). Clarendon Press.
4. Skibba, R. (2018). Einstein, Bohr and the war over quantum theory. *Nature, 555*, 582–584.
5. Chang, D. C. (2021). Review on the physical basis of wave–particle duality: Conceptual connection between quantum mechanics and the Maxwell theory. *Modern Physics Letters B, 35*, 2130004.
6. Fadner, W. L. (1988). Did Einstein really discover '"E=mc2"'? *American Journal of Physics, 56*, 114–122.
7. Hecht, E. (2011). How Einstein confirmed E_0=mc^2. *American Journal of Physics, 79*, 591–600.
8. Penrose, R. (2016). *Fashion, faith, and fantasy in the new physics of the universe.* Princeton University Press.
9. Hossenfelder, S. (2017). Science needs reason to be trusted. *Nature Phys, 13*, 316–317.

Epilogue: The Route of My Quantum Exploration

Although quantum physics has been developed for more than a century, its physical basis is still difficult to explain even today. As a physics student, I was always curious about the foundation of quantum mechanics. When I was in high school, I read some popular science books during the summer holiday, such as "*Introduction to Relativity*" and "*Introduction to Quantum Mechanics*". As a high school student, I thought quantum mechanics was amazing. But at that time, I did not have sufficient physical and mathematical background to ask the right questions; I could only view these quantum phenomena as interesting stories; it was purely for my curiosity.

I had a chance to learn more about quantum mechanics when I entered the National Taiwan University as an undergraduate student in physics. We started to study *Modern physics* in the third year, which introduced the major experimental foundations for building the atomic theory. In the fourth year, I took a course in *Quantum Mechanics*, which introduced the formulism of the quantum theory. But frankly, it was incomprehensible to me. We were taught by an old German professor who spoke English with a heavy German accent. Every class he just kept on writing equations on the blackboard, and we copied what he wrote. Apart from getting a lot of notes, I had very little knowledge about the principle of quantum mechanics.

Three Teachers Who Had Significant Influence on Me

I really started to understand the major concepts in quantum mechanics when I entered Rice University for postgraduate studies. I took a course in *Quantum Mechanics* in the first year. One of the instructors in this course was Prof. William V. Houston, who wrote a textbook entitled "*Principle of Quantum Mechanics*".[1] This is a well-written textbook which explains the basic topics of quantum mechanics very clearly. It uses concise mathematics to demonstrate that mathematical models and experimental

[1] W.V. Houston: *Principle of Quantum Mechanics*. Dover, 1951.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4

evidence can work together to advance our knowledge. After taking this course, I can say that I have some preliminary understanding of quantum mechanics.

During my PhD study, I studied *Advanced Quantum Mechanics* for another year. I started to think about some fundamental questions in quantum physics, but I could not get the answer at that time. These problems have troubled me since then. So, when I have time, I think about these problems, and I keep trying to find a way to solve them. During this process, three physicists whom I came to contact had provided me with significant inspiration.

Prof. William V. Houston. In the development of quantum mechanics, a German physicist, A. Sommerfeld, had made great contributions. Several master figures in quantum mechanics were his students, including W. Heisenberg and W. Pauli. My teacher at Rice University, Professor W.V. Houston, worked with Sommerfeld in his early years. Houston conducted research at the University of Munich where Sommerfeld was his mentor. Later, Houston moved to the University of Leipzig to work with Heisenberg. There, he also collaborated with Felix Bloch, Heisenberg's first graduate student.

Houston's research at the University of Munich was about the motion of electrons in metal. At first, he and Sommerfeld treated the electrons as particles according to the theory at the time, but the results were unsatisfactory. Later, they changed their thinking and regarded the electrons as waves; the movement of the electron in the metal was equivalent to the movement of a wave in a crystal lattice. Then, they got some good results. Sommerfeld was very happy at the time and said with excitement, "*We have finally found out the physical basis of resistance*".[2]Houston's work provided a new way of thinking; it showed that electrons are behaving as physical waves in the metal.

Professor Houston later became a major contributor in the American physics community. After returning to the United States from Germany, he worked for a long time at the California Institute of Technology. He had served as the President of the American Physical Society. Later, he was invited by the Rice University to be their President. When I was a graduate student at Rice, he had retired from the position of President but remained in the Physics Department as a professor. My early study of quantum mechanics was mainly based on his textbook. As a seasoned physicist, Houston was a gentleman-type scholar. I still remember that he often sat in the front row during our departmental seminars. He would ask a lot of questions, but he was very humble and never behaved like an arrogant authority. This really impressed me.

Prof. Harold E. Rorschach, Jr. When I was studying at the Rice University, my research work was mainly experimental physics. Professor H.E. Rorschach was my PhD advisor. My research project at the time was to use spin-echo NMR (nuclear magnetic resonance) to study transport properties of helium-3 atoms at low temperature in the He^3-He^4 superfluid. There was a good tradition at Rice University. Students

[2] H.E. Rorschach, *American Journal of Physics*, **38**, 897–904 (1970).

studying experimental physics were also required to have a deep theoretical understanding on how to explain the experimental observations. So, I was working on two fronts at that time: One was to build a home-made spin-echo NMR spectrometer to do the experiment; the other was to learn the scattering theory in superfluid (also known as "quantum liquid") to explain my experimental results.

In my doctoral dissertation work, helium-3 was treated as a particle to interact with "phonons" and "rotons" in the superfluid. As mentioned above, in the work of Houston and Sommerfeld, they showed that although the electron is a massive particle, its motion inside the metal can be viewed as a wave. On the other hand, in my low-temperature physics research, the "phonons" and "rotons" are different types of excitation waves of condensed matter; but during the theoretical calculations, these excitation waves were treated as individual "particles".

Therefore, in my doctoral dissertation work, I got a deep impression that some massive particles can behave like "waves", while some excitation waves that have nothing to do with real particles can behave like "particles". This clearly showed that in quantum physics, the concepts of "particle" and "wave" are almost interchangeable.

My supervisor, Professor Rorschach, is an excellent physicist. Not only did he guide me experimentally, but he also did a lot of theoretical work. Before supervising my work, he had close collaborations with Felix Bloch at Stanford University. Bloch was the first postgraduate student of Heisenberg, whom I mentioned earlier that worked with Houston in Leipzig. Bloch was one of the pioneers in the development of NMR. In fact, he won the Nobel Prize in Physics in 1952 for his work on nuclear magnetic resonance. So, my work also involved Bloch's theory. Professor Rorschach received his PhD from the Massachusetts Institute of Technology (MIT) and was hired by Houston to Rice University. He later collaborated with both Houston and Bloch.

I was very lucky to have someone like Professor Rorschach as my supervisor. He is a very open-minded person. As a graduate student, when I discussed research work with him, I could express my thoughts freely, without worrying whether he would approve them or not. In fact, there were several occasions when my thinking was different from that of Prof. Rorschach; I would argue with him. He never forced me to give in. This encouraged me to develop a habit of thinking boldly and independently. Professor Rorschach also gave me a lot of freedom on the progress of my work. Although I was supported by scholarship through him, he never pushed me to publish papers.

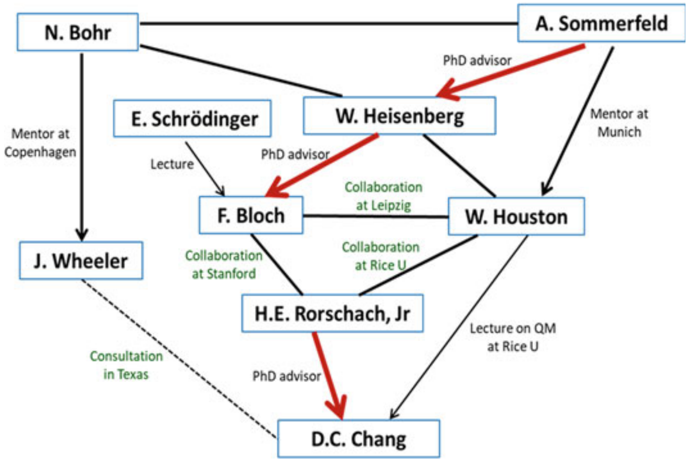
Prof. John A. Wheeler. When I was a graduate student, I was very curious about the physical basis of wave-particle duality. After years of exploration, I got an answer. That is, for particles with or without mass (including electrons and photons), they are quantized excitation waves of the vacuum medium. This was not a mainstream view. At that time, I was an assistant professor, and I was still collaborating with Professor Rorschach. He was quite supportive of my non-mainstream idea. I wrote my idea into an article and submitted it to *Physical Review*. The reviewers first made some specific criticisms. I fully answered them. However, the editor of the journal told me later that, after consulting with an editorial Board member, he decided not to accept my paper for publication. I asked for why; but the editor did not want to give any specific reason. I was very disappointed. As a young physicist, I needed some encouragement. In addition to the support given by Professor Rorschach at that time, another scholar also gave me much-needed encouragement. He was Professor John Wheeler.

Professor Wheeler was an influential scholar in quantum physics. He had studied at the University of Copenhagen under Niels Bohr. He later worked for a long time at Princeton University, applying quantum theory to cosmology. He not only conducted excellent research work, but also trained many famous students (including Richard Feynman). He worked at the University of Texas at Austin in his later years. On one occasion, Rice University invited Professor Wheeler for a week-long visit. At that time, I was an assistant professor in the Department of Physics at Rice University. During that week, I had a lot of contact with Professor Wheeler and specifically asked him to discuss what the nature of particles is in quantum theory. I told him my idea of treating electrons and photons as excitation waves of the vacuum medium. Professor Wheeler did not think my unorthodox ideas were unreasonable, and he encouraged me to continue the work. When he returned to the University of Texas, I kept communicating with him. I sent him the article I wrote for *Physical Review*; he gave me a very encouraging reply. Professor Wheeler was very respected in the physics community. As a young physicist, the support from a famous scientist to my unorthodox ideas was a great inspiration to me. Also, the open-minded spirit of Professor Wheeler gave me a very deep impression.

The Copenhagen School

In quantum mechanics, there was an especially important school called the "*Copenhagen School*". Many scientists who made pioneering contributions to quantum theory were from this School. The three teachers I mentioned above all had a connection with the Copenhagen School.

Prof. Houston was supervised by A. Sommerfeld while working in Munich, and he later worked with W. Heisenberg in Leipzig. Heisenberg and Niels Bohr had a very close working relationship. Heisenberg had worked for a long period at the Bohr Institute in Copenhagen. Sommerfeld's work was also very closely related to Bohr's; he extended the quantum condition in Bohr's original atomic theory to a Bohr-Sommerfeld relation. While in Leipzig, Houston collaborated with Heisenberg's graduate student, Felix Bloch. Later when Houston arrived at Rice University, he hired H. E. Rorschach. Later Rorschach went to Stanford University to work with Bloch. As a result, both Houston and Rorschach had a connection with the Copenhagen School. As for the other teacher mentioned above, Professor Wheeler had studied at the University of Copenhagen in his younger days, where his mentor was Bohr. (The relationship is shown below.)



So, from the academic lineage, I was under the influence of the Copenhagen School. However, my current view of *wave-particle duality* is very different from their view. In quantum mechanics, the orthodox interpretation of the wave function in a quantum equation is the "**Copenhagen Interpretation**". Bohr and Heisenberg mainly proposed this mainstream explanation. They assumed that the particle is a point object, and the quantum wave function just describes the probability for detecting the particle in a particular position in space–time. This is the interpretation that almost all quantum physics textbooks now teach.

When I was a student, all my teachers taught quantum mechanics according to the view of the Copenhagen School. I, however, never fully accepted the Copenhagen interpretation of the quantum wave function. If an electron is regarded as a particle, it is difficult to explain why electrons, like photons, show a wave property in the diffraction experiment. In Schrödinger's days, many physicists called the quantum wave function of a massive particle as "*matter waves*"; they thought matter could behave like a wave. Thus, I believe the quantum wave function in the Schrödinger equation should describe a real physical wave, not a wave of probability.

From Quantum Physics to Bio-Medical Physics

After I finished my PhD work, I developed a strong interest in bio-medical physics. I wanted to use my quantum physics training to study the living system. When I told Professor Rorschach about my plan, he was very supportive. So, I started to collaborate with a young physiologist (CF Hazlewood) at Baylor College of Medicine. We used NMR to study the structure of water molecules in cells. Professor Rorschach was very generous; he allowed me to modify the spin-echo NMR spectrometer which I built for studying liquid helium to study water molecules in biological cells.

This work turned out to be remarkably interesting. We found that the physical state of water molecules in cells is quite different from ordinary water; the NMR relaxation times T_1, T_2, and the spin diffusion coefficient D_s of cellular water were significantly reduced in comparison with pure water, implying that the cellular water is in a more ordering state.

More importantly, the degree of ordering of the cellular water appeared to correlate with the differentiation state of the tissue. Since tumor development is suspected to be associated with de-differentiation, we conducted spin-echo NMR measurements on murine mammary glands in three different morphological states (normal, pre-neoplastic nodule, and tumor). We discovered that the relaxation times T_1 and T_2 (and the spin diffusion coefficient D_s) became significantly longer during tumor development. This finding suggested that NMR can be used as an early diagnostic tool; it can distinguish the tumor or pre-tumor cells from normal tissue.

Our discovery was first reported in the 1972 APS (American Physical Society) March Meeting and became the meeting highlight (see attached *Press release by APS*). This finding was published in Nature (1972) and PNAS (1972).[3] Our finding triggered the development of the MRI (magnetic resonance imaging) technology. One year after we reported our findings, Dr. P. Lauterbur published a short paper in Nature (1973) suggesting that one can use a magnetic field gradient to scan water molecules in different location of a sample. Today, the visualization of tumor from normal tissues by MRI still relies on detecting the difference of relaxation times in the tissue water, as we discovered in 1972.

[3] D. C. Chang, C. F. Hazlewood, B. L. Nichols, and H. E. Rorschach, *Nature* **235**, 170 (1972); C. F. Hazelwood, D. C. Chang, D. Medina, G. Cleveland, and B. L. Nichols, *Proc Natl Acad Sci U S A* **69**, 1478 (1972).

Press release by APS for the 1972 March Meeting:
Using NMR to detect cancer

MAR 30 1972

Nuclear Physics Seen Aiding In Breast Cancer Detection

ATLANTIC CITY — A new approach to early detection of breast cancer by detecting differences in the water of normal and tumorous cells was described here Wednesday by four scientists from Baylor College of Medicine and Rice University.

Delegates to the March meeting of the American Physical Society heard Drs. Donald C. Chang, Harold E. Rohrsbach, B. L. Nichols, and C. F. Hazlewood report how they borrowed an analytical technique from nuclear physics to differentiate between normal and tumorous or pre-tumorous cells.

Experimenting on mice, the scientists noted that water molecules in living cells move about at different speeds, depending on the degree of cell abnormality. Water in a pre-tumor cell module diffuses about 20 per cent faster than in a normal cell and increases to 30 per cent in a tumor cell, they said.

Chang reported his team applied a nuclear technique known as pulsed nuclear magnetic resonance to establish the differences in water cell diffusion. The method involved probing a sample of tissue engulfed in a magnetic field with a pulse of radio frequency energy, vibrating 30 million times per second, and then measuring the echo returned by the hydrogen atoms making up the water in the sample.

The echo signal determines how fast the water moves within the cell, Chang said.

Chang said his measuring technique is about 85 per cent effective in determining whether a test sample is normal or pre-tumorous and 95 per cent accurate in predicting tumors.

At the convention banquet, Dr. Philip M. Morse of Massachusetts Institute of Technology presented the Oliver E. Buckley Solid State Physics Prize to Dr. James C. Phillips of Bell Laboratories and the 1972 American Physical Society High-Polymer Physics Prize to Dr. Anton Peterlin of the Research Triangle Institute.

The five-day meeting ending today brought about 2,000 physicists to Chalfonte-Haddon Hall.

My physics training might give me some advantages over traditional biologists in studying the living system. After the NMR study, I did pioneer work in several new areas, including the development of electroporation for gene transfer,[4] bio-photonics for cell signaling studies and rapid-drug-screening.[5] I also collaborated with Dr. Roger Y. Tsien (2008 Nobel laureate in chemistry) in using GFP for uncovering the signaling mechanism of cell division.[6]

[4] Chang, D.C., Chassy, B.M., Saunders, J.A., and Sowers, A.E. (Eds). *Guide to Electroporation and Electrofusion*, Academic Press, (1992).

[5] Chang, D.C. *Intl J Mod Phys B*, **21**, 4091-4103 (2007).

[6] Li, C.J., Heim, R., Lu, P., Pu, Y.M., Tsien, R.Y. and Chang, D.C. *J Cell Sci.* **112** (10), 1567-1577 (1999).

Returning to the Quest of Truth in Nature

After working for several decades in the study of living systems, I had not forgotten about my original interest in quantum physics. So, in the later years of my career, when I no longer needed to be concerned about grant proposals or publications, I decided to come back to fundamental physics, hoping to satisfy my curiosity about the foundation of quantum mechanics.

Today, there are still many fundamental questions about quantum physics that remain unanswered. In the past one hundred years, the application of quantum physics has made tremendous achievements. Today's cutting-edge technologies, including lasers, IC chips, computers, mobile phones, AI devices, and satellites, are all dependent on it. However, some basic principles of quantum physics are still not clear to scientists. Even worse, many physicists have given up trying to understand the foundation of quantum physics. They think that quantum physics is so amazing that its principles have surpassed human wisdom; we can only use it, but we can never understand it.

After a hundred years of development, the theory of quantum mechanics today is far more complicated than the days of Bohr and Schrödinger. But in terms of basic concepts, I am afraid that today's physicists have not surpassed Bohr's ideas. In fact, there are many fundamental problems in quantum physics still waiting to be resolved today. For example,

- Where do particles come from? Why can particles be created in the vacuum?
- What are the physical properties of the vacuum?
- If particles are excited waves of the vacuum, should all quantum mechanical equations (including Klein–Gordon Equation, Dirac Equation, and Schrödinger Equation) be derived from the physical properties of the vacuum?
- How to resolve the problem of *wave-particle duality*?
- What is the origin of *mass* in quantum physics?

For the past twenty years, I have been working actively to explore these issues. I was able to get some very interesting insights. Some of my findings have been published as research papers. However, in order to give a more comprehensive picture about the true nature of our physical world, I realize that I must summarize my new findings into an easily understood story; that is this book!

Appendix A
The Aether Hypothesis and the Concept of Treating the Vacuum as a Physical Medium

A.1 A Brief History of the Aether Hypothesis

The idea that the vacuum behaves as a wave medium was widely accepted during the study of optics and electromagnetism; it was called the "*Aether hypothesis*". Before the twentieth century, many physicists thought that the space between matters is filled with a hypothetical medium called "aether", which allows light or electromagnetic waves to propagate in it.

In the study of classical physics, it was commonly believed that action in a distance requires a medium to transmit the effect. That means the space cannot be empty; there must be some medium in the space to transmit the physical forces, such as gravity, electric or magnetic forces. Furthermore, in order for a vibrating wave to transmit, a physical medium is required. Since light is a wave, it must have a carrying medium.

An early proposer of the aether concept was René Descartes. According to Whittaker: "*Space is thus, in Descartes's view, a plenum, being occupied by a medium which, though imperceptible to the senses, is capable of transmitting force, and exerting effects on material bodies immersed in it—the aether*" (see E. Whittaker, *A History of the Theories of Aether and Electricity.* Thomas Nelson and Sons Ltd, 1951). Many well-known physicists, including Faraday, Helmholtz, Maxwell, Stokes, and Lorentz, and mathematicians like Cauchy, Poisson, Gauss, and Riemann, had strongly supported such an idea. In fact, they were major contributors to the aether theory.

This light carrying medium (aether) is not only totally transparent; it also has no mass and no friction. As to the detailed physical properties of the aether, there were many different ideas. For example, Hooke and Huygens had independently worked on the refraction of light and developed a wave theory of light. They hypothesized that light consists of a series of vibration of aether. They considered light as a longitudinal wave. However, longitudinal waves have only a single polarization and cannot explain the birefringence phenomenon.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4

Newton also considered aether to be a medium. He favored a picture that "*all space is permeated by an elastic medium or aether, which is capable of propagating vibrations in the same way as the air propagates the vibrations of sound, but with far greater velocity*". The density of aether varies. He thought that aether was composed of very small particles which could interact with light particles and plays an intermediary role between light and matter. And when the light was absorbed by matter, the vibration of aether was responsible for the generation of heat.

Later, Young and Fresnel conducted several well-designed experiments and concluded that light could be a transverse wave rather than a longitudinal wave. To explain this, Young and Fresnel proposed that "*the aether behaves as an elastic solid*", such that the aether can resist the distortion to its body for transmitting the transverse wave.

However, there was a problem to the elastic solid proposal. If aether is a solid, how can the planets, such as the Earth, travel through it without feeling any resistance? Stoke offered a solution. He suggested that the aether was like shoe wax, both rigid enough for transmitting the rapid elastic vibration as of light, but also sufficiently plastic for slow moving bodies such as planets to pass by.

A major contributor to the aether hypothesis was James C. Maxwell. In the nineteenth century, physicists had done a lot of work on how the electric and magnetic force transmit through space. Their works inspired Maxwell to develop the Maxwell theory on electromagnetism. By studying the analogy between electric phenomena and material elasticity, W. Thomson (Lord Kelvin) suggested that the propagation of electric or magnetic force could take the same form as elastic displacement transmitted through an elastic solid. He argued that magnetism is rotational while the electric current is translational. Based on such thinking, Maxwell tried to develop mechanical models of the electromagnetic field. He modeled a dispersive medium composed of atoms of matter embedded in the aether medium. In his final treaty on electromagnetism, he used a model composed of a sea of concealed vortices that rotate about the lines of magnetic force. By using the results from mechanical vortex motion, Maxwell obtained the same equation of motion for the medium. He concluded that light and electric–magnetic phenomena are carried by the same medium.

At the conclusion of Maxwell's famous book: *A Treatise on Electricity and Magnetism,* (Vol. 1), Maxwell wrote: *"If something is transmitted from one particle to another at a distance, what is its condition after it has left the one particle and before it has reached the other? … In fact, whenever energy is transmitted from one body to another in time, there must be a medium or substance in which the energy exists after it leaves one body and before it reaches the other, for energy, as Torricelli remarked, 'is a quintessence of so subtile a nature that it cannot be contained in any vessel except the inmost substance of material things'. Hence all these theories lead to the conception of a medium in which the propagation takes place, and if we admit this medium as an hypothesis, I think it ought to occupy a prominent place in our investigations, and that we ought to endeavour to construct a mental representation of all the details of its action, and this has been my constant aim in this treatise".*

A.2 Why the Aether Hypothesis Was Later Disfavored?

The aether hypothesis, however, became disfavored in the early twentieth century. Its rejection was mainly due to the following reasons:

1. ***The aether hypothesis was not supported by experiments***. In late nineteenth century, several groups (including Michelson and Morley) tried to use optical interferometer to measure the motion between the Earth and the aether. All these experiments failed to detect any movement of the aether. These null results were interpreted to imply that the hypothetical aether does not exist.
2. ***The mechanical properties of this hypothetical aether were full of contradictions***. Aether was supposed to fill all space between matters. In order to do that, the hypothetical aether must be a highly fluidic substance, i.e., either being a gas or liquid. But, aether is supposed to be the medium for transmitting light, which is a high frequency transverse wave; in order to do that, the aether must be a rigid solid. There is a contradiction between these two requirements.
3. ***The aether hypothesis could not explain why large astronomical objects can pass through it***. Since planets like Earth or Mars can move around the Sun without experiencing any resistance from the aether, the aether must be soft enough to allow the planets to pass through it. Also, it should have no friction. This is contradicting to the requirement that aether must be a solid or a fluid.
4. ***The aether hypothesis was revised in an artificial way in order to fit with experimental observations***. In order to explain the null results of the optical interferometer experiments, some scientists tried to modify the aether hypothesis. Lorentz and FitzGerald attempted to explain the null result of the Michelson–Morley experiment by suggesting a length contraction and time dilation during wave propagation. Lorentz suggested that the material body will be contracted in the dimension parallel to the direction of its motion (Lorentz contraction), while the dimension perpendicular to the motion will not be affected. Thus, when the Earth is moving in the aether, everything (including the experimental instrument) in the direction of motion is contracted. Such contraction could explain the null results observed in the interferometer experiment. The physical basis of the Lorentz theory, however, was very complicated and difficult to understand.
5. ***Finally, the aether hypothesis was thought to be unnecessary.*** In 1905, Albert Einstein published the special theory of relativity and showed that one can obtain the mathematics of Lorentz transformation without the assumption of aether. Since Einstein's approach was much simpler and more elegant, it was quickly accepted by the physics community. Since then, the aether hypothesis was gradually abandoned.

A.3 Renewed Interests in the Study of the Vacuum with the Development of Quantum Physics and Cosmology

However, although the aether hypothesis had been disfavored by the mainstream physicists in the past one hundred years, it does not mean that the hypothetical aether is proven non-existent. The existence of aether is still an open question in the minds of many physicists even today. In fact, many major contributors to the relativity theory had revised their ideas and later thought that aether must exist. For example, although Einstein was initially against the aether hypothesis in his 1905 STR paper, he changed his mind later. When Einstein presented a talk entitled "*Ether and the theory of relativity*" at the Leiden University in 1920, he gave a very different view: "*To deny the ether is ultimately to assume that empty space has no physical qualities whatever. The fundamental facts of mechanics do not harmonize with this view.... Recapitulating, we may say that according to the general theory of relativity space is endowed with physical qualities; in this sense, therefore, there exists an ether. According to the general theory of relativity space without ether is unthinkable; for in such space there not only would be no propagation of light, but also no possibility of existence for standards of space and time*".

P. A. M. Dirac was a pioneer in developing the relativistic electron quantum theory. In 1951, he published a letter in *Nature* entitled "*Is there an aether?*" He said: "*Physical knowledge has advanced very much since 1905, notably by the arrival of quantum mechanics, and the situation has again changed. If one reexamines the question in the light of present-day knowledge, one finds that the aether is no longer ruled out by relativity, and good reasons can now be advanced for postulating an aether*". Apparently, although most physicists had given up the aether concept half a century ago, Dirac did not think the aether hypothesis was wrong.

With the development of quantum physics in the first half of the twentieth century, the vacuum can no longer be regarded as an empty space. For example, in quantum electrodynamics, every oscillation mode of radiation is supposed to have a zero-point energy. Such energy is assumed to be a part of the vacuum system. In quantum field theory today, the vacuum is just regarded as the ground state of the quantum system.

Furthermore, not only the empty vacuum model is inconsistent with the current Standard Model of particle physics, it is also totally contradicting the basic assumption of the Standard Model of cosmology today. In fact, the models of vacuum in the current cosmology theories are far more complicated than the aether medium proposed in the nineteenth century.

A.4 How Can the Quantum Wave Model Overcome the Problems of the Previous Aether Hypothesis?

In this book, we introduce the quantum wave model, which hypothesizes that our universe is completely filled with a physical medium called the "*quantum vacuum*". This *quantum vacuum* concept is partially inspired by the earlier aether hypothesis. However, there are significant differences between our model of the *quantum vacuum* and the aether hypothesis of the nineteenth century. For example,

(1) Unlike the *aether* which was supposed to be a medium filling only the space between matters, the *quantum vacuum* is a pre-existing entity that fills all space in our universe.
(2) The *aether* was a hypothetical medium for transmitting electromagnetic radiation only; the *quantum vacuum* discussed here is a medium for all excitation waves, including both radiation and matter waves. In another word, not only the massless photons are excitation waves of the *quantum vacuum*, massive particles (such as electrons) are also excitation waves of the same *quantum vacuum*.

The *quantum vacuum* proposed in this work can overcome most problems encountered by the previous *aether* hypothesis. First, there is no contradicting mechanical requirement for the *quantum vacuum* model. Since the *quantum vacuum* is a pre-existing medium occupying the entire universe, not just filling the space between matters, it does not need to be fluidic.

Second, one can easily explain the lack of dragging effect in the movement of planets. We know all matter are composed of sub-atomic particles. In this wave model, these particles are excitation waves of the *quantum vacuum*. Since there is no friction between the wave and its transmitting medium, there should be no resistance between matter and the *quantum vacuum*. Planets can thus move within the *quantum vacuum* without dragging.

Finally, the null results of the Michelson and Morley experiment can be easily explained. All one needs is to show that the wave equation describing the excitation waves in the *quantum vacuum* is Lorentz invariant. This is indeed the case, since our wave equation is derived from the Maxwell equations, which is known to be Lorentz invariant. In another word, a wave equation being Lorentz invariant is a *mathematical* feature; it does not imply that the vacuum must be an empty space.

Appendix B
The Concept of *Basic Field* and the Lagrangian Formulation of Mechanics

Quantum physics is basically a field theory. So the concept of "*field*" plays a major role. In this work, we use the term "*basic field*" to describe the movement of the vacuum medium during wave excitation. This term needs to be explained very clearly in order for the readers to have a comprehensive understanding of our discussion. How does our concept of "*basic field*" compare to the "*classical fields*" and "*quantum fields*" commonly used in the physics literature? The following is a summary of the similarity and differences between these different field concepts.

B.1 Physical Meaning of the "*Classical Field*"

The classical field is a familiar concept to most people. Examples of such field include the gravitational field and electromagnetic field. The term of "electromagnetic field" was coined by Michael Faraday in the nineteenth century. The classical field is a numerical quantity assigned to every point in space to indicate the action of certain force on the particle. For example, a particle with mass m in a gravitational field G will experience a gravitational force of mG. In an electromagnetic system, if there is an electric field E, a particle with electric charge q will experience an electric force of qE. The classical field is conceptually different from the "*basic field*" used in this work. (However, these two types of fields can be related through certain physical laws. See examples below.)

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4

B.2 Physical Meaning of the "*Quantum Field*" and the Lagrangian Formulation of Mechanics

In the quantum field theory, the "*quantum field*" is not related to force; instead, it is an independent variable in the Lagrangian density that can give the correct equation of motion for a particle using Hamilton's principle. Each type of elementary particle is regarded as an excitation of its own quantum field.

The quantum field concept is developed mainly from the Lagrangian formulation of mechanics based on Hamilton's principle. The following is a short summary of this formulation.

In the study of physics, the main objective is to obtain the appropriate *equation of motion* for a mechanical system, such that the calculation will agree with the experimental observations. In classical mechanics, there are mainly two ways to obtain the equation of motion (see Fig. B.1).

(1) The first way is to conduct a direct calculation on the mechanical system using well-established physical laws of nature, such as Newton's laws or Maxwell equations. This will provide a direct derivation of the equation of motion (such as the wave equation of a photon). Of course, in order to conduct this direct derivation, one must have full knowledge beforehand about the physical laws of nature, which were developed based on a variety of experimental observations.
(2) The second way is to use the Lagrangian formulation. One can start by setting up a "*Lagrangian* (or *Lagrangian density*)" for the mechanical system and then derive the equation of motion through the use of a "*least action principle*" (Hamilton's principle). In this case, the Lagrangian density could be set up either using well-established physical laws of nature, or entirely based on the conjecture of the investigator. In the quantum field theory used today, the Lagrangian density is often determined using "reversed engineering"; that is, the particular form of Lagrangian density is chosen in such a way that would lead to the correct equation of motion. (For example, the Lagrangian density of the Dirac field is chosen to give the Dirac equation.)

In order to give the readers a basic idea of the Lagrangian formulation used in the quantum field theory today, let us review first the Lagrangian formulism used in a simple classical mechanical system which is composed of a number of point masses. This system can be characterized by its Lagrangian $L(q_i, \dot{q}_i)$, which is defined as

$$L = T - V, \tag{B.1}$$

where $T(q_i, \dot{q}_i)$ is its kinetic energy and $V(q_i, \dot{q}_i)$ is its potential energy. With this given L, the action S is defined as

$$S = \int_{t_1}^{t_2} L(q_i, \dot{q}_i)\mathrm{d}t, \tag{B.2}$$

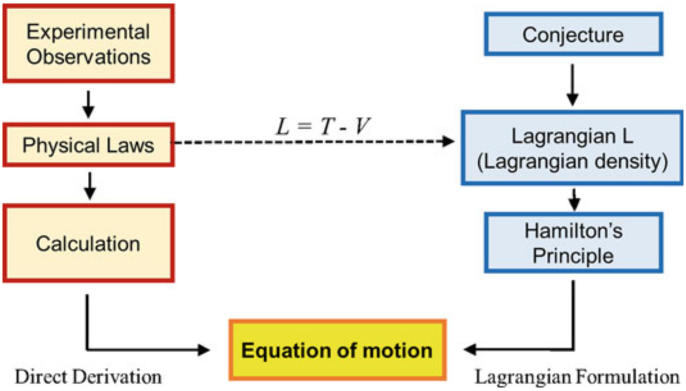


Fig. B.1 Two ways to derive the equation of motion for a mechanical system. The left side is a direct derivation based on well-established physical laws, while the right side is the pathway of derivation using Hamilton's principle (e.g., applying the Euler–Lagrange equation). Examples of the equation of motion include the wave equation of light, the Klein–Gordon equation, the Schrödinger equation, and the Dirac equation

where i is an index representing different particles. The value of S depends on the path of integration in q-space. Hamilton's principle states that the path determined by the equation of motion is corresponding to the path that the action S is at a minimum, i.e.,

$$\delta S = \delta \int_{t_1}^{t_2} L(q_i, \dot{q}_i)\mathrm{d}t = 0. \tag{B.3}$$

It can be shown that, by applying the Hamilton's principle, one can obtain the equation of motion for the system through the use of the *Euler–Lagrange equation*,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0. \tag{B.4}$$

Hamilton's principle may be extended to be applied in a continuous system. In this case, the Lagrangian is the volume integration of the *Lagrangian density* ($\mathcal{L}$), i.e.,

$$L = T - V = \int_0^l \mathcal{L}\mathrm{d}x. \tag{B.5}$$

The action of this system is in the form of

$$S = \int \mathcal{L}\mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t = \int \mathcal{L}\mathrm{d}x^0\mathrm{d}x^1\mathrm{d}x^2\mathrm{d}x^3. \tag{B.6}$$

In the quantum field theory, the Lagrangian density is usually Lorentz invariant and is in the form of

$$\mathcal{L} = \mathcal{L}(\phi, \partial_\mu \phi), \tag{B.7}$$

where ϕ is the *quantum field*. Using Hamilton's principle, the equation of motion of this system can be obtained using the *Euler–Lagrange equation.*

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) = 0. \tag{B.8}$$

In the quantum field theory used in particle physics today, it is thought that all quantum particles are excitations of their own quantum fields. For example, the photon is an excitation of the electromagnetic field; the electron is an excitation of the electron field (or called "Dirac field"); the muon is an excitation of the muon field; and the Higgs particle is an excitation of the Higgs field, and so on.

B.3 Physical Meaning of the "*Basic Field*"

In this work, we use the *basic field* concept to describe the motion of a continuing system, which could be a classical mechanical system or a quantum wave system (see examples below). The *basic field* needs to satisfy two requirements:

(1) ***Physically, the basic field directly or indirectly describes the local movement of the wave medium***.
(2) ***Mathematically, the basic field is an independent variable in the wave equation of a mechanical system***. In the case that the equation of motion is derived using the Lagrangian formulation, the Lagrangian density is constructed based on this basic field.

One may see that, our *basic field* is similar to the *quantum field* used in the current quantum field theory, since both have to satisfy the above Requirement (2). However, our *basic field* is more stringent than the *quantum field,* since it must also satisfy Requirement (1).

Furthermore, in the quantum wave model discussed in this book, one can use the *basic field* concept to derive the correct wave equation through either the direct derivation pathway or the Lagrangian formulation pathway (see Fig. B.1). That is, one can choose a proper *basic field* to derive the equation of motion based on well-established physical laws of nature, or one can use the same *basic field* to set up a Lagrangian density and to derive the equation of motion based on Hamilton's principle. In contrast, the quantum field theory relies only on the Lagrangian formulation to derive the equation of motion; it often cannot use the direct derivation based on well-established physical laws of nature, since such laws are mostly unknown for many quantum particles.

In the quantum field theory today, each quantum particle is associated with its own quantum field. We know nothing about the physical properties of the *quantum field*, except that the quantum particle is an excitation of this field. The *basic field*, on the other hand, has a clear physical meaning; that is, it directly or indirectly represents a local movement of the wave medium.

In the followings, let us use a few examples to explain these different field concepts.

Example 1: Wave Motion in a One-Dimensional Continuous System

A good example of the 1-D continuum system can be a stretched string (see Fig. B.2). We can use the Lagrangian formulation to obtain the wave equation propagating along the string (see Fig. B.2). The distance on the string is marked by the parameter x. The string has a total length of l. For a 1-D continuum, L is an integration of the Lagrangian density ($\mathcal{L}$) along the string,

$$L = T - V = \int_0^l \mathcal{L}\mathrm{d}x \tag{B.9}$$

Let us denote the mass density of the string as ρ. From Newtonian mechanics, we know the kinetic energy (ΔT) for a very small segment of the string (length of Δx) is

$$\Delta T = \frac{1}{2}(\rho\Delta x)\left(\frac{\partial\phi}{\partial t}\right)^2, \tag{B.10}$$

where ϕ is the vertical displacement of the string. For the entire string, the kinetic energy is

$$T = \int_0^l \frac{1}{2}\rho\left(\frac{\partial\phi}{\partial t}\right)^2 \mathrm{d}x \tag{B.11}$$

What about the potential energy V? Again, we will look at a small segment Δx of the string and determine the potential energy in this segment (ΔV). The string is stretched under tension F_1 between two points. For any parts of the string, the deviation from the equilibrium position is very small. As shown in Fig. B.2, the length of the string over Δx is stretched during the wave motion.

It can be shown that, when $\Delta\phi$ is small,

$$\Delta V = \frac{F_1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 \Delta x \tag{B.12}$$

and

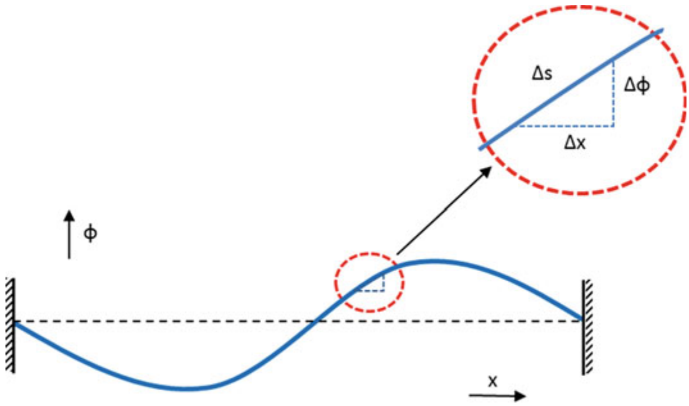


Fig. B.2 Wave propagation in a 1-D continuum system (a stretched string). The wave propagation on a string can be modeled as coupled harmonic oscillations of a string of beads. Here, we denote the vertical displacement of the string as ϕ, ρ is the mass density of the string, and F_1 is the tension of the stretched string

$$V = \int_0^l \frac{F_1}{2}\left(\frac{\partial \phi}{\partial x}\right)^2 \mathrm{d}x. \tag{B.13}$$

Substituting Eqs. (B.11) and (B.13) into Eq. (B.9), we have

$$L = \int_0^l \left[\frac{1}{2}\rho\left(\frac{\partial \phi}{\partial t}\right)^2 - \frac{F_1}{2}\left(\frac{\partial \phi}{\partial x}\right)^2\right]\mathrm{d}x$$

This implies

$$\mathcal{L} = \frac{1}{2}\rho\left(\frac{\partial \phi}{\partial t}\right)^2 - \frac{1}{2}F_1\left(\frac{\partial \phi}{\partial x}\right)^2 \tag{B.14}$$

Applying the Euler–Lagrange equation, one can obtain the wave equation,

$$F_1\frac{\partial^2 \phi}{\partial x^2} - \rho\frac{\partial^2 \phi}{\partial t^2} = 0, \tag{B.15}$$

which can be rewritten as

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{1}{c_1^2}\frac{\partial^2 \phi}{\partial t^2} = 0 \tag{B.16}$$

where $c_1 = \sqrt{F_1/\rho}$. Equation (B.16) is the correct wave equation for a one-dimensional string. Since ϕ denotes the vertical displacement of the string, it satisfies the physical requirement of being a *basic field*. In addition, as shown in Eq. (B.14), ϕ is an independent variable in the Lagrangian density that leads to the correct wave equation, and ϕ also satisfies the mathematical requirement of a *basic field*.

Example 2: Wave Propagation in an Electromagnetic System

In the case of a photon, it is not difficult to identify its *classical field*. From classical electrodynamics, we know the photon is an electromagnetic wave. Indeed, using the Maxwell theory, one can easily derive the wave equations

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = 0,$$

or,

$$\nabla^2\mathbf{H} - \frac{1}{c^2}\frac{\partial^2\mathbf{H}}{\partial t^2} = 0,$$

(where $c = 1\big/\sqrt{\mu_0\varepsilon_0}$ is the speed of light). Thus, the wave equation of a photon directly describes the variation of $\mathbf{E}$ or $\mathbf{H}$**;** they are known as "classical fields". The electric or magnetic field, however, cannot be the *basic field* for a photon, because there is an additional mathematical requirement for the *basic field*. In the Lagrangian formulation of a mechanical system, the Lagrangian density is known to be composed of *the quadratic terms of the first derivatives of the basic field*. Since the Lagrangian density of the electromagnetic field in the vacuum is

$$\mathcal{L} = \frac{1}{2}\left(\varepsilon_o\mathbf{E}^2 - \mu_o\mathbf{H}^2\right); \tag{B.17}$$

$\mathbf{E}$ or $\mathbf{H}$ does not appear to be suitable for playing the role of a *basic field*.

Then, what else can play the role of a *basic field* for the photon? According to the Maxwell theory, $\mathbf{E}$ and $\mathbf{H}$ can be derived from the scalar potential Φ and the vector potential $\mathbf{A}$, such that

$$\begin{cases} \mathbf{B} = \nabla\times\mathbf{A} & \text{(B.18)} \\ \mathbf{E} = -\nabla\Phi - \dfrac{\partial\mathbf{A}}{\partial t} & \text{(B.19)} \end{cases}$$

In the vacuum, the free charge density $\rho_e = 0$ and thus one can set $\nabla\Phi = 0$. Equation (B.19) becomes

$$\mathbf{E} = -\frac{\partial\mathbf{A}}{\partial t}. \tag{B.20}$$

Substituting Eqs. (B.18) and (B.20) into $\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_0}\nabla \times \mathbf{H}$, and using the Coulomb gauge condition $\nabla \cdot \mathbf{A} = 0$, one can easily derive the wave equation

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0, \tag{B.21}$$

where $c = 1\big/\sqrt{\mu_0 \varepsilon_0}$ is the speed of light. The wave function of this equation is **A**; it can be shown that a Lagrangian density composed of **A** will have the right form. Suppose the light wave is traveling along the z axis and the vector potential is along the x-axis, i.e., $\mathbf{A} = A_x \hat{x}$. Using Eqs. (B.17), (B.18), and (B.20), one can show

$$\mathcal{L} = \frac{1}{2}\left[\varepsilon_0 \left|\frac{\partial A_x}{\partial t}\right|^2 - \frac{1}{\mu_0}\left|\frac{\partial A_x}{\partial z}\right|^2\right]. \tag{B.22}$$

It has a similar form as the Lagrangian density of a 1-D string (as shown in Eq. (B.14)). Thus, in the standard teaching of quantum field theory today, **A** is regarded as the *quantum field* of a photon.

In this work, however, we found that the *vector potential* **A** is not a suitable *basic field* for the photon. That is because it fails to satisfy the physical requirement of a *basic field* (i.e., our Requirement (1) in the above); one cannot justify that **A** represents the local movement of the wave medium (which is the vacuum). Instead, we found the *basic field* should be an *electric vector potential* called “**Z**” that represents the displacement of the vacuum medium during wave excitation. (For details, see Chap. 6.)

Appendix C
Helmholtz Decomposition and the Gauge Field

C.1 The Helmholtz Decomposition Theorem

The Helmholtz theorem states that *any sufficiently smooth vector field in a three-dimensional space can be resolved into the sum of an irrotational (curl-free) vector field and a solenoidal (divergence-free) vector field*, i.e.,

$$\mathbf{F} = -\nabla\Phi + \nabla \times \boldsymbol{\Psi}, \tag{C.1}$$

where Φ is called the "*scalar potential*", and $\boldsymbol{\Psi}$ is called the "*vector potential*". The *curl-free component* of a vector field is often referred to as the "*longitudinal component*" and the divergence-free component is referred to as the "*transverse component*".

One can set $\nabla \cdot \boldsymbol{\Psi} = 0$ by choosing a proper *gauge condition.* Since the choice of $\boldsymbol{\Psi}$ is not unique, if the initial choice is $\boldsymbol{\Psi'}$ and $\nabla \cdot \boldsymbol{\Psi'} \neq 0$, one can redefine a new vector potential $\boldsymbol{\Psi}$ such that

$$\boldsymbol{\Psi} = \boldsymbol{\Psi'} + \nabla\chi, \tag{C.2}$$

where χ is an arbitrary scalar function. One can choose χ in such a way that

$$\nabla \cdot \boldsymbol{\Psi} = \nabla \cdot (\boldsymbol{\Psi'} + \nabla\chi) = \nabla \cdot \boldsymbol{\Psi'} + \nabla \cdot \nabla\chi = 0.$$

C.2 Application of the Helmholtz Decomposition Theorem for Analyzing Wave Transmission

The Helmholtz decomposition theorem is highly useful for analyzing different modes of wave transmission in a medium.

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4

Example #1: Wave Excitation in a Mechanical Medium (Elastic Solid)

By using a combination of Newton's Law and the generalized Hooke's Law, one can derive the equation of motion for sound wave transmission in an elastic solid. This wave equation is known as the *Navier equation*,

$$(\lambda_s + \mu_s)\nabla(\nabla \cdot \mathbf{r}) + \mu_s \nabla^2 \mathbf{r} = \rho \frac{\partial^2 \mathbf{r}}{\partial t^2}, \tag{C.3}$$

where λ_s and μ_s are *Lamé's first parameter* and *Lamé's second parameter,* respectively; in some publications μ_s is also called the "*shear modulus*".[1] The principle independent variable here is **r,** which represents the *displacement of the solid element*.

Using the Helmholtz decomposition theorem, the vector **r** can be decomposed into a curl-free longitudinal component ϕ and a divergence-free transverse component $\boldsymbol{\psi}$, i.e.,

$$\mathbf{r} = -\nabla\phi + \nabla \times \boldsymbol{\Psi}, \tag{C.4}$$

By substituting the above equation into the *Navier equation* and recall that ϕ and $\boldsymbol{\psi}$ are independent from each other, one can obtain two separated wave equations,

$$\begin{cases} \nabla^2\phi - \dfrac{1}{c_p^2}\dfrac{\partial^2\phi}{\partial t^2} = 0 & \text{(C.5)} \\ \nabla^2\boldsymbol{\Psi} - \dfrac{1}{c_s^2}\dfrac{\partial^2\boldsymbol{\Psi}}{\partial t^2} = 0, & \text{(C.6)} \end{cases}$$

where $c_p = \sqrt{(\lambda_S + 2\mu_S)/\rho}$ is the speed of the *longitudinal wave* and $c_S = \sqrt{\mu_S/\rho}$ is the speed of the *transverse wave,* respectively.

Example #2: Wave Excitation in the Vacuum Medium

Similarly, one can apply the Helmholtz decomposition theorem to analyze wave excitations in an electrical medium. For the electromagnetic system, it is conventional to use the *(magnetic) vector potential* **A** as the *basic field* to describe the wave equation of an electromagnetic field,

$$\mathbf{B} = \nabla \times \mathbf{A},$$

where **B** is the magnetic flux. In this work, we think **A** is not the most appropriate basic field for describing the movement of the vacuum medium. According to the Maxwell theory, the vacuum behaves more like a dielectric medium. Therefore, it is more appropriate to describe wave transmissions in the vacuum based on the motion of the *electric displacement* **D** instead of the *magnetic flux* **B**.

---

[1] J. Salencon, *Handbook of Continuum Mechanics: General Concepts, Thermoelasticity* (Springer Science & Business Media, 2001).

By applying the Helmholtz decomposition theorem, **D** can be expressed as

$$\mathbf{D} = -\nabla\varphi + \nabla \times \mathbf{Z}, \tag{C.7}$$

In the vacuum, there is no free charge, $\rho_e = 0$, $\nabla \cdot \mathbf{D} = 0$. One can choose $\nabla\varphi = 0$. Equation (C.7) then becomes

$$\mathbf{D} = \nabla \times \mathbf{Z}. \tag{C.8}$$

From Eq. (C.8) and using the relation $\mathbf{D} = \varepsilon_o \mathbf{E}$, we have

$$\mathbf{E} = \frac{1}{\varepsilon_0} \nabla \times \mathbf{Z}. \tag{C.9}$$

Using a combination of Ampère's Law and Faraday's Law, it can be shown that **Z** satisfies the wave equation (see Chap. 6 of the main text),

$$\nabla^2 \mathbf{Z} - \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2} = 0. \tag{C.10}$$

Appendix D
Using the Electric Vector Potential *Z* as a Basic Field to Derive the Maxwell Equations in the Vacuum

In the Standard Model of particle physics, the *quantum field* of the photon is usually assumed to be the *magnetic vector potential* **A**. In fact, one can derive the Maxwell equations based on the variation of **A**. In this work, we propose that the *basic field* for all excitation waves (including photon) of the vacuum should be the *electric vector potential* **Z**. So, one may ask if the complete set of Maxwell's equations in the vacuum can be derived based on the change of **Z**? In the following, we show that this is indeed the case.

The use of **Z** as the *basic field* can help us easily understand the physical meaning of the Maxwell equations. In the vacuum, $\rho = 0,\ \mathbf{J} = 0$,the Maxwell equations are known to be

$$\begin{cases} \nabla \times \mathbf{H} = \dfrac{\partial \mathbf{D}}{\partial t} & (\text{Ampère's Law}) & \text{(D.1)} \\ \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} & (\text{Faraday's Law}) & \text{(D.2)} \\ \nabla \cdot \mathbf{D} = 0 & (\text{Coulomb's Law}) & \text{(D.3)} \\ \nabla \cdot \mathbf{B} = 0 & (\text{Gauss' Law for Magnetism}) & \text{(D.4)} \end{cases}$$

(here, we use the natural unit such that $\mu_0 = 1, \varepsilon_0 = 1,\ c = 1$). In the following, we will show that the entire set of Maxwell equations can be derived by treating **Z** as the *basic field* of the vacuum.

Our starting hypothesis is that the vacuum is a dielectric medium. One can use the Helmholtz decomposition theorem to express the *electric displacement* **D** by an *electric vector potential* **Z**, i.e.,

$$\mathbf{D} = \nabla \times \mathbf{Z}. \tag{D.5}$$

Using the relation $\mathbf{D} = \mathbf{E}$, we have

$$\mathbf{E} = \nabla \times \mathbf{Z} \tag{D.6}$$

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4

Hence, the *electric field* $\mathbf{E}$ is simply a *spatial derivative* of $\mathbf{Z}$. On the other hand, there is evidence suggesting that the *magnetic field* $\mathbf{H}$ is a *time derivative* of the vector potential $\mathbf{Z}$ (see Chap. 6). In fact, we can make this as a formal postulate, i.e.,

$$\mathbf{H} = \frac{\partial \mathbf{Z}}{\partial t} \tag{D.7}$$

Using the 4-vector notation, we can write $Z^{\mu} = (0,\ \mathbf{Z})$. Now, we can define a tensor

$$K^{\mu v} \equiv \partial^{\mu} Z^{v} - \partial^{v} Z^{\mu} = \begin{pmatrix} 0 & H_x & H_y & H_z \\ -H_x & 0 & -E_z & E_y \\ -H_y & E_z & 0 & -E_x \\ -H_z & -E_y & E_x & 0 \end{pmatrix}. \tag{D.8}$$

Using this tensor, one can construct a symmetrical *Lagrangian density*

$$\mathcal{L} = aK_{\mu v}K^{\mu v}, \tag{D.9}$$

(where a is a normalizing constant). Using Hamilton's principle, one can show that the above Lagrangian density will give an equation of motion

$$\partial_{\mu} K^{\mu v} = 0. \tag{D.10}$$

From this equation, one can directly obtain Maxwell Eqs. (D.4) and (D.2). For example, by setting $v = 0$, one can get Eq. (D.4), $\nabla \cdot \mathbf{B} = 0$. By setting $v =$ 1,2,3, one can get Eq. (D.2), Faraday's Law.

The rest of the Maxwell equations can be obtained by using the identity relation,

$$\partial^{\lambda} K^{\mu v} + \partial^{v} K^{\lambda \mu} + \partial^{\mu} K^{v \lambda} = 0 \tag{D.11}$$

For example, when $\lambda, \mu, \nu =$ 1,2,3, one can get Eq. (D.3), Coulomb's Law. When $\lambda, \mu, \nu =$ 0, 1, 2; 0, 1, 3; or 0, 2, 3, one can get Eq. (D.1), Ampère's Law.

Furthermore, using our chosen gauge $\nabla \cdot \mathbf{Z} = 0$, Eq. (D.10) can directly give the excitation wave equation of the vacuum, i.e.,

$$\partial_{\mu} \partial^{\mu} \mathbf{Z} = 0 \tag{D.12}$$

This is identical to the wave equation given in Eq. (6.30).

In summary, we show that the complete set of Maxwell equations can be derived based on the following understandings:

1. The vacuum is thought to behave like a dielectric medium, where the excitation wave is carried by the motion of the *electric displacement* $\mathbf{D}$.

2. Using the Helmholtz decomposition theorem, **D** can be expressed by a divergent-free *electric vector potential* **Z**, which plays the role of the *basic field*. The electric field **E** is a spatial derivative of **Z**, i.e., $\mathbf{E} = \mathbf{D} = \nabla \times \mathbf{Z}$.
3. The magnetic field **H** is a time derivative of **Z**, i.e., $\mathbf{H} = \partial \mathbf{Z}/\partial t$.
4. The entire set of Maxwell equations in the vacuum can be derived using a symmetrical Lagrangian density $\mathcal{L} = aK_{\mu v}K^{\mu v}$, where the tensor $K^{\mu v}$ is composed of Z^{μ} as described in Eq. (D.8).

Appendix E
Was the Relation of Energy-Mass Equivalence Derived from Special Relativity?

In Newtonian mechanics, *mass* and *energy* are two entirely different physical concepts. In the study of modern physics, we now know *mass* and *energy* are related by the relation $E = mc^2$. This new understanding was a great discovery in the twentieth century.

A major challenge is to explain the physical basis for the above discoveries. In the popular science literature, the relation of mass-energy equivalence was often reported to be derived based on the principle of relativity (PR). But a careful examination of the literature indicates that such report is not correct. The reasons are:

(1) The concept of mass-energy equivalence had been suggested by various scientists before the publication of the special theory of relativity (STR) (see attached Table E.1).
(2) Although such concepts were actively promoted by Einstein, his arguments were based on special hypothetical situations (which he called "*thought experiments*") instead of based on first principles.
(3) Many of such *thought experiments* had nothing to do with PR; some might even violate the requirement of Einstein's theory of relativity (see below).

In the following, we will give a detailed summary of literature review to show that the derivation of the mass-energy equivalence relation was not based on special relativity. (Readers interested in this issue may refer to my earlier publication: D. C. Chang, *Mod. Phys. Lett. B* **34**, 2030002, 2020).

E.1 The Concept of Mass-Energy Equivalence Had Been Discussed Actively in Europe Before Einstein

Near the end of the nineteenth century, many European physicists had become aware that the inertial mass can be related to the energy content of an object. For example, in 1881, J. J. Thomson studied the magnetic field generated by a moving charged

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4

sphere. He showed that the field could induce an effective mass on the sphere. In 1889, Oliver Heaviside simplified Thomson's work and suggested that the effective mass should be $m = \frac{4}{3} E/c^2$. Later, Wilhelm Wien and Max Abraham got the same result, which became known as the "electromagnetic mass". The relation of mass-energy equivalence $E = mc^2$ was first mentioned in a paper by Poincare in 1900. In 1904, Fritz Hasenöhrl wrote a series of papers entitled "*On the theory of radiation in moving bodies*", which also provided an elaborated study of the concept relating energy with mass. A detailed account of works regarding the mass-energy converting concept had been reviewed by W. Fadner (*Am. J. Phys.* **56**, 114 (1988)). He showed that there were many discussions on this topic before Einstein first started to address the issue of relating the mass with energy in 1905 (see Table E.1).

E.2 Review of Einstein's First Paper on Deriving the Mass-Energy Equivalence Relation

Einstein clearly recognized the great importance of the mass-energy equivalence concept. In the same year when he published his 1905 paper on the principle of relativity, he published another very short paper entitled "*Does the inertia of a body depend upon its energy-content?*", which described his first derivation of the mass-energy equivalence relation. This paper (here after denoted as the "1905*b* paper") was based on a "*thought experiment*" as summarized in the following:

Consider a setup as illustrated in Fig. E.1. Let an object be placed at the origin of the stationary coordinate system *S*. A system moving along the *x*-axis with speed v is designated as *S'*. The total energy of the object as measured in the frame *S* and *S'* is denoted as E_0 and H_0, respectively.

Now, let this object send out two identical pulses of light in opposite directions making an angle ϕ with *x*-axis. The energy of each of these light pulses is designated

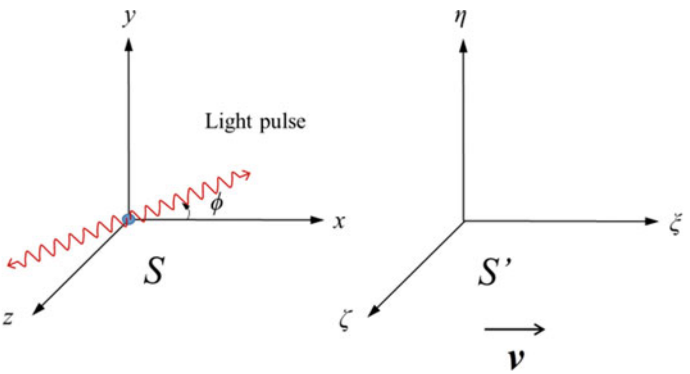


Fig. E.1 Setup of a thought experiment for deriving the mass-energy equivalence relation in Einstein's 1905b paper

$\frac{1}{2}L$ as measured in the S frame. After this light pulse emission, the total energy of the object as measured in the frame S and S' is denoted E_1 and H_1, respectively.

Einstein argued that, based on his theory of relativity, the energy of a light pulse measured in the stationary frame and the moving frame are different. Suppose the energy of the light pulse measured in the S frame is ℓ, then the energy of the light pulse in the S' frame will be

$$\ell^* = \ell \frac{1 - \frac{v}{c}\cos\phi}{\sqrt{1 - v^2/c^2}}. \tag{E.1}$$

Thus, in the above example, the energy of the opposite light pulse measured in the moving frame would become $\frac{1}{2}L\frac{1-\frac{v}{c}\cos\phi}{\sqrt{1-v^2/c^2}}$ and $\frac{1}{2}L\frac{1+\frac{v}{c}\cos\phi}{\sqrt{1-v^2/c^2}}$. Based on this argument, Einstein proposed that

$$\begin{cases} E_0 = E_1 + \frac{1}{2}L + \frac{1}{2}L, & \text{(E.2)} \\ H_0 = H_1 + \frac{1}{2}L\frac{1-\frac{v}{c}\cos\phi}{\sqrt{1-v^2/c^2}} + \frac{1}{2}L\frac{1+\frac{v}{c}\cos\phi}{\sqrt{1-v^2/c^2}} = H_1 + L\frac{1}{\sqrt{1-v^2/c^2}}. & \text{(E.3)} \end{cases}$$

By subtracting Eq. (E.3) from Eq. (E.2), one has

$$H_0 - E_0 - (H_1 - E_1) = L\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right). \tag{E.4}$$

Einstein then argued that the difference between the energy measured in the S and S' frames is mainly due to the kinetic energy K of the object. That is,

$$\begin{cases} H_0 - E_0 = K_0 \\ H_1 - E_1 = K_1 \end{cases} \tag{E.5}$$

Substitute Eq. (E.5) into Eq. (E.4), one has

$$K_0 - K_1 = L\left(\frac{1}{\sqrt{1-v^2/c^2}} - 1\right). \tag{E.6}$$

If $v << c$, one can use the Taylor expansion and ignore the higher-order terms to get

$$K_0 - K_1 = \frac{1}{2}\frac{L}{c^2}v^2. \tag{E.7}$$

From the above equation, one may interpret the decrease in the kinetic energy as a loss of mass (Δm) of the light-emitting object, i.e.,

$$K_0 - K_1 = \frac{1}{2}(\Delta m)v^2. \tag{E.8}$$

By comparing Eqs. (E.7) and (E.8), one may identify the mass loss as

$$\Delta m = \frac{L}{c^2}. \tag{E.9}$$

Einstein thus concluded: "*If a body gives off the energy L in the form of radiation, its mass diminishes by L/c^2*".

The derivation proposed by Einstein in the above *thought experiment* was very simple. However, it is not without problem. For example,

1. Equation (E.9) was not derived based on first principle. Instead, it is just a suggested result demonstrated from a hypothetical special situation.
2. The relation $E = mc^2$ was not an exact derivation; it was an approximation. Equation (E.7) holds only when the speed v is much smaller than c. (This criticism was raised by Planck.)
3. The argument for Eq. (E.5) is incorrect. Strictly speaking, the energy of a particle is now known to be $E^2 = c^2 p^2 + m_0^2 c^4$. The total energy in general thus is not a linear sum of the resting energy and the kinetic energy.
4. This proposed "*thought experiment*" was not realistic. No one can perform such an experiment in reality. One cannot find a physical object that can spontaneously convert part of its mass to emit electromagnetic radiation in opposite directions. If the light-emitting object is an elementary particle, the entire particle will decay (and it cannot emit photons in opposite directions). If the radiation-emitting object is a group of atoms or molecules, the emission of photon(s) can only be due to the release of potential energy from their orbital electrons, not due to a decrease of kinetic energy caused by the loss of the object's rest mass, as suggested in this 1905*b* paper.
5. As pointed out by Herbert Ives, Einstein's 1905 derivation had a problem of using circular logic (H. E. Ives, *J. Opt. Soc. Am.* **42**, 540 (1952)). Ives criticism was also echoed by other investigators.

E.3 Einstein Subsequently Proposed More Thought Experiments to Derive the Mass-Energy Relation

Einstein was not satisfied with his derivation presented in the 1905*b* paper (as described above). He tried to use a different argument based on center of gravity. He proposed a new *thought experiment* in 1906. His argument was summarized in a concise manner in the book *Special Relativity* by A. P. French.

Suppose there is a rigid box with mass M and length L that is floating in space (see Fig. E.2). The box is initially stationary. At $t = 0$, a burst of light wave (with radiation energy E) is emitted from the left end of the box. The radiation carries a momentum *E/c*. Since the total momentum of the system remains zero, the box must acquire a momentum equal to—*E/c*. Hence, the box will recoil with a speed v,

$$v = -\frac{E}{Mc}. \tag{E.10}$$

At a later time, $t = \Delta t$ ($\Delta t = L/c$, provided $v < < c$), the radiation wave reaches the right end of the box and its energy is totally absorbed. It conveys an impulse on the box equal and opposite to the initial burst of momentum. Therefore, it will bring the box to rest again. The net result of this process is to move the box through a distance Δx:

$$\Delta x = v\Delta t = -\frac{EL}{Mc^2}. \tag{E.11}$$

Since this box is an isolated system, one can assume that the center of gravity of the box remains unchanged before and after the radiation process. In order to counter the movement of the rigid box, one must assume that *the radiation should carry with it the equivalent of a mass m*, such that

$$mL + M\Delta x = 0. \tag{E.12}$$

Combining Eqs. (E.11) and (E.12), we have

$$m = \frac{E}{c^2} \text{ or } E = mc^2. \tag{E.13}$$

This suggests that *the radiation wave can behave as a material object which has an effective mass equivalent to its energy divided by* c^2.

One may notice that, although Einstein was able to partially justify the mass-energy equivalence in this *thought experiment*, its derivation was not free of problems. For example,

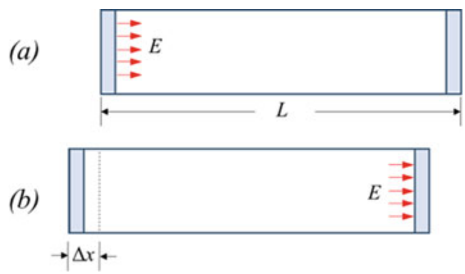


Fig. E.2 Einstein's thought experiment. A hypothetical experiment in which a box recoils from its initial position **a** to a final position **b** as a result of a burst of radiation energy traveling from one end of the box to the other

(1) This *thought experiment* only demonstrated that a radiation wave could carry a mass that is proportional to its radiation energy. It did not show that an object with mass m can convert its mass into an amount of energy $E = mc^2$.
(2) The argument in this *thought experiment* was not based on the principle of relativity. Instead, it was based on the assumption that *the radiation should carry with it the equivalent of a mass m.*
(3) As pointed out by Einstein himself in his later paper (*Ann. Phys.* **23** 371 (1907)), this *thought experiment* violates the basic principle of STR, because it would assume information can be transmitted faster than the speed of light. Otherwise, the entire box cannot move simultaneously when a burst of radiation wave was emitted from its left end at $t = 0$.
(4) It has been pointed out that this *thought experiment* was unrealistic and cannot be carried out in practice. Even if it works, this *thought experiment* can only be regarded as a special case; it was not a general derivation of the mass-energy equivalence relation.

Following the 1905*b* and the 1906 papers, Einstein continued to publish several papers trying to give more convincing arguments about the derivation of the relation of mass-energy equivalence. His major works are summarized in attached Table E.1. These works had been reviewed extensively by various investigators in recent years. Many reviewers were not convinced that Einstein had properly derived the relation of mass-energy equivalence. For example, in a review by T. Rothman, he concluded that, "(Einstein) *was aware of the shortcomings of his derivation (in 1905 and 1906) and wrote a half dozen more papers over the next 40 years trying to patch things up but arguably never succeeded*". ("Was Einstein the first to invent $\mathrm{E} = \mathrm{mc}^2$?," https://www.scientificamerican.com/article/was-einstein-the-first-to-invent-e-mc2/.) A similar view was also given by Max Jammer in Chap. 3 of his book entitled "*Concepts of mass in contemporary physics and philosophy*". A number of physicists had also expressed criticism on Einstein's derivation of the mass-energy equivalence relation.

A more elaborated review of Einstein's attempt to derive the relation of mass-energy equivalence was recently given by Eugene Hecht, who published a series of papers on this subject in *American Journal of Physics* in the last few years. According to Hecht, *"Einstein produced about 18 virtuoso derivations and demonstrations all aimed at establishing the mass-energy principle. We have shown that although each of them gave evidence for the applicability of $E_0 = mc^2$ to a particular set of circumstances, no one derivation, or collection of them taken together, succeeded in providing a definitive proof of its complete generality". "The fact that Einstein continued to create demonstrations of the efficacy of $E_0 = mc^2$ up to 1946 tells us that he knew the definitive proof had not been accomplished*" (E. Hecht, *Am. J. Phys.* **79**, 591 (2011).).

Hence, although Einstein had published many papers on the concept of mass-energy convertibility, his derivations were not really based on the principle of relativity. Instead, his theoretical arguments were based on various hypothetical *thought experiments* which sometimes implied that radiation could behave similarly as matter.

From the recent literature reviews, it is clear that Einstein's attempted derivations of the mass-energy equivalence relation were not successful. That is probably why A. P. French concluded in Chap. 1 of his book "*Special Relativity*" by saying that the general acceptance of Einstein's derivation was not based on the soundness of the theoretical argument. Instead, "*its real vindication (is) in the experimentally observed behavior of particles*".

The following is a list of major publications that had implications on the proposal of the mass-energy equivalence relation, particularly the works of Einstein. This list was compiled mainly based on the reviews of Rothman, Fadner, and Hecht (D. Chang, *J. Mod. Phys.* vol 9, pp 215–240. (2018)).

Table E.1 Historical review on the derivation of the mass-energy equivalence relation

Year	Publication	Major points
1881	J. J. Thomson, *Philos. Mag.* **11**, 229 (1881)	Proposed that a charged conductor in motion increases its mass by $\frac{4}{15}\frac{\mu e^2}{a}$, where μ is magnetic permeability, a is the radius of a charged sphere, e is the electric charge
1889	O. Heaviside, *Philos. Mag.* **27**, 324 (1889)	Simplified Thomson's work and suggested that the effective mass is also proportional to E/c^2
1900	H. Poincaré, *Arch. Néerl. Sci. Exactes Nat.* **2**, 252 (1900)	The relation of mass-energy equivalence $E = mc^2$ was first mentioned in this paper. It is in the form of $\rho = J/c^2$, where ρ is the mass density, J is the energy density
1901 1902	W. Kaufmann, *Göttinger Nachrichten* **2**, 143 (1901); and *Phys. Z.* **4**, 54 (1902)	Reported the first experimental results showing that electron's mass varies with speed
1902	M. Abraham, *Phys. Z.* **4**, 57 (1902)	His study suggested that the effective mass for an electron is $m = (4/3)E/c^2$
1904	H. A. Lorentz, *Proc. R. Neth. Acad. Arts Sci.* **6**, 809 (1904)	Showed that the electron mass parallel to the direction of motion is $m_L = \gamma^3 m$ and the mass perpendicular to the direction of motion is $m_T = \gamma m$, where $\gamma = 1/\sqrt{1 - v^2/c^2}$
1905a	A. Einstein, *Ann. Phys.* **17**, 891 (1905)	Einstein published his famous paper on Special Relativity. He proposed that the mass of an "electron" is not constant. Based on Newton's definition of mass, he derived his speed-dependence relations for longitudinal mass and transverse mass
1905b	A. Einstein, *Ann. Phys.* **18**, 639 (1905)	By proposing a *thought experiment* of an object sending out radiations in opposite directions, he concluded that "*If a body gives off the energy L in the form of radiation, its mass diminishes by L/c*^{2}."
1906	A. Einstein, *Ann. Phys.* **20**, 627 (1906)	He proposed a *thought experiment* using a box with radiation waves transmitting inside. Based on conservation of center of gravity, he derived the mass-energy equivalence relation

(continued)

Table E.1 (continued)

Year	Publication	Major points
1907a	A. Einstein, *Ann. Phys.* **23** 371 (1907)	Proposed another *thought experiment* of a rigid body moving in an electric field. He showed that one can obtain $E_0 = mc^2$ "from a standing point of relativistic electrodynamics"
1907b	A. Einstein, *Jahrbuch der Radioaktivität* **4**, 411 (1907)	Proposed to extend the mass-energy equivalence relation to gravitational mass
1908	M. Planck, *Ann. Phys.* **331**, 1 (1908)	Proposed that the mass change in the absorption and emission of heat energy is $\Delta m = E/c^2$
1909	A. Einstein, *Phys. Z.* **10**, 817 (1909)	Used the same *thought experiment* and argument proposed in 1905b paper to derive "*the inertial mass of a body decreases by* L/c^2 *when the body emits the radiation energy* L". The derivation was made more simple and explicit
1911	A. Einstein, *Ann. Phys.* **340**, 898 (1911)	Proposed a *thought experiment* to show that an increase in gravitational mass is also E/c^2. This paper is also thought to be the author's attempt to develop general relativity
1912	A. Einstein, in *The Collected Papers of Albert Einstein, The Swiss Years: Writings, 1912–1914, Vol. 4,* M. Klein, A. Kox, J. Renn, and R. Schulmann, eds. (Princeton Univ. Press, 1990), pp. 50–55	Proposed a *thought experiment* modified from the 1905b paper to derive the mass-energy equivalence relation. This time with a plate sending out two plane waves in opposite directions, and the plate was treated as a point mass
1913	A. Einstein, in *The Collected Papers of Albert Einstein, The Swiss Years: Writings, 1912–1914, Vol. 4,* M. Klein, A. Kox, J. Renn, and R. Schulmann, eds. (Princeton Univ. Press, 1990), pp. 3–88	Proposed to prove the mass-energy equivalence relation using the stress-energy tensor based on electromagnetic field considerations
1922	A. Einstein, *The Meaning of Relativity* (Princeton Univ. Press, 1922)	Produced a more refined version of his energy–momentum tensor treatment for relativity. This was based on his Stafford Lectures delivered at Princeton during a visit in May 1921
1935	A. Einstein, *Bull. Am. Math. Soc.* **41**, 223 (1935)	Considered a system using two mass points traveling toward each other. Based on the argument of conservation of momentum and conservation of energy, he concluded that one can regard $mc^2(\gamma - 1)$ as the kinetic energy of the particle
1946	A. Einstein, *Technion Yearbook* **5**, 16 (1946)	Einstein's last effort to prove $E_0 = mc^2$. The text was short and simple. The treatment "does not presume the formal machinery of the theory of relativity, but uses only three previously known laws": conservation of momentum, the equation for radiation pressure, and the expression for stellar aberration of light

Appendix F
The Physical Basis of the Four-Dimensional Space–Time by Minkowski

F.1 Is Our World Truly Four-Dimensional?

Before one can answer the above question, one may first ask: What is a four-dimensional world? Generally speaking, the four-dimensional world is composed of three spatial dimensions plus one time dimension. But is time really a 4^{th} dimension?

Our human experience is familiar with the 3-D world, but we do not have direct experience with the 4-D world. We may start with the 3-D world and see what kind of mathematical relationship we should have when we extend the 3-D world to the 4-D world.

The 3-D world is simple. It can be described in the Euclidean system using rectangular coordinates composed of *x, y, z*. For any vector goes from the origin to some position (*x,y,z*), one can write the vector in the form of

$$\vec{\boldsymbol{x}} = x_1\hat{x}_1 + x_2\hat{x}_2 + x_3\hat{x}_3. \tag{F.1}$$

This three-vector could be written in a different form:

$$\vec{x}_\mu = (x_1, x_2, x_3),$$

where $\mu = 1, 2, 3$. We know the 3-axes are perpendicular to each other. Each of these axes represents a dimension; The 3 perpendicular axes x_μ together form a Hilbert space. The amplitude of such a three-vector in the Hilbert space is

$$\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{x}} = x_1^2 + x_2^2 + x_3^2 = \text{constant}. \tag{F.2}$$

This amplitude (the dot product of the three-vector) is invariant under the transformation such as a rotation. That means, in the Hilbert space, the $\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{x}}$ is invariant after undergoing a rotation. If we denote the three-vector $\vec{\boldsymbol{x}}$ after a rotation around the x_3-axis as another three-vector $\vec{\boldsymbol{x}}'$, we will have

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4

$$\vec{\boldsymbol{x}}' \cdot \vec{\boldsymbol{x}}' = x_1'^2 + x_2'^2 + x_3'^2 = \text{constant}. \tag{F.3}$$

Since the constant does not change after the rotation (i.e., invariance of the amplitude of the three-vector), one can show that the components of the three-vectors $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{x}}'$ are connected by the following relations:

$$\begin{cases} x_1' = x_1 \cos\theta + x_2 \sin\theta \\ x_2' = -x_1 \sin\theta + x_2 \cos\theta \end{cases} \tag{F.4}$$

This is the transformation between the original coordinates and the coordinates after the rotation.

Our world is not 4-D in the Hilbert space

What happens if we are living in a four-dimensional world? In this case, there should be a fourth dimension. That means besides x_1, x_2, x_3, we should have a new dimension x_4. In a Hilbert space, a four-vector should be

$$\vec{x}_\mu = (x_1, x_2, x_3, x_4),$$

where $\mu = 1, 2, 3, 4$. If we conduct a rotation on this four-dimensional world, it should give the similar invariant property as we observed in the 3-D world, i.e.,

$$\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{x}} = x_1^2 + x_2^2 + x_3^2 + x_4^2 = \text{constant}, \tag{F.5}$$

and after rotation,

$$\vec{\boldsymbol{x}}' \cdot \vec{\boldsymbol{x}}' = x_1'^2 + x_2'^2 + x_3'^2 + x_4'^2 = \text{constant}. \tag{F.6}$$

However, in reality, we never observe such a physical world which could satisfy the above invariance property. We cannot find a x_4 that could satisfy the Eq. (F.5) and preserve the equation when undergo a rotation. In the real world, there is no such case. This implies that we do not live in a 4-D world, which is generalized from our familiar 3-D world in a Euclidean system. So, at least from a spatial point of view, we do not live in a 4-D world.

F.2 The Time Dimension is a Pseudo-Dimension in Poincaré's Relativity Theory

Then, can we use other justification to make *time* the 4$^{\text{th}}$ dimension? This was first done by Henri Poincaré. Before 1905, Poincaré had developed his own relativity theory.[1] In 1905–1906, he wrote another paper entitled "*On the Dynamics of the Electron*" (The paper was received in the July 1905 and published in January 1906).[2] The invariance is not on the length of the vector itself, but something else. Poincaré compared a *boost* instead of a *rotation.* A *boost* means a transformation from a stationary coordinate system K to a moving coordinate system K'. Based on Michelson-Morley experiment, people already knew at that time that *the equation of motion for light* does not change in these two different inertial frames K and K'. In the K frame, the equation of motion for light looks like

$$x_1^2 + x_2^2 + x_3^2 - (ct)^2 = 0 \text{ (constant).} \tag{F.7}$$

When applying a boost operation $\vec{x} \to \vec{x}'$, the system goes through a Lorentz transformation. According to the Michelson-Morley experiment, *the equation of motion for light* is the same in K' as in K. This means that it is invariant:

$$x_1'^2 + x_2'^2 + x_3'^2 - (ct')^2 = 0 \text{ (constant).} \tag{F.8}$$

By comparing this with the hypothetical four-vectors in Hilbert space, the *time* dimension may look like the x_4. Thus, Poincaré proposed a new four-vector $(\vec{\boldsymbol{x}}, ict)$, where $\vec{\boldsymbol{x}} = x_1\hat{x}_1 + x_2\hat{x}_2 + x_3\hat{x}_3$ and $x_4 = ict$. In short, $(\vec{\boldsymbol{x}}, ict)$ forms a four-vector where

$$\begin{cases} x_1 = x \\ x_2 = y \\ x_3 = z \\ x_4 = ict \end{cases} \tag{F.9}$$

One can also write the four-vector as $\vec{x}_\mu$, where $\mu = 1, 2, 3, 4$.

In Poincaré's theory, this four-vector is invariant in the *boost* transformation from K to K'. From the invariance of the wave equation of light, Eq. (F.7) now looks like Eq. (F.5).

$$\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{x}} = x_1^2 + x_2^2 + x_3^2 + x_4^2 = \text{constant,}$$

and Eq. (F.8) now looks like Eq. (F.6)

[1] H. Poincaré, "La théorie de Lorentz et le principe de réaction," *Arch. Néerl. Sci. Exactes Nat.* **2**, 252–278 (1900).

[2] H. Poincaré, "Sur la dynamique de l'électron," *Rendiconti del Circolo matematico di Palermo* **21**, 129–176 (1906).

$$\vec{\boldsymbol{x}}' \cdot \vec{\boldsymbol{x}}' = x_1'^2 + x_2'^2 + x_3'^2 + x_4'^2 = \text{constant}.$$

Thus, it can be said that we live in a 4-D world based on Poincaré's mathematical framework. However, if one compare Poincaré's mathematical framework with Hilbert space mathematical framework, there is a difference. The 4$^{\text{th}}$ dimension in the Poincaré's mathematical framework is imaginary; $x_4 = ict$ is not a normal dimension. To be more precise, perhaps we should call the 4$^{\text{th}}$ dimension, (i.e., *ict*), the *pseudo-dimension.*

F.3 Minkowski's New Mathematical Framework on the Four-Dimensional Space–Time

When Minkowski worked on this problem, he developed a new mathematical framework using slightly different notations. His purpose was to get rid of the imaginary number *i*. Instead of denoting *ict* as x_4, he denoted that

$$ct \equiv x_0. \tag{F.10}$$

In this case, Minkowski's four-vector became

$$\vec{x}_\mu = (x_0, x_1, x_2, x_3)$$

where $\mu = 0, 1, 2, 3$.

Now, Minkowski developed a new mathematical framework different from the Hilbert space. In this framework, he defined two types of *four-vectors*:

(1) Covariant four-vector:$\mathrm{X}_\mu \equiv (x_0, -\vec{\boldsymbol{x}})$
(2) Contravariant four-vector: $\mathrm{X}^\mu \equiv (x_0, \vec{\boldsymbol{x}})$

where $\mu = 0, 1, 2, 3$, $x_0 \equiv$ ct represent the time dimension, $\vec{\boldsymbol{x}} \equiv (x_1, x_2, x_3)$ represent the three spatial dimensions.

The covariant four-vector and the contravariant four-vector can be related by

$$\mathrm{X}^\mu = g^{\mu v}\mathrm{X}_v, \tag{F.11}$$

where $g^{\mu v}$ is the Minkowski *metric tensor*:

$$g^{\mu v} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{F.12}$$

He further defined that

$$X_\mu \cdot X^\mu = x_0^2 - x_1^2 - x_2^2 - x_3^2 = |X|^2. \tag{F.13}$$

Now, based on the results of the Michelson-Morley experiment, we know *the equation of motion for light* does not change in two different inertial frames K and K'. Thus, it appears that the length of the Minkowski four-vector is invariant under a boost transformation, i.e.,

$$|X|^2 = (ct)^2 - x_1^2 - x_2^2 - x_3^2 = \text{constant}, \tag{F.14}$$

holds for the K frame, while

$$\left|X'\right|^2 = \left(ct'^2\right) - x_1'^2 - x_2'^2 - x_3'^2 = \text{constant}, \tag{F.15}$$

holds for the K' frame. By comparing Eq. (F.13) with Eq. (F.7), one can easily see that *the invariance of the equation of motion for light is equivalent to implying that the amplitude of the Minkowski four-vector is a constant* under a boost operation (i.e., Lorentz transformation).

One advantage of the Minkowski mathematical framework is that it makes the Lorentz transformation look simple. For example, the space–time four-vector is now represented by a 1×4 matrix, the Lorentz transformation is a 4×4 matrix such that one can convert the coordinates in frame K to the coordinates in frame K' by using the following relations

$$\begin{cases} x'^0 = x^0 \cosh\theta - x^3 \sinh\theta \\ x'^1 = x^1 \\ x'^2 = x^2 \\ x'^3 = -x^0 \sinh\theta + x^3 \cosh\theta \end{cases} \tag{F.16}$$

where $\sinh\theta = \frac{\beta}{\sqrt{1-\beta^2}}$, $\cosh\theta = \frac{1}{\sqrt{1-\beta^2}}$, and $\beta = v/c$.

One can see that Eq. (F.13) looks simple and compact. After the boost transformation, it gives a similar result in the K' frame. Because it looks simple and beautiful, it is easy for people to accept it. Also, the Lorentz transformation, Eq. (F.16), now looks similar to the transformation of a rotation as shown in Eq. (F.4) for the spatial coordinates.

The use of the Minkowski mathematical framework thus greatly simplifies the notation involved in the mathematical physics studies of relativity.

F.4 Using Minkowski's 4-Vector to Represent the Energy–Momentum Relation

The Minkowski framework can also be applied to represent energy–momentum 4-vector. As shown in Chap. 7, the energy–momentum relation for a quantum particle is originated from the dispersion relation of the quantum wave function,

$$\omega^2 = \left(k^2 + \ell^2\right)c^2, \tag{F.17}$$

where the wave parameters, ω, $\boldsymbol{k}$, ℓ are related to the particle properties of *energy E*, *momentum* $\boldsymbol{p}$, and *rest mass* m_o. And thus, the above relation becomes

$$E^2 = p^2c^2 + m_0^2c^4. \tag{F.18}$$

This relation can be rearranged to be written as

$$m_0^2c^4 = E^2 - p^2c^2. \tag{F.19}$$

Since the rest mass is a constant, the right-hand side of this equation is just the dot product of a 4-vector (E. $\boldsymbol{p}$) expressed in the Minkowski 4-D framework. This 4-vector is often referred to as the "4-momentum".

So, it is clear that the Minkowski 4-D framework is not unique for its application in STR, it can also fit in with the quantum wave model.

Index

D. C. Chang, *On the Wave Nature of Matter*,
https://doi.org/10.1007/978-3-031-48777-4