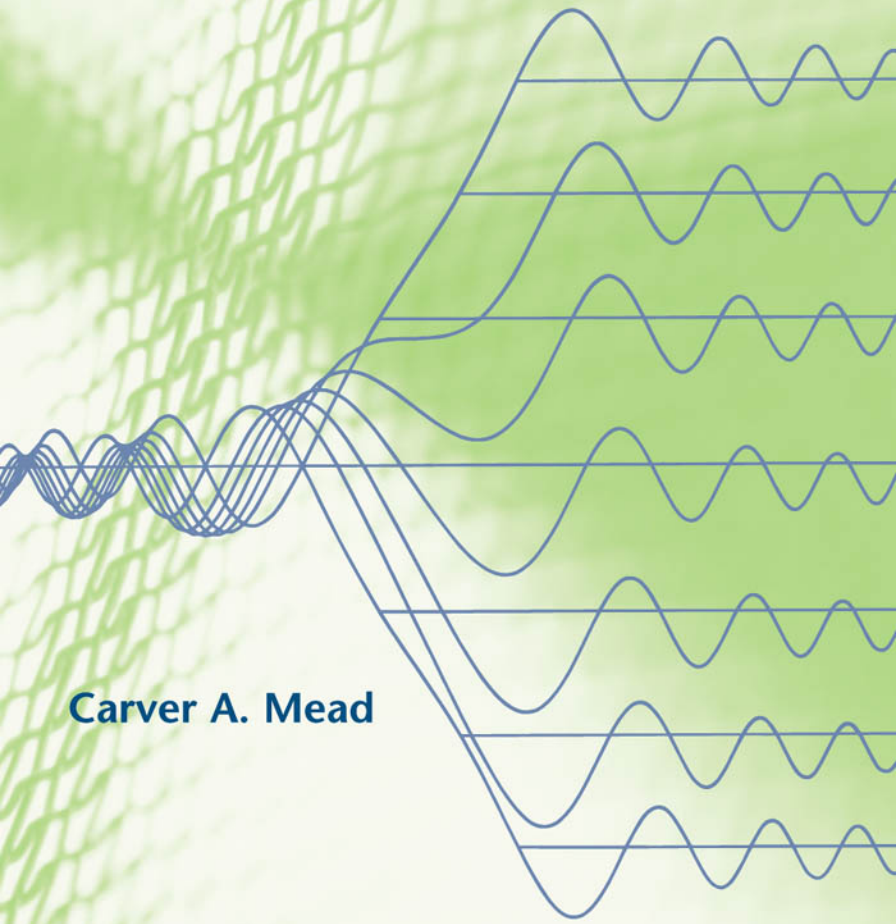
Collective Electrodynamics

Quantum Foundations of Electromagnetism



Carver A. Mead

Collective Electrodynamics

This page intentionally left blank

Collective Electrodynamics

Quantum Foundations of Electromagnetism

Carver A. Mead

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Computer Modern typeface (PostScript Type 1) by the TEX composition system.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Mead, Carver A.
Collective electrodynamics : quantum foundations of electromagnetism / Carver A. Mead.
p. cm.
Includes bibliographical references and index.
ISBN 0-262-13378-4 (hc : alk. paper)
1. Quantum electrodynamics. I. Title.

QC680.M43 2000
530.14′33—dc21

00-028180

Contents

This page intentionally left blank

Foreword

> The formulation of a problem is
> often more essential than its solution,
> which may be merely a matter of
> mathematical or experimental skill.
> To raise new questions, new possibilities,
> to regard old problems from a new angle,
> requires creative imagination and marks real advance in science.
>
> —A. Einstein and L. Infeld[1]

When I was a student, it was commonly understood that one would study a subject until one became an expert; then, one would go out into the world and apply that expertise in one's profession. It went without saying that the expertise itself, as updated through one's experience, would allow the practice of that profession until retirement. The tacit assumption involved in that world view was that the knowledge base evolves slowly, an assumption then already losing validity. Today, we face an explosive growth of knowledge; by any measure, our knowledge base is doubling every few years. How do we, as a human culture, prepare ourselves and our children for this world in which the knowledge base turns over many times within a single human lifetime?

One answer to this dilemma is specialization: One can become an expert in a specialty that is narrow enough to permit one to keep up with the changes as they come along. This is the default solution. In this manner, we can, as it has been said, learn more and more about less and less, until eventually, we know everything about nothing! Specialization, as we all know, has its merits; however, if specialization were to be our only response to rapidly evolving knowledge, I would view our prospects as a culture with deep concern, even with alarm.

In his wonderful book, *The Act of Creation*, Arthur Koestler (3) defines the creative process as starting with the juxtaposition

[1]This quotation appears on page 95 of the popular book, *The Evolution of Physics* (1). This book has recently been reprinted (2); the quotation appears on page 92 of the new version.

of two concepts from separate conceptual spaces. Such a conjunction creates not merely a new idea but an enlargement of the space of ideas, a cross-fertilization that is the very stuff of which innovation is made. If we, by education, by scientific practices, by social norms, restrict the development of individual talents to narrow specializations, we will thereby lose the ability to innovate.

Fortunately, there is, within our culture, an evolution of knowledge over and above the addition of facts and the specialized understanding of those facts. Many phenomena that in the past were seen as separate are now understood to be the same: Fire is a chemical reaction, not a separate element; temperature is energy; light is electromagnetic radiation; molecules are aggregations of atoms; mechanical forces are electromagnetic in origin; . . . Each of these equivalences represents a major *unification* and *simplification* of the knowledge base. Ideas formerly occupying separate conceptual spaces now occupy the same conceptual space. Each unification was made possible by a deeper understanding of existing facts, often triggered by the discovery of a crucial new fact.

It is this unification and simplification of knowledge that gives us hope for the future of our culture. To the extent that we encourage future generations to understand deeply, to see previously unseen connections, and to follow their conviction that such endeavors are noble undertakings of the human spirit, we will have contributed to a brighter future.

Remarks upon acceptance of the
1999 Lemelson–MIT Prize
April 22, 1999
San Francisco, California

Personal Preface

As for the search for truth,
I know from my own painful
searching, with its many blind alleys,
how hard it is to take a reliable step, be it ever so small,
toward the understanding of that which is truly significant.

—Albert Einstein[1]

The material in this little volume has been for me a personal quest that I began nearly fifty years ago. It came about as a direct result of my interactions with Richard Feynman. He and I both arrived at Caltech in 1952—he as a new professor of physics, and I as a freshman undergraduate. My passionate interest was electronics, and I avidly consumed any material I could find on the subject: courses, seminars, books, etc. As a consequence, I was dragged through several versions of standard electromagnetic theory: $\vec{E}$ and $\vec{B}$, $\vec{D}$ and $\vec{H}$, curls of curls, the whole nine yards. The only bright light in the subject was the vector potential, to which I was always attracted because, somehow, it made sense to me. It seemed a shame that the courses I attended didn't make more use of it. In my junior year, I took a course in mathematical physics from Feynman—What a *treat*! This man could *think conceptually* about physics, not just regurgitate dry formalism. After one quarter of Feynman, the class was spoiled for any other professor. But when we looked at the registration form for the next quarter, we found Feynman listed as teaching high-energy physics, instead of our course. Bad luck! When our first class met, however, here came Feynman. "So you're not teaching high-energy physics?" I asked. "No," he replied, "low-energy mathematics." Feynman liked the vector potential, too; for him it was the link between electromagnetism and quantum mechanics. As he put it (5),

In the general theory of quantum electrodynamics, one takes

[1]This quotation was taken from a letter written by Einstein in the year I was born. It appears on page 38 of the wonderful picture book *Essential Einstein* (4). This reference contains many historic photographs of Einstein, each accompanied by a quotation.

> the vector and scalar potentials as fundamental quantities in a set of equations that replace the Maxwell equations.

I learned enough about it from him to know that, some day, I wanted to do all of electromagnetic theory that way.

By 1960, I had completed a thesis on transistor physics and had become a brand-new faculty member in my own right. Fascinated by Leo Esaki's work on tunnel diodes, I started my own research on electron tunneling through thin insulating films. Tunneling is interesting because it is a purely quantum phenomenon. Electrons below the zero energy level in a vacuum, or in the forbidden gap of a semiconductor or insulator, have wave functions that die out exponentially with distance. I was working with insulators sufficiently thin that the wave function of electrons on one side had significant amplitude on the opposite side. The result was a current that decreased exponentially with the thickness of the insulator. From the results, I could work out how the exponential depended on energy. My results didn't fit with the conventional theory, which treated the insulator as though it were a vacuum. But the insulator was *not* a vacuum, and the calculations were giving us important information about how the wave function behaved in the forbidden gap. Feynman was enthusiastic about this tunneling work. We shared a graduate student, Karvel Thornber, who used Feynman's path integral methods to work out a more detailed model of the insulator.

In 1961, Feynman undertook the monumental task of developing a completely new two-year introductory physics course. The first year covered mechanics; although that topic wasn't of much interest to me, it would come up occasionally in our meetings on the tunneling project. When I heard that Feynman was going to do electromagnetic theory in the second year, I got very excited—finally, someone would get it right! Unfortunately, it was not to be. The following quotation from the forward to *Feynman Lectures on Gravitation* (6) tells the story:

> It is remarkable that concurrently with this course on gravitation, Feynman was also creating and teaching an innovative course in sophomore (second-year undergraduate) physics, a course that would become immortalized as the second and third volumes of *The Feynman Lectures on Physics*. Each Monday Feynman would give his sophomore lecture in the morning and the lecture on gravitation after lunch. Later in

> the week would follow a second sophomore lecture and a lecture for scientists at Hughes Research Laboratories in Malibu. Besides this teaching load and his own research, Feynman was also serving on a panel to review textbooks for the California State Board of Education, itself a consuming task, as is vividly recounted in *Surely You're Joking, Mr. Feynman*. Steven Frautschi, who attended the lectures as a young Caltech assistant professor, remembers Feynman later saying that he was "utterly exhausted" by the end of the 1962–63 academic year.

I was another young Caltech assistant professor who attended the gravitation lectures, and I remember them vividly. Bill Wagner (with whom I still communicate over collective electrodynamics material) took notes, and later worked out the mathematical presentation in the written version of the lectures. I also attended many of the sophomore lectures, to which I had mixed reactions. If you read Vol. II of *The Feynman Lectures on Physics* (5), you will find two distinct threads: The first is a perfectly standard treatment, like that in any introductory book on the subject. In his preface, Feynman says of this material:

> In the second year I was not so satisfied. In the first part of the course, dealing with electricity and magnetism, I couldn't think of any really unique or different way of doing it.

There is a second thread, however, of true vintage Feynman—the occasional lectures where he waxed eloquent about the vector potential. Section 15-5 contains a delightful discussion about what a field is and what makes one field more "real" than another.

> What we mean here by a "real" field is this: a real field is a mathematical function we use for avoiding the idea of action at a distance . . . A "real" field is then a set of numbers we specify in such a way that what happens *at a point* depends only on the numbers *at that point* . . . In our sense then, the **A**-field is "real" . . . **E** and **B** are slowly disappearing from the modern expression of physical laws; they are being replaced by **A** and ϕ.

In Chapter 25, he develops the equations of electrodynamics in four-vector form—the approach that I have adopted in this monograph. I can remember feeling very angry with Feynman when I sat in on this particular lecture. Why hadn't he started this way in the first place, and saved us all the mess of a **B** field, which, as

he told us himself, was not real anyway? When I asked him about it, he said something vague, like:

> There are a bunch of classical interactions that you can't get at in any simple way without Maxwell's equations. You need the $v \times \mathbf{B}$ term.

I don't remember his exact words here, only the gist of the discussion. Sure enough, when Vol. II of the lectures was published, the equation $F = q(\mathbf{E} + v \times \mathbf{B})$ in table 15-1 appears in the column labeled "True Always." The equation *is* true for the toy electric motor he shows in Fig. 16-1. It is *not* true in general. For a real electric motor, the **B** field is concentrated in the iron, rather than in the copper in which the current is flowing, and the equation gives the wrong answer by a factor of more than 100! That factor is due to the failure of **B** to be "real," precisely in Feynman's sense. Somehow he had separated science into two worlds: quantum and classical. For him, the vector potential was primary in the quantum world, whereas **E** and **B** were necessary for the classical world. These two worlds had not yet come together.

I was an active researcher in solid-state physics at that time, and I used the quantum nature of electrons in solids every day. Electrodynamics deals with how electrons interact with other electrons. The classical interactions Feynman was talking about were between electrons in metals, in which the density of electrons is so high that quantum interaction is by far the dominant effect. If we know how the vector potential comes into the phase of the electron wave function, and if the electron wave function dominates the behavior of metals, then why can't we do all of electromagnetic theory that way? Why didn't he use his knowledge of quantum electrodynamics to "take the vector and scalar potentials as fundamental quantities in a set of equations that replace the Maxwell equations," as he himself had said? I was mystified; his cryptic answer prodded me to start working on the problem. But every time I thought I had an approach, I got stuck.

Bill Fairbank from Stanford had given a seminar on quantized flux in superconducting rings that impressed me very much. The solid-state physics club was much smaller in those days, and, because I was working in electron tunneling, I was close to the people working on tunneling between superconductors. Their results were breaking in just this same time frame, and Feynman

gave a lecture about this topic to the sophomores; it appears as Chapter 21 in Vol. III of *The Feynman Lectures on Physics* (7). As I listened to that lecture, my thoughts finally clicked: This is how we can make the connection! A superconductor is a quantum system on a classical scale, and that fact allows us to carry out Feynman's grand scheme. But I couldn't get this approach to go all the way through at that time, so it just sat in the back of my mind all these years, vaguely tickling me.

Meanwhile my work on tunneling was being recognized, and Gordon Moore (then at Fairchild) asked me whether tunneling would be a major limitation on how small we could make transistors in an integrated circuit. That question took me on a detour that was to last nearly 30 years, but it also led me into another collaboration with Feynman, this time on the subject of computation. Here's how it happened: In 1968, I was invited to give a talk at a workshop on semiconductor devices at Lake of the Ozarks. In those days, you could get everyone who was doing cutting-edge work into one room, so the workshops were where all the action was. I had been thinking about Gordon Moore's question, and decided to make it the subject of my talk. As I prepared for this event, I began to have serious doubts about my sanity. My calculations were telling me that, contrary to all the current lore in the field, we could scale down the technology such that *everything got better*. The circuits got more complex, they ran faster, and they took less power—WOW! That's a violation of Murphy's law that won't quit! But the more I looked at the problem, the more I was convinced that the result was correct, so I went ahead and gave the talk—to hell with Murphy! That talk provoked considerable debate, and at the time most people didn't believe the result. But by the time the next workshop rolled around, a number of other groups had worked through the problem for themselves, and we were pretty much all in agreement. The consequences of this result for modern information technology have, of course, been staggering.

Back in 1959, Feynman gave a lecture entitled "There's Plenty of Room at the Bottom," in which he discussed how much smaller things can be made than we ordinarily imagine. That talk had made a big impression on me; I thought about it often, and it would sometimes come up in our discussions on the tunneling work. When I told him about the scaling law for electronic devices, Feynman got jazzed. He came to my seminars on the sub-

ject, and always raised a storm of good questions and comments. I was working with a graduate student, Bruce Hoeneisen; by 1971, we had worked out the details of how transistors would look and work when they are a factor of 100 smaller in linear dimension than the limits set by the prevailing orthodoxy. Recently, I had occasion to revisit these questions, and to review the history of what has happened in the industry since those papers were published. I plotted our 1971 predictions alongside the real data; they have held up extremely well over 25 years, representing a factor of several thousand in density of integrated circuit components (8).

Because of the scaling work, I became completely absorbed with how the exponential increase in complexity of integrated circuits would change the way that we think about computing. The viewpoint of the computer industry at the time was an outgrowth of the industrial revolution; it was based on what was then called "the economy of scale." The thinking went this way: A 1000-horsepower engine costs only four times as much as a 100-horsepower engine. Therefore, the cost per horsepower becomes less as the engine is made larger. It is more cost effective to make a few large power plants than to make many small ones. Efficiency considerations favor the concentration of technology in a few large installations. The same must be true of computing. One company, IBM, was particularly successful following this strategy. The "Computing Center" was the order of the day—a central concentration of huge machines, with some bureaucrat "in charge" and plenty of people around to protect the machines from anyone who might want to use them. This model went well with the bureaucratic mindset of the time—a mindset that has not totally died out even today.

But as I looked at the physics of the emerging technology, it didn't work that way at all. The time required to move data is set by the velocity of light and related electromagnetic considerations, so it is far more effective to put whatever computing is required *where the data are located*. Efficiency considerations thus favor the distribution of technology, rather than the concentration of technology. The economics of information technology are the reverse of those of mechanical technology.

I gave numerous talks on this topic, but, at that time, what I had to say was contrary to what the industry wanted to hear. The story is best told in George Gilder's book, *Microcosm* (9).

Feynman had started this line of thought already in his 1959 lecture, and we had a strong agreement on the general direction things were headed. He often came to my group meetings, and we had lively discussions on how to build a machine that would recognize fingerprints, how to organize many thousand little computers so they would be more efficient than one big computer, etc. Those discussions inevitably led us to wonder about the most distributed computer of all: the human brain. Years before, Feynman had dabbled in biology, and I had worked with Max Delbrück on the physics of the nerve membrane, so I knew a bit about nervous tissue. John Hopfield had delved much deeper than either Feynman or I had; and, by 1982, he had a simple model—a caricature of how computation might occur in the brain.

The three of us decided to offer a course jointly, called "Physics of Computation." The first year, Feynman was battling a bout with cancer, so John and I had to go it alone. We alternated lectures, looking at the topic from markedly different points of view. Once Feynman rejoined us, we had even more fun—three totally different streams of consciousness in one course. The three of us had a blast, and learned a lot from one another, but many of the students were completely mystified. After the third year, we decided, in deference to the students, that there was enough material for three courses, each with a more-unified theme. Hopfield taught "Neural Networks," Feynman taught "Quantum Computing," which ended up in the first volume of *Feynman Lectures on Computation* (10), and I taught "Neuromorphic Systems," which ended up in my book on the subject (11).

There is a vast mythology about Feynman, much of which is misleading. He had a sensitive side that he didn't show often. Over lunch one time, I told him how much he had meant to me in my student years, and how I would not have gone into science had it not been for his influence. He looked embarrassed, and abruptly changed the subject; but he heard me, and that was what was important. In those days, physics was an openly combative subject—the one who blinked first lost the argument. Bohr had won his debate with Einstein that way, and the entire field adopted the style. Feynman learned the game well—he never blinked. For this reason, he would never tell anyone when he was working on something, but instead would spring it, preferably in front of an

audience, after he had it all worked out. The only way that you could tell what he cared about was to notice what topics made him mad when you brought them up.

If Feynman was stuck about something, he had a wonderful way of throwing up a smoke screen; we used to call it "proof by intimidation." There is a good example in Vol. II of the *Lectures on Physics* (5), directly related to collective electrodynamics. Section 17-8 contains the following comment:

> we would expect that corresponding to the mechanical momentum $p = mv$, whose rate of change is the applied force, there should be an analogous quantity equal to $\mathcal{L}\mathbf{I}$, whose rate of change is $\mathcal{V}$. We have no right, of course, to say that $\mathcal{L}\mathbf{I}$ is the real momentum of the circuit; in fact it isn't. The whole circuit may be standing still and have no momentum.

Now, this passage does not mean that Feynman was ignorant of the fact that the electrical current $\mathbf{I}$ is made up of moving electrons, that these moving electrons have momentum, and that the momentum of the electrons does not correspond to the whole circuit moving in space. But the relations are not as simple as we might expect, and they do not correspond in the most direct way to our expectations from classical mechanics. It is exactly this point that prevented me, over all these years, from seeing how to do electrodynamics without Maxwell's equations. Feynman was perfectly aware that this was a sticking point, and he made sure that nobody asked any questions about it. There is a related comment in Vol. III of the *Lectures on Physics* (7), Section 21-3:

> It looks as though we have two suggestions for relations of velocity to momentum . . . The two possibilities differ by the vector potential. One of them . . . is the momentum obtained by multiplying the mass by velocity. The other is a more mathematical, more abstract momentum

When Feynman said that a concept was "more mathematical" or "more abstract," he was not paying it a compliment! He had no use for theory devoid of physical content. In the *Lectures on Gravitation*, he says:

> If there is something very slightly wrong in our definition of the theories, then the full mathematical rigor may convert these errors into ridiculous conclusions.

We called that "carrying rigor to the point of *rigor mortis*." At another point, he is even more explicit:

> it is the facts that matter, and not the proofs. Physics can progress without the proofs, but we can't go on without the facts . . . if the facts are right, then the proofs are a matter of playing around with the algebra correctly.

He opened a seminar one time with the statement, "Einstein was a giant." A hush fell over the audience. We all sat, expectantly, waiting for him to elaborate. Finally, he continued, "His head was in the clouds, but his feet were on the ground." We all chuckled, and again we waited. After another long silence, he concluded, "But those of us who are not that tall have to choose!" Amid the laughter, you could see that not only a good joke, but also a deep point, had been made.

Experiments are the ground on which physics must keep its feet—as Feynman knew well. When any of us had a new result, he was all ears. He would talk about it, ask questions, brainstorm. That was the only situation in which I ever personally interacted with him without his combative behavior getting in the way. Down deep, he always wanted to do experiments himself. A hilarious account of how he was "cured" of this craving appears in *Surely You're Joking, Mr. Feynman*. In the end, he had his wish. In 1986, he was asked to join the Rodgers commission to investigate the Challenger disaster. After talking to the technical people, who knew perfectly well what the problem was and had tried to postpone the launch, he was able to devise an experiment that he carried out on national, prime-time TV. In true Feynman style, he sprang it full-blown, with no warning! In his personal appendix to the commission report, he concluded, "For a successful technology, reality must take precedence over public relations, for nature cannot be fooled." The day after the report was released was Caltech's graduation. As we marched together in the faculty procession, "Did you see the headline this morning?" he asked. "No," I replied. "What did it say?" "It said **FEYNMAN ISSUES REPORT**." He paused, and then continued with great glee. "Not **Caltech Professor Issues Report**, not **Commission Member Issues Report**, but **FEYNMAN ISSUES REPORT**." He was a household word, known and revered by all people everywhere who loved truth. His own public relations were all about reality, and were, therefore, okay.

In 1987, one year later, his cancer came back with a vengeance, and he died in February, 1988. Al Hibbs, a former student, colleague, and friend of Feynman's, organized a wake in grand style: bongo drums, news clips, interviews, and testimonials. It was deeply moving—we celebrated the life of this man who had, over the years, come to symbolize not just the spirit of Caltech, but the spirit of science itself. This man had engendered the most intense emotions I have ever felt—love, hate, admiration, anger, jealousy, and, above all, a longing to share and an intense frustration that he would not. As I walked away from Feynman's wake, I felt intensely alone. He was the man who had taught me not only what physics is, but also what science is all about, what it means to *really understand*. He was the only person with whom I could have talked about doing electromagnetism without Maxwell's equations—using the quantum nature of matter as the sole basis. He was the only one who would have understood why it was important. He was the only one who could have related to this dream that I had carried for 25 years. This dream came directly from Feynman, from what he said and from what he scrupulously avoided saying, from the crystal-clear insights he had, and from the topics that had made him mad when I brought them up. But now he was gone. I would have to go it alone. I sobbed myself to sleep that night, but I never shared those feelings with anyone. I learned *that* from him, too.

In 1994, I was invited to give the keynote talk at the Physics of Computation conference. That invitation gave me the kickstart I needed to get going. By the next year, I had made enough progress to ask Caltech for a year relief from teaching so I could concentrate on the new research. In June 1997, the six graduate students working in my lab all received their doctoral degrees, and, for the first time since I joined the faculty, I was a free man. I finished the basic paper on Collective Electrodynamics (12), an expanded version of which appears in the present monograph as Part 1 (p. 9). The memorial volume, *Feynman and Computation* (13), contains reprints of this paper and the scaling paper mentioned previously, along with an earlier version of this preface entitled *Feynman as a Colleague*.

By the end of 1998, I had developed the subject to the point where most of the standard problems in electromagnetic theory could be understood much more easily using this approach than

by using standard textbook methods. Early in 1999, I was notified that I had been chosen to receive the prestigious Lemelson–MIT award for innovation. The ceremony celebrating this award was a gala event at which MIT chairman Alex d'Arbeloff stressed the importance of preparing the students of today to be the innovators of tomorrow. He expressed concern that neither our scientific establishment nor our educational institutions have developed approaches that are adequate to meet this challenge. At that moment, I realized that the work I had been doing was an example of precisely what was needed—the simplification and unification of knowledge. The remarks I made upon receiving the award appear in the foreword to this monograph.

In the end, science is all in how you look at things. Collective Electrodynamics is a way of looking at the way that electrons interact. It is a much simpler way than Maxwell's, because it is based on experiments that tell us about the electrons directly. Maxwell had no access to these experiments. The sticking point I mentioned earlier is resolved in this treatment, in a way that Feynman would have liked. This monograph is dedicated to him in the most sincere way I know: It opens with my favorite quotation, the quotation that defines, for me, what science is all about. In his epilogue, Feynman tells us his true motivation for giving the *Lectures on Physics*:

> I wanted most to give you some appreciation of the wonderful world, and the physicist's way of looking at it, which, I believe, is a major part of the true culture of modern times . . . Perhaps you will not only have some appreciation of this culture; it is even possible that you may want to join in the greatest adventure that the human mind has ever begun.

You succeeded, Dick, and we have—Thanks!

This page intentionally left blank

Acknowledgments

I am indebted to many colleagues who have discussed, argued, and shared many insights with me during the past eight years. Dick Lyon and Sanjoy Mahajan have read and marked more versions of the manuscript than any of us can remember and have provided many of the historic references. Bill Bridges has provided many insights and suggestions from the point of view of electrical engineering practice, as has Bill McClellan. Discussions with David Feinstein started me on the most recent active phase of the endeavor. David Goodstein, Martin Perl, Yaser Abu-Mostafa, Demetri Psaltis, Paul Klein, Paul Tuinenga, Dick Neville, Glen Keller, Mike Godfrey, Jim Eisenstein, Bill Wagner, Christoph Von Der Malsburg, Terry Sejnowski, Rahul Sarpeshkar, Tobi Delbrück, Nick Mascarenhas, Al Barr, Axel Sherer, and others have read and responded to various parts of the manuscript at various stages. Sunit Mahajan and I shared many long hours of quantum-Hall measurements. Hideo Mabuchi and Jeff Kimble shared their splendid experiments and insights about atoms in cavities. Cal Jackson, who has provided the TEXpertise throughout the effort, did the final design, page layout, and typesetting. Bob Turring provided the figures. Lyn Dupré and Dian De Sha edited the manuscript at many stages. Donna Fox provided encouragement as well as administrative support. John Seinfeld, chairman of the division of engineering and applied science at Caltech, has been wonderfully supportive over the past few years. Barbara Smith has been a boundless source of energy, wisdom, and caring.

This page intentionally left blank

Introduction

Foundations of Physics

But the real *glory* of science is that *we can find a way of thinking* such that the law is *evident*.

—R.P. Feynman[1]

It is my firm belief that the last seven decades of the twentieth century will be characterized in history as the dark ages of theoretical physics. Early in this period, a line was drawn between **Classical Physics**, containing mechanics, electricity, and magnetism, and **Modern Physics**, containing relativity and quantum theory. The connection between the two domains was supposed to be Bohr's **Correspondence Principle**: The behavior of a quantum system must approach that of a classical mechanical system in the limit of large quantum numbers. It is the purpose of this monograph to redefine that boundary, and to state a more-correct correspondence principle. As a quantum system contains more and more elements, it exhibits **Collective** behaviors that differ more and more widely from the behaviors of a mechanical system. In the limit of a large number of elements, these behaviors correspond to electromagnetic phenomena. Thus, physics can indeed be divided into two disciplines: the first preoccupied with the behavior of incoherent systems, and the second concerned with coherent quantum phenomena. In what follows, I show that electromagnetism falls squarely in the second category.

Modern science began with mechanics; in many ways, we are all still captive to mechanical ideas. Newton's success in deriving the planetary orbits from the $1/r^2$ law of gravitation became the paradigm. To Niels Bohr, 250 years later, the atom was still a miniature solar system, with a nucleus as the sun and electrons as the planets. But there were problems: An accelerating charge would radiate energy. So the electrons, so busily orbiting the nucleus, would soon radiate away the energy that kept them in orbit, and finally collapse into the nucleus. It became

[1] Page 26-3 of Ref. 14.

clear that the mechanics of the atom are, in some way, totally different from the mechanics of ordinary mechanical systems. Out of the struggle to understand the atom came Quantum Mechanics, a new orthodoxy that replaced the mechanical orthodoxy of the late nineteenth century.[2] Bohr gathered the early contributors into a clan in Copenhagen, encouraged everyone in the belief that they were developing the ultimate theory of nature, and argued vigorously against any opposing views. Quantum Mechanics was enormously successful in calculating the spectral lines of the hydrogen atom, and has become the mantra of physics as we know it today. The story of the quantum revolution is repeated in the popular literature, in science education, and in specialized physics courses. One must solemnly affirm one's allegiance to the Quantum God before one may be admitted to the physics clan.

All this mysticism seems out of place in a discipline as objective as physics, so there must be more to the story. Einstein and Schrödinger both made central contributions to the emerging quantum discipline, but never bought into the orthodoxy of Bohr's Copenhagen clan. There was no disagreement about the numbers the equations gave for the hydrogen energy levels; the dissent was over the conceptual interpretation of the theory. Bohr insisted that the laws of physics, at the most fundamental level, are *statistical* in nature. Einstein and Schrödinger believed in a continuous space–time, that the statistical nature of the experimental results was a result of our lack of knowledge of the state of the system, and that the underlying physical laws can be formulated in a continuous manner. Schrödinger put it this way:

> I no longer regard this [statistical] interpretation as a finally satisfactory one, even if it proves useful in practice. To me it seems to mean a renunciation, much too fundamental in principle, of all attempt to understand the individual process.

The disagreement culminated in a 1927 debate between Bohr and Einstein, refereed by Eherenfest. Bohr was a great debater, and won the contest hands down. A rematch was staged in 1930, and

[2]An excellent exposition of the early contributions is given by Mehra and Rechenberg in their multi-volume *Historical Development of Quantum Theory* (15), and by Jammer in *The Conceptual Development of Quantum Mechanics* (16).

Bohr won again.[3] The word was circulated that Einstein had become senile, and could no longer cope with new ideas. Years later, at the celebration of Einstein's seventieth birthday, several members of the clan used the occasion to roast Einstein once again for his dissent. His comments in reply[4] shed more light on the controversy than all other commentary combined:

> I do not believe that this fundamental concept will provide a useful basis for the whole of physics.
>
> I am, in fact, firmly convinced that the essentially statistical character of contemporary quantum theory is solely to be ascribed to the fact that this [theory] operates with an incomplete description of physical systems.
>
> One arrives at very implausible theoretical conceptions, if one attempts to maintain the thesis that the statistical quantum theory is in principle capable of producing a complete description of an individual physical system.
>
> Roughly stated the conclusion is this: Within the framework of statistical quantum theory there is no such thing as a complete description of the individual system. More cautiously it might be put as follows: The attempt to conceive the quantum-theoretical description as the complete description of the individual systems leads to unnatural theoretical interpretations, which become immediately unnecessary if one accepts the interpretation that the description refers to ensembles of systems and not to individual systems. In that case the whole "egg-walking" performed in order to avoid the "physically real" becomes superfluous. There exists, however, a simple psychological reason for the fact that this most nearly obvious interpretation is being shunned. For if the statistical quantum theory does not pretend to describe the individual system (and

[3]A thoughtful analysis of these debates and developments that followed from them is given by Whitaker (17), and also by Fine (18). Mara Beller (19) gives a good discussion of the philosophical excesses engendered by the triumph of the Copenhagen clan.

[4]The lectures given on this occasion, and Einstein's reply, appear in full in Ref. 20. A thoughtful discussion of the statistical interpretation is given by Ballentine (21), accompanied by an editorial that illustrates the level of controversy that existed in 1970. Davies and Brown published interviews with leading physicsts in *The Ghost in the Atom* (22). These notable figures held views that spanned the entire space of possible opinion, yet many firmly believed that theirs was the only view that any thinking person could possibly hold.

> its development in time) completely, it appears unavoidable to look elsewhere for a complete description of the individual system; in doing so it would be clear from the very beginning that the elements of such a description are not contained within the conceptual scheme of the statistical quantum theory. With this one would admit that, in principle, this scheme could not serve as the basis of theoretical physics. Assuming the success of efforts to accomplish a complete physical description, the statistical quantum theory would, within the framework of future physics, take an approximately analogous position to the statistical mechanics within the framework of classical mechanics. I am rather firmly convinced that the development of theoretical physics will be of this type; but the path will be lengthy and difficult.
>
> If it should be possible to move forward to a complete description, it is likely that the laws would represent relations among all the conceptual elements of this description which, per se, have nothing to do with statistics.

The experiments upon which the conceptual foundations of quantum mechanics were based are extremely crude by modern standards. At that time, the detectors available were Geiger counters, cloud chambers, and photographic film; each of these had a high degree of randomness built into it, and by its very nature could register only statistical results. The atomic sources available were similarly constrained: They consisted of large ensembles of atoms with no mechanism for achieving even the slightest hint of phase coherence in their emitted radiation. It is understandable that the experiments that could be imagined would all be of a statistical sort. In the Schrödinger centenary volume (23), Jon Dorling makes a convincing case for reviving the continuous space–time view of Einstein and Schrödinger, arguing that physics would be in less trouble with their approach than it is now. Greenstein and Zajonc (24) give a delightful modern perspective on conceptual issues raised by the statistical quantum theory, and Gribben (25) presents a thoughtful popular review. It seems to have escaped notice in such discussions that working-level physicists use a totally distinct interpretation[5] to analyze collective systems: The coher-

[5] The formalism involved, called **Second Quantization**, is derived from standard quantum mechanics using arguments from quantum electrodynamics. A nice treatment of the technique can be found in the appendix to Pines and Nozières (26).

ent state is taken as the starting point, and the statistics are put where Einstein would have placed them—in the random thermal excitations out of the perfectly coherent state. The success of this method can be viewed as a vindication of Einstein's conjecture.

Despite the muddle and fuss over theory, the past seventy years have been an age of enlightenment on the experimental front. Of the astounding experimental discoveries made during that period, a number are of particular importance for the present discussion:

1933 Persistent Current in Superconducting Ring
1933 Expulsion of Magnetic Field by Superconductor
1954 Maser
1960 Atomic Laser
1961 Quantized Flux in Superconducting Ring
1962 Semiconductor Laser
1964 Superconducting Quantum Interference Device
1980 Integer Quantum Hall Effect
1981 Fractional Quantum Hall Effect
1996 Bose–Einstein Condensate

Each of these discoveries has made a profound difference in the way we view the physical world. Each represents a coherent, collective state of matter. Each embodies a fundamental quantum principle, which is exhibited on a macroscopic scale. Each has been investigated exclusively by electromagnetic means.

Hindsight is a wonderful thing: We can start at a different place, go at the subject in a completely different way, and build a much clearer and simpler conceptual base. The difficult step with hindsight is to go back far enough to get a really fresh start. I have found it necessary to start not just before the quantum theory, but before electromagnetic theory as it has come to be taught. Collective electrodynamics is the result of asking the question: If we could have looked forward from the mid 1800s with these experimental facts in our hands, would we have built the theory we have today? I have concluded that we would not. This approach does not produce a new theory in the sense that it contains startling new equations, for it does not. The results it derives for standard electromagnetic problems are identical to those found in any text on the subject. To be sure, many embarrassing questions that arise in the standard treatments of electromagnetism have natural answers in the collective context, as I have noted in the preface.

Collective electrodynamics is important in a completely different way, for it shows us that quantities we usually think of as being very different are, in fact, the same, that electromagnetic phenomena are simple and direct manifestations of quantum phenomena. This is the first step toward reformulating quantum concepts in a clear and comprehensible manner, but it is only a first step, and does not claim to be more.

So, how do we get a foothold to start our new approach? It helps to identify some of the confusions that exist in the standard quantum picture. The first confusion, the **Correspondence Principle,** states that the behavior of a quantum system must approach that of a classical mechanical system in the limit of large quantum numbers. Ernst Mach wrote (p. 596 in Ref. 27):

> The view that makes mechanics the basis of the remaining branches of physics, and explains all physical phenomena by mechanical ideas, is in our judgment a prejudice . . . The mechanical theory of nature, is, undoubtedly, in a historical view, both intelligible and pardonable; and it may also, for a time, have been of much value. But, upon the whole, it is an artificial conception.

Classical mechanics is an inappropriate starting point for physics because it is not fundamental; rather, it is the limit of an incoherent aggregation of an enormous number of quantum elements. Feynman wrote (p. 15-8 in Ref. 5):

> There are many changes in concepts that are important when we go from classical to quantum mechanics . . . Instead of forces, we deal with the way interactions change the wavelengths of waves.

To make contact with the fundamental nature of matter, we must work in a coherent context in which the underlying quantum reality has not been corrupted by an incoherent averaging process. Traditional treatments of quantum mechanics universally confuse results that follow from the wave nature of matter with those that follow from the statistical nature of the experiment. In the usual picture, these aspects are inextricably intertwined. Einstein himself had a massive case of this confusion, and it cost him the debate with Bohr. Had he stuck to his hunch that the fundamental laws are continuous, he would have fared better; but to do that, he would have needed a model quantum system in which statistics

plays a vanishingly small role. At that time, no such system was known. Today, we have many such systems, as mentioned above. Of these, none is more accessible than the superconductor itself; it is a quantum system on a grand scale, and, all by itself, provides us strikingly direct access to a near-perfect coherent state. It manifests all the quantum phenomena associated with the wave nature of matter, without the confusion about statistics. Its behavior is, in many ways, simpler than that of an isolated single particle. Of course, the details of how the superconducting state arises in a real solid are complicated; we will not even approach them in this introductory treatment. But, given a superconductor, we can devise a system that we can view as having only one degree of freedom. Its properties are dominated by *known* and *controllable* interactions within the collective ensemble. The dominant interaction is **electromagnetic,** because it couples to the charges of the electrons; and **collective,** because the properties of each electron depend on the state of the entire ensemble. Nowhere in natural phenomena do the basic laws of physics manifest themselves with more crystalline clarity.

This page intentionally left blank

Part 1

Magnetic Interaction of Steady Currents

I feel that it is a delusion to think of the electrons and the fields as two physically different, independent entities. Since neither can exist without the other, there is only *one* reality to be described, which happens to have two different aspects; and the theory ought to recognize this from the start instead of doing things twice.

—Albert Einstein[1]

In atomic theory we have fields and we have particles. The fields and the particles are not two different things. They are two ways of describing the same thing—two different points of view.

—P.A.M. Dirac[2]

1.1 Model System

Our model system, shown in Fig. 1.1, is a loop of superconducting wire—the two ends of the loop are collocated in space and either shorted (a), or insulated (b), depending on the experimental situation. Experimentally, the voltage V between the two ends of the loop in Fig. 1.1b is related to the current I flowing through the loop by

$$L\,I = \int V dt = \Phi \tag{1.1}$$

Two quantities are defined by this relationship: Φ, called the **magnetic flux,**[3] and L, called the **inductance**, which depends on the dimensions of the loop.

[1] In Ref. 28, E.T. Jaynes puts forth a delightful argument for the approach I have adopted in this monograph. The Einstein quotation appears on page 383 of this reference.

[2] This quotation appears on page 1 of Ref. 29.

[3] This definition is independent of the shape of the loop, and applies even to coils with multiple turns. For multiturn coils, what we call the *flux* is commonly referred to as the *total flux linkage*.

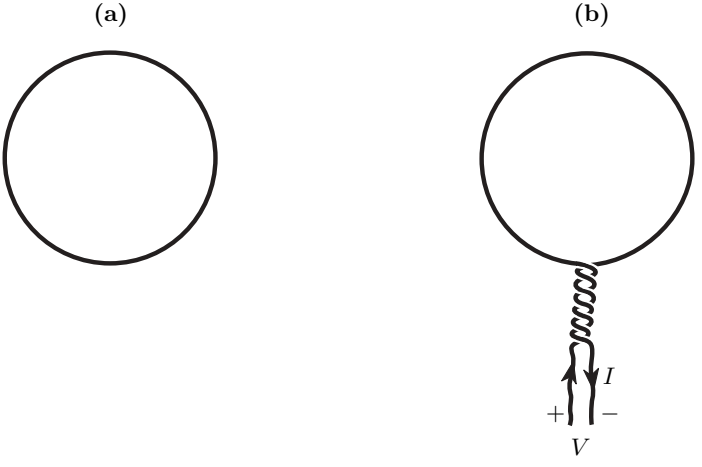


Figure 1.1 Loops of superconducting wire: (a) shorted and (b) with externally applied current.

Current is the flow of charge: $I = dQ/dt$. Each increment of charge dQ carries an energy increment $dW = V\,dQ$ into the loop as it enters.[4] The total energy W stored in the loop is thus

$$\begin{aligned} W &= \int V\,dQ = \int V I\,dt \\ &= L\int \frac{dI}{dt}\,I\,dt = L\int I\,dI = \frac{1}{2}\,L\,I^2 \end{aligned} \tag{1.2}$$

If we reduce the voltage to zero by, for example, connecting the two ends of the loop to form a closed superconducting path, as shown in Fig. 1.1a, the current I will continue to flow *indefinitely:* a **persistent current**. If we open the loop and allow it to do work on an external circuit, we can recover *all of the energy* W.

If we examine closely the values of currents under a variety of conditions, we find the full continuum of values for the quantities I, V, and Φ, except in the case of persistent currents, where only certain, *discrete values* occur for any given loop (30, 31). By experimenting with loops of different dimensions, we find the condition

[4]We use this relation to define the voltage V.

that describes the values that occur experimentally:

$$\Phi = \int V dt = n\,\Phi_0 \tag{1.3}$$

Here, n is any integer, and $\Phi_0 = 2.06783461 \times 10^{-15}$ volt-second is the **flux quantum** or **fluxoid**; its value is accurate to a few parts in 10^9, independent of the detailed size, shape, or composition of the superconductor forming the loop. We also find experimentally that a rather large energy—sufficient to disrupt the superconducting state entirely—is required to change the value of n.

The more we reflect on Eq. 1.3, the more remarkable the result appears. The quantities involved are the voltage and the magnetic flux. These quantities are integrals of the quantities E and B that appear in Maxwell's equations, and are therefore usually associated with the electromagnetic field. Experimentally, we know that they can take on a continuum of values—except under special conditions when the arrangement of matter in the vicinity causes the flux to take on precisely quantized values. In Maxwell's theory, E and B represent the state of strain in a mechanical medium (the ether) induced by electric charge. Einstein had a markedly different view, as illustrated by the opening quotation. At the most-fundamental level, the essence of quantum mechanics lies in the wave nature of matter. Einstein's view suggests that electromagnetic variables are related to the wave properties of the electrons. Quantization is a familiar phenomenon in systems where the boundary conditions give rise to standing waves. The quantization of flux (Eq. 1.3) is a direct manifestation of the wave nature of matter, expressed in electromagnetic variables.

1.2 Matter

To most nonspecialists, quantum mechanics is a baffling mixture of waves, statistics, and arbitrary rules, ossified in a matrix of impenetrable formalism. By using a superconductor, we can avoid the statistics, the rules, and the formalism, and work directly with the waves. The wave concept, accessible to both intuition and common sense, gives us "a way of thinking such that the law is evident." Electrons in a superconductor are described by a **wave function** that has an amplitude and a phase. The earliest treatment of the wave nature of matter is the 1923 wave mechanics of

de Broglie. He applied the 1905 Einstein postulate ($W = \hbar\omega$) to the energy W of an electron wave, and identified the momentum $\vec{p}$ of an electron with the propagation vector of the wave: $\vec{p} = \hbar\vec{k}$. Planck's constant h and its radian equivalent $\hbar = h/2\pi$ are necessary for merely historical reasons—when our standard units were defined, it was not yet known that energy and frequency are the same quantity.

The Einstein–de Broglie relations apply to the collective electrons in a superconductor. The dynamics of the system can be derived from the **dispersion relation** (32) between ω and $\vec{k}$. Both ω and $\vec{k}$ are properties of the phase of the wave function, and do not involve the amplitude—which, in collective systems, is usually determined by some normalization condition. In a superconductor, the constraint of charge neutrality is such a condition. A more-detailed description of the wave function of a large ensemble of electrons is given in the Appendix (p. 115).

The wave function must be *continuous* in space; at any given time, we can follow the phase along a path from one end of the loop to the other: The number of radians by which the phase advances as we traverse the path is the **phase accumulation** φ around the loop.[5] If the phase at one end of the loop changes relative to that at the other end, that change must be reflected in the total phase accumulation around the loop. The **frequency** ω of the wave function at any point in space is the rate at which the phase advances per unit of time. If the frequency at one end of the loop (ω_1) is the same as that at the other end (ω_2), the phase difference between the two ends will remain constant, and the phase accumulation will not change with time. If the frequency at one end of the loop is higher than that at the other, the phase accumulation will increase with time, and that change must be reflected in the rate at which phase accumulates with the distance l along the path. The rate at which phase around the loop accumulates with time is the difference in frequency between the two ends. The rate at which phase accumulates with distance l is the component of the propagation vector $\vec{k}$ in the direction $\vec{dl}$ along the path. Thus, the total phase accumulated around the

[5]The reference for phase angle is, of course, arbitrary. The phase accumulation along a path uses a single reference, and thus has none of the arbitrary nature of the phase at a single point.

loop is

$$\varphi = \int (\omega_1 - \omega_2)\, dt = \oint \vec{k} \cdot \vec{dl} \tag{1.4}$$

We can understand quantization as an expression of the single-valued nature of the phase of the wave function. When the two ends of the loop are connected to an external circuit, the two phases can evolve independently. When the ends are connected to each other, however, the two phases must match up. But the phase is a quantity that has a cyclic nature—*matching up* means being equal modulo 2π. Thus, for a wave that is confined to a closed loop and has a single-valued, continuous phase, the integral of Eq. 1.4 must be $n\,2\pi$, where n is an integer. The large energy required to change n is evidence that the phase constraint is a strong one—as long as the superconducting state stays intact, the wave function remains intact, as well.

These relations tell us that the magnetic flux and the propagation vector will be quantized for a given loop; they do not tell us how the frequency ω in Eq. 1.4 is related to the voltage V in Eq. 1.1. To make this connection, we must introduce one additional assumption: The collective electron system represented by the wave function is made up of elemental charges of magnitude q_0. The voltage V is the difference in the electrostatic potential $\mathcal{V}$ at the two ends of the loop. By the Einstein relation, the energy $q_0\mathcal{V}$ of an elemental charge at potential $\mathcal{V}$ corresponds to a frequency $\omega = q_0\mathcal{V}/\hbar$, which is the time rate of change of the wave-function phase.

1.3 Electrodynamics

Electrodynamics is the interaction of matter via the electromagnetic field. We can formulate our first relation between the electromagnetic quantities V and Φ and the phase accumulation φ of the wave function by comparing Eq. 1.1 with Eq. 1.4. The voltage V is the difference in potential $\mathcal{V}$ between the two ends of the loop.

$$\begin{aligned}\varphi &= \int (\omega_1 - \omega_2)\, dt \\ &= \frac{q_0}{\hbar} \int (\mathcal{V}_1 - \mathcal{V}_2)\, dt = \frac{q_0}{\hbar} \int V dt\end{aligned} \tag{1.5}$$

$$\varphi = \frac{q_0}{\hbar}\, n\, \Phi_0 = n\,(2\pi)$$

From Eq. 1.5, we conclude that $\Phi_0 = h/q_0$.

When we substitute into Eq. 1.5 the measured value of Φ_0 and the known value of h, we obtain for q_0 a value that is exactly twice the charge q_e of the free electron. The usual explanation for this somewhat surprising result is that each state in the superconductor is occupied by a *pair* of electrons, rather than by an individual electron, so the elemental charge q_0 should be $2q_e$, not q_e. Alternatively, we note that the electron is a spin one-half entity, meaning that its wave function matches up with itself modulo π rather than modulo 2π as we had naively assumed. Eq. 1.5 then becomes

$$\varphi = \frac{q_e}{\hbar}\, n\, \Phi_0 = n\,\pi \qquad \Rightarrow \qquad \Phi_0 = \frac{h}{2q_e} \tag{1.6}$$

which is equivalent to choosing $q_0 = 2q_e$. Each of these descriptions expresses a certain aspect of the nature of the collective electron system.

We have established the correspondence between the potential $\mathcal{V}$ and the frequency ω—the time integral of each of these equivalent quantities in a closed loop is quantized. The line integral of the propagation vector $\vec{k}$ around a closed loop also is quantized. We would therefore suspect the existence of a corresponding electromagnetic quantity, whose line integral is the magnetic flux Φ. That quantity is called the **vector potential** $\vec{A}$. The general relations among these quantities, whether or not the loop is closed, are

$$\left.\begin{array}{ll} \text{Phase} & \varphi = \displaystyle\int (\omega_1 - \omega_2)\, dt = \oint \vec{k}\cdot\vec{dl} \\ \text{Flux} & \Phi = \displaystyle\int V\, dt = \oint \vec{A}\cdot\vec{dl} \end{array}\right\} \quad \Phi = \frac{\hbar}{q_0}\,\varphi \tag{1.7}$$

Eq. 1.7 expresses the first set of fundamental relations of collective electrodynamics.[6]

[6]The vector potential was introduced by Maxwell in Art. 405 of his *Treatise on Electricity and Magnetism* (33). The connection between the vector potential and the phase of the electron wave function was introduced by Schrödinger

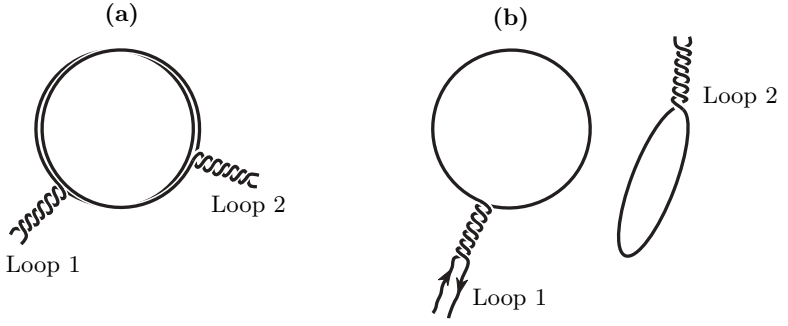


Figure 1.2 Coupled loops: (a) closely coupled; (b) separated in space

1.4 Coupling

Up to this point, we have tentatively identified the phase accumulation and the magnetic flux as two representations of the same physical entity. We assume that "winding up" the wave function with a voltage produces a propagation vector in the superconductor that is related to the motion of the electrons, and that this motion corresponds to a current because the electrons are charged. This viewpoint will allow us to understand the interaction between two coupled collective electron systems. We shall develop these relationships in more detail when we study the current distribution within the wire itself.

Let us consider two identical loops of superconducting wire, the diameter of the wire being much smaller than the loop radius. We place an extremely thin insulator between the loops, which are then superimposed on each other as closely as allowed by the insulator. In this configuration, shown in Fig. 1.2a, both loops can be described, to an excellent approximation, by the same path in space, despite their being electrically distinct. As we experiment with this configuration, we make the following observations:

1. When the two ends of the second loop are left open, the second loop's presence has no effect on the operation of the first

in the fourth of his epic papers, entitled *Quantization as a Problem of Proper Values* (34). These papers are reprinted, together with Schrödinger's *Lectures on Wave Mechanics*, in Ref. 35. The vector potential is introduced on page 119 of this reprint volume. A useful commentary on the history can be found on page 81 of Ref. 36. Tonomura (37) gives a wonderful modern account of experiments with electron waves.

loop: The relationship between a current flowing in the first loop and the voltage observed between the ends of the first loop follows Eq. 1.1, with exactly the same value of L as that observed when the second loop is absent.
2. The voltage observed between the two ends of the second loop under open conditions is almost exactly equal to the voltage observed across the first loop.
3. When the second loop is shorted, the voltage observed across the first loop is nearly zero, independent of the current.
4. The current observed in the second loop under shorted conditions is nearly equal to that flowing in the first loop, but is of the opposite sign.

Similar measurements can be performed when the loops are separated, as shown in Fig. 1.2b. These experiments allow us to observe how the coupling between the loops depends on their separation and relative orientation:

5. For a given configuration, the voltage observed across the second loop remains proportional to the voltage across the first loop. The constant of proportionality, which is nearly unity when the loops are superimposed, decreases with the distance between the loops.
6. The constant of proportionality decreases as the axes of the two loops are inclined with respect to each other, goes to zero when the two loops are orthogonal, and reverses when one loop is flipped with respect to the other.

Observation 1 tells us that the presence of electrons in the second loop does not *per se* affect the operation of the first loop. The voltage across a loop is a direct manifestation of the phase accumulation around the loop. Observation 2 tells us that current in a neighboring loop is as effective in producing phase accumulation in the wave function as current in the same loop. The ability of current in one location to produce phase accumulation in the wave function of electrons in another location is called **magnetic interaction.** Observation 6 tells us that the magnetic interaction is vectorial in nature. After making these and other similar measurements on many configurations, involving loops of different sizes and shapes, we arrive at the proper generalization of Eqs. 1.1 and 1.7:

$$\int V_1\,dt = \oint \vec{A}\cdot d\vec{l}_1 = \Phi_1 = L_1\,I_1 + M\,I_2$$
$$\int V_2\,dt = \oint \vec{A}\cdot d\vec{l}_2 = \Phi_2 = M\,I_1 + L_2\,I_2 \tag{1.8}$$

Here, the line elements $d\vec{l}_1$ and $d\vec{l}_2$ are taken along the first and second loops, respectively. The quantity M, which by observation 6 can be positive or negative depending on the configuration, is called the **mutual inductance**; it is a measure of how effective the current in one loop is at causing phase accumulation in the other. When $L_1 = L_2 = L$, the magnitude of M can never exceed L. Observations 1 through 4 were obtained under conditions where $M \approx L$. Experiments evaluating the mutual coupling of loops of different sizes, shapes, orientations, and spacings indicate that each element of wire of length dl carrying the current $\vec{I}$ makes a contribution to $\vec{A}$ that is proportional both to $\vec{I}$ and to the inverse of the distance r from the current element to the point at which $\vec{A}$ is evaluated:

$$\vec{A} = \frac{\mu_0}{4\pi}\int \frac{\vec{I}}{r}\,dl \quad \Rightarrow \quad \vec{A} = \frac{\mu_0}{4\pi}\int \frac{\vec{J}}{r}\,d\text{vol} \tag{1.9}$$

The constant μ_0 is called the **permeability of free space.** The second form follows from the first if we visualize a distribution of current as being carried by a large number of wires of infinitesimal cross section, and the **current density** $\vec{J}$ as being the number of such wires per unit area normal to the current flow. The $1/r$ form of the integrand of Eq. 1.9 is called the **Green's function**; it tells us how the vector potential is generated by currents everywhere in space. (It is perhaps more correct to say that the vector potential is a bookkeeping device for evaluating the effect at a particular point of all currents everywhere in space.)

We can express Eq. 1.2 in a way that gives us additional insight into the energy stored in the coil:

$$W = \int V\,dQ = \int V\,I\,dt = \int I\,d\Phi \tag{1.10}$$

Eq. 1.10 is valid for any $\vec{A}$; it is not limited to the $\vec{A}$ from the current in the coil itself. The integrals in Eq. 1.10 involve the

entire coil; from them, we can take a conceptual step and, using our visualization of the current density, imagine an **energy density** $\vec{J} \cdot \vec{A}$ ascribed to every point in space:

$$W = \int \vec{I} \cdot \vec{A}\, dl = \int \vec{J} \cdot \vec{A}\, d\text{vol} \tag{1.11}$$

The concept of energy density will be developed in greater depth in Part 3 (p. 49).

1.5 Integral and Differential Forms

Ernst Mach, the great master of classical mechanics, commented on the power of an integral law of interaction (p. 323 in Ref. 27):

> It cannot be denied that when we can command all the phenomena taking place in a medium, together with the large masses contained in it, by means of a single complete picture, our concepts are on an entirely different plane from what they are when only the relations of these isolated masses as regards acceleration are known . . . The present tendencies in the development of physics are entirely in this direction.

It is clear that this statement is as true for electrodynamics as it is for gravitation—that is why we introduced the basic laws of collective electrodynamics in integral form. There is a reason, however, that physicists often prefer to work with differential equations: When the currents are not known, but rather are the result of mutual interactions with all other electrons, the equivalent formulation by way of a differential equation provides a more tractable approach. Fortunately for the interactions of collective electrodynamics, there is a one-to-one relation between the integral form and the corresponding differential form:

$$\vec{A} = \frac{\mu_0}{4\pi} \int \frac{\vec{J}}{r}\, d\text{vol} \quad \Rightarrow \quad \nabla^2 \vec{A} = -\mu_0 \vec{J} \tag{1.12}$$

The $1/r$ in the integrand on the left is the Green's function for the differential form on the right. An excellent discussion of the relations involved is given in Chapter 7 of Morse and Feshbach (38). In mathematical circles, Eq. 1.12 is known as Green's Representation Theorem; it is stated with full rigor in Gradshteyn and Ryzhik (p. 1122 of Ref. 39).

Aside from mathematical niceties, we can understand how these relations work by considering the distribution of $\vec{A}$ from a highly localized current element $\vec{dI} = \vec{J}d\text{vol}$ located at the origin. The vector potential is everywhere in the direction of the current, so the problem reduces to a scalar relation between the magnitude J of the current, and the magnitude A of the vector potential. From the integral form, we know that once we are outside the current element itself, A dies off as the inverse of the distance r between the current element and the point of measurement, independent of direction. We use spherical coordinates to express the fact that the form of A does not depend on direction, so the differential equation describing the situation is

$$\nabla^2 A = \frac{1}{r}\frac{\partial^2(rA)}{\partial r^2} = -\mu_0 J \tag{1.13}$$

To make the problem easy, let us assume that the current element is a tiny sphere of radius r_0, filled with current density of magnitude J all flowing in one direction. The integral over the volume is $J4\pi r_0^3/3$. Far enough from the sphere, the distance to every point inside the sphere can be taken as equal to the distance r to the center, so the integral form gives $A = \mu_0 J r_0^3/3r$.

We can also find the solution for the differential form. In the region outside the sphere, $J = 0$, so $\nabla^2 A = 0$, which has solution $A \propto 1/r$. Because Eq. 1.12 is *linear* in the current, we can build up a solution by taking each current element, multiplying by its inverse distance from the point of measurement, and adding up such elementary vector contributions from all of space, which is exactly the meaning of the integral. So the form of the solution to the differential equation is the same as the form of the Green's function. What about the magnitude?

Eq. 1.13 tells us that the second derivative of the vector potential is proportional to the current density. Because there can be no physical system that contains an infinite current density, we know that both A and its derivative are continuous at all points, including at $r = r_0$. Inside the sphere, the solution to Eq. 1.13 is $A = A_0 - \mu_0 J r^2/6$, which can be verified by direct substitution. The continuity of derivative gives $-A(r_0)/r_0 = -2\mu_0 J r_0/6$, from which we obtain the solution outside the sphere:

$$A = A(r_0)\,\frac{r_0}{r} \qquad A(r_0) = \mu_0 J r_0^2/3 \tag{1.14}$$

in agreement with the result from the integral form. The continuity of value gives us $A(r_0) = A_0 - \mu_0 J r_0^2/6$. Substituting $A(r_0)$ from these two expressions gives $A_0 = \mu_0 J r_0^2/2$ for the vector potential at the origin. We can obtain A_0 more directly from the integral form by observing that $d\text{vol} = 4\pi r^2 dr$:

$$A_0 = \frac{\mu_0}{4\pi} \int_0^{r_0} \frac{J}{r} 4\pi r^2 dr = \frac{\mu_0 J r_0^2}{2} \tag{1.15}$$

The latter procedure also serves to eliminate any residual concerns that we may have been harboring about the possible divergence of A at $r = 0$. Thus, we have two methods for evaluating the vector potential from an arbitrary distribution of current; in showing their equivalence in this simple case, we have seen how each approach complements the other.

1.6 Electrodynamic Momentum

Feynman commented on the irrelevance of the concept of force in a quantum context. At the fundamental level, we can understand the behavior of a quantum system using only the wave properties of matter. But we experience forces between currents in every encounter with electric motors, relays, and other electromagnetic actuators. How do these forces arise from the underlying quantum reality? We can make a connection between the classical concept of force and the quantum nature of matter through the concept of *momentum*. Using the de Broglie postulate relating the momentum $\vec{p}$ of an electron to the propagation vector $\vec{k}$ of the wave function, and identifying the two integrands in Eq. 1.7, the electrodynamic momentum of an elemental charge is

$$\vec{p} = \hbar \vec{k} = q_0 \vec{A} \tag{1.16}$$

We shall now investigate the electrodynamic momentum in one of our loops of superconducting wire. There is an electric field E along the loop, the line integral of which is the voltage V between the ends. From a classical point of view, Newton's law tells us that the force $q_0 E$ on a charge should be equal to the time rate of change of momentum. From Eq. 1.16,

$$q_0 \vec{E} = \frac{\partial \vec{p}}{\partial t} = q_0 \frac{\partial \vec{A}}{\partial t} \quad \Rightarrow \quad V = \oint \vec{E} \cdot \vec{dl} = \frac{\partial \Phi}{\partial t} \tag{1.17}$$

Integrating the second form of Eq. 1.17 with respect to time, we recover Eq. 1.7, so the classical idea of inertia is indeed consistent with the quantum behavior of our collective system. Electrodynamic inertia acts exactly as a classical mechanical inertia: It relates the integral of a force to a momentum, which is manifest as a current. We note that for any system of charges that is overall charge neutral, as is our superconductor, the net electromagnetic momentum is zero. For the $-q\vec{A}$ of each electron, we have a canceling $+q\vec{A}$ from one of the background positive charges. The electric field that accelerates electrons in one direction exerts an equal force in the opposite direction on the background positive charges. We have, however, just encountered our first big surprise: We recognize the second form of Eq. 1.17, which came from Newton's law, as the integral form of one of Maxwell's equations!

We would expect the total momentum P of the collective electron system to be the momentum per charge times the number of charges in the loop. If there are η charges per unit length of wire that take part in the motion, integrating Eq. 1.16 along the loop gives

$$P = \eta q_0 \oint \vec{A} \cdot \vec{dl} = \eta q_0 \Phi = \eta q_0 L\, I \tag{1.18}$$

The current I is carried by the η charges per unit length moving at velocity v; therefore, $I = \eta q_0 v$, and Eq. 1.18 becomes

$$P = L\,(\eta q_0)^2\, v \tag{1.19}$$

The momentum is proportional to the velocity, as it should be. It is also proportional to the size of the loop, as reflected by the inductance L. Here, we have our second big surprise: Instead of scaling linearly with the number of charges that take part in the motion, the momentum of a collective system scales as the *square* of the number of charges! We can understand this collective behavior as follows: In an arrangement where charges are constrained to move in concert, each charge produces phase accumulation, not only for itself but for all the other charges as well. So the inertia of *each charge* increases linearly with the number of charges moving in concert. The inertia of the ensemble of coupled charges must therefore increase as the square of the number of charges.

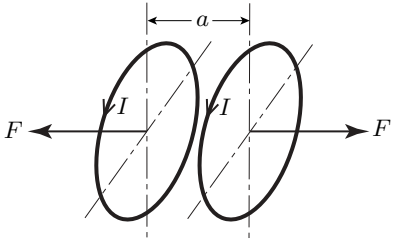


Figure 1.3 Two loops carrying persistent currents in the same direction require a mechanical force F to keep them separate

1.7 Forces on Currents

In our experiments on coupled loops, we have already seen how the current in one loop induces phase accumulation in another loop; the relations involved are captured in Eq. 1.8. In any situation where we change the coupling of collective systems by changing the spatial arrangement, mechanical work may be involved. Our model system for studying this interaction consists of two identical shorted loops of individual inductance L_0, each carrying a persistent flux Φ, as shown in Fig. 1.3. As long as the superconducting state retains its integrity, the cyclic constraint on the wave function guarantees that the flux Φ in each loop will be constant, independent of the coupling between loops. Because M enters symmetrically in Eq. 1.8, if the current I started out the same in both loops, it will continue to be the same throughout the experiment. Hence, L_0 and Φ will remain constant, whereas M and I will be functions of the spatial arrangement of the loops: M will be large and positive when the loops are brought together with their currents flowing in the same direction, and will be large and negative when the loops are brought together with their currents flowing in opposite directions. From Eq. 1.8, $\Phi = (L_0+M)I$. Substituting Φ into Eq. 1.10, and noting that the total energy of the system is twice that for a single coil,

$$W = 2\int I\, d\Phi = (L_0 + M)I^2 = \frac{\Phi^2}{(L_0 + M)} \tag{1.20}$$

The force F_x along some direction x is defined as the rate of change of energy with a change in the corresponding coordinate:

$$F_x = \frac{\partial W}{\partial x} = -\left(\frac{\Phi}{L_0 + M}\right)^2 \frac{\partial M}{\partial x} \tag{1.21}$$

The negative sign indicates an attractive force because the mutual inductance M increases as the coils—whose currents are circulating in the same direction—are moved closer. It is well known that electric charges of the same sign *repel* each other. We might expect the current, being the spatial analog of the charge, to behave in a similar manner. However, Eq. 1.20 indicates that the total energy of the system *decreases* as M increases. How does this *attractive* interaction of currents circulating in the same direction come about?

The electron velocity is proportional to I. As M is increased, the electrons in both loops *slow down* because they have more inertia due to their coupling with electrons in the other loop. This effect is evident in Eq. 1.20, where $I = \Phi/(L_0 + M)$. Thus, there are two competing effects: the decrease in energy due to the lower velocity, and the increase in energy due to the increase in inertia of each electron. The energy goes as the square of the velocity, but goes only linearly with the inertia, so the velocity wins. The net effect is a decrease in energy as currents in the same direction are coupled; hence, an attractive force. We can see how the force law discovered in 1823 by Ampère arises naturally from the collective quantum behavior. A classical mechanical argument would give the opposite *sign* for the effect.

1.8 Multiturn Coils

The interaction in a collective system scales as the square of the number of electrons moving in concert. Thus, we might expect the quantum scaling laws to be most clearly manifest in the properties of closely coupled multiturn coils, where the number of electrons is proportional to the number of turns. We can construct an N-turn coil by connecting in series N identical, closely coupled loops. In this arrangement, the current through all loops is equal to the current I through the coil, and the voltage V across the coil is equal to the sum of the individual voltages across the loops. If A_0 is the vector potential from the current in one loop, we expect the vector potential from N loops to be NA_0, because the current in each loop contributes. The flux integral is taken around N turns, so the path is N times the length l_0 of a single turn. The total flux integral is thus:

$$\Phi = \int V\,dt = \int_0^{Nl_0} NA_0 \cdot dl = N^2 L_0 I \tag{1.22}$$

From Eq. 1.22, we conclude that an N-turn closely coupled coil has an inductance $L = N^2 L_0$. Once again, we see the collective interaction scaling as the square of the number of interacting charges. We noted that collective quantum systems have a correspondence limit markedly different from that of classical mechanical systems. When two classical massive bodies, each having a separate inertia, are bolted together, the inertia of the resulting composite body is simply the sum of the two individual inertias. The inertia of a collective system, however, is a manifestation of the *interaction*, and cannot be assigned to the elements separately. This difference between classical and quantum systems has nothing to do with the size scale of the system. Eq. 1.22 is valid for large as well as for small systems; it is valid where the total phase accumulation is an arbitrary number of cycles—where the granularity of the flux due to $\hbar$ is as small as might be required by any correspondence procedure. Thus, it is clear that *collective quantum systems do not have a classical correspondence limit.*

It is instructive to work out the magnitude of the electron inertia in a concrete case. A small superconducting magnet has 10^4 turns of NbTi wire approximately 0.1 mm in diameter. The magnet is 7 cm long, and just under 5 cm in diameter, and produces a peak field of 7 tesla at a current of 40 amperes. The magnet weighs about 0.5 kilograms, and has a measured inductance of approximately 0.5 henry. There are of the order of 10^{28} electrons per cubic meter in the wire, or 10^{20} electrons per meter length of wire, corresponding to approximately 10 coulombs of electronic charge per meter of wire. At 40 amperes, these electrons move at a velocity $v \approx 4$ m/sec. The total length l of wire is about 10^3 meters, so the total electronic charge in the magnet is about 10^4 coulombs. Using these values,

$$A \approx \frac{\Phi}{l} = \frac{L\,I}{l} \approx 0.02 \quad \frac{\text{V sec}}{\text{meter}} \tag{1.23}$$

The electromagnetic momentum p_{el} of an electron is just this vector potential multiplied by the electronic charge; from this, we can infer an electromagnetic mass m_{el} for each electron:

$$\begin{aligned} p_{\text{el}} = qA = 3.2 \times 10^{-21} \quad \frac{\text{coulomb V sec}}{\text{meter}} &= \; m_{\text{el}} v \\ \Rightarrow \qquad m_{\text{el}} \;&\approx\; 10^{-21}\ \text{kg} \end{aligned} \tag{1.24}$$

For comparison, the mass of a free electron is approximately 10^{-30} kg, and the rest mass of a proton is a factor of 1800 larger than that of an electron. The electromagnetic mass of an electron in our magnet is thus a factor of 10^9 larger than the rest mass of a free electron. The total inertia of the electron system in the magnet is much larger that the actual mass of all the atoms making up the magnet. It is curious that the electromagnetic momentum has been largely ignored in introductory treatments of the subject, in light of its large role in real situations. For almost all problems involving currents in wires, the electron density is so high, and the requirement for charge neutrality enforced so strongly, that the momentum of the collective, interacting system is overwhelmingly larger than that calculated by adding the momenta of the free particles moving at the same velocity. For this reason, people often speak of the "momentum in the field" instead of recognizing the collective nature of the system.

1.9 Total Momentum

To understand how our simplistic approach has taken us this far, we must understand the current distribution within the superconductor itself. We saw that the vector potential made a contribution to the momentum of each electron; this we called the electrodynamic momentum, $\vec{p}_{\rm el} = q\vec{A}$. Up to this point, we have been able to neglect any additional contribution to electron momentum. To understand the current distribution in the superconductor, we must include the contribution of the mass m of an electron moving with velocity $\vec{v}$: $\vec{p}_{\rm mv} = m\vec{v}$. The total momentum is the sum of these two contributions:

$$\hbar\vec{k} = \vec{p} = \vec{p}_{\rm el} + \vec{p}_{\rm mv} = q_0\vec{A} + m\vec{v} \tag{1.25}$$

The velocity $\vec{v} = (\hbar\vec{k} - q_0\vec{A})/m$ is thus a direct measure of the imbalance between the total momentum $\hbar\vec{k}$ and the electrodynamic momentum $q_0\vec{A}$: When these two quantities are matched, the velocity is zero. The current density J is just the motion[7] of

[7]Here we are using the most naive relation between current density and momentum. To do better is a matter of some delicacy, treated nicely by Pines and Nozières (26). These considerations affect the magnitude of the skin depth, discussed in the following sections.

$\mathcal{N}$ elementary charges per unit volume: $J = q_0 \mathcal{N} \vec{v}$. We can thus express Eq. 1.25 in terms of the wave vector $\vec{k}$, the vector potential $\vec{A}$, and the current density $\vec{J}$:

$$\vec{J} = \frac{q_0 \mathcal{N}}{m} (\hbar \vec{k} - q_0 \vec{A}) \tag{1.26}$$

1.10 Current Distribution

We are now in position to investigate how current distributes itself inside a superconductor. If $\vec{A}$ were constant throughout the wire, the motion of the electrons would be determined by the common wave vector $\vec{k}$ of the collective electron system; and we would expect the persistent current for a given flux to be proportional to the cross-sectional area of the wire, and, thus, the inductance L of a loop of wire to be inversely related to the cross section of the wire. When we perform experiments on loops of wire that have identical paths in space, however, we find that the inductance is only a weak function of the wire diameter, indicating that the current is not uniform across the wire, and, therefore, that $\vec{A}$ is far from constant. If we make a loop of superconducting *tubing*, instead of wire, we find that it has exactly the same inductance as does a loop made with wire of the same diameter, indicating that current is flowing at the *surface* of the loop, but is not flowing throughout the bulk.

Before taking on the distribution of current in a wire, we can examine a simpler example. In a simply connected bulk superconductor, the single-valued nature of the wave function can be satisfied only if the phase is everywhere the same: $\vec{k} = 0$. Any phase accumulation induced through the $\vec{A}$ vector created by an external current will be canceled by a screening current density $\vec{J}$ in the opposite direction, as we saw in observations 3 and 4. To make the problem tractable, we consider a situation where a vector potential A_0 at the surface of a bulk superconducting slab is created by distant currents parallel to the surface of the slab. The current distribution perpendicular to the surface is a highly localized phenomenon, so it is most convenient to use the differential formulation of Eq. 1.9. We suppose that conditions are the same at all points on the surface, and, therefore, that A changes in only the x direction, perpendicular to the surface, implying that

$\nabla^2 A = \partial^2 A/\partial x^2$. Using J from Eq. 1.26 we obtain

$$\nabla^2 A = \frac{\partial^2 A}{\partial x^2} = -\mu_0 \vec{J} = \frac{\mu_0 q_0^2 \mathcal{N}}{m} A \tag{1.27}$$

The solution to Eq. 1.27 is

$$A = A_0\, e^{-x/\lambda} \qquad \lambda^2 = \frac{m}{\mu_0 q_0^2 \mathcal{N}} \tag{1.28}$$

The particular form of Eq. 1.28 depends on the geometry, but the qualitative result is always the same, and can be understood as follows: The current is the *imbalance* between the wave vector and the vector potential. When an imbalance exists, a current proportional to that imbalance will flow such that it cancels out the imbalance. The resulting screening current dies out exponentially with distance from the source of imbalance. The distance scale at which the decay occurs is given by λ, the **screening distance**, **penetration depth,** or **skin depth.** For a typical superconductor, $\mathcal{N}$ is of the order of $10^{28}/\mathrm{M}^3$, so λ should be a few tens of nanometers. Experimentally, simple superconductors have $\lambda \approx 50$ nanometers—many orders of magnitude smaller than the macroscopic wire thickness that we are using.

1.11 Current in a Wire

At long last, we can visualize the current distribution within the superconducting wire itself. Because the skin depth is so small, the surface of the wire appears flat on that scale, and we can use the solution for a flat surface. The current will be a maximum at the surface of the wire, and will die off exponentially with distance into the interior of the wire. We can appreciate the relations involved by examining a simple example: A 10-cm-diameter loop of 0.1-mm-diameter wire has an inductance of 4.4×10^{-7} henry (p. 193 in Ref. 40). A persistent current of 1 ampere in this loop produces a flux of 4.4×10^{-7} volt-second, which is 2.1×10^8 flux quanta. The electron wave function thus has a total phase accumulation of 2.1×10^8 cycles along the length of the wire, corresponding to a wave vector $k = 4.25 \times 10^9\,\mathrm{M}^{-1}$. Due to the cyclic constraint on the wave function, this phase accumulation is shared by all electrons in the wire, whether or not they are carrying current.

In the region where current is flowing, the moving mass of the electrons contributes to the total phase accumulation. The 1-ampere current results from a current density of 6.4×10^{10} amperes per square meter flowing in a thin "skin" $\approx \lambda$ thick, just inside the surface. This current density is the result of the 10^{28} electrons per cubic meter moving with a velocity of $v \approx 20$ meters per second. The mass of the electrons moving at this velocity contributes $mv/\hbar = 1.7 \times 10^5\,\mathrm{M}^{-1}$ to the total wave vector of the wave function, which is less than one part in 10^4 of that contributed by the vector potential. That small difference, existing in about 1 part in 10^6 of the cross-sectional area, is enough to bring $\vec{k}$ and $\vec{A}$ into balance in the interior of the wire.

In the interior of the wire, the propagation vector of the wave function is matched to the vector potential, and the current is therefore zero. As we approach the surface, A decreases slightly, and the difference between k and $Aq_0/\hbar$ is manifest as a current. At the surface, the value and radial slope of A inside and outside the wire match, and the value of A is still within one part in 10^4 of that in the center of the wire. So our simplistic view—that the vector potential and the wave vector are two representations of the same quantity—is precisely true in the center of the wire, and is nearly true even at the surface. The current $\vec{I}$ is not the propagation vector $\vec{k}$ of the wave, but, for a fixed configuration, $\vec{I}$ is proportional to $\vec{k}$ by Eqs. 1.9 and 1.26. For that reason, we could deduce the electromagnetic laws relating current and voltage from the quantum relations between wave vector and frequency.

1.12 Summary

We took to heart Einstein's belief that the electrons and the fields are two aspects of the same reality, and have been able to treat the macroscopic quantum system and the electromagnetic field as elements of a unified subject. We heeded Mach's advice that classical mechanics was not the place to start, followed Feynman's directive that interactions change the wavelengths of waves, and saw that there is a correspondence limit more appropriate than the classical-mechanics version used in traditional introductions to quantum theory. We found Newton's law masquerading as one of Maxwell's equations. We were able to derive a number of important results using only the simplest properties of waves, the

Einstein postulate relating frequency to energy, the de Broglie postulate relating momentum to wave vector, and the discrete charge of the electron. It thus appears possible to formulate a unified, conceptually correct introduction to both the quantum nature of matter and the fundamental laws of electromagnetic interaction without using either Maxwell's equations or standard quantum formalism.

This page intentionally left blank

Part 2

Propagating Waves

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.

—James Clerk Maxwell[1]

There is no such concept as "the" field,
an independent entity with degrees of freedom of its own.

—J.A. Wheeler and R.P. Feynman[2]

2.1 A Brief History

In Part 1 (p. 9), we found that the vector potential $\vec{A}$ forms the natural link between the quantum nature of matter and the realm of electromagnetic phenomena. Although the four canonical equations that bear Maxwell's name are expressed in terms of the magnetic field $\vec{B}$ and the electric field $\vec{E}$, Maxwell himself derived many of his fundamental results using $\vec{A}$, which he also called the **electromagnetic momentum.** In his *Treatise on Electricity and Magnetism*, he describes his method:

> In this way we have pointed out the existence of the electrodynamic momentum $\vec{A}$ as a vector whose direction and magnitude vary from one part of space to another, and from this we have deduced, by a mathematical process, the magnetic induction $\vec{B}$, as a derived vector.[3]

Maxwell developed the basic form of the magnetic interaction, Eq. 1.9, in both its differential and integral forms (Art. 617 of Ref. 33). As the opening quotation indicates, however, Maxwell was deeply devoted to the ether as the medium in which electromagnetic phenomena take place. $\vec{E}$ and $\vec{B}$ were to him the degrees

[1]The Maxwell quotation appears in an impassioned plea for the existence of an ether in Article 866 of his *Treatise on Electricity and Magnetism* (33).
[2]The Wheeler–Feynman quotation is from their 1949 paper (41).
[3](Art. 606 in Ref. 33) I have taken the liberty of substituting $\vec{A}$ and $\vec{B}$ for the German symbols used in Maxwell's original text.

of freedom of the ether that are coupled to the charge in ordinary matter. This devotion led him to include an additional term, the displacement current, in his definition of current density. In his own words,

> One of the chief peculiarities of this treatise is the doctrine which it asserts, that the true electric current, that on which the electromagnetic phenomena depend, is not the same thing as the current of conduction, but that the time variation of the electric displacement must be taken into account in estimating the total movement of electricity.[4]

Because of this addition, he narrowly missed the simple and beautiful four-vector form of the electromagnetic laws on which we concentrate in this present treatment. This slip dealt the discipline a severe blow from which it has yet to fully recover.

In 1865, Maxwell showed solutions for the vector potential that propagated at the velocity of light, and he asserted that light is an electromagnetic phenomenon. He took this result as strong evidence for the existence of an ether:

> In several parts of this treatise, an attempt has been made to explain electromagnetic phenomena by means of mechanical action transmitted from one body to another by means of a medium occupying the space between them . . . the combination of the optical with the electrical evidence will produce a conviction of the reality of the medium similar to that which we obtain, in the case of other kinds of matter, from the combined evidence of the senses. (Art. 781 in Ref. 33)

Oliver Heaviside and Heinrich Hertz—both thoroughly imbued with the spirit of the ether—went Maxwell one better. They criticized his extensive use of the vector potential, and developed the streamlined presentation of Maxwell's equations that is in use today. In his 1905 paper, Einstein called these expressions the Maxwell–Hertz equations. For Heaviside (42),

> the question of the propagation of, not merely the electric potential Ψ but the vector potential $\mathbf{A}$ . . . when brought forward, proves to be one of a metaphysical nature . . . the electric force $\mathbf{E}$ and the magnetic force $\mathbf{H}$ . . . actually represent the state of the medium everywhere . . . Granting this, it is perfectly obvious that in any case of propagation, since it is the physical state that is propagated, it is $\mathbf{E}$ and $\mathbf{H}$ that are propagated.

[4] (Art. 610 in Ref. 33) German symbols omitted in this quotation.

Hertz[5] had a similar criticism:

> I may mention the predominance of the vector potential in [Maxwell's] fundamental equations. In the construction of the new theory the potential served as a scaffolding . . . it does not appear to me that any . . . advantage is attained by the introduction of the vector potential in the fundamental equations; furthermore, one would expect to find in these equations relations between physical magnitudes which are actually observed, and not between magnitudes which serve for calculation only.

Hertz's splendid experiments in 1888 demonstrated the propagation of electromagnetic waves, and were taken by the community at large as an unequivocal confirmation of Maxwell's theory. It was 70 years later that Aharonov and Bohm (44) suggested a configuration in which $\vec{A}$ could be observed in the phase of the wave function even in the absence of $\vec{B}$; the experiment was carried out soon after (45, 46, 47), and confirmed their prediction. Feynman (Sec.15-5 in Ref. 5) analyzes the Aharonov–Bohm effect, ending with a delightful discussion of what makes a field "real." That section concludes with the statement, "In our sense then, the $\vec{A}$ field is 'real.' " Modern experiments demonstrating the Aharonov–Bohm effect are described in Ref. 37. We can only speculate that the comments of Hertz and Heaviside would have been quite different had these researchers known, as we do today, that $\vec{A}$ is the magnitude that is actually observed, whereas $\vec{B}$ (or $\vec{H}$) is more of a metaphysical concept.

Meanwhile, the growing popularity of the ether as the medium in which electromagnetic waves propagate led several groups to search for more direct evidence of its existence. In 1887, Michelson and Morley succeeded in carrying out an optical experiment sufficiently sensitive that it could detect the earth's motion through the ether. The experiment failed to detect any motion, at any time of day or night. The same year, W. Voigt (48) published a little-known paper in which he showed that when the Cartesian coordinates of a reference system are x, y, z, t, and those of a frame of reference moving at velocity v in the x direction (with respect to our reference system) are x', y', z', t', then Maxwell's equations,

[5] (p. 196 in Ref. 43) This source is an excellent English translation of all Hertz's relevant papers.

in space free of charges and currents, are not altered by a transformation of the form:

$$\begin{aligned} x' &= \gamma\,(x - vt) & x &= \gamma\left(x' + vt'\right) \\ y' &= y & y &= y' \\ z' &= z & z &= z' \\ t' &= \gamma\left(t - vx/c^2\right) & t &= \gamma\left(t' + vx'/c^2\right) \end{aligned} \tag{2.1}$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$. This transformation was reinvented in 1892 by H.A. Lorentz, and is today called the **Lorentz transformation**. Lorentz derived his result independently, but in 1909 referred[6] to Voigt's paper:

> which to my regret escaped my notice all these years. The idea of the transformations used above might therefore have been borrowed from Voigt, and the proof that it does not alter the equations for the free ether is contained in his paper.

In his famous 1905 relativity paper (51), Einstein put forth his thesis:

> failure of attempts to detect a motion of the earth relative to the "light medium," lead to the conjecture that not only in mechanics, but in electrodynamics as well, the phenomena do not have any properties corresponding to the concept of absolute rest, but that in all coordinate systems in which the mechanical equations are valid, also the same electrodynamic and optical laws are valid, as has already been shown for quantities of the first order. We shall raise this conjecture (whose content will be called "the principle of relativity" hereafter) to the status of a postulate and shall introduce, in addition, the postulate, only seemingly incompatible with the former one, that in empty space light is always propagated with a definite velocity V which is independent of the velocity of the emitting body. These two postulates suffice for arriving at a simple and consistent electrodynamics of moving bodies on the basis of Maxwell's theory for bodies at rest. The introduction of a "light ether" will prove superfluous, inasmuch as in accordance with the concept to be developed here, no "space at absolute rest" endowed with special properties will be introduced, nor

[6] A great deal of historical discussion, with many valuable references, is given in Refs. 49, 50.

> will a velocity vector be assigned to a point of empty space at which electromagnetic processes are taking place.

Thus, in one stroke, Einstein did away with the ether that gave rise to Maxwell's equations in the first place—the ether that made the vector potential an unacceptable metaphysical concept, the ether that has ever since haunted us with the specter of a magnetic field that does not connect with the quantum nature of matter. Newton's laws had to be revised, but Maxwell's equations—$\vec{B}$ field and all—survived! Rindler's comment (52) is particularly apt:

> There is an element of irony in the fact that the discipline which had always accepted a relativity principle should have to be amended in the light of the new relativity, whereas classical electro-magnetic theory, which so firmly committed itself to the existence of a preferred frame, should, in fact, come through unscathed.

In a sweeping generalization, Einstein went on to extend his principle of relativity to all the laws of physics:

> The laws governing the changes of the state of any physical system do not depend on which one of two coordinate systems in uniform translational motion relative to each other these changes of the state are referred to.

In a Voigt–Lorentz–Einstein world, space and time coordinates are not separable. For that reason, the relativistic nature of physical law finds its most natural expression in the language of four-vectors.

2.2 Four-Vectors

A **four-vector U** (denoted by boldface type) has a **spatial part** $\vec{U}$, which behaves as an ordinary three-vector, and a **time part** U_t, which is a scalar:

$$\mathbf{U} = \left[\vec{U},\, U_t\right] \tag{2.2}$$

The **sum** of two four vectors is another four-vector, each component of which is the sum of the corresponding components of the constituent vectors:

$$\mathbf{A} + \mathbf{B} = \left[\vec{A} + \vec{B},\, A_t + B_t\right] \tag{2.3}$$

The **scalar product** of two four-vectors is defined as follows:[7]

$$\mathbf{A} \bullet \mathbf{B} = \vec{A} \cdot \vec{B} - A_t B_t \tag{2.4}$$

An important special case is the scalar product of a four-vector with itself:

$$\mathbf{U} \bullet \mathbf{U} = \mathbf{U}^2 = \left\| \vec{U} \right\|^2 - U_t^2 \tag{2.5}$$

Unlike the square magnitude of an ordinary three-vector, the square of a four-vector can be positive, negative, or zero. The **four-gradient operator** $\square$ is defined as follows:[8]

$$\square = \left[\nabla, -\frac{1}{c}\frac{\partial}{\partial t} \right] \tag{2.6}$$

The four-gradient operator and all four-vectors $\mathbf{U}$ transform[9] according to the Lorentz transformation:

$$\begin{aligned}
U'_x &= \gamma \left(U_x - \frac{v}{c} U_t \right) & U_x &= \gamma \left(U'_x + \frac{v}{c} U'_t \right) \\
U'_y &= U_y & U_y &= U'_y \\
U'_z &= U_z & U_z &= U'_z \\
U'_t &= \gamma \left(U_t - \frac{v}{c} U_x \right) & U_t &= \gamma \left(U'_t + \frac{v}{c} U'_x \right)
\end{aligned} \tag{2.7}$$

The scalar product of a four-vector with itself, with another four-vector, or with the four-gradient operator is a **Lorentz invariant**: It has the same value in any inertial frame of reference. This property follows directly from the Lorentz transformation, and serves to define the class of four-vectors. The following are known to be four-vectors:

---

[7] There are almost as many sign conventions in relativity theory as there are authors. We use the simplest one, which is similar to that used by Feynman (Chap. 17 in Ref. 14; Chap. 25 in Ref. 5).

[8] Many authors use the box symbol for the quadratic operator. We follow the convention of Feynman (14, 5) and of Morse and Feshbach (38).

[9] We have chosen the sign of the time coordinate of each of our four-vectors such that it corresponds to what tensor-based relativity treatments call a **contravariant vector** U^μ.

1. The **interval** between two events, $\mathbf{R_{1,2}} = [\vec{r}_{1,2},\, c(t_2 - t_1)]$, which contains the ordinary distance vector $\vec{r}$ and the time difference[10]
2. The **four-potential,** $\mathbf{A} = \left[\vec{A},\, \mathcal{V}/c\right]$, which contains the vector potential $\vec{A}$ and the scalar (electrostatic) potential $\mathcal{V}$
3. The **current density,**[11] $\mathbf{J} = \left[\vec{J},\, c\rho\right]$, which contains the actual current density $\vec{J}$ and charge density ρ
4. The **propagation vector,** $\mathbf{k} = \left[\vec{k},\, \omega/c\right]$, which contains the ordinary propagation vector $\vec{k}$ and the frequency ω of a relativistic wave

In this convention, the propagation four-vector is the four-gradient of the phase φ of the wave

$$\mathbf{k} = \Box\varphi \quad \Rightarrow \quad \omega = -\frac{\partial\varphi}{\partial t} \tag{2.8}$$

Thus, the phase increment

$$\begin{aligned} d\varphi = \mathbf{k} \bullet d\mathbf{R} &= \vec{k} \cdot d\vec{r} - \omega dt \\ &= \frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy + \frac{\partial\varphi}{\partial z}dz + \frac{\partial\varphi}{\partial t}dt \end{aligned} \tag{2.9}$$

is a Lorentz invariant (see Fig. 21.1 in Ref. 53). A wave with positive $\vec{k}$ and positive ω propagates in the $\vec{k}$ direction. We shall return to this sign convention when we consider the propagation of a step function.

2.3 The Riemann–Sommerfeld Equation

Maxwell (p. 490 in Ref. 33) quotes an important precursor to the four-vector expression of electrodynamics as follows:

> In a memoir presented to the Royal Society of Göttingen in 1858, but afterwards withdrawn, and only published in

[10] The Lorentz transformation (Eq. 2.7) of this vector reduces to Eq. 2.1 in the special case where $x = x' = 0$ when $t = t' = 0$.

[11] Note that $\vec{J}$ does not contain displacement current. The current density four-vector is often called the **four-current.**

> Poggendorff's *Annalen*, bd.cxxxi. pp. 237–263, in 1867, after the death of the author, Bernhard Riemann deduces the phenomena of the induction of electric currents from a modified form of Poisson's equation
>
> $$\frac{d^2V}{dx^2} + \frac{d^2V}{dy^2} + \frac{d^2V}{dz^2} + 4\pi\rho = \frac{1}{a^2}\frac{d^2V}{dt^2}$$
>
> where V is the electrostatic potential, and a a velocity.
>
> This equation is of the same form as those which express the propagation of waves and other disturbances in elastic media. The author, however, seems to avoid making explicit mention of any medium through which the propagation takes place.

The four-vector generalization of Riemann's equation was formulated by Sommerfeld (54) shortly after Einstein's 1905 paper introduced the special theory of relativity:

$$\Box^2\mathbf{A} = \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{A} = -\mu_0\mathbf{J} \tag{2.10}$$

Treatments of this subject invariably go to great lengths to show the compatibility of this expression with the Maxwell equations. Because $\vec{A}$ appears in the phase of the wave function, there is no need for the magnetic field $\vec{B}$, and hence no need for the Maxwell equations. Thus, Eq. 2.10 (or its equivalent Green's function integral form), with the relation that we shall develop between **A** and the electron wave function, forms the complete basis for collective electrodynamics.

Eq. 2.10 is really four equations—one for each component of the four-vector **A**. The fourth of these, involving the potential $\mathcal{V}$, is identical to Riemann's equation. When the time-derivative term is negligible compared to the ∇^2 term, Riemann's equation reduces to **Electrostatics,** and the spatial components of Eq. 2.10 reduce to the case considered in Part 1 (p. 9), thus quantifying the notion of **slowly varying** used there. Both of these special cases were given in Maxwell's original treatment. However, the $\vec{J}$ used by Maxwell included the displacement current, unlike the spatial components of the four-current.

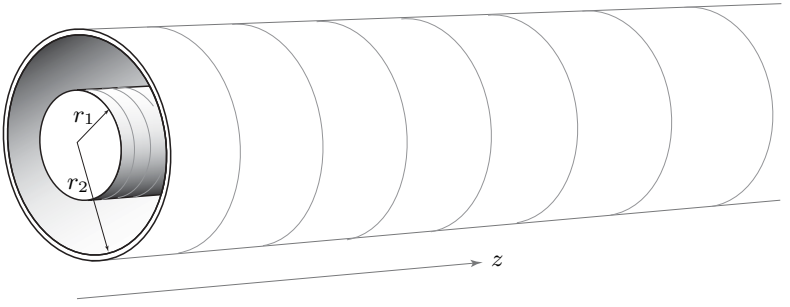


Figure 2.1 Coaxial transmission line

2.4 Model System

As our first example in which the $\partial^2/\partial t^2$ term in Eq. 2.10 cannot be neglected compared to the ∇^2 term, we consider the propagation of signals in a coaxial transmission line. The line, shown in Fig. 2.1, consists of a superconducting wire of radius r_1 inside a superconducting tube of inner radius r_2. The space between the conductors is assumed to be vacuum. Because the currents and charges are not known *a priori*, we use the Riemann–Sommerfeld differential form relating $\mathbf{A}$ and $\mathbf{J}$. In cylindrical coordinates, where we assume that there is no angular dependence, and that $\mathbf{A}$ and $\mathbf{J}$ have only z and t components, Eq. 2.10 becomes

$$\frac{\partial^2 \mathbf{A}}{\partial r^2} + \frac{1}{r}\frac{\partial \mathbf{A}}{\partial r} + \frac{\partial^2 \mathbf{A}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} \tag{2.11}$$

In the region between the coaxial conductors, where $\mathbf{J}$ is zero, functions of the form

$$\mathbf{A} = \mathbf{A_0} \ln\left(\frac{r_0}{r}\right) f(t \pm z/c) \tag{2.12}$$

solve Eq. 2.11. Thus, a four-potential waveform $f(t)$ will propagate in the z direction with velocity c, *independent of the waveform.* Choosing a value for r_0 is equivalent to choosing a reference for the potential $\mathbf{A}$. Eq. 2.12 is completely symmetrical with respect to a change in the sign of z, of t, or of both. A general solution is the sum of terms of all four types, each with a particular coefficient.

2.5 General Boundary Condition

Eq. 2.12 can be considered a solution of Eq. 2.10 only if it is consistent with the net charge and current density at the surfaces of the conductors. In the final two sections of Part 1 (p. 9), we found that the current density—being the imbalance between the wave vector and the vector potential—dies out exponentially with depth into the superconductor. That analysis is valid for the four-vector quantities as long as the $\partial^2/\partial t^2$ term in Eq. 2.10 is small compared with the ∇^2 term perpendicular to the surface. In other words, the time variation of the propagated signal should be slow compared with the velocity of light divided by the skin depth λ. Just inside the surface of the conductor, the four-current density is high, and the ∇^2 term in Eq. 2.10 is enormous in the direction s normal to the surface. Neglecting the other terms,

$$-\mu_0 \mathbf{J} = \nabla^2 \mathbf{A} \approx \frac{\partial^2 \mathbf{A}}{\partial s^2} \tag{2.13}$$

We obtain the relation between four-current and four-potential by integrating Eq. 2.13 with respect to s through the surface layer:

$$\left.\frac{\partial \mathbf{A}}{\partial s}\right|_{\text{surf}} = -\mu_0 \int_{\text{bulk}}^{\text{surf}} \mathbf{J}\, dr = -\mu_0 \mathbf{J}_s \tag{2.14}$$

where the coordinate s is directed from the interior to the exterior of the superconductor. Eq. 2.14 is the general boundary condition at a superconducting surface; as with all such four-vector relations, it encompasses both electric and magnetic cases:

$$\begin{aligned} \left.\frac{\partial \vec{A}}{\partial s}\right|_{\text{surf}} &= -\mu_0 \vec{J}_s \\ \left.\frac{\partial V}{\partial s}\right|_{\text{surf}} &= -\mu_0 c^2 Q_s \end{aligned} \tag{2.15}$$

2.6 Coaxial Boundary Condition

When Eq. 2.14 is applied to the coaxial line, $s = r$ for the inner conductor, and $s = -r$ for the outer conductor. For either con-

ductor of the coaxial line, $\partial \mathbf{A}/\partial r = -\mathbf{A}_0/r$ from Eq. 2.12, so from Eq. 2.14 we obtain the boundary conditions

$$\begin{aligned} \mathbf{A}_0 &= \frac{\mu_0}{2\pi}\, 2\pi r \mathbf{J}_s \qquad \text{inner} \\ -\mathbf{A}_0 &= \frac{\mu_0}{2\pi}\, 2\pi r \mathbf{J}_s \qquad \text{outer} \end{aligned} \tag{2.16}$$

In other words, the quantity $2\pi r \mathbf{J}_s$ for the two conductors is equal in magnitude and opposite in sign. The z component of this quantity is the total current I, and the t component is the charge Q per unit length of the particular conductor. Signals $f(t \pm z/c)$ propagating along the line are represented by equal and opposite charges and currents in any cross-section of the two conductors. It is conventional to use the outer conductor as the reference ($\mathbf{A} = 0$), corresponding to $r_0 = r_2$ in Eq. 2.12. We shall henceforth adopt that convention, and determine the potential of the inner conductor relative to that reference. For a particular value of z and t, Eq. 2.12 thus becomes

$$\mathbf{A}(r_1) = \frac{\mu_0}{2\pi}\, 2\pi r \mathbf{J}_s \ln\left(\frac{r_2}{r_1}\right) \tag{2.17}$$

The t component of Eq. 2.17 is thus

$$V = \frac{\mu_0 c^2}{2\pi}\, Q \ln\left(\frac{r_2}{r_1}\right) \tag{2.18}$$

We can then define a **capacitance** C per unit length:

$$C = \frac{Q}{V} = \frac{2\pi}{\mu_0 c^2 \ln\left(r_2/r_1\right)} \tag{2.19}$$

In a similar manner, the z component of Eq. 2.17 is

$$A_z(r_1) = \frac{\mu_0 I}{2\pi} \ln\left(\frac{r_2}{r_1}\right) \tag{2.20}$$

In Part 1 (p. 9), we defined the magnetic flux Φ by the integral of A around a closed path. In the coaxial structure, we choose the

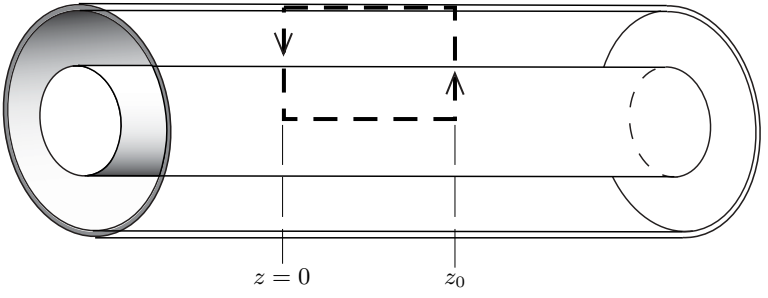


Figure 2.2 Path of integration for flux in Eq. 2.21

path shown in Fig. 2.2. Starting at $z = 0$, the path traverses a distance z_0 along the center of the inner conductor where the vector potential is $A_z(r_1)$; outward in the r direction into the outer conductor, where the vector potential is 0; back within the outer conductor to $z = 0$; and then inward in the $-r$ direction to the starting point. Because $A_r = 0$, the radial segments make no contribution, so the value of this integral is just $\Phi = z_0 A_z(r_1)$. From Eq. 2.20, the inductance L per unit length is thus

$$L = \frac{\Phi}{Iz_0} = \frac{\mu_0 \ln(r_2/r_1)}{2\pi} \tag{2.21}$$

2.7 Current–Voltage Relation

We have seen that a signal propagating along the coaxial structure is represented by a four-potential. The circuit variables (voltage and current), being manifestations of the components of the same four-potential, must be related. We can determine this relation by finding the relation between the charge density ρ and the current density $\vec{J}$ that results from the **conservation of charge.** In four-vector notation, charge conservation is expressed as

$$\Box \bullet \mathbf{J} = \nabla \cdot \vec{J} + \frac{\partial \rho}{\partial t} = 0 \tag{2.22}$$

The positive sign results from one negative sign in the scalar product and another in the definition of the four-gradient. If we assume that the four-current is confined to a thin skin at the surface, that the thickness of the skin is negligible compared to the macroscopic

dimensions of the conductors, and that therefore $\vec{J}$ flows in the z direction only, Eq. 2.22 becomes

$$\frac{\partial J_s}{\partial z} = -\frac{\partial \rho_s}{\partial t} \tag{2.23}$$

Any solution of the form $f(t \pm z/c)$ (Eq. 2.12) gives us

$$J_s = c\rho_s \qquad \Rightarrow \qquad I = c\,Q \tag{2.24}$$

This result is indeed as remarkable as it is easy to remember: *The current is just the net charge moving at the velocity of light!*

Substituting Eq. 2.24 into Eqs. 2.18 and 2.20, we obtain the relation of the vector and scalar potentials:

$$V = c\,A_z(r_1) \tag{2.25}$$

2.8 Electron Interaction

At this point, it is worthwhile to reflect on how we got here. We started with a relativistic expression for the potentials $\vec{A}$ and $\mathcal{V}$, in which the currents and charges play the role of sources. We found a solution for these relations in the region where no sources are present. We then derived what the sources *must be* at the surfaces of the conducting boundaries to generate these solutions. At first blush, this procedure seems backward: Why not start with the sources and derive the potentials? In Part 1 (p. 9), we found that either procedure will provide a solution. In most circumstances, it is the *interaction* of the collective electron system that dominates the behavior, rather than the configuration at a particular location: Charges move about such that their behavior is consistent with that of all other charges. The potentials $\vec{A}$ and $\mathcal{V}$ are not degrees of freedom of their own—they are, as their name implies, the *potential for interaction* of the collective degrees of freedom of the electron system. They represent the net effect of all charges in the system on an infinitesimal charge at a given point. The solutions of Eq. 2.10 in a region where $\mathbf{J} = 0$ are just the possible modes of interaction of charges on the boundaries, constrained by the shape of the region. If the charges actually present on the boundaries can fulfill the requirements of a particular solution, that solution will represent a possible mode of behavior of the system. The superconducting nature of the boundaries provides us the additional

information necessary to establish boundary conditions on the solutions (Eqs. 2.18 and 2.20). If a solution is indeed correct, it must be consistent with the Einstein–de Broglie relation for the electron wave function developed in Part 1 (p. 9). We can visualize these relations in the present context as follows: In the interior of the superconductor, the wave function's phase φ advances with time according to its frequency $-\partial\varphi/\partial t = \omega = q_0\mathcal{V}/\hbar$, and advances with distance according to its wave vector $\nabla\varphi = \vec{k} = q_0\vec{A}/\hbar$. Stated in four-vector notation, these relations take the following form:

$$\mathbf{k} = \frac{q_0}{\hbar}\mathbf{A} \tag{2.26}$$

By combining this expression with the boundary conditions given in Section 2.5 (p. 40), we can understand the potentials in the following way: The vector potential represents the propagation vector of the electron wave function at which there is zero current, and the scalar potential represents the frequency of the electron wave function at which there is zero charge.

2.9 Propagation of Step Function

The relationship between the four-potential and the phase of the electron wave function is illustrated most clearly by a specific example: a coaxial line driven by a voltage step $V_0\,u(t)$ applied by a signal source to the center conductor of the line at $z = 0$. We assume that the initial voltage along the inner conductor is zero, and visualize the unit step function $u(t)$ as rising smoothly from 0 to 1 within a **rise time** that is short compared with the time required for a signal to propagate a macroscopic distance along the line, but that is long compared with r_2/c. According to Eq. 2.12, the step travels in the positive z direction with velocity c:

$$V(z,t) = V_0\,u(t - z/c) \tag{2.27}$$

In the region $z < ct$, the voltage has reached its steady value V_0. From Eqs. 2.25 and 2.26, the electron wave function in that region has

$$\begin{aligned} \omega &= \frac{q_0}{\hbar}\,V_0 \\ k_z &= \frac{\omega}{c} = \frac{q_0}{\hbar}\,\frac{V_0}{c} \end{aligned} \tag{2.28}$$

After a time t has passed, the wave front will have advanced a distance $z = ct$, and the total phase accumulation in the region behind the wave front will have increased by ωt. The phase accumulation per unit of additional distance behind the wavefront is just ω/c. Thus, k_z is consistent with ω only if the wave front propagates with velocity c.

We can derive the same result by examining the phase of the wave function in the neighborhood of the wave front. In the region $z > ct$, the step has not yet arrived, and therefore $\mathbf{A} = 0$, $\omega = 0$, and $k_z = 0$. The wave function in the neighborhood of the step evolves as follows: As the step arrives at a point on the line, V rises and the phase begins to advance. As the potential rises further, the rate of advance increases. From Eq. 2.26, we can express the phase increment $d\varphi$ at some point on the wave front in four-vector notation:

$$d\varphi = \mathbf{k} \bullet d\mathbf{R} = k_z dz - \omega dt \tag{2.29}$$

For a positive voltage, the energy $\hbar\omega$ of a negative electron is negative relative to its reference value in the undisturbed portion of the line. Because $\omega = -\partial\varphi/\partial t$, the phase of the electron wave function advances for positive voltages. The electron wave function thus has positive phase behind the wave front. The spatial derivative $k_z = \partial\varphi/\partial z$ within the wave front is thus negative, corresponding to motion of electrons in the $-z$ direction, as required to charge the uncharged portion of the line to a positive voltage. Because $\partial\varphi/\partial t$ is opposite in sign to $\partial\varphi/\partial z$, the contours of constant phase propagate in the $+z$ direction. From Eq. 2.25, both the spatial and temporal components of $\mathbf{A}$ are given by the voltage V:

$$d\varphi = \frac{q_0}{\hbar}\left\{\frac{V}{c}\,dz - V\,dt\right\} \tag{2.30}$$

If the potential is of the form $V = f(t - z/c)$, then a point on the wave front of a given voltage corresponds to a point of stationary phase. Thus, the wave front propagates in the same manner as the zero crossings of the electron wave function. We have thus come full circle: The Einstein–de Broglie relations for the electron wave function, expressed in a relativistically invariant form, *require* that a wave front propagate with the ve-

locity of light.[12] It has become clear that the propagation of the electromagnetic wave and the propagation of the electron wave function are, as Einstein posited, two views of the same reality.

2.10 Waveguides

The coaxial structure we have considered is a particularly simple example of a transmission line; we chose it to illustrate the relativistic nature of the electron wave function as clearly as possible. There are, however, many transmission lines in which electromagnetic signals travel at velocities slower than c: for example, a rectangular superconducting tube, elongated in the z direction. The internal dimensions of the tube are w in the x direction, and h in the y direction. For the interior of the tube, $\mathbf{J} = 0$, and Eq. 2.10 becomes

$$\frac{\partial^2 \mathbf{A}}{\partial x^2} + \frac{\partial^2 \mathbf{A}}{\partial y^2} + \frac{\partial^2 \mathbf{A}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0 \tag{2.31}$$

The simplest nontrivial propagating solution of Eq. 2.31 is of the form

$$\mathbf{A} = \mathbf{A_0}\ \cos k_x x\ e^{ik_z z}\ \cos \omega t \tag{2.32}$$

with $k_x = \pi/w$. Substituting Eq. 2.32 into Eq. 2.31, we obtain

$$k_z^2 = \frac{\omega^2}{c^2} - \left(\frac{\pi}{w}\right)^2 \tag{2.33}$$

Modes of this kind have no definite velocity of propagation of the kind we found for the coaxial line (Eq. 2.12). The velocity

[12] In the electromagnetic limit, the contribution of the electron mass to the total phase accumulation is negligible compared with that due to the electromagnetic interaction of the collective electron system. In this limit, the propagation velocity approaches c. In Hertz's early experiments, there was a discrepancy between the measured velocity of propagation in space and that along a transmission line. A lively debate ensued between Hertz and other researchers. In the end, Hertz was "amply satisfied by the experiments which MM. Sarasin and de la Rive have carried out ... (see Archives de Genève, **29**, pp. 358 and 441). These experiments have proved the equality of the velocity in air and in wires, and have thus established the full agreement between experiment and theory." (p. 14 in Ref. 43)

of propagation of a sinusoidal signal depends on the signal's frequency. Below a certain frequency ω_c, called the **Cutoff Frequency**, the structure does not support propagating solutions at all. From Eq. 2.33

$$\omega_c = \frac{c\pi}{w} \tag{2.34}$$

For $\omega < \omega_c$, k^2 is negative, and hence k is imaginary, representing a signal that dies out exponentially with z. Solutions of this kind are called **Evanescent Waves**. Above the cutoff frequency, we can define two velocities: the **Phase Velocity** $v_p = \omega/k_z$, which indicates how fast the contours of constant phase travel, and the **Group Velocity** $v_g = \partial\omega/\partial k_z$, which indicates how fast the signal amplitude travels. Differentiating Eq. 2.33 with respect to k_z, we obtain

$$\frac{\omega}{k_z}\frac{\partial\omega}{\partial k_z} = v_p v_g = c^2 \tag{2.35}$$

The phase velocity is always faster than c, and the group velocity is slower than c; the product of the two velocities is equal to c^2. The group velocity describes the *net motion* of the charge, and it is sometimes useful to think of it as "the velocity of the electrons." At the fundamental level, however, there is only one velocity associated with the electron wave function: the velocity of light c. Lower velocities are the result of restrictions on the freedom of the electron wave function, an extreme example being electrons in a block of material, as described in the Appendix (p. 115). In that case, the electron wave function forms a standing wave, and there is no net propagation of charge. In the case of the rectangular waveguide, a certain part of the propagation vector was tied up in the x direction, and was therefore unable to fully contribute to propagation in the z direction. Detailed description of the mode behaviors of several useful waveguide configurations are given in (40).

2.11 Summary

The behavior of collective electron systems is dominated by the interaction of each element with all the others. It would seem, at first, that determining the modes of a given system should be a nasty, complex, nonlinear problem. Two circumstances conspire to make the task tractable. First, the principle of linear superposition holds for the four-potential, if the sources are known.

Second, the four-current is confined to a thin skin at the surface of the superconductor. The first circumstance does not help us directly, because we do not know the distribution of the four-current. However, the second circumstance allows us to derive simple boundary conditions relating the propagation four-vector of the electrons in the superconductor to the four-potential at the surface. These conditions allow us to invert the problem, and to solve the linear equations for the four-potential, given the shape of the superconducting boundaries. In this way, the seemingly intractable, nonlinear self-consistency problem associated with the electron wave function is, to an excellent approximation, converted to a boundary-value problem associated with the linear differential equation for the four-potential. Solutions that are consistent with the configuration of the superconducting boundaries are possible behaviors of the system. This view provides the conceptual basis for the historic success of the boundary-value approach to electromagnetic problems.

As Einstein (28) said,

> there is only *one* reality to be described, which happens to have two different aspects; and the theory ought to recognize this from the start instead of doing things twice.

By following his advice, we found that the behavior of the system can be expressed in terms of the phase of the wave function, of the four-potential, or of the circuit variables (voltage and current). These three sets of variables do indeed represent aspects of the same reality—they all represent the same underlying degrees of freedom.

Part 3

Electromagnetic Energy

I believe that we should adhere to the
strict validity of the energy principle
until we shall have found important
reasons for renouncing this guiding star.

—Albert Einstein[1]

3.1 Energy in Electromagnetic Systems

In Part 1 (p. 9), we found that the interaction of a current with itself, or with other currents, can be expressed in terms of the kinetic energy of the electrons. This interaction energy scales quadratically with the number of interacting electrons. In Part 2 (p. 31), we found that the modes of behavior of regions bounded by superconducting walls can be expressed in terms of the phase of the electron wave function; the electromagnetic four-potential; and the circuit variables, voltage, and current. Each of these sets of variables is useful in clarifying certain aspects of system behavior, although all represent the same underlying degrees of freedom. A natural expression of the electrostatic interaction of charges results from the relativistic invariance of the underlying four-vector representation. The energy associated with electrostatics is most naturally expressed as a potential energy. We shall find that the concepts of kinetic energy, potential energy, and the conservation of total energy have a natural relativistic representation, and retain their utility across all three sets of variables.

Maxwell, in his original treatment of the subject, had a conceptually correct formulation of the kinetic and potential energies associated with an electromagnetic ensemble. His treatise contains the following equations:[2]

[1] This quotation appeared as part of the debate between Einstein and Ritz, in which Einstein pointed out that using retarded-only potentials violated the conservation of energy. The English version of this paper can be found on page 357 of Ref. 55. I discuss this issue at length in Part 4 (p. 73).

[2] I have expressed these equations (found in Art. 87 of Ref. 56 and Art. 636 of Ref. 33, respectively), and also the equations found in subsequent quotations in the notation used in Part 2 (p. 31).

$$W = \frac{1}{2}\sum q_n V_n \qquad \text{Potential Energy}$$
$$W = \frac{1}{2}\int \vec{A}\cdot\vec{J}\, d\text{vol} \qquad \text{Kinetic Energy} \tag{3.1}$$

After deriving the second of these equations, Maxwell showed that it is equivalent to a spatial integral involving the magnetic field. Because he was firmly committed to the ether as the medium in which all electromagnetic phenomena takes place, he chose the "energy in the field" as primary. His comment on the choice is instructive:

> The electrokinetic energy of the system may therefore be expressed either as an integral to be taken where there are electric currents, or as an integral to be taken over every part of the field in which magnetic force exists. The first integral, however, is the natural expression of the theory which supposes the currents to act upon each other directly at a distance, while the second is appropriate to the theory which endeavors to explain the action between the currents by means of some intermediate action in the space between them. As in this treatise we have adopted the latter method of investigation, we naturally adopt the second expression as giving the most significant form to the kinetic energy.

That choice has carried over into essentially all modern treatments, in spite of the vigorous discussion surrounding the Aharnov and Bohm paper (44), and the experimental verification that it is indeed $\vec{A}$, not $\vec{B}$, that connects with the phase of the electron wave function. Feynman's delightful discussion of which field is "real" (Sec. 15-5 in Ref. 5) ends with the statement:

> $\vec{E}$ and $\vec{B}$ are slowly disappearing from the modern expression of physical laws; they are being replaced by $\vec{A}$ and $\mathcal{V}$.

Sommerfeld introduces the Lorentz-invariant quantity

$$\mathcal{S} = \mathbf{J}\bullet\mathbf{A} = \left(\vec{J}\cdot\vec{A} - \rho\mathcal{V}\right) \tag{3.2}$$

which he calls the **Schwarzschild invariant** (p. 269 in Ref. 54). Sommerfeld's definition is accompanied with the following comment:

> K. Schwarzschild, Göttinger Nachr. 1903. See in particular the first of the three papers . . . Note the date of publication 1903! Thus Schwarzschild arrived intuitively at the correct postulate of the theory of invariants six years ahead of Minkowski.[3]

Feynman, Morinigo, and Wagner (p. 32 of Ref. 6) make the following comment:

> The guts of electromagnetism are contained in the specification of the interaction between a current and the field as $\mathbf{J} \cdot \mathbf{A}$.

The Schwarzschild invariant (Eq. 3.2) is the difference between two terms: the first is twice the Maxwell *kinetic* energy, and the second is twice the Maxwell *potential* energy. When integrated over all four coordinates of space–time, this quantity has the units of energy×time; it is called the **action** of the system. When divided by $\hbar$, this integral is dimensionless. For an isolated free particle, the phase accumulation of the wave function along a path can, under certain conditions, be associated with this integral (32). As we have shown in Part 1 (p. 9), for a collective system, the magnetic flux integral is the phase accumulation of the wave function multiplied by $\hbar/q$. From our point of view, the action is flux×charge, rather than energy×time. We shall find the Schwarzschild invariant useful for determining the relative magnitude of potential and kinetic energies, rather than for evaluating the phase of the wave function.

In a collective system, the sum of the potential and kinetic energy terms, when integrated over all space, represents the total electrodynamic energy. As Maxwell indicated, if the accounting is done this way, there is no additional "energy in the field" for which to account. Both forms of energy are the result of interaction; they are both zero in the absence of four-current, or in the absence of four-potential (from other four-current); only where the two overlap is there energy of interaction. Although the total energy is not a Lorentz invariant, we shall find that its utility as a concept is not thereby compromised in the slightest.

The factor of 2 discrepancy between Maxwell's energy density and that which appears in the Schwarzschild invariant can be understood as follows: If a current element $\vec{j}$ is introduced into a

[3]W. G. Wagner called this reference to my attention, noting that it was published two years before Einstein's paper on special relativity.

vector potential $\vec{A}$ generated by other currents, the energy of the space containing current element $\vec{j}$ is increased by the Maxwell energy $\vec{A} \cdot \vec{j}/2$. However, by the reciprocal nature of magnetic interaction, current element $\vec{j}$ increases the vector potential at the source currents enough to increase their energy by an exactly equal amount. So, when current element $\vec{j}$ is added to the system, the total energy of the system is increased by $\vec{A} \cdot \vec{j}$. If we integrate over the whole system, the Maxwell expression gives the correct total energy:

$$\begin{aligned} W &= W_{\text{kin}} + W_{\text{pot}} \\ W_{\text{kin}} &= \frac{1}{2} \int \vec{J} \cdot \vec{A} \, d\text{vol} \\ W_{\text{pot}} &= \frac{1}{2} \int \rho \mathcal{V} \, d\text{vol} \end{aligned} \tag{3.3}$$

If we integrate over the volume of a small four-current element in a large applied four-potential, we use twice the Maxwell expression to obtain the total energy associated with that element.

3.2 Elementary Examples

3.2.1 Inductor For the inductor that we studied in Part 1 (p. 9), in the presence of steady currents, $\mathcal{V} = 0$, so all of the energy is kinetic:

$$W_{\text{kin}} = \frac{1}{2} \int \vec{A} \cdot \vec{J} \, d\text{vol} \tag{3.4}$$

where the integral is taken over the region of the wire in which current is flowing. We can make use of the fact that current is flowing along the length of the wire to divide the volume integral into two parts: the first with respect to area elements da in a cross section of the wire, and the second with respect to area elements $d\vec{l}$ along the length of the wire:

$$W_{\text{kin}} = \frac{1}{2} \int \int J \, da \; \vec{A} \cdot d\vec{l} = \frac{1}{2} \; I \int \vec{A} \cdot d\vec{l} \tag{3.5}$$

where we have made use of the fact that I is independent of l. We recognize in the second form of the integral our old friend, the

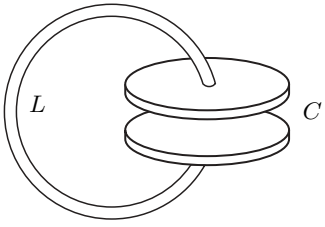


Figure 3.1 Coherent quantum resonator

magnetic flux Φ:

$$W_{\text{kin}} = \frac{1}{2}\, I\, \Phi = \frac{\Phi^2}{2L} = \frac{LI^2}{2} \tag{3.6}$$

in agreement with Eq. 1.2.

3.2.2 Capacitor We can conduct a similar evaluation of the potential energy associated with an electrostatic problem. For example, a capacitance C stores a total charge $Q = CV$ on a conducting electrode held at potential V with respect to the reference electrode. In this case, $A = 0$, and $J = 0$, so $\mathcal{W} = \mathcal{V}\rho/2$, which we integrate over the volume of the electrode:

$$W_{\text{pot}} = \frac{1}{2}\int V\rho\, d\text{vol} = \frac{1}{2}VQ = \frac{CV^2}{2} \tag{3.7}$$

3.2.3 Resonator For the superconducting structure shown in Fig. 3.1, composed of a capacitor in parallel with an inductor, the current flowing out of the capacitor is $I = -C\,\partial V/\partial t$. This same current flows through the inductor: $V = L\,\partial I/\partial t$. The time derivative of the total energy is

$$\begin{aligned}\frac{\partial}{\partial t}\Big(W_{\text{kin}} + W_{\text{pot}}\Big) &= \frac{\partial}{\partial t}\left(\frac{LI^2}{2}\right) + \frac{\partial}{\partial t}\left(\frac{CV^2}{2}\right)\\ &= LI\frac{\partial I}{\partial t} + CV\frac{\partial V}{\partial t}\\ &= VI - IV = 0\end{aligned} \tag{3.8}$$

Energy can shift between kinetic and potential forms, but the total energy does not change with time. The principle involved,

conservation of energy, is much more general than is indicated by this simple example. The total energy is conserved for any particular observer, but it is not a Lorentz invariant: Its magnitude depends on the motion of the observer.

It is particularly instructive to view this example from the perspective of the electron wave function. If we use one electrode of the capacitor as a reference, the wave function phase in the other electrode is advancing at a rate $\omega = -qV$. The total phase accumulation in the inductor is thus

$$\varphi = \int \omega \, dt = \int \vec{k} \cdot d\vec{l} \tag{3.9}$$

where $\vec{k}$ is positive in the direction leading from the reference node to the node in question. For the simplest mode of the system, k is the same at any point along the length l of the wire forming the inductor, and hence

$$\frac{dk}{dt} \propto \omega \tag{3.10}$$

The propagation vector k determines the rate at which charge propagates along the wire to the node in question

$$\frac{dQ}{dt} \propto k \tag{3.11}$$

The electrostatic interaction in the capacitor relates potential directly to the charge: $V \propto Q$ and, therefore, $\omega \propto -Q$. Gathering up these facts, we obtain the usual differential equation

$$\frac{d^2\omega}{dt^2} \propto -\omega \tag{3.12}$$

The proportionality constants are embodied in the capacitance C and inductance L of the elements. Thus, as we found in Part 1 (p. 9) and Part 2 (p. 31), it is possible to reason about the behavior of electromagnetic systems directly, using only the phase of the wave function, of which the circuit and field variables are a convenient abstraction.

3.3 Energy of Propagating Solution

The Schwarzschild invariant allows us to make relativistically correct statements about the difference between kinetic and potential energy. For a single signal propagating on the coaxial transmission line treated in Part 2 (p. 31), we can evaluate the two energy densities by substituting Eqs. 2.24 and 2.21 into Eq. 3.2:

$$\mathcal{W} = \frac{1}{2}\left(\vec{A}\cdot\vec{J} - \rho\mathcal{V}\right) = \frac{1}{2}\left(\frac{\mathcal{V}}{c}\cdot\rho c - \rho\mathcal{V}\right) = 0 \tag{3.13}$$

At any point along the line, at any given time, *the kinetic and potential energies are exactly equal*. Observers moving at different velocities will disagree about the magnitude of either energy density, but will always agree that the two energy densities are equal. We can use the inductance per unit length from Eq. 2.21 in Eq. 3.6 and the capacitance per unit length from Eq. 2.19 in Eq. 3.7 to express the energy per unit length of the coaxial line in terms of the circuit variables:

$$W_{\text{kin}} = W_{\text{pot}} \quad \Rightarrow \quad LI^2 = CV^2 \tag{3.14}$$

The equality of kinetic and potential energy is a property of a single propagating solution; when the behavior requires a superposition of more than one solution, the two energies are no longer equal, as we shall see in several examples.

3.4 Characteristic Impedance

Eqs. 3.13 and 3.14 apply to a single solution of the form $f(t \pm z/c)$. That form represents a signal propagating in one direction, carrying energy as it goes. The energy propagating along the line is provided by the input signal generator driving the line at $z = 0$, with voltage $V(0,t)$ and current $I(0,t)$. Substituting $I = Qc$ from Eq. 2.24 into Eq. 2.19, we obtain the relationship of V to I, in the form of the characteristic impedance Z_0 of the line:

$$Z_0 = \frac{V}{I} = \frac{\mu_0 c}{2\pi}\ln\frac{r_2}{r_1} = \frac{\ln\left(r_2/r_1\right)}{2\pi}R_{\text{FS}} \tag{3.15}$$

The line looks like a resistor to the signal generator: The voltage is always in phase with the current, independent of the function $f(t)$.

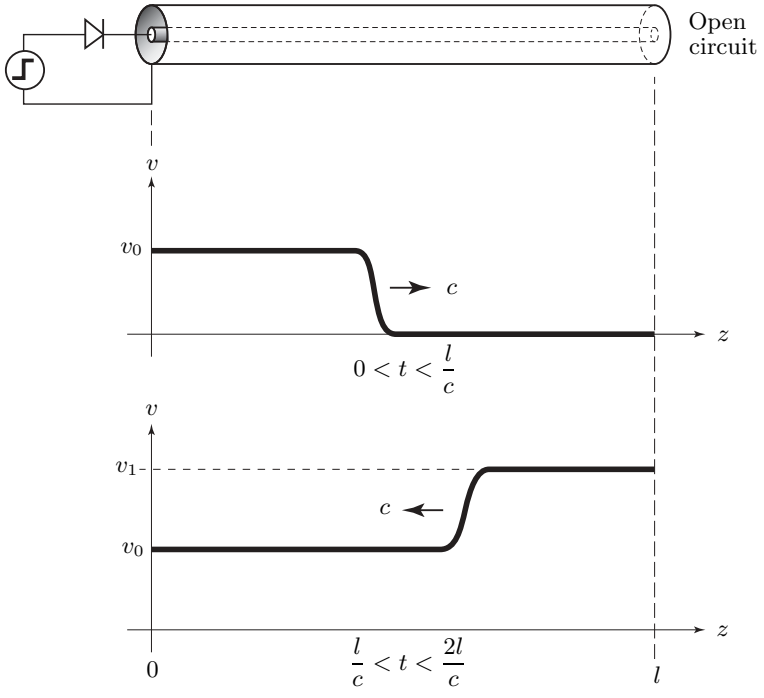


Figure 3.2 Signals propagating on finite coaxial line

The quantity $R_{\mathrm{FS}} = \mu_0 c = 376.7303135$ ohms is a fundamental unit of resistance, called the **impedance of free space**.

We can also derive the relationship between current and voltage from the kinetic and potential energies, which, for a single solution, are equal. From Eq. 3.14,

$$LI^2 = CV^2 \qquad \Rightarrow \qquad Z_0 = \frac{V}{I} = \sqrt{\frac{L}{C}} \tag{3.16}$$

Using the capacitance and inductance per unit length from Eqs. 2.19 and 2.21, we recover Eq. 3.15. Once again, these expressions are valid for only a single solution traveling away from the source.

3.5 Finite Coaxial Line

In Part 2 (p. 31), we saw that the energy associated with a signal propagating in a transmission line can be expressed in terms of the propagation four-vector of the electron wave function, the four-potential, or the voltage and current in the line. Eqs. 3.13 and 3.14

show us that kinetic and potential energies are equal for a single propagating solution. We have not as yet seen the conservation of energy at work in the context of a propagating wave. A particular example that illustrates the relations involved is shown in Fig. 3.2: a coaxial line of length l, driven at one end. We use a signal source to apply a voltage step $V_0 u(t)$ to the line at $z = 0$ through a diode, so the source can supply current to the line, but the line cannot push current back into the source. The initial voltage along the line is zero, and the end of the line at $z = l$ is an open circuit. The initial phase of the behavior is the propagation of a step function, described in Part 2 (p. 31): The step travels in the positive z direction with velocity c:

$$V(z,t) = V_0\, u(t - z/c) \qquad t < \frac{l}{c} \tag{3.17}$$

The open-circuit boundary condition at $z = l$ requires that $I = 0$, but that requirement cannot be satisfied by the solution of Eq. 3.17 for $t > l/c$. The source is supplying a current $I = V_0/Z_0$, which is traveling in the z direction at the speed of light. As we found in Eq. 3.14, this moving charge has kinetic energy just equal to the potential energy of the line charged to a voltage V_0. When the step reaches the end of the line, that kinetic energy keeps the charge moving; because it cannot leave the line, however, the charge starts piling up at the end. That accumulation of charge increases the voltage at the end of the line, creating an electric field in the direction that slows down the oncoming charge. When the potential at the end of the line has reached some new voltage V_1, all the kinetic energy at $z = l$ has been converted to potential energy, the charge has been brought to rest, and therefore $I = 0$. For $z < l$, the charge is still moving, so the piling up of charge continues, creating a new step, a **reflection**, moving back toward $z = 0$. At $t = 2l/c$, the step from V_0 to V_1 reaches the input; the diode reverse biases, terminating input current flow; the entire line is now at V_1; and all charge is stationary. What is V_1? At $t = l/c$, the initial step has just reached the end, and a charge $Q = Il/c$ has entered the line, just sufficient to charge the line to V_0. By $t = 2l/c$, an additional charge Q has entered the line, making the total charge $2Q$. Twice the charge implies twice the voltage, so $V_1 = 2V_0$.

Let's see whether the energy checks out: At $t = 0$, both V and I are zero, so both the potential and kinetic energies are zero. At $t = l/c$, the total energy delivered to the line by the signal source is $IV_0 l/c = QV_0$, of which half is potential energy $QV_0/2$, and the other half, by Eq. 3.14, is kinetic energy. By $t = 2l/c$, twice the energy has been delivered to the line, all of which now resides in potential energy, so the potential energy is four times what it was at $t = l/c$. Because the potential energy scales as the square of the voltage, the voltage must be $2V_0$. So, everything checks out. But, what about Eq. 3.13? How can we have a situation where we have potential energy and no kinetic energy? We derived that equation for a *single solution* of the form $f(t - z/c)$. But for $t > l/c$, the potentials on the line must be described by a superposition of two solutions, one propagating to the right, and the other propagating to the left. Because it is derived from the current vector, the kinetic energy of two superposed solutions does not just add up: When the currents from the two solutions cancel out, the net current is zero, and so is the associated kinetic energy. A similar situation holds for potential energy: When two solutions superpose such that the voltages cancel, the net voltage is zero, and so is the potential energy.

3.6 Reflection and Transmission

The open-circuit at $z = l$ in the previous section is a special case of a discontinuity in the transmission line. A more general case is that of an abrupt change in the dimensions of the line at $z = 0$, as shown in Fig. 3.3. According to Eq. 3.15, such a discontinuity can cause a discontinuity in the characteristic impedance of the line, from Z_0 in the region $z < 0$ to Z_1 in the region $z > 0$. We start with a solution $f(t - z/c)$—a step function, for example—approaching the origin from the $-z$ direction. As long as this **incident wave** has not reached the discontinuity, we have $V = 0$ and $I = 0$ everywhere in advance of the oncoming wave. Once the wave front has reached the discontinuity, however, we can no longer satisfy the boundary conditions in both regions with a single solution. In general, there is a **reflected wave** from any kind of discontinuity, and a **transmitted wave** propagating into the region of different characteristic impedance. Therefore, we try a solution of the form

$$\begin{aligned} V &= f(t - z/c) + \alpha f(t + z/c) \qquad z < 0 \\ V &= \beta\, f(t - z/c) \qquad\qquad\qquad\quad z > 0 \end{aligned} \tag{3.18}$$

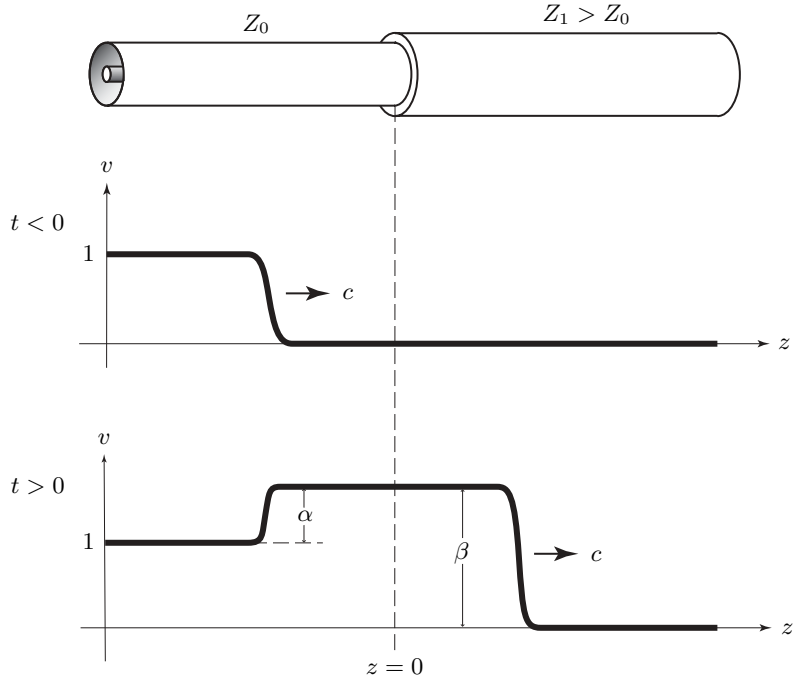


Figure 3.3 Reflection and transmission of a signal at a discontinuity

In each section of the line, the current for each individual solution is equal to V/Z. The current is, however, a vector quantity: It is positive for a positive-voltage signal traveling in the $+z$ direction, and negative for a positive-voltage signal traveling in the $-z$ direction. Thus, the current contributed by the reflected wave will be in the direction opposite from that contributed by the incident wave:

$$\begin{aligned} I &= \frac{1}{Z_0}\Big(f(t - z/c) - \alpha f(t + z/c)\Big) \qquad z < 0 \\ I &= \frac{\beta}{Z_1} f(t - z/c) \qquad z > 0 \end{aligned} \tag{3.19}$$

At the discontinuity ($z = 0$), two conditions must hold: 1. the voltage for the $z > 0$ solution must be the same as that for the $z < 0$ solution, and 2. the current for the $z > 0$ solution must be the same as that for the $z < 0$ solution. From Eqs. 3.18 and 3.19,

these conditions become

$$\begin{aligned} &1. \quad 1+\alpha = \beta \\ &2. \quad (1-\alpha)Z_1 = \beta Z_0 \end{aligned} \tag{3.20}$$

From this, we can determine α and β:

$$\begin{aligned} \alpha &= \frac{Z_1 - Z_0}{Z_1 + Z_0} \\ \beta &= \frac{2Z_1}{Z_1 + Z_0} \end{aligned} \tag{3.21}$$

Because α determines the amplitude and sign of the reflected wave, it is called the **reflection coefficient** of the discontinuity. Similarly, β determines the amplitude of the transmitted wave, and hence is called the **transmission coefficient.** For an open circuit, as at $z = l$ in the previous section, $Z_1 = \infty$, so $\alpha = 1$, indicating that the reflected wave is equal to the incident wave. For a short circuit, $Z_1 = 0$, so $\alpha = -1$, indicating that the reflected wave is equal in magnitude and opposite in sign to the incident wave, and $\beta = 0$, indicating that no wave is transmitted. For $Z_1 = Z_0$, $\alpha = 0$, indicating no reflected wave, and $\beta = 1$, indicating that the transmitted wave is equal to the incident wave, as it must be.

We can also use the conservation of energy to derive the general relation for transmission and reflection coefficients. A signal carries total energy V^2/Z per unit time. Before the reflection, $V = 1$, so the incident signal has energy $1/Z_0$. After the reflection, the reflected signal has energy α^2/Z_0, and the transmitted signal carries energy β^2/Z_1. Conservation of energy requires that the sum of the energies in the transmitted and reflected signals be equal to the energy in the original incident signal:

$$\frac{1}{Z_0} = \frac{\alpha^2}{Z_0} + \frac{\beta^2}{Z_1} \tag{3.22}$$

This expression, with the conservation of charge (the second half of Eq. 3.20), results in Eq. 3.21.

It is instructive to understand these relations in terms of the behavior of the electron wave function. The Einstein relation tells

us that voltage is just another name for the rate of change of the wave-function phase. Continuity of the phase requires that the frequency at one point on the line not be different from that at an immediately adjacent point. Thus, the continuity of voltage across the discontinuity is just a result of the continuity of phase. Similarly, the continuity of current is a direct result of the conservation of charge. So the behavior of the transmission line at a discontinuity is, once again, a direct manifestation of the quantum properties of the collective electron system.

3.7 Infinite-Response Solutions

The behavior described for the finite line provided us with an example of finite response: Conditions in the line reach a steady state after a finite time has elapsed. Finite sections of transmission line are also capable of supporting modes of behavior that do not die out with time—the constant persistent current in the closed loop considered in Part 1 (p. 9) being the simplest example. Fig. 3.4 shows a section of coaxial line of length l, shorted at both ends. This configuration provides an example of a closed loop with interesting time-dependent behaviors. At the center of the structure, we make a small opening in the outer conductor, thus gaining access to the inner conductor. We imagine, for the moment, that we can inject a short pulse of current into the inner conductor at $t = 0$, and observe the ensuing voltage response without creating a discontinuity. During the pulse, one-half of the total current $2I_0$ that we inject flows in the $+z$ direction; the other half flows in the $-z$ direction. Thus, two voltage pulses of amplitude $V_0 = I_0 Z_0$ propagate away from the center, one toward each end. At $t = l/2c$, when the pulses reach their respective ends of the line, they are reflected with a coefficient -1, according to Eq. 3.20, and begin propagating in the opposite direction with amplitude $-V_0$. At time $t = l/c$, the pulses encounter each other at the center of the line, and the voltages add, resulting in an observed pulse of amplitude $-2V_0$. The individual pulses continue on to the opposite ends of the line, where, at $t = 3l/2c$, they are once again reflected with coefficient -1. At this time, each pulse has amplitude $+V_0$, and is propagating in its original direction. At $t = 2l/c$, the pulses overlap at the center of the line, resulting in a pulse of amplitude $+2V_0$, and we have arrived at our initial condition. The behavior continues, as long

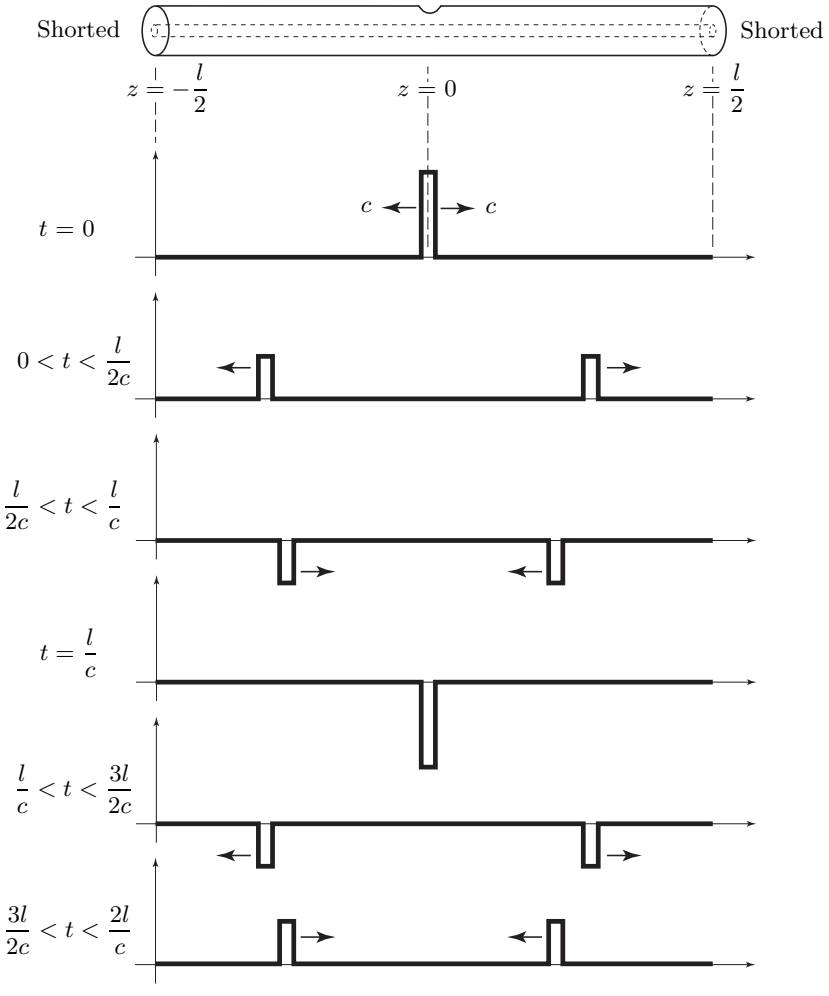


Figure 3.4 Signals propagating on finite section of coaxial line shorted at both ends

as it is not disturbed,[4] with period $2l/c$. Because the period of the behavior is twice the transit time, a structure of this kind—shorted at both ends—is called a **half-wave line.** It is one of the basic **resonant structures** that we can construct from a fi-

[4]In this discussion, we assume an ideal superconductor. In real superconductors, a time-varying current of the kind considered here has a long, but finite, lifetime. Ramo and colleagues provide a good discussion (40).

nite section of transmission line. As we noted in connection with Eq. 1.12, a description of the solution requires a superposition of four types of signals, in this case, positive and negative pulses traveling in both the $+z$ and $-z$ directions. The change in sign of V corresponds to the time reversal of the wave function. Before the signal is injected, the phase accumulation around a closed path along the inner conductor and back along the outer conductor is zero. The signal does not destroy the superconducting state, so the total phase accumulation around this path remains zero. The phase accumulation in one pulse is positive; that in the other pulse is negative. These interactions, described in terms of voltage and current, map directly onto the phase of the wave function by Eqs. 2.24, 2.21, and 2.22. When two pulses overlap each other, the propagation four-vector of the superposition is simply the sum of propagation four-vectors of the two individual pulses.

3.8 Local Conservation of Energy

We can understand the behavior at the ends of the line and in the region of pulse overlap by applying the principle of conservation of energy. Consider the superposition of two signals: I_1, V_1 and I_2, V_2. If the signals are propagating in the same direction, they can be summed into one signal and treated as a single solution. If they are propagating in opposite directions, the relative signs of I and V will be reversed, as we saw in the previous example. The total energy per unit length of the superposition is

$$\begin{aligned} W &= \frac{L}{2}(I_1 + I_2)^2 + \frac{C}{2}(V_1 + V_2)^2 \\ &= \frac{L}{2}\Big(I_1^2 + I_2^2 + 2I_1 I_2\Big) + \frac{C}{2}\Big(V_1^2 + V_2^2 + 2V_1 V_2\Big) \qquad (3.23) \\ &= \frac{L}{2}\Big(I_1^2 + I_2^2\Big) + \frac{C}{2}\Big(V_1^2 + V_2^2\Big) \end{aligned}$$

The cross terms cancel because $CV_1V_2 = -LI_1I_2$ (from Eq. 3.16) and because of the sign reversal just mentioned. Thus, the total energy of the superposition is exactly the same as the sum of the total energies of the two individual signals. We can see this

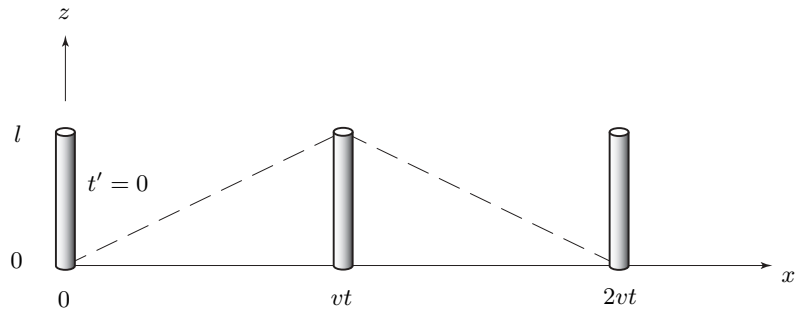


Figure 3.5 A signal propagating in a finite coaxial line, as viewed by a moving observer.

principle in action in the behavior of the half-wave line: Each of the two original pulses had kinetic energy $LI_0^2/2$ and an equal potential energy $CV_0^2/2$ per unit length. The total energy in each pulse was therefore $\tau c L I_0^2$, where τ is the pulse duration. When a positive-voltage pulse traveling in the $+z$ direction reaches the end of the line, it initiates a reflected negative-voltage pulse traveling in the $-z$ direction. The *currents* due to both of these superposed solutions are in the same direction, resulting in a current $2I_0$ in the region of overlap. The *voltages* due to the two solutions are, however, of the opposite sign, and therefore cancel in the region of overlap. When one-half of the pulse has disappeared into the end and reemerges as a negative shadow of its original self, the voltage is zero everywhere due to the complete cancellation of the two half-pulses. The potential energy is thus zero everywhere, but the kinetic energy in the region of overlap is $2LI_0^2$ per unit length—four times its original value. The total energy in the overlap region $\tau c/2$ in width is therefore $\tau c L I_0^2$, exactly the same as that in the original pulse.

The pulses also overlap in the center of the line—this time with the same sign, so the voltage in the region of overlap is $2V_0$. The two pulses are, however, traveling in opposite directions, so their currents cancel. In the region of total overlap of width τc, the kinetic energy is zero, and the potential energy is $2CV_0^2$ per unit length, or $2\tau c C V_0^2$ in total—exactly the same as the total energy in the separate pulses. So energy is indeed conserved, not only before and after the interaction, but also locally, during the interaction process.

3.9 Lorentz Transformation Revisited

Einstein wrote (20)

> a light signal, which is reflected back and forth between the ends of a rigid rod, constitutes an ideal clock

The propagation of a step function along the coaxial transmission line gives us a mechanism for relating time and distance. If we start a step function propagating in a section of transmission line of length l, shorted at both ends, the step will bounce back and forth between the two shorted ends. Each transit of the step from one end to the other requires a time $t = l/c$, or $t = 2l/c$ for a round trip. The arrival of the step at either end can be used as the "tick" of a clock, our basic time reference. Likewise, the length l of the transmission line serves as our "meter stick," the unit of length. The first experiment we perform is to compare the period t of our stationary clock to the period t' of the same clock viewed by an observer moving with constant velocity v in the $-x$ direction. The situation is illustrated in Fig. 3.5. The basic tenet of relativity is that electromagnetic waves propagate with velocity c, independent of the velocity of their source. In the time t' required for one transit as viewed by the moving observer, the step propagates a distance l in the z direction, and a distance vt' in the x direction. Because the x coordinate is orthogonal to the z coordinate, the total distance d traveled by the step is

$$d^2 = l^2 + v^2 t'^2 \tag{3.24}$$

We observe that $d = ct'$ and $l = ct$, so Eq. 3.24 can be written

$$c^2 t'^2 = c^2 t^2 + v^2 t'^2 \tag{3.25}$$

Combining terms and dividing by c^2, we obtain

$$t'^2 \left(1 - \frac{v^2}{c^2}\right) = t^2 \tag{3.26}$$

or

$$t' = \frac{t}{\sqrt{1 - v^2/c^2}} = \gamma\, t \tag{3.27}$$

where the expression

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{3.28}$$

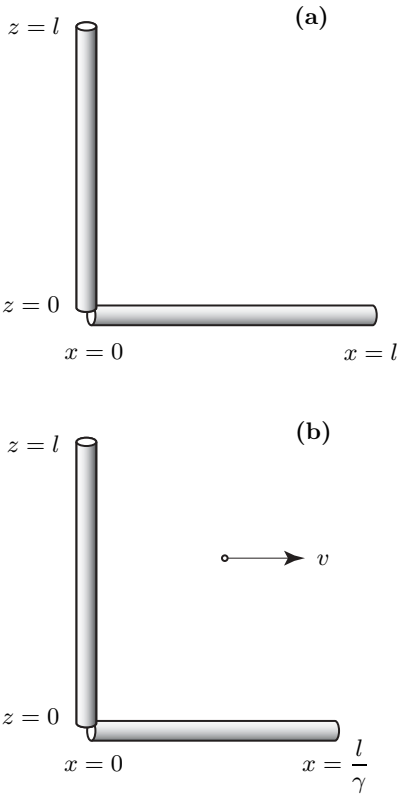


Figure 3.6 Two clocks: one parallel and one perpendicular to the direction of relative motion. View (a) is the configuration seen by an observer in the same coordinate system as the clocks. View (b) is the configuration seen by an observer moving with velocity $-v$ with respect to the clocks.

which occurs throughout relativity theory, is called the **Lorentz factor**; it is the factor by which moving clocks run slower than stationary clocks. The round-trip time required for the step to return to its original position in the moving frame is $2t'$.

For our second experiment, we require two clocks, as shown in Fig. 3.6. The first has one end located at the origin of the stationary (unprimed) frame of reference, and its direction of propagation is oriented parallel to the z axis, as before. The second clock also has one end located at the origin of the stationary frame of reference, but its direction of propagation is oriented along the x axis, parallel to the direction of motion of the moving observer. In the stationary frame of reference, the steps of both clocks are observed to arrive at the origin simultaneously. In the moving frame of reference, the arrival of steps at the origin is also simultaneous, but slower, as we saw in Eq. 3.27. An observer in the moving (primed)

frame infers that the step in the second moving clock travels in the $+x$ direction from the origin for a time t'_1 to the opposite end of the line, which is then at position $(l' + vt'_1) = ct'_1$. The observer infers that

$$t'_1 = \frac{l'}{c - v} \tag{3.29}$$

The step is reflected and travels in the $-x$ direction for a time t'_2 until it reaches the starting end of the line, having traveled a distance $(l' - vt'_2) = ct_2$. The moving observer infers that

$$t'_2 = \frac{l'}{c + v} \tag{3.30}$$

At time $2t' = t'_1 + t'_2$, the steps in both clocks have returned to their common origin, which is then at position $x = vt'$.

$$t' = \frac{t'_1 + t'_2}{2} = \frac{l'c}{c^2 - v^2} \tag{3.31}$$

Using $t = lc$ in Eq. 3.27 to eliminate t' from Eq. 3.31, we conclude

$$l' = l\sqrt{1 - v^2/c^2} = \frac{l}{\gamma} \tag{3.32}$$

The distance scale along the direction of motion is contracted by the Lorentz factor.

3.10 The Lorentz Transformation

We have derived how the time and distance scales in two frames of reference, moving with respect to each other, are related. We are now in a position to derive the actual coordinates in the moving frame in terms of those of the stationary frame. The stationary observer sees the x-directed clock with its origin at $x = 0$ and its other end at $x = l$, keeping time with round-trip period $t = 2l/c$. The moving observer sees the same clock with its origin at $x' = 0$ at time $t' = 0$, and both ends moving in the x' direction with velocity v. To the moving observer, any given distance l in the stationary frame appears as a distance $l' = l/\gamma$ in the moving frame. The moving observer applies this logic to the distance x from the origin of the stationary system, and thus infers

$$x' = \frac{x}{\gamma} + vt' \tag{3.33}$$

from which he infers

$$x = \gamma \left(x' - vt' \right) \tag{3.34}$$

Einstein's theory is based on two postulates, the first that electromagnetic waves propagate at velocity c in any inertial frame of reference, and the second that there is no preferred frame of reference. The second postulate is really one of symmetry, for any observer has the right to consider themselves stationary and others as moving. In the situation analyzed above, the observer in the primed system is certainly entitled to view the clocks as moving with velocity v in her frame of reference. She then writes Eq. 3.34 as

$$x' = \gamma \left(x + vt \right) \tag{3.35}$$

We can eliminate x from Eqs. 3.34 and 3.35 to obtain an expression for t:

$$t = \gamma \left(t' - \frac{vx'}{c^2} \right) \tag{3.36}$$

The complete Lorentz transformation for a four-vector

$$\mathbf{U} = \left[\vec{U}, U_t \right] \tag{3.37}$$

is

$$\begin{aligned} U'_x &= \gamma \left(U_x - \frac{v}{c} U_t \right) \\ U'_y &= U_y \\ U'_z &= U_z \\ U'_t &= \gamma \left(U_t - \frac{v}{c} U_x \right) \end{aligned} \tag{3.38}$$

Two four-vectors that transform according to Eq. 3.38 have the property that their dot product is the same in all inertial frames. We can prove this statement as follows:

$$\begin{aligned} \mathbf{U}' \cdot \mathbf{V}' &= U'_x V'_x + U'_y V'_y + U'_z V'_z - U'_t V'_t \\ &= \gamma^2 \left(U_x - \frac{v}{c} U_t \right) \left(V_x - \frac{v}{c} V_t \right) \end{aligned}$$

$$
\begin{aligned}
&+ U_y V_y + U_z V_z - \gamma^2 \left(U_t - \frac{v}{c} U_x\right) \left(V_t - \frac{v}{c} V_x\right) \\
&= \gamma^2 \left(1 - \frac{v^2}{c^2}\right) (U_x V_x - U_t V_t) + U_y V_y + U_z V_z \\
&= U_x V_x + U_y V_y + U_z V_z - U_t V_t = \mathbf{U} \cdot \mathbf{V}
\end{aligned}
\tag{3.39}
$$

3.11 Resistance in Coherent Systems

The resistance R_{FS} derived in Section 3.4 (p. 55) arises from the charge traveling at the velocity of light away from the source; it is a direct result of the relativistically invariant statement that the kinetic and potential energies are exactly equal ($\mathcal{W} = 0$ in Eq. 3.3). Although we are not, at present, in a position to determine the class of situations for which this statement can be made, we might imagine that there are configurations where the applied vector potential $\mathbf{A}_0$ is much larger than what is due to current in the sample under study, and where $\mathcal{W} = 0$. In such a configuration, $\vec{J} \cdot \vec{A} = \rho \mathcal{V}$. We consider, for example, a uniform current density J that is due to the motion of a uniform charge density ρ, flowing in a strip of thickness z_0 and height y_0, lying in the xy plane. If both the $\vec{J}$ and $\vec{A}$ are in the x direction, and A_x is a function of y, then the potential will depend on y:

$$
\rho\Big(\mathcal{V}(y_1) - \mathcal{V}(y_0)\Big) = J_x\Big(A_x(y_1) - A_x(y_0)\Big) \tag{3.40}
$$

An alternative but much more laborious way to obtain this result is to notice that the current density J_x is just the charge density ρ moving at velocity v. Due to the vector potential A_x, the potential $\mathcal{V}'$ in the frame of reference of the moving electrons is modified from the potential $\mathcal{V}$ in the laboratory frame via the Lorentz transformation Eq. 3.38:

$$
\mathcal{V}' = \frac{\mathcal{V} - v A_x}{\sqrt{1 - v^2/c^2}} \approx \mathcal{V} - v A_x \tag{3.41}
$$

where the approximation is valid for $v \ll c$. The electrons are constrained by the sample to move in the x direction. Any potential difference in the y direction will accelerate electrons in the y direction, thus causing an accumulation of charge on the top and bottom surfaces of the sample. This charge accumulation generates

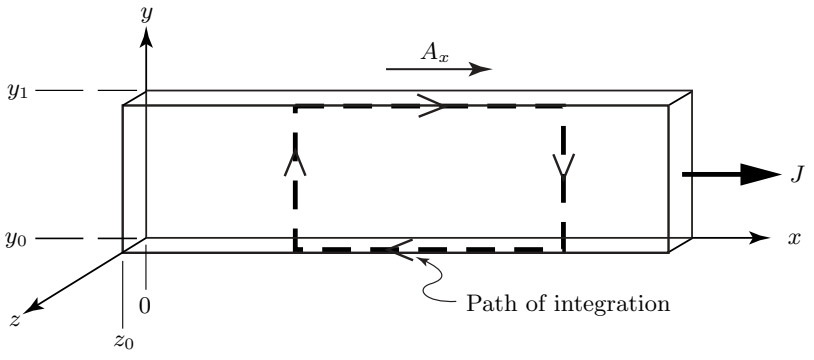


Figure 3.7 Flux integral for quantum-Hall system.

an electric field in the direction opposite to the original y-directed field, until the electrons experience zero y-directed electric field in their own frame of reference:

$$\Big(\mathcal{V}'(y_1) - \mathcal{V}'(y_0)\Big) = 0 \tag{3.42}$$

This condition implies a potential difference in the laboratory frame of reference:

$$\mathcal{V}(y_1) - \mathcal{V}(y_0) = v\Big(A_x(y_1) - A_x(y_0)\Big) \tag{3.43}$$

Multiplying both sides of Eq. 3.43 by ρ, we recover Eq. 3.40. We notice that this entire derivation was merely a tedious reconstruction of a relation that followed directly from the Schwarzschild invariant.

The dependence of $\vec{A}$ on position is called a **magnetic field**. The potential gradient transverse to the direction of current flow created by a magnetic field is called the **Hall effect**. We can multiply both sides of Eq. 3.40 by the volume $x_0 y_0 z_0$ of a segment of the strip that is x_0 in length:

$$QV = I\Big(A_x(y_1) - A_x(y_0)\Big)x_0 \tag{3.44}$$

where Q is the total charge in the segment, V is the potential difference between the top and bottom of the segment, and I is the total current in the sample. A flux integral Φ may be defined

for the segment, as shown in Fig. 3.7. The path for the flux integral is taken along the top of the segment ($y = y_1$) in the $+x$ direction, down the right face of the segment, back along the bottom of the segment ($y = y_0$) in the $-x$ direction, and up the left face of the segment to the starting point. Because A has only an x component, the left and right faces make no contribution to the integral. If the value of A_x is independent of x, which is true when x_0 is small enough, the top segment contributes $x_0 A_x(y_1)$ to the integral, and the bottom segment contributes $-x_0 A_x(y_0)$ to the integral. Hence,

$$\Phi = \Big(A_x(y_1) - A_x(y_0)\Big)x_0 \tag{3.45}$$

Thus, Eq. 3.44 can be written

$$QV = I\Phi \tag{3.46}$$

In a coherent electron system, both Φ and Q are quantized. On the basis of these considerations alone, we would expect that, under the proper experimental conditions, the ratio of voltage transverse to current would be quantized, and Eq. 3.46 could be written

$$\frac{V}{I} = \frac{n\Phi_0}{mq_0} = \frac{n}{m}\,\frac{h}{q_0^2} \tag{3.47}$$

where n and m are integers. In extremely pure semiconductors at cryogenic temperatures, the mutual quantization of charge and flux is an extremely strong effect, allowing the ratio h/q_0^2 to be determined to a few parts in 10^9. When q_0 is equal to the electronic charge q_e, the resistance defined by Eq. 3.46 with $n = m$ is $R_{\mathrm{vK}} = 25812.8056$ ohms; this is called the **quantized Hall resistance,** or the **von Klitzing resistance**, after its discoverer (57), and is now the international standard of resistance. Quantization observed for $m \geq n$ is called the **integer quantum Hall effect**, and that observed for $m < n$ is called the **fractional quantum Hall effect**. Excellent reviews with extensive references can be found in (58) and in (59).

Thus, there are two ways that collective coherent quantum systems exhibit a voltage proportional to current: The first is in an open, propagating system; the second is in a closed, quantized system. In both cases, the proportionality results from the exact

equality of kinetic and potential energies. In neither case is energy dissipated, unlike the usual conception of resistance as having a dissipative effect. The ratio of these two fundamental resistances is a dimensionless number. For historical reasons, the ratio that was first recognized, called the **fine-structure constant** α, differs from the resistance ratio by a factor of 2:

$$\frac{R_{\mathrm{FS}}}{R_{\mathrm{vK}}} = \frac{\mu_0 c\, q_e^2}{h} = 2\,\alpha \tag{3.48}$$

3.12 Summary

The behavior of collective electron systems is dominated by the interaction of each element with all of the others. The interaction energy scales with the square of the number of electrons. The degrees of freedom of the system can be expressed in terms of the phase of the wave function, of the four-potential, or of the circuit variables (voltage and current). These three sets of variables are projections of the same reality onto three different screens—each represents the same underlying degrees of freedom. To illustrate the power of maintaining these parallel views, we have treated a number of configurations using more than one approach. We have found that the concepts of kinetic and potential energy, and the conservation of total energy are useful in unifying these three representations. The energy-density scalar represents all the energy in the system; no additional "field energy" is required. For certain systems, the equality of kinetic and potential energies allowed us to draw far-reaching conclusions concerning the quantized Hall effect and the fine-structure constant.

Part 4

Radiation in Free Space

The failure of the conventional quantum field theory to deal with the problem of infinities is well known and has remained with us despite numerous attempts to resolve it. The success of the alternative approach described here reminds us that the problem is not purely local but we need to include cosmological considerations.

—F. Hoyle and J. Narlikar[1]

4.1 Integral and Differential Four-Vector Forms

In Part 1 (p. 9), we derived the integral form of magnetic interaction, using the 3-dimensional differential form as a starting point. We can carry out a similar physically based procedure in four dimensions, using the Riemann–Sommerfeld differential form in four-vector notation:

$$\Box^2 \mathbf{A} = -\mu_0 \mathbf{J} \tag{4.1}$$

For spherically symmetrical situations, Eq. 4.1 can be written

$$\Box^2 \mathbf{A} = \left(\nabla^2 - \frac{\partial^2}{\partial t^2} \right) \mathbf{A} = \frac{1}{r} \frac{\partial^2 (r\mathbf{A})}{\partial r^2} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} \tag{4.2}$$

As before, we imagine the current as confined to a small sphere around the origin. In the region outside the sphere, where no current is present, Eq. 4.2 has solutions of the form

$$\mathbf{A}(r,t) = \frac{r_0}{r}\ \mathbf{A}(r_0, t \pm r/c) \tag{4.3}$$

We can determine how $\mathbf{A}(r_0)$ is related to $\mathbf{J}$ for rapidly varying currents. To do so, we suppose that the four-current in our tiny

[1]These authors have published a series of papers and books based on the Wheeler–Feynman papers. This quotation appears on the first page of the preface to their recent monograph on the subject (60), through which the other material may be traced.

spherical volume has a time dependence $\mathbf{J}(t)$ that changes slowly on the time scale of r_0/c. Eq. 1.14 applies to this situation, so Eq. 4.3 becomes

$$\mathbf{A}(r,t) = \frac{\mu_0}{4\pi r}\ \frac{4\pi r_0^3}{3}\ \mathbf{J}(t \pm r/c) \tag{4.4}$$

The quantity $\mathbf{A}(r_0)$ is the four-potential, measured at the boundary of our volume element, *due to the four-current element alone.* Since we can make r_0 as small as we wish, we can consider $4\pi r^3/3$ as our volume element, and compute the contribution of this current element to the total potential at some distance $r \gg r_0$ outside of the volume element:

$$d\mathbf{A}(r,t) = \frac{\mu_0}{4\pi}\ \frac{\mathbf{J}(t \pm r/c)}{r}\ d\text{vol} \tag{4.5}$$

When $r \gg r_0$, the detailed shape of the volume element is immaterial, as long as r is measured to the center. Because of the linearity of Eq. 4.1, we can add up as many of these elementary contributions as we wish to obtain the total vector potential:

$$\mathbf{A}(r,t) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(t \pm r/c)}{r}\ d\text{vol} \tag{4.6}$$

An elegant mathematical proof of this four-dimensional Green's function is given in Sommerfeld (54). We describe the time relation in Eq. 4.6 by saying that the element where $\mathbf{A}$ is measured is on the **Light Cone** of the current element $\mathbf{J}$. This "cone" is, of course, a three-dimensional structure in four-dimensional space-time. The universe on our light cone is, literally, the universe that we see. We can view the light cone as being all points that are local in four-space; that is, for which the square of the interval from here and now is zero. From Section 2.2 (p. 37), the interval squared contains the spatial distance r and the time difference Δt:

$$\mathbf{R^2} = \mathbf{R} \cdot \mathbf{R} = r^2 - c^2\ \Delta t^2 \tag{4.7}$$

Eq. 4.6 indicates that the four-potential at a given point in space-time is determined by all four-currents separated from that point by zero interval. This definition of locality is the only one that makes sense in four-space.

4.2 Retarded and Advanced Potentials

The English-language statement of Eq. 4.6 is that the four-potential at any point in space-time is the sum of all four-currents on its light cone, each weighted inversely by its distance from the point. In this formulation, all directions along the light cone are treated equally. In particular, there is no distinction between the "retarded potentials," generated by currents in the past, and the "advanced potentials," generated by currents in the future. It is our common experience, however, that we send a radar pulse at a particular time, and receive the return echo at a later time. There is no evidence of an echo arriving before the pulse has been transmitted. How does this asymmetry in the direction of time come about?

The entire matter has an interesting history. The integral form was first introduced by Ludvig Lorentz in 1867 (61). In that paper, Lorentz simply assumed that only the retarded solution is possible:

> the action in the point (xyz) at the moment t does not depend on the *simultaneous* condition in the point $(x'y'z')$ but on the condition in which it was at the moment $t - r/a$; that is, so much time in advance as is required to traverse the distance r with the constant velocity a.

In 1908, Ritz (62) made a strong case for adopting only the retarded solutions:

> to eliminate solutions which are physically impossible, we must adopt *a priori* the retarded potentials.

Einstein[2] responded critically:

> In contrast to Mr. Ritz, I regard the forms containing retarded functions as merely auxiliary mathematical forms. The reason I see myself compelled to take this view is first of all that those forms do not subsume the energy principle, while I believe that we should adhere to the strict validity of the energy principle until we shall have found important reasons for renouncing this guiding star.

After defining Eq. 4.6 with the negative sign as f_1, and with a positive sign as f_2, he elaborates:

[2]See Ref. 63 for the original German reference. The English translation of this and the following Einstein quotations can be found in Einstein's *Collected Papers* (55).

> Putting $f(x, y, z, t) = f_1$ amounts to calculating the electromagnetic effect at the point x, y, z from those motions and configurations of the electric quantities that took place *prior to* the instant t. Putting $f(x, y, z, t) = f_2$ we are determining the above electromagnetic effect from the motions and configurations that take place *after* the instant t.
>
> In the first case the electric field is calculated from the totality of the processes producing it, and in the second case from the totality of the processes absorbing it. If the whole process occurs in a (finite) space bounded on all sides, then it can be represented in the form $f = f_1$ as well as in the form $f = f_2$. If we consider a field that is emitted from the finite into the infinite, we can, naturally, use only the form $f = f_1$, precisely because the totality of the absorbing processes *is not taken into consideration.* But here we are dealing with a misleading paradox of the infinite. Both kinds of representation can always be used, regardless of how distant the absorbing bodies are imagined to be. Thus, one cannot conclude that the solution $f = f_1$ is more special than the solution $a_1 f_1 + a_2 f_2$, where $a_1 = a_2 = 1$.

The spirited debate between Einstein and Ritz led to a joint publication (64), in which the bases of their respective views were clarified:

> In the special case in which an electromagnetic process *remains restricted to a finite space*, the process can be represented by $f = f_1$ as well as in the form $f = f_2$ and in other forms.
>
> While Einstein believes that one could restrict oneself to this case without *substantially* limiting the generality of the consideration, Ritz considers this restriction not to be possible in *principle*. If one takes this standpoint, then experience compels one to consider the representation by means of retarded potentials as the only one possible, if one is inclined to the view that the fact of irreversibility of radiation processes must already find its expression in the fundamental equations. Ritz considers the restriction to the form of retarded potentials as one of the roots of the second law, while Einstein believes that irreversibility is exclusively due to reasons of probability.

The dominance of Maxwell's theory in Einstein's relativity led to a widespread belief that action-at-a-distance theories are incapable of providing solutions in accord with the rapidly growing body of experiment. Despite numerous problems of the kind mentioned in

the preface, the Heaviside–Hertz form of Maxwell's equations was taught as gospel, without any mention of the conceptual problems they raised, or of the alternative approaches that had been advanced. In the 1940s, Wheeler and Feynman published two papers (65, 41) outlining an action-at-a-distance approach for free particles that gave results in accord with experiment. In their words:

> (1) There is no such concept as "the" field, an independent entity with degrees of freedom of its own.
>
> (2) There is no action of an elementary charge upon itself and consequently no problem of an infinity in the energy of the electromagnetic field.
>
> (3) The symmetry between past and future in the prescription for the fields is not a mere logical possibility, as in the usual theory, but a postulational requirement.

The second item was required because they assumed that the interacting entities were point particles. As we have seen in Section 1.5 (p. 18) and Section 4.1 (p. 73), the Green's function for collective systems is totally free of singularities, and cannot, by its very nature, generate infinities. Wheeler (66) gives a wonderful account of Feynman's student days, and of the genesis of the ideas in the two papers. In his words,

> The startling conclusion that Dick Feynman and I reached, which I still believe to be correct, is that if there were only a few chunks of matter in the universe . . . the future would, indeed, affect the past. What prevents this violation of common sense and experience is the presence in the universe of a nearly infinite number of other objects containing electric charge, all of which can participate in a grand symphony of absorption and reemission of signals going both forward and backward in time.

While working on the second of the two papers, Wheeler and Feynman chatted with Einstein about their approach. Wheeler reports the following Einstein comment:

> I have always believed that electrodynamics is completely symmetric between events running forward and events running backward in time. There is nothing fundamental in the laws that makes things run in only one direction. The one-way flow of events that is observed is of statistical origin. It comes about

because of the large number of particles in the universe that can interact with each other.

In 1980, Bernard Schutz showed that the radiation reaction can be obtained using time-symmetric laws if one averages over a random ensemble of initial conditions, but no physical interpretation of this averaging process was given. The treatment given below can be viewed as a way of deriving Schutz's initial conditions.

In the early 1980s, John Cramer (67) extended the Wheeler–Feynman approach to atomic transitions. His *Transactional Interpretation of Quantum Mechanics* showed how the mathematical formalism of the usual quantum theory can be interpreted in terms of direct interaction between emitter and absorber. This important paper seems to have been overlooked in most discussions of the subject, except for that of Gribben (25), in which it plays a central role.

It is curious that Einstein, for whom Minkowski's four-dimensional space-time played such a central role, and who was generally convinced that time should play a symmetric role in fundamental equations of physical law, took a strong position for local causality in three-space. The whole matter is clouded by the ongoing debate about quantum mechanics. The Copenhagen clan had come up with a mathematical formalism that could yield statistical results in accord with experiment—*that* was agreed by all. Were there more-fundamental physical laws beneath the statistical rules? Einstein thought there were, but Bohr said no, there was no deeper level to be understood. Einstein was still fuming about the whole thing in 1935, when he coauthored a paper with Podolsky and Rosen (68) that took up the crusade once again. The idea went as follows: Suppose an atom emits two photons simultaneously, as part of the same fundamental transition. The two photons travel in opposite directions, and have opposite polarizations. Once the photons have traveled far away, we measure the polarization of one, and thereby know the polarization of the other. So far, so good. But there are several ways to measure polarization. If we determine that one photon is right-hand circularly polarized, we know the other is left-hand. We can also ask if the first photon is linearly polarized along the y axis, in which case we know that the other is linearly polarized along the x axis. Now

suppose we wait until after the photons have been emitted, and then quickly change the apparatus with which we measure the polarization of one of them. Einstein was convinced that there is no way that the other photon can "know" that the first measurement had changed. When the experiment was carried out,[3] it was indeed found to be true that the state of one photon depends on how the other is measured, a result that was immediately touted as a triumph for the Copenhagen view. From our present perspective, however, this result has nothing to do with which interpretation of quantum mechanics we use. Cramer's transactional interpretation gives an even clearer picture of what is going on: The two photons are part of the same transaction, local in space-time. In three-space, we describe them by solutions propagating both forward and backward in time, as so elegantly put forth by Wheeler and Feynman. By now there are an enormous number of experiments for which forward and backward propagating solutions are by far the most economical explanation. These experiments effectively preclude the interpretation of photons as little bullets moving at the velocity of light and carrying energy as they go. But Einstein did not live to see these experiments completed, and his feelings about them were all tangled up with the statistical interpretation. So, although he strongly believed that time can run in both directions in fundamental physical law, he still held out for local causality in three-space. Causality is, however, not an idea that has roots in the fundamental laws of physics, but rather in thermodynamics.

4.3 Thermodynamics

Thermodynamics is based on two postulates, called the first and second **Laws of Thermodynamics**. The first law is just the conservation of energy. The second law involves the monotonic increase of the entropy of a system. There has been an enormous amount of controversy over the second law, from the time it was initially proposed by Boltzmann up to the present day. In a recent, thoughtful review, Sklar (69) concludes:

[3]This type of experiment was suggested by John Bell, who participated in the conference that generated *The Quantum Challenge* (24). This reference contains a modern account of the experiments, with many citations to the original literature.

> Even making the appropriate allowances for the weakened, statistical, sense in which we wish to hold the Second Law true, no amount of reliance upon the subtleties of dynamics or of the constitution of systems allows us to extract coherently from the theory an explanatory understanding of the temporal asymmetry of the world.

In the same volume, Price (70) offers the following interpretation:

> A century or so ago, Ludwig Boltzmann and others attempted to explain the temporal asymmetry of the second law of thermodynamics. The hard-won lesson of that endeavor—a lesson still commonly misunderstood—was that the real puzzle of thermodynamics is not why entropy always increases with time, but why it was ever so low in the first place.

It is not my intention to indulge in speculation about thermodynamics in general, or about the origin of the universe as we experience it. I am content to offer a single, simple example in which totally time-symmetric laws give a time-asymmetric behavior. This example supports the views of Wheeler and Price. The example is that of an otherwise lossless superconducting resonator suspended in space.

4.4 Radiation Damping

In our investigation of radiative coupling, we use a superconducting resonator as a model system. The resonator is a coherent quantum system, interacting within itself in a purely electromagnetic manner. In this sense, it can be viewed as a "giant atom." As a model system, however, it is much simpler than either an atom or a free particle. Its lowest mode of oscillation has a single degree of freedom, the configuration of which is known to astounding precision. Its orientation in space is known and controlled. Its phase can be measured to extreme accuracy. We can build such a resonator from a superconducting loop and a capacitor, as described in Section 3.2.3 (p. 53). If we suspend the resonator in free space, far from any other matter, we obtain the following experimental results:

1. If the resonator is initialized to zero amplitude, its average amplitude of oscillation increases with time until it fluctuates around a mean amplitude V_0.
2. If the resonator is initially oscillating at an amplitude that is

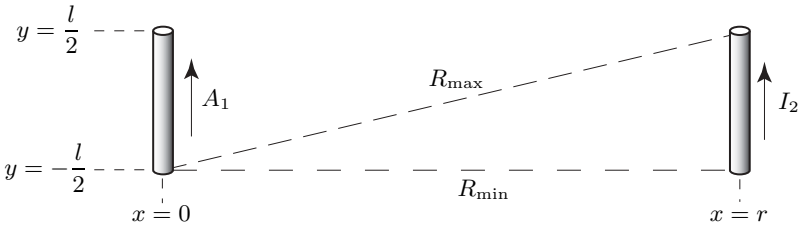


Figure 4.1 Coupling of two segments of thin superconducting wire

large compared with V_0, the amplitude decreases with time until it fluctuates around V_0.

3. For large amplitudes, the rate of decrease of amplitude is proportional to the amplitude, leading to an exponential damping of the oscillation.
4. The final approach to V_0 from either direction is, on average, exponential, with the same time constant as the large-amplitude decay.
5. The value of V_0 is dependant on the frequency ω of the resonator; the dependence on frequency is that of a black body at ≈ 3 kelvin.

4.5 Coupled Loops

As a prelude to understanding the damping of a small resonator in free space, we consider the coupling of two elements of thin wire separated by a distance R, the first $d\vec{l}_1$ being that in which $\vec{A}_1$ is to be evaluated, and the second $d\vec{l}_2$ carrying current I_2. When the current is flowing in a thin wire, we can integrate Eq. 4.6 over the cross-section of the wire:

$$A_1(t) = \frac{\mu_0}{8\pi} \int \frac{I_2(t \pm R/c)}{R}\, dl_2 \tag{4.8}$$

To be definite, we next consider two straight segments of thin wire, as shown in Fig. 4.1. Each segment of length l is running in the y direction. Both segments are centered at $y = 0, z = 0$. The first segment is at $x = 0$, and the second segment is at $x = r$. The flux integral along the first segment is

$$\Phi_1(t) = \frac{\mu_0}{8\pi} \int A_1(t)\, dl_1 = \frac{\mu_0}{8\pi} \int \int \frac{I_2(t \pm R/c)}{R}\, dl_2 dl_1 \tag{4.9}$$

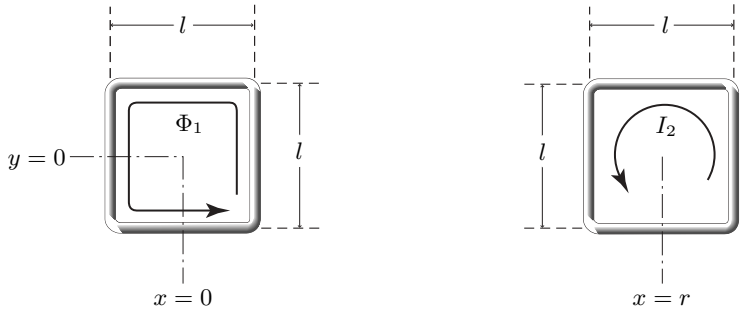


Figure 4.2 Flux Φ_1 due to current I_2 in superconducting loop that is far away

In general, both I_2 and R in Eq. 4.9 will be functions of y_2, and no further simplification is possible. If the distance r is large enough, however, the distance from any part of one wire to any part of the other may be taken as equal, which we can understand as follows: The distance between the lower end of one wire and the corresponding end of the other wire is r: the minimum value of R. The distance between the lower end of one wire and the opposite end of the other wire is $\sqrt{r^2 + l^2} = r\sqrt{1 + l^2/r^2}$: the maximum value of R. Expanding the radical, we find the maximum difference $\Delta R \approx l^2/2r$ between these two values of R_{12}. As long as $r > l^2/2d$, the spread in distance is limited by the diameter d of the wire, rather than by geometric factors. This simplification is called the **Far Field** or **Plane Wave** approximation. In the plane-wave approximation, A_1 is independent of y_1. A particularly simple case occurs when the plane wave approximation applies and, in addition, when the current at any particular time is the same at all values of y_2. In this happy circumstance, the integral over y_2 reduces to multiplication by l, as does the integral over y_1.

$$\Phi_1(t) \approx \frac{\mu_0}{8\pi} \frac{l^2}{r} \, I_2(t \pm r/c) \tag{4.10}$$

We can use Eq. 4.10 to determine the phase accumulation around a square loop that is due to the current in another identical square loop located far away in space, as shown in Fig. 4.2. For simplicity, we confine our analysis to the case where the dimensions of the loops are small, and the distance between them is large, when compared with a wavelength of the frequency at

which they are driven. Because the loops are small compared to the wavelength, the current in every element $\vec{dl}$ of the loop is the same. Because the loops are far apart, the far-field approximation is valid. Suppose both loops are in the plane $z = 0$, and both are squares with sides of length l parallel to the x and y axes. The center of the first loop is at $x = 0$, and that of the second at $x = r$. In the far-field limit, the effect of the sides parallel to the x axis cancels out, so we can concentrate on the vector potential A in the segments at $x = \pm l/2$ created by the current I_2 in the segments at $x = r \pm l/2$. Furthermore, the $1/R$ factor is arbitrarily close to $1/r$ if $r \gg l$. In addition, the integration with respect to l_2 becomes a multiplication by l. The vector potential in the segment at $x = l/2$ is the sum of a positive contribution due to the current in its corresponding segment at distance r, and a negative contribution due to the opposing current in the segment at distance $r - l$.

$$A_1\left(\frac{l}{2}, t\right) = \frac{\mu_0}{8\pi}\frac{lI_2(t \pm r/c)}{r} - \frac{\mu_0}{8\pi}\frac{lI_2\left(t \pm \frac{r-l}{c}\right)}{r} \tag{4.11}$$

The vector potential in the segment at $x = -l/2$ is the sum of a negative contribution, due to the current in its corresponding segment at distance r, and a positive contribution, due to the opposing current in the segment at distance $r + l$.

$$A_1\left(-\frac{l}{2}, t\right) = \frac{\mu_0}{8\pi}\frac{lI_2\left(t \pm \frac{r+l}{c}\right)}{r} - \frac{\mu_0}{8\pi}\frac{lI_2(t \pm r/c)}{r} \tag{4.12}$$

Both segments contribute to the phase accumulation; the integration along the path around the origin corresponds to a multiplication by l for the first segment, and by $-l$ for the second. The total phase accumulation around the first loop due to current in the second loop is thus the sum of three contributions:

$$\Phi_1\left(t \pm \frac{r}{c}\right) = \frac{\mu_0 l^2}{8\pi r}\Big(- I_2(t - l/c) + 2I_2(t) - I_2(t + l/c)\Big) \tag{4.13}$$

where we have now taken the time at loop 2 as our reference. It is noteworthy that the expression on the right side of Eq. 4.12 is totally independent of the sign of r/c on the left side. In other words, the coupling between the two loops is exactly the same, whether

retarded or advanced potentials are used. Any loop couples to any other on its light cone, whether past or future. We notice that the total flux in Eq. 4.13 has the form of an approximation to the second derivative of the current with respect to time. Using Taylor's series, we obtain

$$\Phi_1\left(t \pm \frac{r}{c}\right) \approx -\left(\frac{\mu_0 l^4}{8\pi r c^2}\right) \cdot \frac{\partial^2 I_2}{\partial t^2} \tag{4.14}$$

These approximations are valid only when the current changes slowly compared with the time l/c, and when $r \gg l \gg d$. We shall henceforth adopt the convention that time is to be measured relative to the common light cone through the center of the loop, and not carry along explicit reference to the time of origin. Under this convention, Eq. 4.14 becomes

$$\Phi_1 \approx -\left(\frac{\mu_0 l^4}{8\pi r c^2}\right) \cdot \frac{\partial^2 I_2}{\partial t^2} \tag{4.15}$$

The total phase accumulation in a loop is the sum of that due to its own current, and that due to currents in other loops far away.

4.6 Resonator

We consider the case of one square loop in space, considered in the previous section. A particular loop has a capacitor C connected across its terminals, thereby forming a resonator, as analyzed in Section 3.2.3 (p. 53). The voltage across the terminals is $V = \partial\Phi/\partial t$. The voltage across the capacitor opposes current flow through the loop, so the sign of current I through the loop is opposite to the sign of current through the capacitor: $-I = C\,\partial V/\partial t$. Eliminating V, we have

$$I = -C\,\frac{\partial^2 \Phi}{\partial t^2} = -C\,\ddot{\Phi} \tag{4.16}$$

Our limitation on the frequency allows us to assume that the current is the same everywhere in the loop, and, therefore, has the same magnitude as the current in the capacitor, but of opposite sign. The flux Φ is the sum of two terms: LI, due to current flowing in the loop itself, and ϕ, due to currents far away, as given by Eq. 4.15.

$$\Phi = LI + \phi \tag{4.17}$$

Substituting Eq. 4.16 into Eq. 4.17, we obtain

$$LC\,\ddot{\Phi} + \Phi = \phi \tag{4.18}$$

We can use Eq. 4.15 to express ϕ in terms of the current I_2 in a distant resonator:

$$\phi \approx -\left(\frac{\mu_0 l^4}{8\pi r c^2}\right) \cdot \frac{\partial^2 I_2}{\partial t^2} \tag{4.19}$$

Using Eq. 4.16, we can express ϕ in terms of the flux Φ_2 in the distant resonator:

$$\phi \approx \left(\frac{\mu_0 l^4 C}{8\pi r c^2}\right) \cdot \frac{\partial^4 \Phi_2}{\partial t^4} \tag{4.20}$$

4.7 Two Coupled Resonators

We obtain the equation for two identical coupled resonators by substituting ϕ from Eq. 4.20 into Eq. 4.18:

$$LC\,\frac{\partial^2 \Phi_1}{\partial t^2} + \Phi_1 = \left(\frac{\mu_0 l^4 C}{8\pi r c^2}\right) \cdot \frac{\partial^4 \Phi_2}{\partial t^4} \tag{4.21}$$

The equations governing the behavior of two coupled resonators are thus

$$\begin{aligned} \tau^2\,\frac{\partial^2 \Phi_1}{\partial t^2} + \Phi_1 &= \delta\tau^4\,\frac{\partial^4 \Phi_2}{\partial t^4} \\ \tau^2\,\frac{\partial^2 \Phi_2}{\partial t^2} + \Phi_2 &= \delta\tau^4\,\frac{\partial^4 \Phi_1}{\partial t^4} \end{aligned} \tag{4.22}$$

where

$$\delta = \frac{\mu_0 l^4}{8\pi r c^2 L^2 C} \qquad \text{and} \qquad \tau^2 = LC \tag{4.23}$$

Assuming solutions of the form $\Phi = e^{i\omega t}$, Eq. 4.22 becomes

$$\begin{aligned} \left(1 - \tau^2\,\omega^2\right)\Phi_1 &= \delta\tau^4\,\omega^4 \Phi_2 \\ \left(1 - \tau^2\,\omega^2\right)\Phi_2 &= \delta\tau^4\,\omega^4 \Phi_1 \end{aligned} \tag{4.24}$$

The result of eliminating the flux variables is an expression for the frequencies of the normal modes of the system:

$$\left(1-\tau^2\,\omega^2\right)^2=\left(\delta\tau^4\,\omega^4\right)^2 \tag{4.25}$$

Substituting Eq. 4.25 into Eq. 4.24, we conclude:

$$\begin{aligned}\Phi_1 &= +\,\Phi_2 = e^{i\omega_1 t} \qquad \text{Even Mode}\\ \Phi_1 &= -\,\Phi_2 = e^{i\omega_2 t} \qquad \text{Odd Mode}\end{aligned} \tag{4.26}$$

Because δ is small, the roots of Eq. 4.25 will be very near $\omega_0 = 1/\tau$, which we can use as the first term in our expansion, which assumes real ω^2:

$$\begin{aligned}\omega_1^2 &\approx \omega_0^2\,(1\,-\,\delta) \qquad \text{Even Mode}\\ \omega_2^2 &\approx \omega_0^2\,(1\,+\,\delta) \qquad \text{Odd Mode}\end{aligned} \tag{4.27}$$

In the odd mode, Φ_1 is antiphase to Φ_2, whereas in the even mode, the two fluxes are in phase. As we saw in the static case, currents coupled in the same sense increase the phase accumulation, while currents coupled in the opposite sense reduce the phase accumulation. The phase accumulation is reflected directly in the flux per unit current, and therefore in the electrodynamic inertia. The larger the inertia, the lower the resonant frequency. For this reason, the even mode has a frequency lower than ω_0, and the odd mode has a frequency higher than ω_0.

The modes given by Eq. 4.27 are **normal modes**; they correspond to stationary states of the system. Once the system is oscillating in one of these modes, it will continue to do so forever. To understand energy transfer between the resonators, we can use mixtures of the normal modes. Any weighted sum of the normal modes is also a legal behavior of the system. The sum of the two normal modes with equal weight is particularly instructive:

$$\begin{aligned}2\Phi_1 &= e^{i\omega_1 t} + e^{i\omega_2 t}\\ 2\Phi_2 &= e^{i\omega_1 t} - e^{i\omega_2 t}\end{aligned} \tag{4.28}$$

This behavior can be expressed in terms of the natural frequency ω_0 of the uncoupled system:

$$\begin{aligned}\Phi_1 &= e^{i\omega_0 t}\cos\delta\omega_0 t = a_1(t)e^{i\omega_0 t} \\ \Phi_2 &= e^{i(\omega_0 t-\pi/2)}\sin\delta\omega_0 t = a_2(t)\,e^{i(\omega_0 t-\pi/2)}\end{aligned} \tag{4.29}$$

Here, a_1 and a_2 are the (signed) amplitudes of oscillation in the two resonators, which vary slowly with time.

4.8 Energy Coupling

According to the viewpoint of collective electrodynamics, the energy radiated from an antenna cannot appear as an excitation of space itself, because space has no degrees of freedom. Any energy leaving one resonator is transferred to some other resonator, somewhere in the universe. That transfer is experienced as a damping of the oscillation in the source resonator in question. The energy in a single resonator alternates between the kinetic energy of the electrons (inductance), and the potential energy of the electrons (capacitance). At the point in the oscillation where there is no voltage across the capacitor, the current is at its maximum, and all of the energy is in the inductance. The shift between kinetic and potential energy within a resonator happens twice per cycle, at the frequency $\omega_0 = 1/\sqrt{LC}$. The peak energy in an inductor is $W = a^2/2L$. For each resonator in isolation, the total energy is constant, equal to W. Even when the resonators are coupled, as long as $\delta \ll 1$, a_1 and a_2 change slowly with time, and W is still a good measure of the energy in each resonator. With the two resonators coupled, the energy shifts back and forth between the two resonators in such a way that the total energy is constant, given by

$$W_{\text{tot}} = \frac{a_1^2}{2L} + \frac{a_2^2}{2L} = \frac{\left(\sin^2\delta\omega_0 t + \cos^2\delta\omega_0 t\right)}{2L} = \frac{1}{2L} \tag{4.30}$$

The conservation of energy holds despite an arbitrary separation between the resonators; it is a direct result of the symmetry of the advanced and retarded potentials. There is no energy "in transit" between them. From Eq. 4.29 we can ascertain the phase relation-

ships involved in the energy transfer:

$$\begin{aligned} \frac{\partial a_1}{\partial t} &= -\,\delta\omega_0 a_2 \\ \frac{\partial a_2}{\partial t} &= \delta\omega_0 a_1 \end{aligned} \tag{4.31}$$

But these relations are not symmetric. Anything resonator 1 can do to resonator 2, resonator 2 can do to resonator 1. Why is one derivative positive and the other negative? The apparent asymmetry is due to our choice of the relative phase of the oscillation. If we had chosen the convention that resonator 2 was leading in phase by 90°, a_2 would have been $-\sin\delta\omega t$, and the relative signs in Eq. 4.31 would have been reversed. From Eq. 4.29

$$\begin{aligned} \frac{\partial W_1}{\partial t} &= \frac{a_1}{L}\frac{\partial a_1}{\partial t} = \frac{-\,\delta\omega_0 a_1 a_2}{L} \\ \frac{\partial W_2}{\partial t} &= \frac{a_2}{L}\frac{\partial a_2}{\partial t} = \frac{\delta\omega_0 a_1 a_2}{L} \end{aligned} \tag{4.32}$$

Because the amplitudes a_1 and a_2 can be of either sign, the energy can flow in either direction. We can understand the energy transfer only when we know the amplitudes and phases of the two oscillations, not from a knowledge of the energies alone.

4.9 Interaction with Arbitrary Phase

In the previous section, we explored the transfer of energy between two identical resonators oscillating with a particular phase relationship. The next step in generalizing these results is to consider the coupled resonators from the electromagnetic energy point of view developed in Part 3 (p. 49). We begin by writing the total instantaneous energy in a single resonator:

$$\begin{aligned} W &= \frac{1}{2}\int \vec{J}\cdot\vec{A}\ d\text{vol} + \frac{1}{2}\int \rho V\ d\text{vol} \\ &= \frac{1}{2}\left(I\Phi + QV\right) = \frac{C}{2}\left(-\ddot{\Phi}\Phi + \dot{\Phi}^2\right) \end{aligned} \tag{4.33}$$

where we have used Eq. 4.16 for the final form. Substituting Eq. 4.18 into Eq. 4.33, we obtain

$$W = \frac{\Phi^2}{2L} - \frac{\Phi\phi}{2L} + \frac{C}{2}\dot{\Phi}^2 \tag{4.34}$$

Taking the derivative, and using Eq. 4.18 once again, we obtain the rate of energy transfer from one resonator to the other:

$$\begin{aligned} \dot{W} &= \frac{\Phi\dot{\Phi}}{L} - \frac{\dot{\Phi}\phi}{2L} - \frac{\Phi\dot{\phi}}{2L} + C\dot{\Phi}\ddot{\Phi} \\ &= \frac{\Phi\dot{\Phi}}{L} - \frac{\dot{\Phi}\phi}{2L} - \frac{\Phi\dot{\phi}}{2L} - \frac{\dot{\Phi}\Phi}{L} + \frac{\dot{\Phi}\phi}{L} \\ &= \frac{1}{2L}\left(\dot{\Phi}\phi - \Phi\dot{\phi}\right) \end{aligned} \tag{4.35}$$

It is interesting that the final result is made up of contributions from both the potential and kinetic energies, because the flux appears in both energies. For sinusoidal signals with frequencies $\omega \approx \omega_0$ whose amplitudes, a_1 for resonator 1 and a_2 for resonator 2, vary slowly with time, we have

$$\begin{aligned} \Phi &= a_1 \sin\omega t & \phi &= \delta a_2 \sin(\omega t + \theta) \\ \dot{\Phi} &\approx \omega a_1 \cos\omega t & \dot{\phi} &\approx \omega\delta a_2 \cos(\omega t + \theta) \end{aligned} \tag{4.36}$$

where we have made use of the fact that $\phi = \delta\Phi_2$ at $\omega = \omega_0$ from Eq. 4.24. From Eq. 4.23, we observe that δ is inversely proportional to the distance between the resonators. The energy stored in a resonator is $a^2/2L$, and therefore $\dot{W} = \dot{a}a/L$. Thus, Eq. 4.35 becomes

$$\begin{aligned} \dot{a}_1 &= \delta\omega_0\ a_2 \sin\theta \\ \dot{a}_2 &= -\delta\omega_0\ a_1 \sin\theta \end{aligned} \tag{4.37}$$

in agreement with Eq. 4.31 for the case of $\theta = -90^\circ$ considered there. The general case of Eq. 4.35 is thus

$$\begin{aligned} \dot{W}_1 &= \delta\omega_0\ a_1 a_2 \sin\theta \\ \dot{W}_2 &= -\delta\omega_0\ a_1 a_2 \sin\theta \end{aligned} \tag{4.38}$$

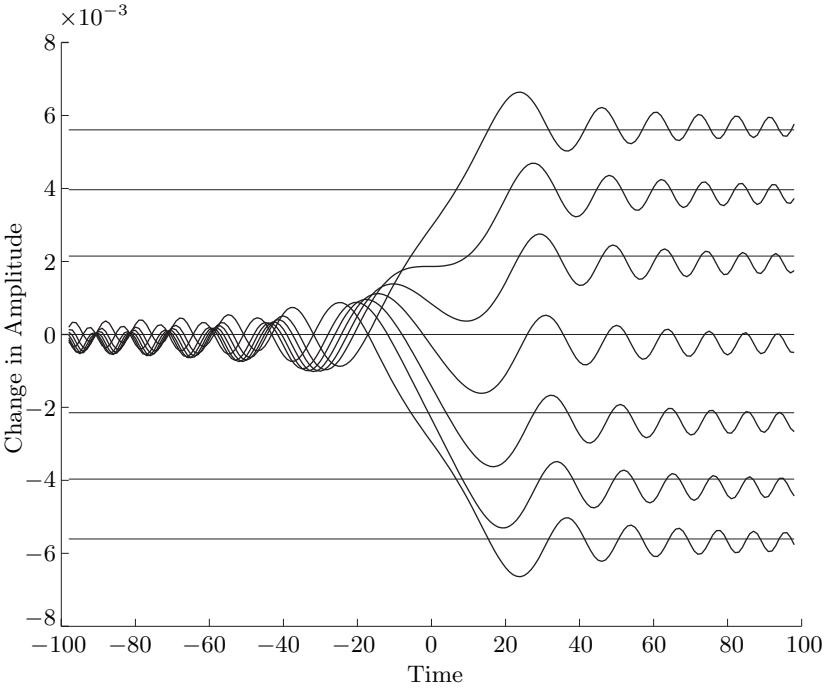


Figure 4.3 Amplitude of coupled resonator as given by Eq. 4.40. The horizontal lines represent the energy increments given by Eq. 4.41.

4.10 Generalized Energy Transfer

In the previous sections, we explored simple examples that illustrate how energy conservation works in a theory where space itself has no degrees of freedom. In a more realistic case, the universe contains a truly enormous number of resonators, no two of which are exactly equivalent. How does a single resonator behave in an inhomogeneous universe full of other matter? In the examples considered in the preceding sections, we have assumed that the two resonators have had constant and equal resonant frequencies. In the real universe, no two resonators have identical resonant frequencies for long; however, it is a common occurrence that two frequencies will cross, and that energy will be exchanged between the resonators during the crossing. From the viewpoint of collective electrodynamics, this exchange of energy is the microscopic origin of the thermodynamic behavior of the universe as we observe it. Accordingly, we consider the case in which the change

in frequency with time is represented by a time-varying phase: $\theta = \epsilon t^2 + \varphi$, in which case Eq. 4.37 becomes

$$\begin{aligned} \dot{a}_1 &= \delta\omega_0 \; a_2 \sin\left(\epsilon t^2 + \varphi\right) \\ \dot{a}_2 &= -\delta\omega_0 \; a_1 \sin\left(\epsilon t^2 + \varphi\right) \end{aligned} \tag{4.39}$$

As our first approximation, we assume that the change in amplitude Δa experienced by either resonator is small compared to its amplitude before the crossing, and therefore that we may assume a_1 and a_2 on the right hand side of Eq. 4.39 to be approximately constant for the purposes of computing $\dot{a}_1$ and $\dot{a}_2$.

$$\begin{aligned} \Delta a_1 &= \delta\omega_0 \; a_2 \int \sin\left(\epsilon t^2 + \varphi\right) \; dt \\ \Delta a_2 &= -\delta\omega_0 \; a_1 \int \sin\left(\epsilon t^2 + \varphi\right) \; dt \end{aligned} \tag{4.40}$$

We can see from Fig. 4.3 that these integrals oscillate around zero, except in the region where the argument is near zero, where the entire lasting contribution to the integral takes place. For that reason, we may approximate the integrals by extending the limits to $\pm\infty$, thereby obtaining a standard definite integral:

$$\begin{aligned} \Delta a_1 &= \delta\omega_0 \; a_2 \sqrt{\frac{\pi}{\epsilon}} \; \sin\left(\varphi + \frac{\pi}{4}\right) \\ \Delta a_2 &= -\delta\omega_0 \; a_1 \sqrt{\frac{\pi}{\epsilon}} \; \sin\left(\varphi + \frac{\pi}{4}\right) \end{aligned} \tag{4.41}$$

We can see immediately from Eq. 4.41 that the amplitude increment due to a crossing can be of either sign, depending on the relative phase φ between the two resonators at the crossing.

4.11 Random Universe

In a random universe, any particular phase is equally likely for any given crossing. A particular resonator is therefore equally likely to receive either an increment or a decrement due to a given crossing. Let us consider two crossings, one with phase $\pi/4$, and one with $5\pi/4$; these phases correspond to the maximum positive and

negative increments Δa:

$$\begin{aligned} \Delta a_1 &= \pm\delta\omega_0 \; a_2\sqrt{\frac{\pi}{\epsilon}} \\ \Delta a_2 &= \mp\delta\omega_0 \; a_1\sqrt{\frac{\pi}{\epsilon}} \end{aligned} \tag{4.42}$$

These expressions are correct to only the first order, due to our use of the initial values of a in computing Δa. As a second iteration, we use $a + \Delta a/2$ in the right-hand sides of Eq. 4.42:

$$\begin{aligned} \Delta a_1 &= \pm\delta\omega_0 \; \sqrt{\frac{\pi}{\epsilon}}\left(a_2 - \Delta \; a_1\sqrt{\frac{\pi}{\epsilon}}\right) \\ \Delta a_2 &= \mp\delta\omega_0 \; \sqrt{\frac{\pi}{\epsilon}}\left(a_1 + \Delta \; a_2\sqrt{\frac{\pi}{\epsilon}}\right) \end{aligned} \tag{4.43}$$

The energy stored in a resonator is $W = a^2/2L$. After the crossing, the energy is

$$2LW = (a + \Delta a)^2 = a^2 + 2a\Delta a + (\Delta a)^2 \tag{4.44}$$

Because the phase φ is randomly distributed, the second term will average to zero over many crossings. If we substitute Eq. 4.43 into Eq. 4.44 for both resonators, we obtain

$$\begin{aligned} \Delta W_1 &= \frac{\pi\delta^2}{2L\epsilon}\left(a_2^2 - a_1^2\right) \\ \Delta W_2 &= \frac{\pi\delta^2}{2L\epsilon} \; \left(a_1^2 - a_2^2\right) \end{aligned} \tag{4.45}$$

In a random universe, there is no first-order effect in which energy flows from the high-amplitude resonator to the low-amplitude resonator; there is, however, a second-order effect in which energy flows, on the average, from the high-amplitude resonator to the low-amplitude resonator. The rate of energy flow is proportional to the difference in energies, and to the inverse square of the distance. In collective electrodynamics, this effect is the origin of the radiation damping of resonators coupled to matter in the rest of

a random universe. Our analysis has thus established a microscopic mechanism for the damping of an isolated resonator. Our approach is consistent with Einstein's view that the second law of thermodynamics is not a reflection of any asymmetry in the fundamental laws, but is of statistical origin. We are in agreement with Wheeler and Feynman's assertion that "Radiation is concluded to be a phenomenon as much of statistical mechanics as of pure electrodynamics." The statistics, in the present theory, come from the random phase of excitation of each of the myriad of resonators in the universe, and from the constant shifting of the resonator frequencies with respect to each other. This randomness averages the first-order contribution to energy transfer between resonators to zero, but a finite second-order contribution, proportional to the difference in energy of excitation, remains. The direction of energy transfer is determined by the second-order nature of the effect, and by the phase of oscillation induced in one resonator by another.

With regard to how and when we choose the positive and negative signs for time in our solutions, our approach is on precisely the same footing with conventional field theory. We have created no additional problems there, nor have we eliminated any. In both cases, solutions of advanced and retarded nature are permissible. In both cases, we choose the solutions that satisfy the constraints of the problem at hand. In neither case do we have a complete understanding of which solutions occur in nature, and why. The development given above is but one small step toward such an understanding.

We are, however, in a position to make one additional observation on the subject: When we transmit a radar signal, we employ aspects of the universe that are the furthest from thermodynamic equilibrium. Metals are refined, electric power systems are constructed, high-power transmitters are constructed, and electronic logic circuits control the signal sequence that is transmitted, and how the echo is detected. Our receiving apparatus uses irreversible thermodynamic principles to amplify the received signal. Reversing the relative time of transmission and reception does not create a time-reversed copy of the entire physical situation.

4.12 Cosmological Properties of Radiation Damping

On the basis of the foregoing results, we may draw certain conclusions concerning the makeup of the universe. These conclusions are

in agreement with some, but not all, of those reached by Wheeler and Feynman, or by Hoyle and Narlikar (60).

1. The spectral density of distant resonators acting as absorbers is, of necessity, identical to that of those resonators producing the local random field, because they are the same resonators. By Eq. 4.45, a local resonator approaches an equilibrium state of excitation with energy equal to the mean energy of the far-away matter at that frequency. If the approach is upward from zero, it is called noise excitation. If the approach is downward from a large amplitude oscillation, it is called damping. The time constant of that approach is independent of the direction of approach. These statements contain the essence of the fluctuation-dissipation theorem (71).
2. The present theory, in and of itself, does not connect the direction of energy flow to either the forward or backward light cone of the emitting resonator. The usual assumption, that of emission into the future and absorption from the past, if true, must find its roots outside the considerations presented here.
3. The rate of energy flow given by Eq. 4.45 is proportional to δ^2, and therefore to the inverse square of the distance. A spherical shell of the universe of thickness dR at a distance R from a local resonator has volume $d\text{vol} = 4\pi R^2 dR$. If matter is, on the average, evenly distributed in the universe, each shell of material makes equal contribution to the radiation damping. This result is implicit in the Wheeler–Feynman papers, but was never stated.
4. The cosmic microwave background, with an excitation spectrum consistent with a temperature of $\approx$3K, cannot be ascribed to degrees of freedom of the radiation field. It must originate from matter on either the forward or backward light cone. Its presence in the universe as we see it strongly suggests the presence of widely distributed cold matter.
5. The arrow of time introduced by radiation damping is not related, in any direct way, to the expansion of the universe. It is strictly the result of the predominance of cold matter in the universe, together with the second-order nature of resonator coupling, the randomness in the frequencies and phases of resonators, and the continual shifting of resonator frequencies.

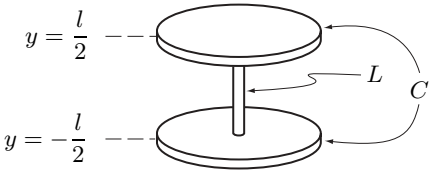


Figure 4.4 Superconducting electric dipole resonator

4.13 Electric Dipole

The radiating structures we considered in Part 4 (p. 73) were coupled entirely by the vector potential; the scalar potential played no role. In the more general case, both scalar and vector potentials contribute to the coupling. We now consider an important special case of this more general kind of coupling. The coupling between two loops considered in Part 4 (p. 73) is called **magnetic dipole coupling**. It is characterized by its proportionality to the second derivative of the current with respect to time (Eq. 4.15). A much stronger coupling can be obtained between two straight sections of wire, as given by Eq. 4.10. Referred to the light cone, the flux ϕ in section 1 due to the current in section 2 becomes

$$\phi \approx \frac{\mu_0}{8\pi} \frac{l^2}{r} I_2 \tag{4.46}$$

We can imagine a resonator configuration for which this type of coupling is realizable: Two parallel capacitor plates of capacitance C are connected by a straight section of superconducting wire of inductance L between their centers. Such a configuration, shown in Fig. 4.4, is called an **electric dipole**. Because there are charges at the two ends of the dipole, we can have a contribution to the electric coupling from the scalar potential $\mathcal{V}$ as well from the magnetic coupling ϕ from the vector potential $\vec{A}$. By symmetry, the charge at the top of the dipole ($y = l/2$) is equal in magnitude and opposite in sign to that at bottom ($y = -l/2$). These charges are equidistant to any point in the x-z plane. By the argument given in Section 4.5 (p. 82), the electrostatic coupling approaches zero for other dipoles in the x-z plane at large distances.

4.14 Coupled Electric Dipoles

For two identical y-directed electric dipole resonators in the x-z plane, separated by a distance R that is large compared with

a wavelength, the magnetic coupling equations corresponding to Eq. 4.22 are

$$\begin{aligned} \frac{\partial^2 \Phi_1}{\partial t^2} + \omega_0^2 \Phi_1 &= \delta \frac{\partial^2 \Phi_2}{\partial t^2} \\ \frac{\partial^2 \Phi_2}{\partial t^2} + \omega_0^2 \Phi_2 &= \delta \frac{\partial^2 \Phi_1}{\partial t^2} \end{aligned} \tag{4.47}$$

where the coupling term is evaluated from Eq. 4.46.

$$\delta = \frac{\mu_0 l^2}{8\pi RL} \tag{4.48}$$

For $\Phi = e^{i\omega t}$ with real ω, the solutions to Eq. 4.47 are $\Phi_1 = \pm\Phi_2$, and the natural frequencies ω_1 and ω_2 for the two cases are given by

$$\begin{aligned} \omega_1^2(1-\delta) &= \omega_0^2 \\ \omega_2^2(1+\delta) &= \omega_0^2 \end{aligned} \tag{4.49}$$

where $\omega_0 = 1/\sqrt{LC}$. For small δ, Eq. 4.49 is equivalent to Eq. 4.27. Taking the square root of Eq. 4.49 and expanding the $1 \pm \delta$ term, we obtain

$$\begin{aligned} \omega_1 &\approx \omega_0 \left(1 + \frac{\delta}{2}\right) \\ \omega_2 &\approx \omega_0 \left(1 - \frac{\delta}{2}\right) \end{aligned} \tag{4.50}$$

Thus, in the electric dipole case, as well as in the magnetic dipole case considered in Part 4 (p. 73), as long as δ is small, we can view δ as the fractional "splitting" of the two normal mode frequencies of the coupled system. We can, therefore, compare the coupling strength of electric dipole resonators with that of magnetic dipole resonators by comparing the δ in Eq. 4.48 with that in Eq. 4.23:

$$\frac{\delta_{\text{electric}}}{\delta_{\text{magnetic}}} = \frac{LCc^2}{l^2} = \left(\frac{\lambda_0}{l}\right)^2 \tag{4.51}$$

where λ_0 is the wavelength at the resonant frequency. In other words, electric dipole coupling is stronger than magnetic dipole

coupling by the square of the ratio of the wavelength to the size of the element. For radiators that are small compared with a wavelength, this ratio can be enormous. For example, an atom half a nanometer in diameter radiates visible light of 500 nanometer wavelength. In this case, electric dipole coupling is a million times stronger than magnetic dipole coupling.

4.15 General Electromagnetic Coupling

From Eq. 2.26, we have ascertained that the vector potential represents the propagation vector of the electron wave function at which there is zero current, and the scalar potential represents the frequency of the electron wave function at which there is zero charge. These two conditions allow us to evaluate coupling between two or more structures where the full four-potential contributes. We use the coupling between two electric dipoles to illustrate the method. In the special case where the two y-directed electric dipoles both lie on the y axis, the charges at the ends of the dipoles are maximally separated, and the electric contribution to the coupling through the scalar potential $\mathcal{V}$ is maximum. In addition, there is the magnetic coupling through the flux ϕ. We first write the total flux in terms of the local current I and the flux ϕ from the distant resonator:

$$\Phi = \phi + LI \tag{4.52}$$

Because the vector potential represents the phase accumulation of the electron wave function when there is no current, $\Phi = \phi$ when $I = 0$.

Similarly, we write the voltage V between the capacitor plates in terms of the difference in scalar potential from the distant resonator, and a local term proportional to the charge on the capacitor. This voltage is the rate of change of flux in the inductor:

$$V = \dot{\Phi} = \mathcal{V}_{\text{top}} - \mathcal{V}_{\text{bot}} + \frac{Q}{C} \tag{4.53}$$

The scalar potential represents the time rate of change of the wave function phase when there is no charge. Thus, the difference in scalar potential between the two capacitor plates represents the rate of change of phase accumulation along the wire when there is no charge on the capacitor.

We pause here to remark upon how naturally the two coupling terms emerge from our view on the relationship between the electron wave function and the electromagnetic variables.[4] The method of evaluating coupling terms used in this special case is, in fact, a general one. The local circuit equations can always be expressed in the form of Eqs. 4.52 and 4.53: One set contains the magnetic flux from local and distant sources, and the local current; and a second set contains the time rate of change of flux, the scalar potential, and the local charge. Setting the local current to zero determines the condition where the local flux is equal to the integral of the vector potential from distant sources; setting the charge to zero determines the condition where the time derivative of the local flux (the voltage) is equal to the difference in scalar potential from distant sources. The vector and scalar potentials from distant sources are obtained from Eq. 4.6.

4.16 Radiation Pattern

We can now return to the problem at hand, determining the equations for two coupled electric-dipole resonators in the far-field limit. Differentiating Eq. 4.53, we obtain:

$$\ddot{\Phi} = \dot{\mathcal{V}}_{\text{top}} - \dot{\mathcal{V}}_{\text{bot}} + \frac{I}{C} \tag{4.54}$$

Substituting Eq. 4.54 into Eq. 4.52, we eliminate I and thereby

[4] In traditional treatments, these terms are derived from Maxwell's equations, using an additional term in the electric field. This additional term, the gradient of a scalar, is compatible with Maxwell's equations in their traditional Heaviside–Hertz form, but is not required or suggested by them. The scalar so introduced is then identified with the scalar potential. In this backhanded manner, a theory based on electric and magnetic fields works its way back to using the potentials that Heaviside and Hertz worked so hard to eliminate. The coupling between two radiating structures is then evaluated using some form of reciprocity theorem. A good presentation of the traditional approach is given by Stutzman and Thiele (72). Other treatments, for example that of Ramo et al. (40), use these potentials explicitly, but still require a redefinition of the electric field of the form $E = \dot{A} + \nabla\mathcal{V}$. This condition, although natural enough to be passed off as obvious, amounts to an additional ad hoc assumption within the traditional framework. The electric field so obtained is then integrated along the receiving structure to obtain the total potential, including both electric and magnetic coupling.

obtain a second-order differential equation in Φ:

$$\Phi - \phi = LC\left(\ddot{\Phi} - \dot{\mathcal{V}}_{\text{top}} + \dot{\mathcal{V}}_{\text{bot}}\right) \tag{4.55}$$

which we can arrange in canonical form:

$$LC\,\ddot{\Phi} - \Phi = LC\left(\dot{\mathcal{V}}_{\text{top}} - \dot{\mathcal{V}}_{\text{bot}}\right) - \phi \tag{4.56}$$

The total coupling from far-away resonators is represented by the right-hand side of Eq. 4.56: The first term is the electrostatic coupling, the second term is the magnetic coupling; both can be evaluated from Eq. 4.6. Because the vector potential acts along the length of the wire, because the resonators are short compared with a wavelength, and because the resonators are far enough apart that the far-field approximation is valid, Eq. 4.46 can be used for the ϕ term in Eq. 4.56. For the $\mathcal{V}$ terms, however, we must separately evaluate the contribution from the charges at $R \pm l/2$. Evaluating the t component of Eq. 4.6 for the forward light cone,

$$\begin{aligned}
\mathcal{V}(t, l/2) &= \frac{\mu_0 c^2}{4\pi R} Q(t + R/c, R + l/2) \\
&\quad + Q(t + R/c - l/c, R - l/2) \\
\mathcal{V}(t, -l/2) &= \frac{\mu_0 c^2}{4\pi R} Q(t + R/c, R - l/2) \\
&\quad + Q(t + R/c + l/c, R + l/2)
\end{aligned} \tag{4.57}$$

where the c^2 arises because the t component of $\mathbf{J}$ is $c\rho$ and the t component of $\mathbf{A}$ is $\mathcal{V}/c$. We can simplify Eq. 4.57 by using the fact that, at a particular time, the charges on the capacitor plates of the source resonator are equal and opposite. Taking the charge $Q/2$ as being that on the top plate ($y = R + l/2$), and $-Q/2$ as being that on the bottom plate ($y = R - l/2$), and using the forward light cone as our time reference, we obtain

$$\begin{aligned}
\mathcal{V}_{\text{top}} &= \frac{\mu_0 c^2}{8\pi R}\Big(Q(t) - Q(t - l/c)\Big) \\
\mathcal{V}_{\text{bot}} &= \frac{\mu_0 c^2}{8\pi R}\Big(-Q(t) + Q(t + l/c)\Big)
\end{aligned} \tag{4.58}$$

which becomes

$$\mathcal{V}_{\text{top}} - \mathcal{V}_{\text{bot}} = \frac{\mu_0 c^2}{8\pi R}\Big(2Q(t) - Q(t - l/c) - Q(t + l/c)\Big) \tag{4.59}$$

The same result is obtained on the backward light cone. Differentiating and recognizing that $\dot{Q} = I$, we obtain:

$$\dot{\mathcal{V}}_{\text{top}} - \dot{\mathcal{V}}_{\text{bot}} = \frac{\mu_0 c^2}{8\pi R}\Big(2I(t) - I(t - l/c) - I(t + l/c)\Big) \tag{4.60}$$

Using Taylor's series to approximate a second derivative, we obtain:

$$\dot{\mathcal{V}}_{\text{top}} - \dot{\mathcal{V}}_{\text{bot}} \approx -\frac{\mu_0 c^2}{8\pi R}\frac{l^2}{c^2}\,\ddot{I} = -\frac{\mu_0 l^2}{8\pi R}\,\ddot{I} \tag{4.61}$$

Substituting into Eq. 4.56, we obtain

$$LC\,\ddot{\Phi} - \Phi \approx -\frac{LC\mu_0 l^2}{8\pi R}\,\ddot{I} - \phi \tag{4.62}$$

where ϕ is obtained from Eq. 4.46:

$$\begin{aligned} LC\,\ddot{\Phi} - \Phi &\approx -\frac{LC\mu_0 l^2}{8\pi R}\,\ddot{I} - \frac{\mu_0}{8\pi}\,\frac{l^2}{R}\,I \\ &\approx -\frac{\mu_0 l^2}{8\pi R}\Big(LC\,\ddot{I} + I\Big) \end{aligned} \tag{4.63}$$

As noted at the beginning of this derivation, the total coupling from far-away resonators is represented by the right-hand side of Eq. 4.63, the first term being the electrostatic coupling, and the second term being the magnetic coupling. Because the source resonator is very far away, we can evaluate I from ordinary circuit considerations, as if the coupling were zero:

$$V = L\dot{I} \qquad I = -C\dot{V} \tag{4.64}$$

from which we obtain

$$I = -LC\ddot{I} \tag{4.65}$$

We therefore conclude that the coupling between electric dipole resonators aligned along their axes, as given by the right-hand side

of Eq. 4.63, is identically zero. The electric coupling is exactly equal to and opposite in sign from the magnetic coupling. We noted in the previous section that the electric coupling between two y-directed dipoles in the xz plane is zero, and therefore that the total coupling is equal to the magnetic coupling. For dipoles at intermediate angles, the electric coupling cancels some, but not all, of the magnetic coupling. We can evaluate the angular dependence as a function of angle in the far-field limit by noting that each factor l/c in the l^2/c^2 in Eq. 4.61 represents the delay between the potentials from the two ends of one dipole with respect to the line joining the two dipoles. The general form of Eq. 4.61 thus becomes

$$\begin{aligned} \dot{\mathcal{V}}_{\text{top}} - \dot{\mathcal{V}}_{\text{bot}} &\approx -\frac{\mu_0 c^2}{8\pi R}\frac{l^2}{c^2}\cos\theta_1 \cos\theta_2 \ \ddot{I} \\ &= -\frac{\mu_0 l^2}{8\pi R}\cos\theta_1 \cos\theta_2 \ \ddot{I} \end{aligned} \tag{4.66}$$

where θ_1 and θ_2 are the angles between the axes of the two dipoles and the line joining them. For dipoles that are small compared with a wavelength, the magnetic coupling ϕ does not depend on time delay. One l of the l^2 in Eq. 4.46 represents the integral of the current along the source dipole, which has no angular dependence. The second l represents the distance along which the $\vec{A}$ vector is integrated. Because it is the component of the $\vec{A}$ vector from dipole 2 along the axis of dipole 1 that determines the magnitude of the flux integral, the general form of Eq. 4.46 is

$$\phi \approx \frac{\mu_0}{8\pi}\frac{l^2}{R}\cos\left(\theta_1 - \theta_2\right) \ I_2 \tag{4.67}$$

The general far-field form of Eq. 4.63, including the total coupling strength as a function of the angle between the axes of the two dipoles and the line joining them, is, therefore:

$$\begin{aligned} LC\,\ddot{\Phi} - \Phi &\approx -\frac{\mu_0 l^2}{8\pi R}\Big(LC\ddot{I}\cos\theta_1\cos\theta_2 \ + I\cos\left(\theta_1 - \theta_2\right)\Big) \\ &= -\frac{\mu_0 l^2 \ I}{8\pi R}\Big(-\cos\theta_1\cos\theta_2 \ + \cos\left(\theta_1 - \theta_2\right)\Big) \\ &= \frac{\mu_0 l^2 \ I}{8\pi R}\sin\theta_1 \sin\theta_2 \end{aligned} \tag{4.68}$$

of which Eqs. 4.47 and 4.63 are special cases. The angular dependence of the right-hand side of Eq. 4.68 is called the **radiation pattern** of the dipole, each factor of $\sin\theta$ normally being attributed to its respective dipole. For a practical antenna, the radiation pattern is one of the key design objectives. Many books on antenna theory are totally occupied with such questions, and have developed elaborate shorthand methods to aid the calculations. The point of our abbreviated treatment is to illustrate that any such calculation, involving electric and magnetic coupling of structures of any shape, can be carried out easily from first principles using the direct interaction of electron wave functions given by Eq. 4.6.

Part 5

Electromagnetic Interaction of Atoms

> The wave function Ψ for an electron in an atom does not, then, describe a smeared-out electron with a smooth charge density. The electron is either here, or there, or somewhere else, but wherever it is, it is a point charge.
>
> —R.P. Feynman[1]

> Concepts which have proved useful for ordering things easily assume so great an authority over us, that we forget their terrestrial origin and accept them as unalterable facts ... The road of scientific progress is frequently blocked for long periods by such errors.
>
> —Albert Einstein[2]

> There are no quantum jumps, nor are there particles!
>
> —H.D. Zeh[3]

At long last, we are in a position to treat the interaction of atoms through electromagnetic coupling—the topic that gave rise to quantum mechanics in the first place. We have outlined some of the debate between Einstein and Bohr in earlier parts of this manuscript. Einstein was firmly of the belief that statistics have no place in the fundamental laws of physics, while Bohr believed that only statistical information is meaningful in quantum theory. Unfortunately, much of the debate centered around the uncertainty relation, which, from our present point of view, is not about statistics at all, but results from the wave nature of matter. At the time, there were no compelling experiments where the wave nature of

[1]The Feynman quotation appears on page 21-6 of Ref. 7, in the section in which he discusses the two kinds of momentum. This section contains a number of misconceptions, as I have mentioned in the preface; it certainly does not represent Feynman at his best.

[2]This quotation is from his 1916 obituary for Mach, quoted on page 15 of Fine's book, *The Shaky Game* (18).

[3]This quotation is the title of Zeh's manifesto (73).

matter was manifest in a non-statistical manner. During that entire period, there was spirited debate in and around the apparently discontinuous nature of quantum transitions, a retrospective of which is Schrödinger's paper, *Are there quantum jumps?* (74). Under the pressure of Bohr's constant verbal abuse, Schrödinger finally gave up his quest for a theory that would be continuous in both space and time. It was much later that this quest was put on a firm foundation, most notably by Barut,[4] Zeh, and their collaborators. I will not attempt to review the development of these ideas here because an outstanding reference (76) is available. Instead, I shall describe continuous quantum transitions in a simple and intuitively appealing way by extending the notions of collective electrodynamics to the wave function of a single electron. We shall require only the most rudimentary techniques of conventional quantum theory, applied in the sense described by Barut(75) and Zeh (73).

5.1 The Two-State System

Let us consider a simple two-state system. The system has a single positive charge around which there are two eigenstates, labeled 1 and 2, that an electron can occupy. In State 1, the electron has wave function Ψ_1 and energy E_1; in State 2, it has wave function Ψ_2 and energy $E_2 > E_1$:

$$\Psi_1 = R_1 e^{i\frac{E_1}{\hbar}t} \qquad \Psi_2 = R_2 e^{i\frac{E_2}{\hbar}t} \tag{5.1}$$

We visualize the wave function as an abstraction of a totally continuous matter wave, and take R_1 and R_2 as real functions of the space coordinates. We can interpret all of the usual operations involving the wave function as methods for computing properties of this continuous distribution. The only particularly quantal assumption involved is that the wave function obeys a **normalization condition**:

$$\int \Psi^* \Psi = 1 \tag{5.2}$$

This condition assures that the total charge will be a single electronic charge, and the total mass will be a single electronic mass.

[4]The most accessible treatment is given by Barut and Dowling (75).

The first moment $< x >$ of the electron distribution along some coordinate x is:

$$< x > = \int \Psi^* x \Psi \tag{5.3}$$

where the integral is taken over all space where the wave function is non-vanishing. This moment, when multiplied by the electronic mass, gives the center of mass; and, when multiplied by the electronic charge, gives the electric dipole moment of the charge distribution. For the purposes of estimating radiative coupling between two atoms, the moment $< x >$ can be identified with the length l of the dipole resonator discussed in Part 4 (p. 73). We notice that there is no need to introduce the notion of probability at any point in the discussion. The issue of probability comes up when we consider disordered systems where the phases of individual interactions are unknown. Statistical quantum texts are preoccupied with the question of which quantities are "measurable" and which are not—that question does not appear in the present discussion.

From Eqs. 5.3 and 5.1, the first moment for either eigenstate is:

$$< x_i > = \int \Psi_i^* x \Psi_i = \int R_i^* x R_i = \int R_i^2 x \tag{5.4}$$

For symmetrical wave functions, the integral vanishes, and the dipole moment is zero. Even if the wave function is not symmetrical, the dipole moment, and all higher moments as well, are independent of time. By their very nature, eigenstates are stationary states—none of their spatial attributes are functions of time. The notion of stationarity is the quantum answer to the original question about atoms: why doesn't the electron orbiting the nucleus radiate its energy away? Even Schrödinger was ambivalent about this point. In the notes for a 1955 lecture (77), he wrote:

> Then I don't know why the pure proper modes should be any more "stationary" than any superposition, why the system should have any preference for remaining in such a state rather than . . .

In his 1917 book, *The Electron*, R.A. Millikan[5] anticipates the solution in his comment about the

> ...apparent contradiction involved in the non-radiating electronic orbit—a contradiction which would disappear, however, if the negative electron itself, when inside the atom, were a ring of some sort, capable of expanding to various radii, and capable, only when freed from the atom, of assuming essentially the properties of a point charge

Ten years before the statistical quantum theory was put in place, Millikan had clearly seen that a continuous, symmetric electronic charge distribution would not radiate, and that the real problem was the assumption of a point charge. The continuous wave nature of the electron implies a continuous charge distribution. That smooth charge distribution can propagate around the nucleus and thereby generate a magnetic moment, as observed in many atoms. The smooth propagation around the nucleus does not imply radiation.

5.2 Transitions

In order to radiate electromagnetic energy, the charge distribution must change with time. Because the spatial attributes of the system in an eigenstate are stationary, the system in an eigenstate cannot radiate energy. Because the eigenstates of the system form a complete basis set, any behavior of the system can be expressed by forming a linear combination of eigenstates. We consider the simplest such combination:

$$\Psi = A\Psi_1 + B\Psi_2 \tag{5.5}$$

The spatial attributes of this mixed state are functions of time. Computing the electric dipole moment using the definitions of Ψ_1 and Ψ_2 from Eq. 5.1 in Eq. 5.3, we obtain:

$$< x > = A^2 \int \Psi_1^* x \Psi_1 + B^2 \int \Psi_2^* x \Psi_2 \tag{5.6}$$

[5]See Ref. 78. Millikan was the first researcher to directly observe and measure the quantized charge on the electron with his famous oil-drop experiment, for which he later received the Nobel prize. This reference contains a fascinating discussion of these experiments, and a wonderful section contrasting the corpuscular and the ether theories of radiation.

$$+ AB \int (\Psi_1^* x \Psi_2 + \Psi_2^* x \Psi_1)$$
$$= \text{constant} \; + \; ABd \; \cos \omega t$$

where $\omega = (E_2 - E_1)/\hbar$. The dipole moment of a superposition oscillates with a frequency that corresponds to the difference in energy between its two states, whereas the dipole moment of a pure eigenstate is not a function of time. Eq. 5.6 makes it abundantly clear why the system can stay in an eigenstate but not in a superposition: An oscillating dipole moment radiates energy; a stationary dipole moment does not. The factor $d = 4 \int R_1 R_2 x$ is a measure of the maximum strength of the oscillating dipole moment. If one R is an even function of x and the other is an odd function of x, then this factor is nonzero, and the transition is said to be **electric dipole allowed**. If both R_1 and R_2 are either even or odd functions of x, then this factor is zero, and the transition is said to be **electric dipole forbidden**. Even in this case, some other moment of the distribution generally will be nonzero, and the transition can proceed by magnetic dipole, magnetic quadrapole, or other higher-order moments. For now, we will assume that the transition is electric dipole allowed. The atom in a superposition state is thus a very small quantum resonator, with properties much like those of the electric dipole resonator treated in Section 4.13 (p. 95).

5.3 Transitions in Macroscopic Quantum Resonator

There is an essential difference between the atomic and the macroscopic quantum resonators. In the macroscopic dipole resonator, the electron wave function is constrained by the fixed physical shape of the superconductor. Because the capacitor breaks the superconducting loop, the phase accumulation from one end to the other can take on any one of a continuum of values. In the case of the atom, however, the wave function is constrained to come back on itself, exactly like the wave function in a closed superconducting loop. Each eigenstate corresponds to a wave function having a different integral phase around the loop. With any given integer number of half-cycles around the loop, the phase maps back on itself, and the properties of the system are stationary. The only way to get something between an integer number of half-cycles is to "break" the superconducting state, thereby combining two

eigenstates into a superposition. The charge density of this superposition oscillates with time, and thus radiates energy.

In a simple superconducting loop, such as the one described in Section 1.1 (p. 9), there is no simple way to mix two states of different phase accumulation. If, however, we insert a "weak link" in the superconductor at one point in the loop, we can arrange to develop additional phase shift across the link. In this way, we can generate transient states with a phase accumulation that is not an integral number of cycles. A full treatment of this situation is well beyond the scope of the present discussion.[6] For the moment, we will simply assume that we can form a state Ψ with nonintegral phase accumulation by a simple superposition of two eigenstates Ψ_1 and Ψ_2, which have phase accumulations $n_1\Phi_0$ and $n_2\Phi_0$, respectively. From Eq. 5.1,

$$\begin{aligned} \Psi_1 &= R\, e^{i(n_1\Theta + E_1 t/\hbar)} \\ \Psi_2 &= R\, e^{i(n_2\Theta + E_2 t/\hbar)} \end{aligned} \tag{5.7}$$

where R is a function that confines the wave function to within the superconductor, Φ_0 is the flux quantum, Θ is the angle around the loop, and the energies are related to the fluxes and the inductance L of the loop. Using the Eq. 3.6 for the energies,

$$\begin{aligned} E_1 &= n_1^2\, \frac{\Phi_0^2}{2L} = n_1^2\, E_0 \\ E_2 &= n_2^2\, \frac{\Phi_0^2}{2L} = n_2^2\, E_0 \end{aligned} \tag{5.8}$$

A superposition of these two states is:

$$\Psi = \Psi_2 + \Psi_2 = R\left(A\, e^{i(n_1\Theta + E_1 t/\hbar)} + B\, e^{i(n_2\Theta + E_2 t/\hbar)}\right) \tag{5.9}$$

[6]Phenomena associated with structures of this type were first discussed by Brian Josephson (79, 80). Devices with two weak links are called Superconducting Quantum Interference Devices (SQIDs). A nice account is given by Tinkham (Chapter 6 of Ref. 81).

The electron density is given by Eq. 5.2:

$$\begin{aligned}\Psi^*\Psi &= \text{constant} + 2ABR^2\left(e^{i(\Delta n\,\Theta+\omega\,t)} + e^{-i(\Delta n\,\Theta+\omega\,t)}\right) \\ &\quad + O(2n) \\ &= \text{constant} + 4ABR^2\cos\left(\Delta n\;\Theta + \omega\,t\right) + O(2n)\end{aligned} \tag{5.10}$$

where $\omega = (E_2 - E_1)/\hbar$, $\Delta\, n = n_2 - n_1$, and $O(2n)$ indicates high-order terms that are of no interest in the present discussion. For the special case of $\Delta n = 1$, the time-varying part of the electron density is:

$$\text{Density} = 4ABR^2\cos\left(\Theta + \omega\,t\right) \tag{5.11}$$

In other words, the net charge density is a sinusoidal function of angle around the loop, rotating with angular frequency ω. This rotating dipole couples to any other dipole on its light cone. The rotating four-potential of this distribution is called **circularly polarized radiation**. Along either the x or y axis, the dipole behaves exactly like that of Eq. 5.6.

The operational significance of the phase constraint around the loop is that the dipole moment is not fixed by the physical length of the resonator, but depends instead on the state of the system. In an eigenstate, the dipole moment is zero. As the superposition state develops, the dipole moment increases, reaching a maximum when the upper and lower states make equal contributions to the total wave function ($A = B$). Therefore, the problem of two coupled quantum systems in a superposition state is essentially nonlinear in character. This nonlinearity leads to the appearance of rapid transitions between eigenstates, which in traditional treatments of quantum mechanics were taken to be discontinuous. We will now trace the continuous trajectory of the state of two radiatively coupled atoms through such a "quantum jump."

5.4 Radiation Transfer Between Atoms

We have developed a detailed description of the energy-transfer process between macroscopic quantum resonators in Part 4 (p. 73). We are now in a position to understand the radiative transfer between two identical atomic systems. Let us consider one atom in an initial state that is nearly the eigenstate of the upper energy level, but that has been ever so slightly perturbed by a tiny addi-

tion of the wave function of the lower energy level. Such a state can be written:

$$A_1^2 = \epsilon \qquad B_1^2 = 1 - \epsilon \tag{5.12}$$

In addition, we suppose that the first atom is coupled to a second, identical atom in the lower energy level, and perturbed in a way that is perfectly matched to the perturbation of the first atom:

$$A_2^2 = 1 - \epsilon \qquad B_2^2 = \epsilon \tag{5.13}$$

According to Eq. 5.6, the dipole moment of each system will oscillate with a small amplitude proportional to $\sqrt{\epsilon}$. Because Eq. 5.6 is symmetric in A and B, if $A_1 = B_2$ and $A_2 = B_1$, the amplitude of oscillation of the two atoms is identical. Now, let us suppose that the relative phase of the two systems is also that considered in Eq. 4.68. The two atoms act like two small dipole resonators, and energy is radiatively transferred from the first to the second. The frequency of the radiation is $\omega = (E_2 - E_1)/\hbar$. By the Hellman–Feynman theorem, the total energy E of a system whose wave function is a mixture of eigenstates is:

$$E = A^2 E_1 + B^2 E_2 = A^2 E_1 + \left(1 - A^2\right) E_2 \tag{5.14}$$

The radiative coupling decreases the energy of the first atom, thus decreasing B_1 and increasing A_1, and concomitantly increasing B_2 and decreasing A_2, as required by Eq. 5.14. We can solve explicitly for the amplitude of the oscillation as a function of time for this transition. The rate of energy transfer will, according to Eq. 4.68, be proportional to $l^2 \propto \epsilon$. From Eq. 5.6, dAB is the amplitude of the oscillation of both systems. The rate of energy loss of the first atom and of energy gain of the second, both due to radiation, is proportional to the square of this amplitude. For the first atom,

$$-\frac{\partial E}{\partial t} \propto B_1^2 \left(1 - B_1^2\right) \tag{5.15}$$

Solving Eq. 5.14 for B_1^2, we obtain:

$$B^2 = \frac{E - E_1}{E_2 - E_1} \tag{5.16}$$

In other words, B_1^2 is just the energy of the first atom, normalized to the transition energy, and using E_1 as its reference energy. Differentiating Eq. 5.16 and substituting into Eq. 5.15, we obtain:

$$-\frac{\partial B_1^2}{\partial t} \propto B_1^2 \left(1 - B_1^2\right) \tag{5.17}$$

The solution[7] of Eq. 5.17 is of the form:

$$B_1^2 = \frac{1}{e^{\alpha t} + 1} \qquad A_1^2 = \frac{1}{e^{-\alpha t} + 1} \tag{5.18}$$

Following the same set of steps for the second atom, we obtain:

$$A_2^2 = \frac{1}{e^{\alpha t} + 1} \qquad B_2^2 = \frac{1}{e^{-\alpha t} + 1} \tag{5.19}$$

where

$$\alpha = \frac{1}{\tau} \propto (dq)^2 \tag{5.20}$$

Here, α is called the **decay constant** of the transition, and τ is called the **lifetime** of the excited state Ψ_2.

We have thus derived the natural, continuous form of a "quantum jump." We can understand the transition from the following simple model: In the beginning of a transition, some perturbation couples two atoms, and puts both in a mixed state with exactly the same difference of energies and exactly the right phase. One of these atoms must start in the upper energy level, and the other in the lower energy level. Once the coupled mixed state starts to develop, it becomes self-reinforcing. The energy transferred from one atom to the other causes an increase in the minority state of the superposition, thus increasing the dipole moment of both states, and thereby increasing the coupling and, hence, the rate of energy transfer. This self-reinforcing behavior gives the transition its initial exponential character. Once the transition is fully under way, the two states are nearly equally represented in the superposition, and the coupled system closely resembles the coupled resonators analyzed in Part 4 (p. 73). Once the transition has run its course, each atom settles into its final eigenstate. H.A.

[7]The author thanks Rahul Sarpeshkar for pointing out this solution.

Lorentz,[8] for whom the Lorentz transformation was named, had this description of the transition:

> It is, perhaps, more satisfactory to suppose that on the occasion of a quantum jump the atom itself is transformed into a vibrator. This would imply that the jump is by no means instantaneous, but that the atom passes from the first stationary state to the state of a vibrator, and acquires the second stationary state only at the moment at which, radiating as a vibrator, it has reached the energy of the second stationary state, and then passes into that state and ceases to radiate.

The rate of energy transfer from atom 1 to atom 2 is obtained by substituting Eq. 5.18 into Eq. 5.17.

$$\frac{\partial B_1^2}{\partial t} = \frac{e^{\alpha t}}{(e^{\alpha t}+1)^2} \tag{5.21}$$

By the time the transition is complete, B_1 has changed from unity to zero; hence, by Eq. 5.16, the first atom has transferred a total amount of energy

$$\Delta E = \Delta B_1 \, (E_2 - E_1) = E_2 - E_1 = \hbar\omega \tag{5.22}$$

to the second. This electromagnetic transfer of a certain quantity of energy that is proportional to frequency is called a **photon**. Our description differs from Lorentz's only in detail. The passage from the initial stationary state to the state of a vibrator is accomplished smoothly; the superposed state develops an oscillating dipole moment as it loses energy. The coupled system also approaches its final state smoothly as both atoms lose all dipole moment and each settles into its stationary state.

We can thus answer Schrödinger's rhetorical question, "Are there quantum jumps?" Indeed, there are quantum jumps, but they are not discontinuities. They may look discontinuous because of the nonlinear, self-reinforcing nature of a quantum transition; but at the fundamental level, everything can be followed in a smooth and continuous way, and we can view nature as being continuous in both space and time. This picture of nature is what Einstein wanted most. But to arrive at this picture, we had to give

[8]This quotation appears on page 151 of Ref. 82.

up the one-way direction of time, and allow coupling to everything on the light cone. This, too, was okay with Einstein. So why was he so hung up on local causality? Why do all the textbooks state that the coupling of states unified by a light cone is a violation of relativity? In science, as in all other aspects of human endeavor, each age has its blind spots, and what is impossible to comprehend in one generation seems natural and obvious to another. So, after only one generation, Zeh could say, "There are no quantum jumps, nor are there particles." The coherent world has continuous wave functions, and abrupt-but-continuous quantum transitions.

It has often been said, as reflected in the Feynman quotation at the beginning of this section, that one cannot obtain the correct answers for the energy levels of the hydrogen atom by using a continuous charge distribution. As it turns out, these pronouncements are simply wrong. Throughout this derivation, we have treated the electron as a wave, continuous in space, carrying a continuous charge density with it. Ed Jaynes (83, 84) set the stage for this approach in 1958; Barut[9] showed that one obtains the correct energy levels as well as all the other effects that had been touted as triumphs of traditional quantum electrodynamics. Arriving at the correct results required taking into account the interaction of the electron with itself, exactly as we have done in the case of the superconducting loop. The electron wave function depends on the potential; the potential depends on the charge density that is determined by the wave function. Thus, we have an inherently nonlinear problem, as Jaynes had already pointed out. The nonlinearity of the problem poses some *computational* issues, but no *conceptual* issues. Conceptually, the continuous charge distribution of the electron wave function is much more understandable than are the multiple levels of infinity that result from point charges (23). By using a continuous, self-interacting quantum system as the conceptual basis for electromagnetic theory, we have advanced one step closer to a simpler and more unified discipline.

[9]See Barut and Dowling (75) and the earlier references cited therein.

This page intentionally left blank

Appendix

The Fermi Sphere

It is typical of modern physicists that
they will erect skyscrapers of theory
upon the slender foundations
of outrageously simplified models.

—J.M. Ziman[1]

A.1 Monolithic Solid

In a solid, the most appropriate electronic states are traveling waves described by a wave function ψ:

$$\psi = e^{i(\omega t - k \cdot r)} \tag{A.1}$$

where, for example, in Cartesian coordinates:

$$k \cdot r = k_x x + k_y y + k_z z \tag{A.2}$$

The propagation vector or wave vector k has three components: k_x, k_y, and k_z; its "length" $|k|$ is given by $|k|^2 = k_x^2 + k_y^2 + k_z^2$.

We can think of the solid as a "box" with sides perpendicular to the coordinate axes, and having dimensions l_x, l_y, and l_z. The electrons are confined to the solid, so the wave function must be zero outside the box. Only certain values of k satisfy the boundary conditions that require that the wave function vanish at the edges of the box. We write these conditions as:

$$k_x\, l_x = \pi\, n_x \qquad k_y\, l_y = \pi\, n_y \qquad k_z\, l_z = \pi\, n_z \tag{A.3}$$

By the Pauli exclusion principle,[2] two electrons with the same spin are not allowed in the same state. The first electrons we add to a solid fill the lowest energy states, which are the states of lowest k (longest wavelength). We notice that for every state with wave vector k there is another state of precisely the same energy

[1] A delightful discussion of the Fermi Surface is given by Ziman (85). The opening quotation can be found on page 8 of that reference.

[2] An extended discussion of this topic, with many useful references, is given in Ref. 86.

with wave vector $-k$. Thus, at any state of "fill," the electron system will have a remarkable kind of symmetry: Every electron in a state k will have a matching electron in state $-k$, and the two together form a standing wave. As long as states are occupied symmetrically, neither the individual pairs of electrons nor the entire collective state has any net propagation. This symmetry is not dependent on any symmetry properties of the solid, but is a result of the time-reversal symmetry of the electron wave function. The collective wave function is symmetric in a spatial as well as a temporal sense. Each of the individual component wave functions has a different k; by construction, these wave functions are spatially orthogonal. The sum of a very large number of them becomes spatially extraordinarily smooth and featureless.

At some point, we will have added just enough electrons to neutralize the positive charges in the solid; it will then be electrically neutral, and we can stop. For a typical metal, we will have added of the order of 10^{23} electrons per cubic centimeter. The electrons fill up all the states to a certain energy, called the **Fermi energy** or **Fermi level** E_f.

A.2 The k-Lattice

A construction that is universally used to visualize the relation between energy and wave vector takes place in a coordinate system whose axes are k_x, k_y, and k_z. The remarkable property of states expressed in this "k-space" can be seen from Eq. A.3—*the states form a periodic lattice in k space.* In general, the spacing of the lattice will be different in the three principal directions. The volume of one unit cell in k-space is:

$$\begin{aligned} \text{unit cell volume in } k \text{ space} &= \frac{\pi}{l_x} \cdot \frac{\pi}{l_y} \cdot \frac{\pi}{l_z} \\ &= \frac{\pi^3}{\text{volume in real space}} \end{aligned} \tag{A.4}$$

A.3 Fermi Sphere

A given energy corresponds to a surface in k-space. In particular, the most important energy, the Fermi energy, forms the **Fermi surface**. Each state lying within the Fermi surface contains two electrons of opposite spin; those outside the Fermi surface are empty (at zero temperature).

For simplicity, we assume that the solid is totally symmetric: Any direction is equivalent to any other.[3] For our idealized symmetrical solid, the energy surfaces are spheres. The value of wave vector k_f, corresponding to electrons at the Fermi level, is just the "radius" of the Fermi sphere in k-space. That sphere must contain one state for every pair of electrons we have added. The total number of electrons is just the density n of electrons multiplied by the volume of the solid in real space. Multiplying both sides of Eq. A.4 by the number of electron pairs, we obtain

$$\text{Volume in } k \text{ space} = \frac{\pi^3}{2}\left(\text{Electron density in real space}\right) \quad \text{(A.5)}$$

where the factor of 2 comes from counting electrons instead of electron pairs:

$$\frac{4}{3}\pi k_f^3 = \frac{\pi^3}{2}\, n \qquad \Rightarrow \qquad k_f^3 = \frac{3}{8}\pi^2\, n \qquad \text{(A.6)}$$

It is significant that this result does not depend on the size or shape of the solid, but only on the density of the electrons.

A.4 Ring Geometry

The monolithic solid we have considered thus far has helped us understand the nature of electronic states in a solid, but it is limited by exactly the time-reversal symmetry that we found so remarkable—the collective state has no net motion in any direction. We can, however, create a geometry in which the general nature of the states developed above remain valid, but which allows net motion of the electron system: This geometry is a loop of wire. The quantization conditions in the two directions, y and z, perpendicular to the length are identical to those that led to Eq. A.3. Along the length of the wire (x direction), we have a *cyclic* constraint instead of a boundary condition—the wave function must come back in phase with itself after a trip around the loop. If the length of the loop is l_x, the condition is $k_x l_x = 2\pi n_x$. This condition is really the same as that of Eq. A.3, where the "round trip" distance is $2l_x$—across the solid and back.

[3]This approximation is equivalent to assuming the positive charge is "spread out" in a uniform manner. Such a construction is called "Jellium" (26).

A.5 Collective States

The simplest non-stationary state in this geometry can be generated by *translating the Fermi sphere one lattice unit* $\epsilon = 2\pi/l_x$ *in the* k_x *direction.* This operation corresponds to multiplying each individual wave function by a factor $e^{i\epsilon x}$. The resulting collective wave function is no longer featureless, but takes on the periodicity $e^{i\epsilon x}$ that is shared by all of its components. It is in this way that an enormous, complex, macroscopic ensemble can manifest quantum effects that we normally associate with a system of atomic dimensions. We need make no assumptions about the detailed character of the individual states, nor about the interaction among the states, other than that the quantization occurs as a result of the single-valued nature of the wave function in a spatially constrained geometry.

Because of the periodic nature of the lattice in k-space, we can construct states of higher net propagation vector by translating the Fermi sphere several lattice units rather than just one. In a ring geometry, the state formed by translating the Fermi sphere m lattice units corresponds to a magnetic flux of $mh/2q$.

A.6 Classes of Solids

The preceding discussion applies to any solid. It is well known, however, that solids can be classed as normal metals, superconductors, and semiconductors. All of these distinctions have to do with the details of electron states around the Fermi energy. Normal metals have a continuous distribution of states in the energy range around the Fermi energy, whereas superconductors and pure semiconductors have an **energy gap**—a region of energy where there are no electron states. In a semiconductor, the energy gap is fixed with respect to the lattice, so no states of net propagation are possible without exciting electrons past the energy gap. In a superconductor, however, the energy gap is the result of collective interaction (87, 88) among the electrons. For that reason, the energy gap translates in k-space along with the Fermi sphere. Thus, the energy gap prevents electrons from being scattered into lower energy states outside the Fermi sphere, and ensures the existence of a stable macroscopic quantum state that shares a common displacement in k-space.[4]

[4]Fritz London (89) seems to have been the first to recognize the superconductor as a quantum state of macroscopic dimensions. A fascinating history of superconductivity is recounted by Dahl (90).

Epilogue

> Assuming the success of efforts to accomplish a complete physical description, the statistical quantum theory would, within the framework of future physics, take an approximately analogous position to the statistical mechanics within the framework of classical mechanics. I am rather firmly convinced that the development of theoretical physics will be of this type, but the path will be lengthy and difficult.
>
> —Albert Einstein[1]

The early pioneers saw the quantum nature of matter indirectly, obscured by the statistical nature of the systems that exhibited it: black-body radiation, spectral lines in glowing gasses, darkened grains in photographic emulsion. It is a tribute to their insight and persistence that they were able to construct a formalism that still gives useful results to the present day. Much has been written on the brilliance of their contributions, and I will not repeat those kudos here. Much has also been written on the total failure of the Copenhagen statistical interpretation to provide a satisfactory conceptual framework for understanding quantum phenomena. Each of us who make our living working with ordinary, every-day, hands-on quantum systems has had to develop our own intuitive understanding of real quantum phenomena. By the orthodoxy of the age, any discussion of this intuitive understanding was prohibited. As more examples of collective, coherent quantum systems were developed, the actual understanding of each system departed more fundamentally from the statistical view. The emission of photons by glowing gasses was the original example around which the Copenhagen interpretation formed. Lasers, in which glowing gasses emit coherent light, are put forth as the quintessential example of quantum theory in action. But the statistical interpretation misses the very reason that the laser is interesting. In his epic textbook on lasers, Tony Siegman has this to say:[2]

[1] This quotation appears on page 672 of Ref. 20.

[2] This quotation appears in a delightful discussion of the importance of coherence, entitled *Coherence and "Photons,"* on page 33 of Ref. 91.

> We have hardly mentioned *photons* yet in this book. Many descriptions of laser action use a photon picture . . . in which billiard-ball-like photons travel through the laser medium. Each photon, if it strikes a lower-level atom, is absorbed and causes the atom to make a "jump" upward. On the other hand, a photon, when it strikes an upper-level atom, causes that atom to drop down to the lower level, releasing another photon in the process. Laser amplification then appears as a kind of photon avalanche process.
>
> Although this picture is not exactly incorrect, we will avoid using it to describe laser amplification and oscillation, in order to focus from the beginning on the *coherent* nature of the stimulated transition process. The problem with the simple photon description . . . is that it leaves out and even hides the important *wave* aspects of the laser interaction process . . . the whole stimulated transition process should be treated, not as a "photon process" but as a coherent wave process.

This monograph has shown how the traditional discipline of electromagnetism is most economically viewed as the study of coherent electron wave functions. These wave functions are continuous in space, representing a charge distribution that is distributed in space, and that interacts with itself. Any theory based on this view is essentially nonlinear in nature; that the nonlinear theory gives the correct energy levels for the hydrogen atom has been shown only recently (75).

The fundamentally nonlinear and nonlocal nature of atomic transitions was difficult for early workers. After their brilliant exposition of the oscillating dipole moment, Rice and Teller (92) provided this description:

> It may be seen that, although the electron does not "move" in a stationary state, motion can be readily produced as soon as an electron is in a superposition of two stationary states.

But, rather than follow this line of thought, they immediately adopted the party line:

> In this case, however, we can make only probability statements as to the energy of the electron.

From our present perspective, it is hard to imagine how the second sentence could possibly follow from the first. Fortunately, they were kind enough to share with us the conceptual problem:

> It seems very tempting to identify the frequency of the electronic motion with the frequency of the light emitted or absorbed. But we must not consider emission or absorption of light as a consequence of this electron oscillation, but rather as a phenomenon corresponding to it. Otherwise, we would obtain the result that a hydrogen atom which at the moment is with certainty in the first excited state does not move, does not oscillate, and does not emit light.

So the conceptual problem was not how energy is radiated by the oscillating wave function, but how the transition can be initiated in the first place. It is by now a common experimental fact that an atom, if sufficiently isolated from the rest of the universe, can stay in an excited state for an arbitrarily long period. It is also true that achieving that degree of isolation was not possible until the last few years. The mechanism for initiating an atomic transition is not present in the isolated atom; it is the direct result of coupling with the rest of the universe.

Ironically, the quantum nature of radiation was an invention of Einstein. In his 1905 paper (p. 86 in Ref. 55), he introduces this possibility:

> According to the assumption to be contemplated here, when a light ray is spreading from a point, the energy is not distributed continuously over ever-increasing spaces, but consists of a finite number of energy quanta that are localized in points in space, move without dividing, and can be absorbed or generated only as a whole.

His viewpoint reflected the then-universal belief, that the radiation itself contains the degrees of freedom in which the energy resides. The following year, in the paper that would eventually win him the Nobel prize (p. 192 in Ref. 55) he applied this line of reasoning to the photoelectric effect. Referring to his earlier (1905) paper, he comments:

> By a route described in that study, I was led to the view that light . . . can only be absorbed or emitted in quanta of energy

This statement contains no reference to where the energy is located. In 1909, Einstein gave an overview talk (p. 379 in Ref. 55) that clarified his views even further:

> According to the prevailing theory, an oscillating ion produces an outwardly propagated spherical wave. The opposite process

> does not exist as *an elementary process*. It is true that the inwardly propagated spherical wave is mathematically possible; however, its approximate realization requires an enormous amount of emitting elementary structures. Thus, the elementary process of light radiation as such does not possess the character of reversibility. Here, I believe, our wave theory is off the mark.
>
> In other words, the elementary radiation process seems to proceed such that it does not, as the wave theory would require, distribute and scatter the energy of the primary electron in a spherical wave propagating in all directions. Rather, it seems that at least a large part of this energy is available at some location . . . *the elementary process of radiation seems to be* **directed**.

From these statements, and from his 1905 paper, it might seem that Einstein would continue to pursue the idea of light quanta as little bullets carrying energy. Instead, he reached a tentative conclusion very much in keeping with the views put forward in this monograph:

> Would it not be possible to replace the hypothesis of light quanta by another assumption that would also fit the known phenomena? If it is necessary to modify the elements of the theory, would it not be possible to retain at least the equations for the propagation of radiation and conceive only the elementary processes of emission and absorption differently than they have been until now?

By 1909, Einstein had already moved past the naive "bullet" view of quanta that overtook physics in the late 1920s. Can there be any wonder that he was frustrated with the Copenhagen clan?

It is precisely Einstein and Schrödinger's program that is carried out in Part 5 (p. 103). Initiating a transition requires that signals propagate forward and backward in time, what Einstein called "the character of reversibility." The inwardly propagated spherical wave that Einstein saw as a mere mathematical possibility is, in reality, composed of the advanced waves from all atoms on the light cone that are responding to the atom in question, the same waves that caused radiation damping of our resonator in Part 4 (p. 73). In a time-symmetric universe, an isolated system does not exist. The electron wave function in an atom is particularly sensitive to coupling with other electrons; it is coupled either to

far-away matter in the universe or to other electrons in a resonant cavity or other local structure. In the initial parts of this monograph, we were able to ignore coupling to far-away matter because we used a collective structure in which there are 10^{23} electrons, arranged in such a way that the collective properties intrinsic to the structure scaled as the square of the number of electrons. In this way, we could see the basic laws clearly, and could include coupling to far-away matter as a perturbation.

Once we were able to derive the behavior of a macroscopic quantum resonator, we used only the most rudimentary assumptions of standard quantum theory to connect our treatment to the behavior of atomic transitions. We required the highly nonlinear, self-reinforcing nature of the quantum transition to make the transaction "directed" in Einstein's sense. Both the time-symmetric nature of quantum coupling and the nonlinear, self-reinforcing nature of the quantum transition are essential to make an Einstein-Schrödinger theory viable. The possibility of carrying out such a program was foreseen with considerable clarity by Dorling (23), and the connection with standard quantum formalism was articulated by Cramer (67).

Bohr was adamant that the only role of theory in science is to calculate certain "observables." Einstein foresaw a quantum theory that could be "understood" as well as provide an algorithm for obtaining certain numerical results. Dirac, originally accepting Bohr's line of reasoning, later thought better of it (93):

> Some physicists may be happy to have a set of working rules leading to results in agreement with observation. They may think that this is the goal of physics. But it is not enough. One wants to understand how Nature works.

Statistical quantum mechanics has never helped us understand how nature works; in fact, it actively impedes our understanding by hiding the coherent wave aspects of physical processes. It has forced us to wander seventy years in the bewilderness of "principles"—complementarity, correspondence, and uncertainty. We have seen that complementarity and uncertainty are natural attributes of any wave theory. Correspondence to classical mechanics was the root cause of the worst conceptual nightmares. The idea of a point particle brought with it infinite energies that must be "renormalized" away. Degrees of freedom in the vacuum

brought even more infinities, and made a sensible theory of gravitation impossible. The path has been, as Einstein predicted, lengthy and difficult; the challenge now is to put all of that behind us, and to start anew.

Following the tradition of Einstein and Schrödinger, the pioneers in this new endeavor, Jaynes, Cramer, Barut, Zeh, and others, have given us a great foundation: They have shown that traditional criticisms of this new approach are groundless. They have put us in a position to finally settle the Einstein–Bohr debate—with a resounding victory for Einstein.

Starting in the 1960s, astounding experimental demonstrations of numerous coherent, collective systems have provided a new conceptual base for Dirac's "understanding how Nature works." They give us Feynman's "way of thinking such that the law is evident." We have followed the line of reasoning suggested by these experiments far enough to see that electromagnetism is a direct expression of the coherent electron wave function. There is, as Einstein said, "only one reality to be described," so we have built a theory that "recognized this from the start instead of doing things twice."

I opened this monograph by acknowledging my debt of gratitude to Dick Feynman. I close by recognizing the enormous wealth of insight we have inherited from Einstein, only a minuscule fraction of which has, as yet, entered the consciousness of physics as a whole.

References

1. Einstein, A., and L. Infeld (1938) *Evolution of Physics* (Simon & Schuster, New York).
2. Einstein, A., and L. Infeld (1966) *Evolution of Physics* (Simon & Schuster, New York).
3. Koestler, A. (1964) *The Act of Creation* (Hutchison, London).
4. Einstein, A. (1995) *Essential Einstein*, A.B. Eddington, Ed. (Pomegranate, Rohnert Park, CA).
5. Feynman, R.P., R.B. Leighton, and M. Sands (1964) *The Feynman Lectures on Physics* (Addison-Wesley, Reading), Vol. II.
6. Feynman, R.P., F.B. Morinigo, and W.G. Wagner (1995) *Feynman Lectures on Gravitation* (Addison-Wesley, Reading).
7. Feynman, R.P., R.B. Leighton, and M. Sands (1965) *The Feynman Lectures on Physics* (Addison-Wesley, Reading), Vol. III.
8. Mead, C.A. (1994) *Journal of VLSI Signal Processing* 8, 9–25.
9. Gilder, G. (1989) *Microcosm: The Quantum Revolution in Economics and Technology* (Simon & Shuster, New York).
10. Feynman, R.P. (1996) *Feynman Lectures on Computation / Richard P. Feynman*, A.J.G. Hey and R.W. Allen, Eds. (Addison-Wesley, Reading).
11. Mead, C.A. (1989) *Analog VLSI and Neural Systems* (Addison-Wesley, Reading).
12. Mead, C.A. (1997) *Proceedings National Academy Science USA* 94, 6013–6018.
13. Feynman, R.P. (1999) *Feynman and Computation*, A.J.G. Hey, Ed. (Addison-Wesley, Reading).
14. Feynman, R.P., R.B. Leighton, and M. Sands (1963) *The Feynman Lectures on Physics* (Addison-Wesley, Reading), Vol. I.
15. Mehra, J., and H. Rechenberg (1982) *The Historical Development of Quantum Theory* (Springer-Verlag, New York).
16. Jammer, M. (1966) *The Conceptual Development of Quantum Mechanics* (McGraw-Hill, New York).

17. Whitaker, A. (1996) *Einstein, Bohr and the Quantum Dilemma* (Cambridge University Press, Cambridge).
18. Fine, A. (1996) *The Shaky Game* (University of Chicago Press, Chicago).
19. Beller, M. (1998) *Physics Today* Sept., 29–34.
20. Einstein, A. (1970) *Albert Einstein, Philosopher-Scientist*, P.A. Schilpp, Ed. (Open Court, Peru, IL), pp. 3–94 & 665–688.
21. Ballentine, L.E. (1970) *Reviews of Modern Physics* 42, 357–381.
22. Davies, P.C.W., and J.R. Brown (1986) *The Ghost in the Atom* (Cambridge University Press, Cambridge).
23. Dorling, J. (1987) *Schrödinger, Centenary Celebration*, C.W. Kilmister, Ed. (Cambridge University Press, Cambridge), pp. 16–40.
24. Greenstein, G., and A.G. Zajonc (1997) *The Quantum Challenge* (Jones & Bartlett, Sudbury, MA).
25. Gribben, J. (1995) *Schrödinger's Kittens and the Search for Reality* (Little, Brown & Co., Boston).
26. Pines, D., and P. Nozières (1989) *The Theory of Quantum Liquids* (Addison-Wesley, Reading), Vol. I.
27. Mach, E. (1960) *The Science of Mechanics* (Open Court, La Salle, IL).
28. Jaynes, E.T. (1990) *Complexity, Entropy, and the Physics of Information*, W.H. Zurek, Ed. (Addison-Wesley, Reading), pp. 381–403.
29. Dirac, P.A.M. (1966) *Lectures on Quantum Field Theory* (Yeshiva University & Academic Press, New York).
30. Deaver, B.S., and W.M. Fairbank (1961) *Physical Review Letters* 7, 43–46.
31. Doll, R., and N. Nabauer (1961) *Physical Review Letters* 7, 51–52.
32. Feynman, R.P., and A.R. Hibbs (1965) *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York).
33. Maxwell, J.C. (1953) *A Treatise on Electricity and Magnetism* (Dover, New York), Vol. 2.
34. Schrödinger, E. (1926) *Annalen der Physik* 81, 109.
35. Schrödinger, E. (1982) *Collected Papers on Wave Mechanics* (Chelsea, New York).

36. O'Raifeartaigh, L. (1997) *The Dawning of Gauge Theory* (Princeton University Press, Princeton).
37. Tonomura, A. (1998) *The Quantum World Unveiled by Electron Waves* (World Scientific, Singapore).
38. Morse, P.M., and H. Feshbach (1953) *Methods of Theoretical Physics I* (McGraw-Hill, New York).
39. Gradshteyn, I.S., and I.M. Ryzhik (1994) *Integrals, Series, and Products*, A. Jeffrey, Ed. (Academic Press, London).
40. Ramo, S., J.R. Whinnery, and T. van Duzer (1994) *Fields and Waves in Communication Electronics* (John Wiley, New York).
41. Wheeler, J.A., and R.P. Feynman (1949) *Reviews of Modern Physics* 21, 425–435.
42. Heaviside, O. (1889) *Philosophical Magazine* Jan., 30.
43. Hertz, H. (1900) *Electric Waves* (Macmillan, London).
44. Aharonov, Y., and D. Bohm (1959) *Physical Review* 115, 485–491.
45. Werner, F.G., and D.R. Brill (1960) *Physical Review Letters* 4, 344.
46. Chambers, R.G. (1960) *Physical Review Letters* 5, 3–5.
47. Boersch, H., H. Hamisch, D. Wohlleben, and K. Grohmann (1960) *Zeitschrift fur Physik* 159, 397.
48. Voigt, W. (1887) *Göttinger Nachrichten*, page 45.
49. O'Rahilly, A. (1965) *Electromagnetic Theory* (Dover, New York), Vol. 1.
50. O'Rahilly, A. (1965) *Electromagnetic Theory* (Dover, New York), Vol. 2.
51. Einstein, A. (1905) *Annalen der Physik* 17, 891–921.
52. Rindler, W. (1960) *Special Relativity* (Oliver and Boyd, London).
53. Misner, C., K. Thorne, and J.A. Wheeler (1973) *Gravitation* (Freeman, New York).
54. Sommerfeld, A. (1952) *Electrodynamics* (Academic Press, New York).
55. Einstein, A. (1989) *The Collected Papers of Albert Einstein*, Anna Beck and Peter Havas, Eds. (Princeton University Press, Princeton), Vol. 2.
56. Maxwell, J.C. (1953) *A Treatise on Electricity and Magnetism* (Dover, New York), Vol. 1.

57. von Klitzing, K., G. Dorda, and M. Pepper (1980) *Physical Review Letters* 45, 494–497.
58. Prange, R.E., and S.M. Girvin (Second Edition, 1990) *The Quantum Hall Effect* (Springer-Verlag, New York).
59. Chakraborty, T., and P. Pietiläinen (1995) *The Quantum Hall Effects* (Springer-Verlag, Heidelberg).
60. Hoyle, F., and J.V. Narlikar (1996) *Cosmology and Action at a Distance Electrodynamics* (World Scientific, Singapore).
61. Lorentz, L. (1867) *Philosophical Magazine* 34, 287–301.
62. Ritz, W. (1908) *Physikalische Zeitschrift* 9, 903–907.
63. Einstein, A. (1909) *Physikalische Zeitschrift* 10, 185–193.
64. Ritz, W., and A. Einstein (1909) *Physikalische Zeitschrift* 10, 323–324.
65. Wheeler, J.A., and R.P. Feynman (1945) *Reviews of Modern Physics* 17, 157–181.
66. Wheeler, J.A., and K. Ford (1998) *Geons, Black Holes, and Quantum Foam* (W.W. Norton, New York), pp. 163–168.
67. Cramer, J.G. (1986) *Reviews of Modern Physics* 58, 647–687.
68. Einstein, A., B. Podolsky, B., and N. Rosen (1935) *Physical Review* 47, 777–780.
69. Sklar, L. (1995) *Time's Arrow Today*, S.F. Savitt, Ed. (Cambridge University Press, Cambridge), pp. 191–216.
70. Price, H. (1995) *Time's Arrow Today*, S.F. Savitt, Ed. (Cambridge University Press, Cambridge), pp. 66–94.
71. Callen, H., and T. Welton (1951) *Physical Review* 83, 34–45.
72. Stutzman, W.I., and G.A. Thiele (1981) *Antenna Theory and Design* (John Wiley, New York).
73. Zeh, H.D. (1993) *Physics Letters A* 172, 189–192.
74. Schrödinger, E. (1952) *British Journal for the Philosophy of Science* 3, 109–123 & 233–247.
75. Barut, A.O., and J.P. Dowling (1990) *Physical Review A* 41, 2284–2294.
76. Giulini, D., E. Joos, C. Kiefer, J. Kupsch, I.O. Stamatescu, and H.D. Zeh (1996) *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer-Verlag, Heidelberg).

77. Schrödinger, E. (1995) *The Interpretation of Quantum Mechanics* (Ox Bow Press, Woodbridge, CN), pp. 109.
78. Millikan, R.A. (1917) *The Electron* (University of Chicago Press, Chicago).
79. Josephson, B.D. (1962) *Physics Letters* 1, 251.
80. Josephson, B.D. (1965) *Advances in Physics* 14, 419.
81. Tinkham, M. (1996) *Introduction to Superconductivity* (McGraw-Hill, New York).
82. Lorentz, H.A. (1927) *Problems in Modern Physics*, H. Bateman, Ed. (Ginn, Boston).
83. Jaynes, E.T. (1958) *Some Aspects of Maser Theory. Microwave Laboratory Report No. 502* (Stanford University, Stanford).
84. Jaynes, E.T., and F.W. Cummings (1963) *Proceedings of the IEEE* Jan., 89–109.
85. Ziman, J. (1962) *Electrons in Metals* (Taylor & Francis, London).
86. Duck, I., and E.C.G. Sudarshan (1997) *Pauli and the Spin-Statistics Theorem* (World Scientific, Singapore).
87. Nozières, P., and D. Pines (1989) *The Theory of Quantum Liquids* (Addison-Wesley, Reading), Vol. II.
88. deGennes, P. (1966) *Superconductivity of Metals and Alloys* (Addison-Wesley, Reading).
89. London, F. (1950) *Superfluids* (John Wiley, New York), Vol. 1.
90. Dahl, P.F. (1992) *Superconductivity* (American Institute of Physics, New York).
91. Siegman, A. (1986) *Lasers* (University Science Books, Sausalito).
92. Rice, F.O., and E. Teller (1949) *The Structure of Matter* (John Wiley, New York).
93. Dirac, P.A.M. (1981) *Proceedings of the Conference on Pertubative Quantum Chromodynamics* (American Institute of Physics, New York), Vol. 74, pp. 129–130.

This page intentionally left blank

Index