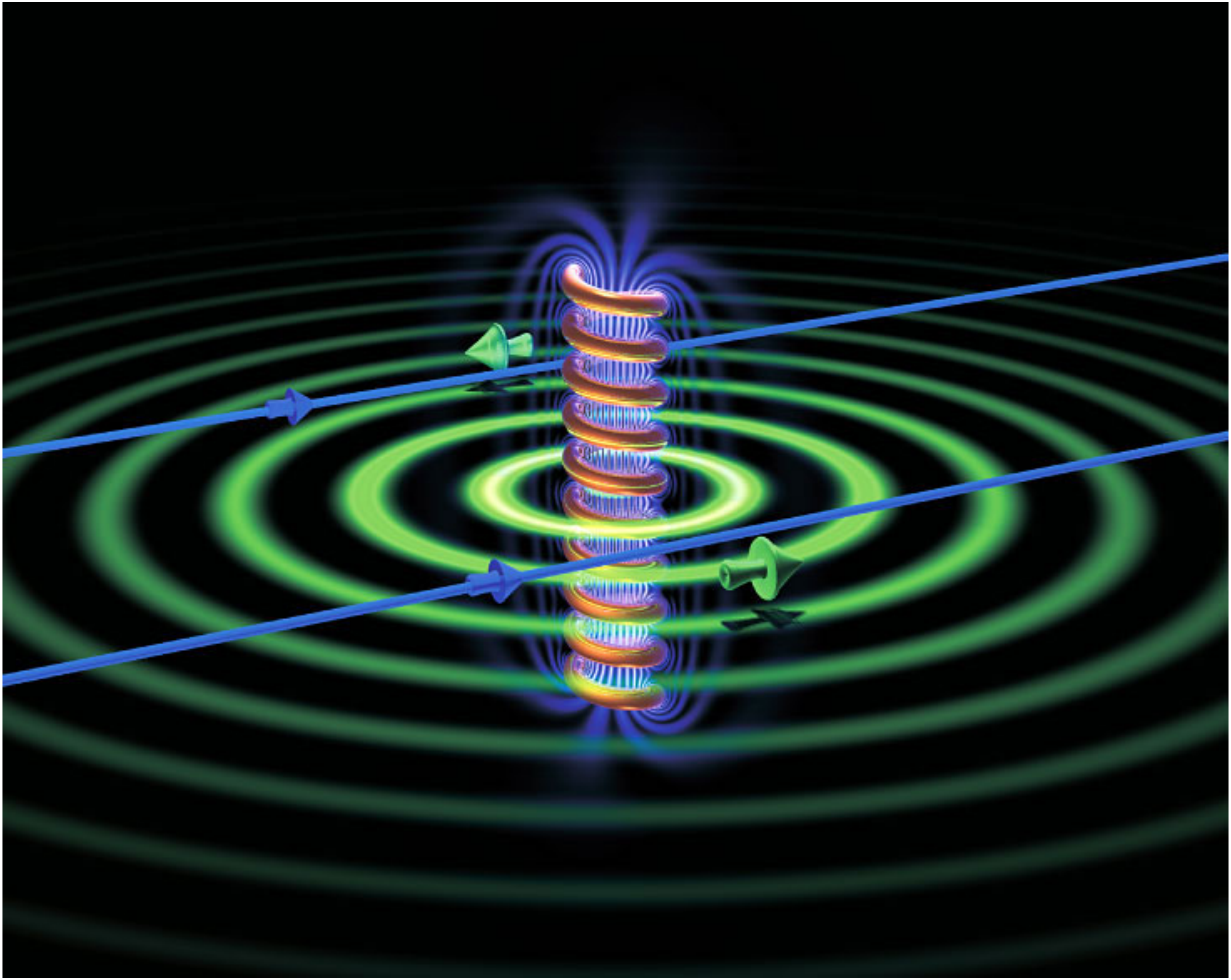


Physics 541 May **2, 2012**

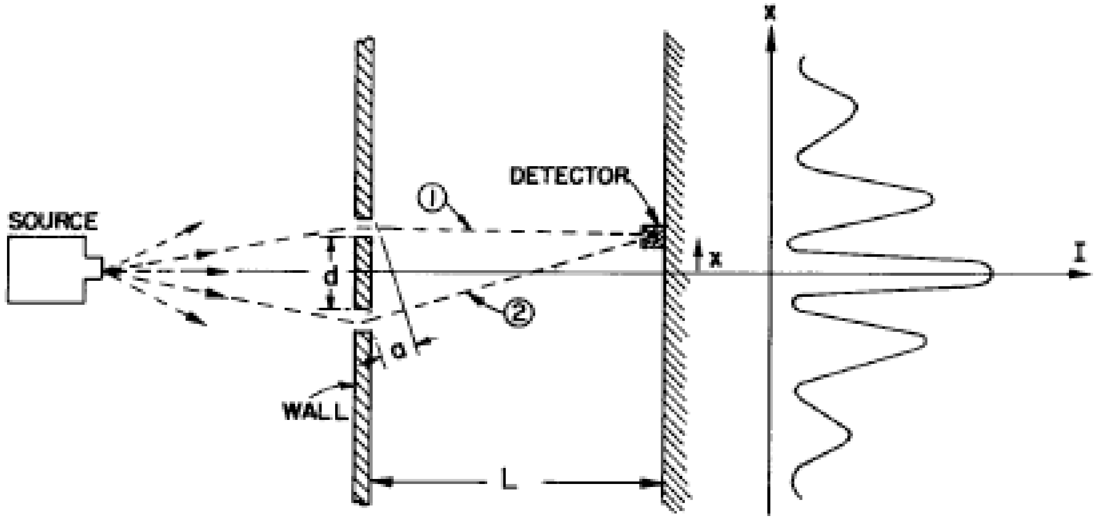
The Aharonov-Bohm Effect:

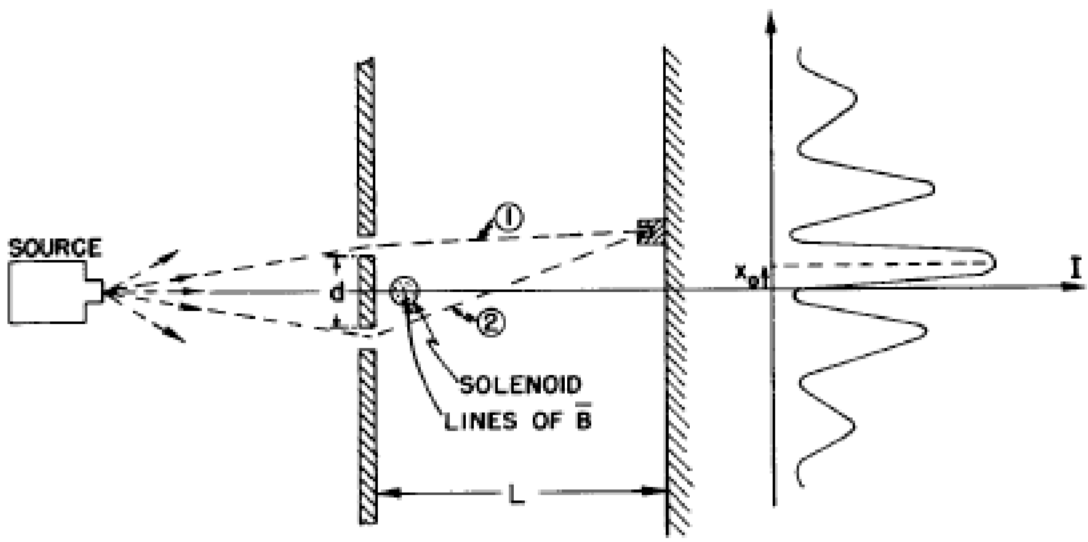
Topology
The Vector Potential
and
Gauge Transformations

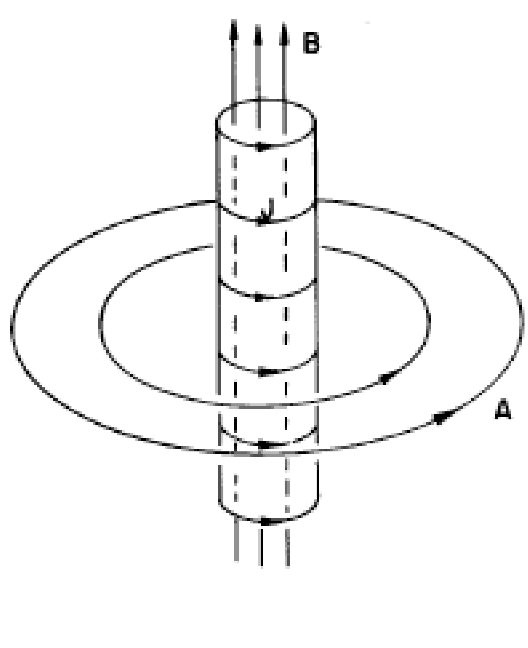
The Aharonov-Bohm Effect

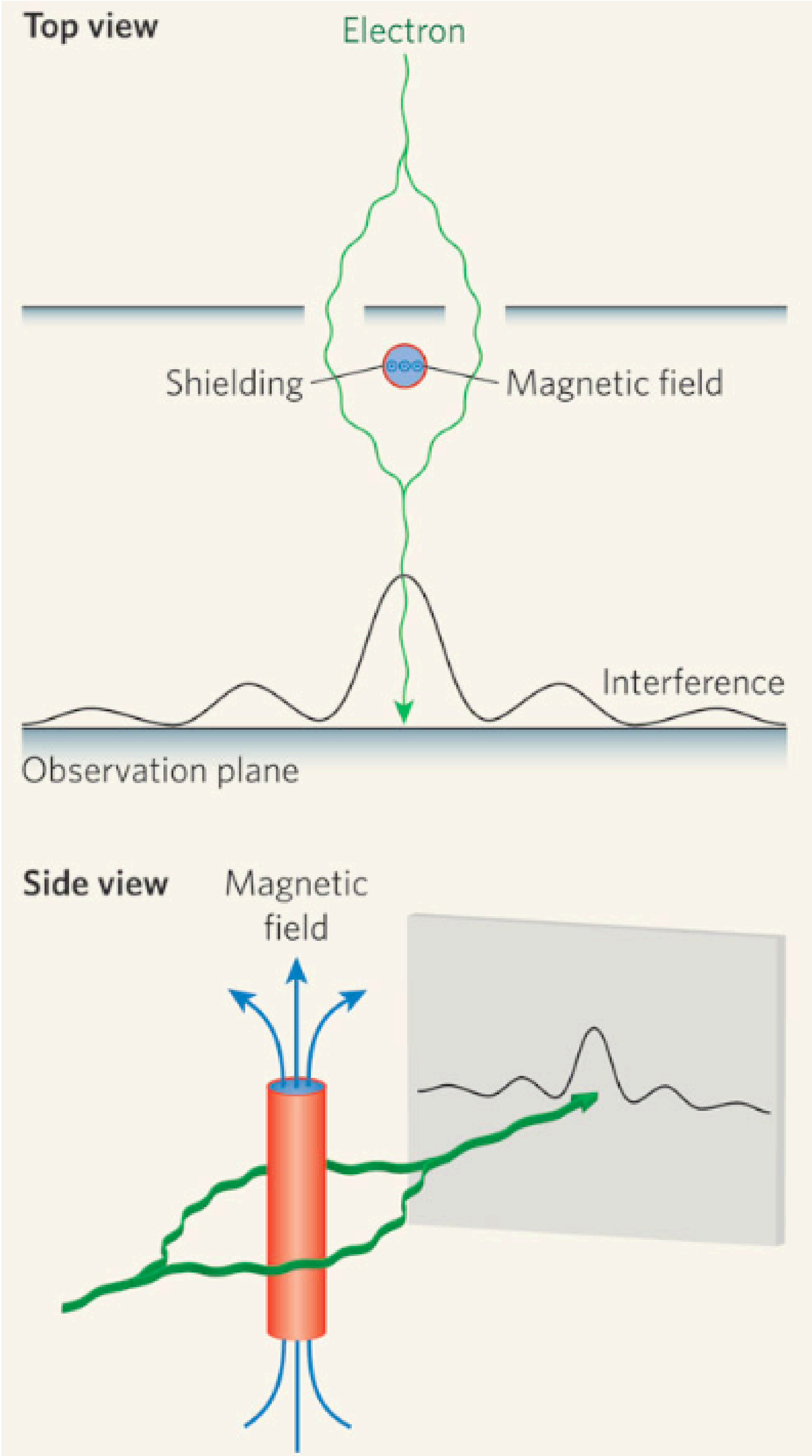
An electron moving in a region where E and B are zero, but A is not exhibits physical effects.

Therefore A is real whereas E and B are not.


x
DETECTOR
SOURCE
1
x
d
I
2
a
WALL
L


1
SOURCE
d
2
I
SOLENOID
LINES OF $\overline{B}$
L


B
A



Top view
Electron
Shielding
Magnetic field
Interference
Observation plane
Side view
Magnetic
field

Experiments:

http://www.youtube.com/watch?v=sT6OyzJ8Oqw Single electron

http://www.youtube.com/watch?v=ZJ-0PBRuthc Akira Tonomura
http://www.hitachi.com/rd/research/em/doubleslit-f1.html
http://www.hitachi.com/rd/research/em/doubleslit.html
http://www.hitachi.com/rd/research/em/abe.html

Simulations:

http://www.youtube.com/watch?v=OgDPK5MLVnE Wolfram AB Demo Video

http://www.physics.brocku.ca/faculty/Sternin/teaching/mirrors/qm/abe/index.html Brock University

http://rugth30.phys.rug.nl/quantummechanics/ab.htm University of Groningen

People

http://www.youtube.com/watch?NR=1&feature=endscreen&v=SvyD2o7w24g David Bohm

http://www.youtube.com/watch?v=YJGOhl8iK3o Yakir Aharonov
http://www.youtube.com/watch?v=6qxRq3Nncpw

Conference:

http://www.youtube.com/results?search_query=%2250+years+of+the+Aharonov-Bohm+Effect%22&page=1

http://www.youtube.com/watch?v=YJGOhl8iK3o Yakir Aharonov

http://www.youtube.com/watch?v=WnsrDFSjcZ0&feature=relmfu CN Yang

This formula corresponds to the result we found for the electrostatic energy:

$$U = \tfrac{1}{2}\int \rho\phi\, dV. \tag{15.21}$$

So we may if we wish think of A as a kind of potential energy for currents in magnetostatics. Unfortunately, this idea is not too useful, because it is true only for static fields. In fact, neither of the equations (15.20) and (15.21) gives the correct energy when the fields change with time.

15–4 *B* versus *A*

In this section we would like to discuss the following questions: Is the vector potential merely a device which is useful in making calculations—as the scalar potential is useful in electrostatics—or is the vector potential a "real" field? Isn't the magnetic field the "real" field, because it is responsible for the force on a moving particle? First we should say that the phrase "a real field" is not very meaningful. For one thing, you probably don't feel that the magnetic field is very "real" anyway, because even the whole idea of a field is a rather abstract thing. You cannot put out your hand and feel the magnetic field. Furthermore, the value of the magnetic field is not very definite; by choosing a suitable moving coordinate system, for instance, you can make a magnetic field at a given point disappear.

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. If we have a charged particle at the position P, it is affected by other charges located at some distance from P. One way to describe the interaction is to say that the other charges make some "condition"—whatever it may be—in the environment at P. If we know that condition, which we describe by giving the electric and magnetic fields, then we can determine completely the behavior of the particle—with no further reference to how those conditions came about.

In other words, if those other charges were altered in some way, but the conditions at P that are described by the electric and magnetic field at P remain the same, then the motion of the charge will also be the same. 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*. We do not need to know any more about what's going on at other places. It is in this sense that we will discuss whether the vector potential is a "real" field.

You may be wondering about the fact that the vector potential is not unique—that it can be changed by adding the gradient of any scalar with no change at all in the forces on particles. That has not, however, anything to do with the question of reality in the sense that we are talking about. For instance, the magnetic field is in a sense altered by a relativity change (as are also E and A). But we are not worried about what happens if the field *can* be changed in this way. That doesn't really make any difference; that has nothing to do with the question of whether the vector potential is a proper "real" field for describing magnetic effects, or whether it is just a useful mathematical tool.

We should also make some remarks on the usefulness of the vector potential A. We have seen that it can be used in a formal procedure for calculating the magnetic fields of known currents, just as ϕ can be used to find electric fields. In electrostatics we saw that ϕ was given by the scalar integral

$$\phi(1) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(2)}{r_{12}}\, dV_2. \tag{15.22}$$

From this ϕ, we get the three components of E by three differential operations. This procedure is usually easier to handle than evaluating the three integrals in the vector formula

$$E(1) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(2)e_{12}}{r_{12}^2}\, dV_2. \tag{15.23}$$

First, there are three integrals; and second, each integral is in general somewhat more difficult.

The advantages are much less clear for magnetostatics. The integral for $\boldsymbol{A}$ is already a vector integral:

$$\boldsymbol{A}(1) = \frac{1}{4\pi\epsilon_0 c^2}\int \frac{\boldsymbol{j}(2)\,dV_2}{r_{12}}, \tag{15.24}$$

which is, of course, three integrals. Also, when we take the curl of $\boldsymbol{A}$ to get $\boldsymbol{B}$, we have six derivatives to do and combine by pairs. It is not immediately obvious whether in most problems this procedure is really any easier than computing $\boldsymbol{B}$ directly from

$$\boldsymbol{B}(1) = \frac{1}{4\pi\epsilon_0 c^2}\int \frac{\boldsymbol{j}(2) \times \boldsymbol{e}_{12}}{r_{12}^2}\,dV_2. \tag{15.25}$$

Using the vector potential is often more difficult for simple problems for the following reason. Suppose we are interested only in the magnetic field $\boldsymbol{B}$ at one point, and that the problem has some nice symmetry—say we want the field at a point on the axis of a ring of current. Because of the symmetry, we can easily get $\boldsymbol{B}$ by doing the integral of Eq. (15.25). If, however, we were to find $\boldsymbol{A}$ first, we would have to compute $\boldsymbol{B}$ from *derivatives* of $\boldsymbol{A}$, so we must know what $\boldsymbol{A}$ is at all points in the *neighborhood* of the point of interest. And most of these points are off the axis of symmetry, so the integral for $\boldsymbol{A}$ gets complicated. In the ring problem, for example, we would need to use elliptic integrals. In such problems, $\boldsymbol{A}$ is clearly not very useful. It is true that in many complex problems it is easier to work with $\boldsymbol{A}$, but it would be hard to argue that this ease of technique would justify making you learn about one more vector field.

We have introduced $\boldsymbol{A}$ because it *does* have an important physical significance. Not only is it related to the energies of currents, as we saw in the last section, but it is also a "real" physical field in the sense that we described above. In classical mechanics it is clear that we can write the force on a particle as

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}), \tag{15.26}$$

so that, given the forces, everything about the motion is determined. In any region where $\boldsymbol{B} = 0$ even if $\boldsymbol{A}$ is not zero, such as outside a solenoid, there is no discernible effect of $\boldsymbol{A}$. Therefore for a long time it was believed that $\boldsymbol{A}$ was not a "real" field. It turns out, however, that there are phenomena involving quantum mechanics which show that the field $\boldsymbol{A}$ is in fact a "real" field in the sense we have defined it. In the next section we will show you how that works.

15–5 The vector potential and quantum mechanics

There are many changes in what concepts are important when we go from classical to quantum mechanics. We have already discussed some of them in Vol. I. In particular, the force concept gradually fades away, while the concepts of energy and momentum become of paramount importance. You remember that instead of particle motions, one deals with probability amplitudes which vary in space and time. In these amplitudes there are wavelengths related to momenta, and frequencies related to energies. The momenta and energies, which determine the phases of wave functions, are therefore the important quantities in quantum mechanics. Instead of forces, we deal with the way interactions change the wavelength of the waves. The idea of a force becomes quite secondary—if it is there at all. When people talk about nuclear forces, for example, what they usually analyze and work with are the energies of interaction of two nucleons, and not the force between them. Nobody ever differentiates the energy to find out what the force looks like. In this section we want to describe how the vector and scalar potentials enter into quantum mechanics. It is, in fact, just because momentum and energy play a central role in quantum mechanics that $\boldsymbol{A}$ and ϕ provide the most direct way of introducing electromagnetic effects into quantum descriptions.

We must review a little how quantum mechanics works. We will consider again the imaginary experiment described in Chapter 37 of Vol. I, in which elec-

Feynman's Paradox

A paradox is a situation which gives one answer when analyzed one way, and a different answer when analyzed another way, so that we are left in somewhat of a quandary as to actually what would happen. Of course, in physics there are never any real paradoxes because there is one correct answer; at least we believe that nature will act in only one way (and that is the *right way*, naturally). So a paradox in physics is only a confusion in our understanding.

ably chosen $\boldsymbol{B}$, can cause the electron to keep moving on its assumed orbit. In the betatron this transverse force causes the electron to move in a circular orbit of constant radius. We can find out what the magnetic field at the orbit must be by using again the relativistic equation of motion, but this time, for the transverse component of the force. In the betatron (see Fig 17–4), $\boldsymbol{B}$ is at right angles to $\boldsymbol{v}$, so the transverse force is qvB. Thus the force is equal to the rate of change of the transverse component p_t of the momentum:

$$qvB = \frac{dp_t}{dt}. \tag{17.8}$$

When a particle is moving in a *circle*, the rate of change of its transverse momentum is equal to the magnitude of the total momentum times ω, the angular velocity of rotation (following the arguments of Chapter 11, Vol. I):

$$\frac{dp_t}{dt} = \omega p, \tag{17.9}$$

where, since the motion is circular,

$$\omega = \frac{v}{r}. \tag{17.10}$$

Setting the magnetic force equal to the transverse acceleration, we have

$$qvB_{\text{orbit}} = p\,\frac{v}{r}, \tag{17.11}$$

where B_{orbit} is the field at the radius r.

As the betatron operates, the momentum of the electron grows in proportion to B_{av}, according to Eq. (17.7), and if the electron is to continue to move in its proper circle, Eq. (17.11) must continue to hold as the momentum of the electron increases. The value of B_{orbit} must increase in proportion to the momentum p. Comparing Eq. (17.11) with Eq. (17.7), which determines p, we see that the following relation must hold between B_{av}, the average magnetic field *inside* the orbit at the radius r, and the magnetic field B_{orbit} at the orbit:

$$\Delta B_{\text{av}} = 2\,\Delta B_{\text{orbit}}. \tag{17.12}$$

The correct operation of a betatron requires that the average magnetic field inside the orbit increase at twice the rate of the magnetic field at the orbit itself. In these circumstances, as the energy of the particle is increased by the induced electric field the magnetic field at the orbit increases at just the rate required to keep the particle moving in a circle.

The betatron is used to accelerate electrons to energies of tens of millions of volts, or even to hundreds of millions of volts. However, it becomes impractical for the acceleration of electrons to energies much higher than a few hundred million volts for several reasons. One of them is the practical difficulty of attaining the required high average value for the magnetic field inside the orbit. Another is that Eq. (17.6) is no longer correct at very high energies because it does not include the loss of energy from the particle due to its radiation of electromagnetic energy (the so-called synchrotron radiation discussed in Chapter 36, Vol. I). For these reasons, the acceleration of electrons to the highest energies—to many billions of electron volts—is accomplished by means of a different kind of machine, called a *synchrotron*.

17–4 A paradox

We would now like to describe for you an apparent paradox. A paradox is a situation which gives one answer when analyzed one way, and a different answer when analyzed another way, so that we are left in somewhat of a quandary as to actually what should happen. Of course, in physics there are never any real paradoxes because there is only one correct answer; at least we believe that nature will act in only one way (and that is the *right way*, naturally). So in physics a paradox is only a confusion in our own understanding. Here is our paradox.

Imagine that we construct a device like that shown in Fig. 17–5. There is a thin, circular plastic disc supported on a concentric shaft with excellent bearings, so that it is quite free to rotate. On the disc is a coil of wire in the form of a short solenoid concentric with the axis of rotation. This solenoid carries a steady current I provided by a small battery, also mounted on the disc. Near the edge of the disc and spaced uniformly around its circumference are a number of small metal spheres insulated from each other and from the solenoid by the plastic material of the disc. Each of these small conducting spheres is charged with the same electrostatic charge Q. Everything is quite stationary, and the disc is at rest. Suppose now that by some accident—or by prearrangement—the current in the solenoid is interrupted, without, however, any intervention from the outside. So long as the current continued, there was a magnetic flux through the solenoid more or less parallel to the axis of the disc. When the current is interrupted, this flux must go to zero. There will, therefore, be an electric field induced which will circulate around in circles centered at the axis. The charged spheres on the perimeter of the disc will all experience an electric field tangential to the perimeter of the disc. This electric force is in the same sense for all the charges and so will result in a net torque on the disc. From these arguments we would expect that as the current in the solenoid disappears, the disc would begin to rotate. If we knew the moment of inertia of the disc, the current in the solenoid, and the charges on the small spheres, we could compute the resulting angular velocity.

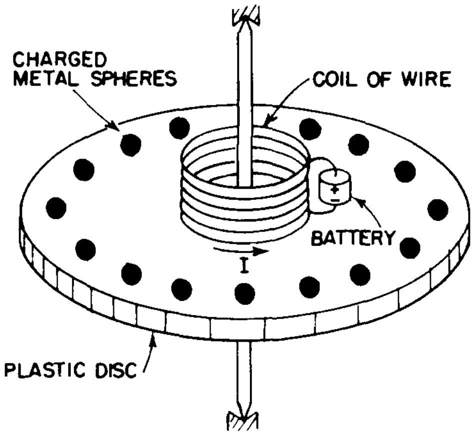


Fig. 17–5. Will the disc rotate if the current I is stopped?

But we could also make a different argument. Using the principle of the conservation of angular momentum, we could say that the angular momentum of the disc with all its equipment is initially zero, and so the angular momentum of the assembly should remain zero. There should be no rotation when the current is stopped. Which argument is correct? Will the disc rotate or will it not? We will leave this question for you to think about.

We should warn you that the correct answer does not depend on any nonessential feature, such as the asymmetric position of a battery, for example. In fact, you can imagine an ideal situation such as the following: The solenoid is made of superconducting wire through which there is a current. After the disc has been carefully placed at rest, the temperature of the solenoid is allowed to rise slowly. When the temperature of the wire reaches the transition temperature between superconductivity and normal conductivity, the current in the solenoid will be brought to zero by the resistance of the wire. The flux will, as before, fall to zero, and there will be an electric field around the axis. We should also warn you that the solution is not easy, nor is it a trick. When you figure it out, you will have discovered an important principle of electromagnetism.

17–5 Alternating-current generator

In the remainder of this chapter we apply the principles of Section 17–1 to analyze a number of the phenomena discussed in Chapter 16. We first look in more detail at the alternating-current generator. Such a generator consists basically of a coil of wire rotating in a uniform magnetic field. The same result can also be achieved by a fixed coil in a magnetic field whose direction rotates in the manner described in the last chapter. We will consider only the former case. Suppose we have a circular coil of wire which can be turned on an axis along one of its diameters. Let this coil be located in a uniform magnetic field perpendicular to the axis of rotation, as in Fig. 17–6. We also imagine that the two ends of the coil are brought to external connections through some kind of sliding contacts.

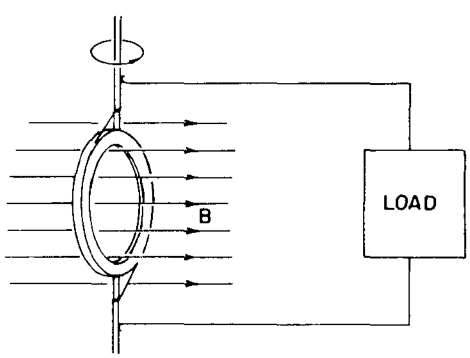


Fig. 17–6. A coil of wire rotating in a uniform magnetic field—the basic idea of the ac generator.

Due to the rotation of the coil, the magnetic flux through it will be changing. The circuit of the coil will therefore have an emf in it. Let S be the area of the coil and θ the angle between the magnetic field and the normal to the plane of the coil.*

* Now that we are using the letter A for the vector potential, we prefer to let S stand for a Surface area.

The Vector Potential

EM Field Carries Momentum

Vector potential A => Photons

V. SUMMARY AND CONCLUSION

The spherical orrery is a useful device for demonstrating and investigating principles of celestial mechanics. The physics of this device is more closely analogous to celestial mechanics than that of an earlier cylindrical orrery in which particles orbit a rod. Phenomena that are easily investigated include Kepler's laws, precession, adiabaticity, molecular drag, and collisions. The use of videotape allows the phenomena to be shown to large audiences. An improved vacuum will allow longer orbital decay times which will facilitate the investigation of resonant perturbations applied for long periods. This type of perturbation is responsible for the Kirkwood gaps in the asteroid belt, much of the fine structure in Saturn's rings, and may also lead to dynamical chaos.

ACKNOWLEDGMENTS

CS acknowledges support from the National Science Foundation Research Experiences for Undergraduates Program. DA and SR acknowledge support from National Aeronautics and Space Administration (NASA). KV and BW acknowledge support from the Research Corporation Cottrell College Science Award Program and the NASA Joint Venture Program. The authors acknowledge valuable discussions on orbital mechanics with Mihály Horányi and initial experimental work by Greg Newton.

[1]T. Biewer, D. Alexander, S. Robertson, and B. Walch, "Electrostatic orrery for celestial mechanics," Am. J. Phys. **62**, 821–827 (1994).

[2]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge U.P., Cambridge, 1986), p. 553.

[3]H. Winter and H. W. Ortjohann, "Simple demonstration of storing macroscopic particles in a Paul trap," Am. J. Phys. **59**, 807–812 (1991).

[4]For a recent review of particle traps, see D. A. Church, "Collision measurements and excited-level lifetime measurements on ions stored in Paul, Penning, and Kingdon ion traps," Phys. Rep. **228**, 253–358 (1993).

[5]J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton, U.P., Princeton, 1987), Chaps. 2 and 3.

[6]J. A. Burns, "Elementary derivation of the perturbation equations of celestial mechanics," Am. J. Phys. **44**, 944–949 (1976); errata, **45**, 545 (1976).

[7]M. Horányi, J. A. Burns, M. Tatrallyay, and J. G. Luhmann, "Toward understanding the fate of dust lost from the Martian satellites," Geophys. Res. Lett. **17**(6), 853–856 (1990).

[8]The uncertainty in the charge-to-mass ratio is determined by the measured spread in charge (±18%), the measured spread in particle diameter (±9%), and the spread in wall thickness which is not known.

[9]B. Walch, M. Horányi, and S. Robertson, "Charging of dust grains in plasma with energetic electrons," Phys. Rev. Lett. **75**, 838–841 (1995).

[10]Model EDC1000, Electrim Corporation, Box 2074, Princeton, NJ 08543.

[11]Model FX-620, Sony Corporation. It is important for the camera to have a manual focusing ability.

Thoughts on the magnetic vector potential

Mark D. Semon
Physics Department, Bates College, Lewiston, Maine 04240

John R. Taylor
Department of Physics, Box 390, University of Colorado, Boulder, Colorado 80309

(Received 13 May 1996; accepted 25 July 1996)

We collect together several ideas that we have found helpful in teaching the magnetic vector potential **A**. We argue that students can be taught to visualize **A** for simple current distributions and to see **A** as something with physical significance beyond its bare definition as the "thing whose curl is **B**." © *1996 American Association of Physics Teachers.*

I. INTRODUCTION

Despite the beautiful symmetry between electric and magnetic fields, the ways in which we teach these two concepts could scarcely be more different. In our introductory physics courses ("freshman physics" in a typical American college), students acquire a reasonable understanding of the electric field **E** and the electrostatic potential ϕ; by contrast, their understanding of the magnetic field **B** is hazy at best, and they probably do not meet the magnetic potential **A** at all. By the end of their next course in electromagnetism ("junior E and M"), students generally have a reasonable understanding of the magnetic field **B** and have met the magnetic potential as a mathematical artifact used to express **B** as $\mathbf{B}=\nabla\times\mathbf{A}$. Nevertheless, they still have almost no idea of what **A** really is, much less any picture of what **A** looks like in even the simplest situations. Even after a graduate course in electrodynamics, many students probably could not say much more about **A** than that it is the vector whose curl is **B**.

In this paper, we focus on the vector potential **A** and argue that there is much that can be said to improve students' understanding of it. Many of the ideas we discuss have appeared before (often in this journal), and some are hinted at in some of the popular textbooks. Nevertheless, it seems clear from the textbooks and from our discussions with numerous colleagues that these ideas are not widely recognized and are certainly not incorporated into most courses in electromagnetism. Given the increasing importance of the vector potential in modern physics (superconductivity, the

 © 1996 American Association of Physics Teachers

Aharonov–Bohm effect, Josephson junctions, SQUIDS, etc.) anything we can say to help our students master the concept seems worth emphasizing.

Perhaps the most obvious difficulty in teaching the vector potential is that it requires a knowledge of vector calculus. One can define the scalar potential ϕ as the potential energy per unit charge and give a remarkably good feeling for ϕ without ever using vector calculus. Typical introductory courses convey a good sense of equipotential surfaces for simple charge distributions and of the rate of change of ϕ as $-\mathbf{E}$ without ever mentioning the equation $\mathbf{E}=-\nabla\phi$. By contrast, it is very hard to teach the vector potential until our students understand the meaning of the curl (as in $\mathbf{B}=\nabla\times\mathbf{A}$). For this reason alone, the vector potential is beyond the reach of almost all freshman physics courses. Obviously, we cannot deny this problem, but we do believe that there is much we can do to improve students' understanding of **A** once they do meet it in junior E and M or in a graduate course.

A second obstacle to our students' understanding of the vector potential is the still prevalent view that **A** is merely a mathematical fiction whose only role is to express **B** as $\nabla\times\mathbf{A}$. Curiously, the founder of our subject, Maxwell himself,[1,2] advocated in 1865 a quite opposite view, which we shall echo, that the vector potential can be seen as a stored momentum per unit charge in much the same way that ϕ is the stored energy per unit charge. Indeed, one of Maxwell's several names for the vector potential was "electromagnetic momentum." An equivalent view, that **A** can be seen as the appropriate field momentum per unit charge, was stated by Thomson[3] in 1904 and was forcefully advocated by Konopinski[4] in an article in this journal in 1978. (There is a long history of distinguished articles on related topics in this journal.[5])

The modern view, that **A** is an artifact devoid of physical significance seems to originate with Heaviside (in 1886),[6] who described the potentials as "highly artificial quantities," and Hertz (in 1893),[7] who disparaged the components of **A** as "magnitudes which serve for calculation only." These views can be found in almost any modern textbook on electromagnetism. Perhaps the strongest statement is due to Rohrlich,[8] who says:

> These functions, known as potentials, have no physical meaning and are introduced solely for the purpose of mathematical simplification of the equations.

Similar statements can be found, for example, in Refs. 9 and 10.

We do not claim that the Maxwell–Thomson view of **A** as stored momentum per unit charge is of immense practical value (although we do offer some examples to show the insight that it can contribute). Nonetheless, we do argue that, by giving physical meaning to an otherwise rather abstract notion, this view can help students to feel more at home with and better understand the undeniably important concept of the vector potential.

A third difficulty in teaching the vector potential is that much of its importance appears only later in more advanced subjects which the students of junior E and M have often not studied: In relativity, **A** combines with ϕ to form the four-potential $A=(\mathbf{A},\phi/c)$, just as the momentum **p** combines with the energy E to form the four-momentum $p=(\mathbf{p},E/c)$. In the Lagrangian mechanics of a charged particle, the generalized, or canonical, momentum turns out to be $\mathbf{p}=m\mathbf{v}+q\mathbf{A}$, and, under the appropriate conditions, it is **p** (rather than $m\mathbf{v}$) which is conserved. That is, $q\mathbf{A}$ is the quantity that must be added to $m\mathbf{v}$ to give the "proper" conserved momentum, just as $q\phi$ is the quantity that must be added to $\frac{1}{2}mv^2$ to give the "proper" conserved energy. In quantum theory, **A** (as opposed to **B**) is the fundamental quantity in the Schrödinger equation for a charged particle and in the interactions of quantum electrodynamics.

If your students haven't studied these more advanced subjects, then these arguments for the importance of **A** will carry less weight. Nonetheless, some students have studied relativity, Lagrangian mechanics, or even quantum mechanics before meeting the vector potential, and for such students these ideas are well worth exploring. Even if your students have *not* studied these subjects, the arguments can at least be mentioned.

In Sec. II, we describe two ways to help our students visualize, and even calculate, the vector potential for a number of steady current distributions. While most students can readily calculate and visualize the scalar potential ϕ of various charge distributions, very few can do the same for the vector potential of any current distributions. Any tricks to help them do this seem well worth emphasizing. In particular, the formal analogy between $\mu_0\mathbf{J}$ as the source of **B** (as in $\nabla\times\mathbf{B}=\mu_0\mathbf{J}$) and **B** as the source of **A** (as in $\nabla\times\mathbf{A}=\mathbf{B}$) allows one to find the vector potential in several situations by taking advantage of the well-known B fields of certain current distributions. Although this point is mentioned briefly in the fine textbooks[11] of Griffiths and of Barger and Olsson, and is clearly stated in a recent article of Carron,[12] it seems not to be as widely appreciated as it deserves.

In Sec. III, we review the main arguments for the Maxwell–Thomson view that **A** is the stored momentum per unit charge, that is, that **A** does for momentum what the scalar potential ϕ does for energy. Finally, in Sec. IV, we give some examples of problems that can be solved and perhaps better understood using this way of viewing **A**.

To conclude this introduction, we need to discuss the consistency of the view of **A** as stored momentum with the requirements of gauge invariance. Since **A** is not uniquely defined, one is bound to be a bit suspicious of the claim that **A** can be interpreted as stored momentum. We shall address this objection at the appropriate points throughout the paper, but it may be helpful to summarize the situation now: First, we note that many familiar physical quantities are not uniquely defined (potential energy, the energy flow vector, the Lagrangian, etc.) but are nevertheless physically significant. In the case of the vector potential, there are many different choices of **A**, all corresponding to the same electromagnetic fields, and we shall see that each different choice of **A** (that is, each different gauge) defines a different generalized momentum $\mathbf{p}=m\mathbf{v}+q\mathbf{A}$. In some gauges the generalized momentum may be conserved and in others it may not. As one might expect, the most convenient choice of gauge is usually one in which **p** *is* conserved. Since conservation laws are generally associated with symmetries, this means finding a gauge where **A** has the same symmetries as the underlying problem.

II. ON VISUALIZING AND CALCULATING THE VECTOR POTENTIAL

By the time they are in a junior E and M course, most of our students have a reasonable picture of the way the magnetic field **B** circles around the current that produces it, and, with the help of Ampere's law, they can calculate the B field

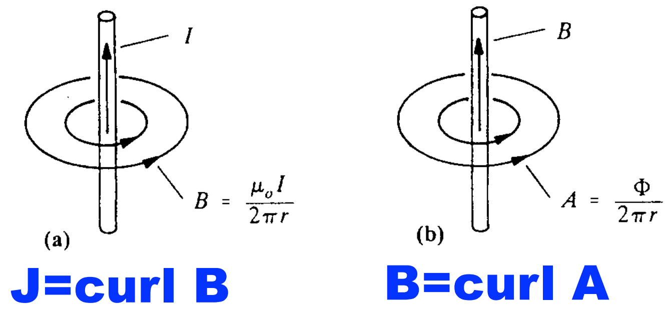


Fig. 1. (a) The B field of a current confined inside a long straight wire circulates around the wire with magnitude given by Ampere's law as $B=\mu_0 I/2\pi r$. (b) The B field of a long solenoid is confined inside the solenoid. Comparing with (a), we can immediately conclude that the vector potential outside the solenoid circulates with magnitude $A=\Phi/2\pi r$.

for several simple current distributions. On the other hand, they usually have very little idea how to visualize or calculate the vector potential for any current distributions.

One approach to finding any magnetostatic **A**, which leans nicely on the students' experience in electrostatics, is to note that, in the gauge where $\nabla\cdot\mathbf{A}=0$, the vector potential **A** satisfies

$$\nabla^2\mathbf{A}=-\mu_0\mathbf{J} \tag{1}$$

(just as the scalar potential satisfies $\nabla^2\phi=-\rho/\epsilon_0$). The solution of this equation is well known to be

$$\mathbf{A}(\mathbf{r})=\frac{\mu_0}{4\pi}\int dV'\,\frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{2}$$

(Again, it helps to emphasize the parallel between this and the corresponding result for ϕ.) We mention the well-known result (2) because it makes clear that the contribution of each $\mathbf{J}(\mathbf{r}')$ to $\mathbf{A}(\mathbf{r})$ is in the direction of $\mathbf{J}(\mathbf{r}')$. For example, if **J** has the same direction everywhere (as with the current in a long straight wire), then the same is true of **A**, and **A** has the same direction as **J**. Similarly, Eq. (2) implies that if **J** is axially symmetric and points in circles around its axis of symmetry (as with the current in a circular loop or solenoid), then **A** has these same two properties.

A second way to find **A** is to recognize that (in the gauge with $\nabla\cdot\mathbf{A}=0$) **A** is determined by the two equations

$$\nabla\times\mathbf{A}=\mathbf{B} \quad \text{and} \quad \nabla\cdot\mathbf{A}=0. \tag{3}$$

Comparing these with the two Maxwell equations for **B**,

$$\nabla\times\mathbf{B}=\mu_0\mathbf{J} \quad \text{and} \quad \nabla\cdot\mathbf{B}=0, \tag{4}$$

we see that **B** can be regarded as the "source" of **A** in just the same way that $\mu_0\mathbf{J}$ is the source of **B**. This analogy means that all the students' hard-won experience using Ampere's law to find **B** for given **J** can be applied immediately to the problem of finding **A** for given **B**. For example, most students in junior E and M are familiar with the way **B** tends to circle around its source current **J**; in just the same way, it follows that the vector potential **A** tends to circle around its corresponding **B**.

A. Example: Vector potential for a solenoid

As a first illustration of this approach to finding **A**, recall that, for a steady current in a long straight cylindrical wire, the B field circulates outside the wire, as shown in Fig. 1(a). Using the analogy between Eqs. (3) and (4), we can immediately find the potential **A** corresponding to the B field of a long cylindrical solenoid, as shown in Fig. 1(b). This field is uniform inside the cylinder and zero outside. Therefore, exactly as **B** circulates around the current **I** in Fig. 1(a), so the vector potential **A** must circulate around **B** in Fig. 1(b).

Quantitatively, the students all know from Ampere's law that the B field outside the current of Fig. 1(a) is

$$B=\frac{\mu_0 I}{2\pi r} \quad \text{[outside wire]}, \tag{5}$$

where I is the total current in the wire. It immediately follows that the vector potential outside the solenoid of Fig. 1(b) must be

$$A=\frac{\Phi}{2\pi r} \quad \text{[outside solenoid]}, \tag{6}$$

where Φ is the total flux of B inside the solenoid. This is surely a most economical and transparent derivation of the vector potential outside a solenoid—a configuration that occurs in the Aharonov–Bohm effect and many other important applications.

We can use the same argument to find the vector potential *inside* the solenoid. If the current in Fig. 1(a) is uniform inside the wire, then we can use Ampere's law, $\oint\mathbf{B}\cdot d\mathbf{r}=\mu_0 I(\text{inside})$, with a circular path of radius r, to show that the B field inside the wire is $B=\mu_0 Ir/2\pi a^2$, where a is the radius of the wire. In exactly the same way, Eq. (3) implies an "Ampere's law" for **A**, namely, $\oint\mathbf{A}\cdot d\mathbf{r}=\Phi(\text{inside})$, and we can see that the vector potential inside the solenoid of Fig. 1(a) circulates around the axis with magnitude

$$A=\frac{\Phi r}{2\pi a^2}=\frac{Br}{2}. \tag{7}$$

If we bear in mind that r in (7) denotes the perpendicular distance out from the axis and we let $a\to\infty$, this result gives the vector potential for a uniform B field:

$$\mathbf{A}=\frac{1}{2}\,\mathbf{B}\times\mathbf{r}, \tag{8}$$

a result of great importance, which appears, for example, in the quantum theory of an atom in a uniform magnetic field (the Zeeman and Paschen–Back effects).

B. Another example: Vector potential for a long straight wire

Figure 2(a) shows a current circulating uniformly around the surface of a long conducting cylinder, that is, a solenoid. The corresponding B field is known (from Ampere's law) to be zero outside the cylinder and uniform, directed along the cylinder, on the inside, as shown in Fig. 2(a). It immediately follows that a B field circulating uniformly around the surface of a cylinder corresponds to a vector potential **A** that is zero outside and uniform inside, as shown in Fig. 2(b). The circulating B field of Fig. 2(b) is produced by a uniform current flowing up a cylinder of radius r and back down a coaxial cylinder of slightly larger radius $r+dr$, as shown in Fig. 2(c). Thus the vector potential of Fig. 2(b) is the potential of a uniform current in the coaxial cable of Fig. 2(c).

It is easy to write down quantitative expressions for the fields involved in this example. The B field inside the sole-

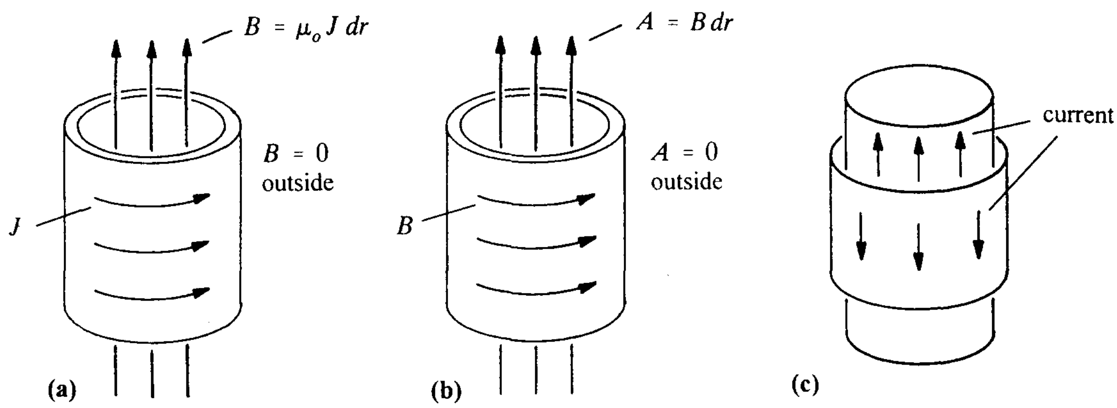


Fig. 2. (a) A current circulating around a long cylinder (or solenoid) produces a B field that is uniform inside the cylinder and zero outside. (b) It immediately follows that a B field circulating around the surface of a cylinder corresponds to a potential **A** that is uniform inside the cylinder and zero outside. (c) The B field of (b) is produced by a current flowing up one cylinder and down a second, coaxial cylinder of slightly larger radius.

noid of Fig. 2(a) is well known to be $\mu_0 nI$, where n is the number of turns per unit length. In terms of the current density J this is

$$B=\mu_0 J\ dr \quad \text{[inside cylinder]}, \tag{9}$$

where dr denotes the thickness of the conducting cylinder. Therefore, the vector potential of Fig. 2(b) is

$$A=B\ dr \tag{10}$$

$$=\frac{\mu_0 I}{2\pi r}\,dr \quad \text{[inside cylinder]}, \tag{11}$$

and $A=0$ outside. Here, dr is the small separation between the two coaxial conductors, and Eq. (11) follows from (10) because the field between the two cylinders is, according to Ampere's law, $B=\mu_0 I/2\pi r$.

From the result (11) we can easily find the vector potential for a single cylindrical wire. Consider, first, two coaxial cylinders at $r=a$ and $r=b$, not necessarily close together. We can regard this arrangement as the superposition of many coaxial pairs, starting at $r=a$ and ending at $r=b$, in which each pair is close together. To find A at any r between a and b, we note that those coaxial pairs inside r contribute nothing to A. Thus the total vector potential at any r between the two cylinders is the integral of (11) from r to b:

$$A(r)=\frac{\mu_0 I}{2\pi}\,[\ln(b)-\ln(r)]. \tag{12}$$

To find the potential for a single wire, we cannot simply let the outer radius b in (12) tend to infinity because the term $\ln(b)$ diverges. However, if we fix b at a value larger than the values of r in which we are interested, then $\ln(b)$ is just a constant, which we can drop, to give

$$A(r)=-\frac{\mu_0 I}{2\pi}\ln(r), \quad [\text{for } a<r<b]. \tag{13}$$

This is the vector potential for two coaxial cylinders of radii a and b. However, we know that the corresponding B field for $a<r<b$ is the same as that outside a single wire of radius a. Since (13) is independent of b, we can now let $b\to\infty$, and we conclude that the vector potential **A** outside a single wire is parallel to the current [the direction predicted in connection with Eq. (2)] and has magnitude given by[13] (13).

One can find other examples of currents for which the known form of **B** for a given $\mu_0\mathbf{J}$ lets one write down **A** for a given **B**. For instance, the vector potential for two antiparallel current sheets and for a toroidal solenoid can both be found easily in this way. (The latter is discussed in detail in Ref. 12.)

III. THE VECTOR POTENTIAL AS "ELECTROMAGNETIC MOMENTUM"

Whenever we define a new concept, we need to say as much as possible—beyond the bare definition—to show our students what the concept really *is*. This is desirable in its own right, but, equally important, it gives the students a context within which to place and understand the new concept. Thus, beyond defining **A** as the "thing whose curl is **B**," we need to say as much as possible about its physical significance. We wish to argue that the Maxwell–Thomson view that **A** is the stored momentum per unit charge supplies this needed physical meaning. Throughout this section we shall consider the motion of a single charge q in an applied electromagnetic field, given by potentials ϕ and **A**.

As a first indication that **A** is at least a candidate for stored momentum per unit charge, we can point out that the units of **A** are precisely those of [momentum/charge]. (Verifying this makes a nice exercise for your students in handling the units of magnetic field.) Another point that is easily made is that in relativity, **A** is related to ϕ exactly as momentum is related to energy. Even students who have not studied relativity formally are almost certainly aware that momentum and energy combine to form the four-momentum $p=(\mathbf{p},E/c)$. Thus we can at least tell them that (as they will learn later) **A** and ϕ combine to form the four-potential $A=(\mathbf{A},\phi/c)$, with **A** in the "momentum slot" and ϕ in the "energy slot."

Perhaps the most compelling argument that **A** is somehow connected with momentum comes from the Lagrangian mechanics of a charged particle in an electromagnetic field. If our students have already learned about Lagrangians, then we can show them that the Lagrangian

Gauge Transformations

ϕ determines E

and

A determines B

But ϕ and A are not unique

Gauge Transformations produce the different choices of ϕ and A that give the same E and B

Lots of Gauges

Coulomb Gauge

Lorenz Gauge

Axial Gauge

Temporal Gauge

Velocity Gauge

Kirchhoff Gauge

Landau Gauge

Feynman Gauge

t' Hooft Gauge

Unitary Gauge

The advantages are much less clear for magnetostatics. The integral for $\boldsymbol{A}$ is already a vector integral:

$$\boldsymbol{A}(1) = \frac{1}{4\pi\epsilon_0 c^2} \int \frac{\boldsymbol{j}(2)\, dV_2}{r_{12}}, \tag{15.24}$$

which is, of course, three integrals. Also, when we take the curl of $\boldsymbol{A}$ to get $\boldsymbol{B}$, we have six derivatives to do and combine by pairs. It is not immediately obvious whether in most problems this procedure is really any easier than computing $\boldsymbol{B}$ directly from

$$\boldsymbol{B}(1) = \frac{1}{4\pi\epsilon_0 c^2} \int \frac{\boldsymbol{j}(2) \times \boldsymbol{e}_{12}}{r_{12}^2}\, dV_2. \tag{15.25}$$

Using the vector potential is often more difficult for simple problems for the following reason. Suppose we are interested only in the magnetic field $\boldsymbol{B}$ at one point, and that the problem has some nice symmetry—say we want the field at a point on the axis of a ring of current. Because of the symmetry, we can easily get $\boldsymbol{B}$ by doing the integral of Eq. (15.25). If, however, we were to find $\boldsymbol{A}$ first, we would have to compute $\boldsymbol{B}$ from *derivatives* of $\boldsymbol{A}$, so we must know what $\boldsymbol{A}$ is at all points in the *neighborhood* of the point of interest. And most of these points are off the axis of symmetry, so the integral for $\boldsymbol{A}$ gets complicated. In the ring problem, for example, we would need to use elliptic integrals. In such problems, $\boldsymbol{A}$ is clearly not very useful. It is true that in many complex problems it is easier to work with $\boldsymbol{A}$, but it would be hard to argue that this ease of technique would justify making you learn about one more vector field.

We have introduced $\boldsymbol{A}$ because it *does* have an important physical significance. Not only is it related to the energies of currents, as we saw in the last section, but it is also a "real" physical field in the sense that we described above. In classical mechanics it is clear that we can write the force on a particle as

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}), \tag{15.26}$$

so that, given the forces, everything about the motion is determined. In any region where $\boldsymbol{B} = 0$ even if $\boldsymbol{A}$ is not zero, such as outside a solenoid, there is no discernible effect of $\boldsymbol{A}$. Therefore for a long time it was believed that $\boldsymbol{A}$ was not a "real" field. It turns out, however, that there are phenomena involving quantum mechanics which show that the field $\boldsymbol{A}$ is in fact a "real" field in the sense we have defined it. In the next section we will show you how that works.

15–5 The vector potential and quantum mechanics

There are many changes in what concepts are important when we go from classical to quantum mechanics. We have already discussed some of them in Vol. I. In particular, the force concept gradually fades away, while the concepts of energy and momentum become of paramount importance. You remember that instead of particle motions, one deals with probability amplitudes which vary in space and time. In these amplitudes there are wavelengths related to momenta, and frequencies related to energies. The momenta and energies, which determine the phases of wave functions, are therefore the important quantities in quantum mechanics. Instead of forces, we deal with the way interactions change the wavelength of the waves. The idea of a force becomes quite secondary—if it is there at all. When people talk about nuclear forces, for example, what they usually analyze and work with are the energies of interaction of two nucleons, and not the force between them. Nobody ever differentiates the energy to find out what the force looks like. In this section we want to describe how the vector and scalar potentials enter into quantum mechanics. It is, in fact, just because momentum and energy play a central role in quantum mechanics that $\boldsymbol{A}$ and ϕ provide the most direct way of introducing electromagnetic effects into quantum descriptions.

We must review a little how quantum mechanics works. We will consider again the imaginary experiment described in Chapter 37 of Vol. I, in which elec-

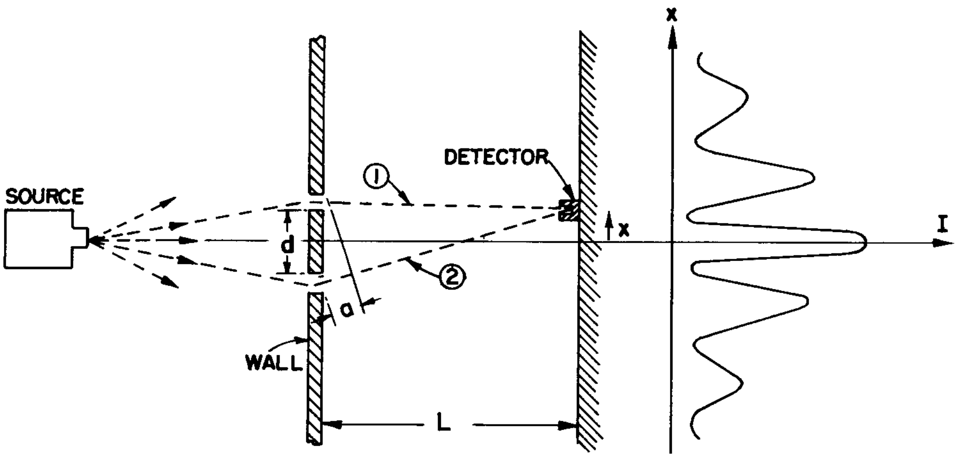


Fig. 15–5. An interference experiment with electrons (see also Chapter 37 of Vol. I).

trons are diffracted by two slits. The arrangement is shown again in Fig. 15–5. Electrons, all of nearly the same energy, leave the source and travel toward a wall with two narrow slits. Beyond the wall is a "backstop" with a movable detector. The detector measures the rate, which we call I, at which electrons arrive at a small region of the backstop at the distance x from the axis of symmetry. The rate is proportional to the probability that an individual electron that leaves the source will reach that region of the backstop. This probability has the complicated-looking distribution shown in the figure, which we understand as due to the interference of two amplitudes, one from each slit. The interference of the two amplitudes depends on their phase difference. That is, if the amplitudes are $C_1e^{i\Phi_1}$ and $C_2e^{i\Phi_2}$, the phase difference $\delta = \Phi_1 - \Phi_2$ determines their interference pattern [see Eq. (29.12) in Vol. I]. If the distance between the screen and the slits is L, and if the difference in the path lengths for electrons going through the two slits is a, as shown in the figure, then the phase difference of the two waves is given by

$$\delta = \frac{a}{\lambdabar}. \tag{15.27}$$

As usual, we let $\lambdabar = \lambda/2\pi$, where λ is the wavelength of the space variation of the probability amplitude. For simplicity, we will consider only values of x much less than L; then we can set

$$a = \frac{x}{L}\,d$$

and

$$\delta = \frac{x}{L}\frac{d}{\lambdabar}. \tag{15.28}$$

When x is zero, δ is zero; the waves are in phase, and the probability has a maximum. When δ is π, the waves are out of phase, they interfere destructively, and the probability is a minimum. So we get the wavy function for the electron intensity.

Now we would like to state the law that for quantum mechanics replaces the force law $\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B}$. It will be the law that determines the behavior of quantum-mechanical particles in an electromagnetic field. Since what happens is determined by amplitudes, the law must tell us how the magnetic influences affect the amplitudes; we are no longer dealing with the acceleration of a particle. The law is the following: the phase of the amplitude to arrive via any trajectory is changed by the presence of a magnetic field by an amount equal to the integral of the vector potential along the whole trajectory times the charge of the particle over Planck's constant. That is,

$$\text{Magnetic change in phase} = \frac{q}{\hbar}\int_{\text{trajectory}} \boldsymbol{A}\cdot d\boldsymbol{s}. \tag{15.29}$$

If there were no magnetic field there would be a certain phase of arrival. If there is a magnetic field anywhere, the phase of the arriving wave is increased by the integral in Eq. (15.29).

Although we will not need to use it for our present discussion, we mention that the effect of an electrostatic field is to produce a phase change given by the *negative* of the *time* integral of the scalar potential ϕ:

$$\text{Electric change in phase} = -\frac{q}{\hbar}\int \phi\, dt.$$

These two expressions are correct not only for static fields, but together give the correct result for *any* electromagnetic field, static or dynamic. This is the law that replaces $\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$. We want now, however, to consider only a static magnetic field.

Suppose that there is a magnetic field present in the two-slit experiment. We want to ask for the phase of arrival at the screen of the two waves whose paths pass through the two slits. Their interference determines where the maxima in the probability will be. We may call Φ_1 the phase of the wave along trajectory (1). If $\Phi_1(B = 0)$ is the phase without the magnetic field, then when the field is turned on the phase will be

$$\Phi_1 = \Phi_1(B = 0) + \frac{q}{\hbar}\int_{(1)} \boldsymbol{A} \cdot d\boldsymbol{s}. \tag{15.30}$$

Similarly, the phase for trajectory (2) is

$$\Phi_2 = \Phi_2(B = 0) + \frac{q}{\hbar}\int_{(2)} \boldsymbol{A} \cdot d\boldsymbol{s}. \tag{15.31}$$

The interference of the waves at the detector depends on the phase difference

$$\delta = \Phi_1(B = 0) - \Phi_2(B = 0) + \frac{q}{\hbar}\int_{(1)} \boldsymbol{A} \cdot d\boldsymbol{s} - \frac{q}{\hbar}\int_{(2)} \boldsymbol{A} \cdot d\boldsymbol{s}. \tag{15.32}$$

The no-field difference we will call $\delta(B = 0)$; it is just the phase difference we have calculated above in Eq. (15.28). Also, we notice that the two integrals can be written as *one* integral that goes forward along (1) and back along (2); we call this the closed path (1–2). So we have

$$\delta = \delta(B = 0) + \frac{q}{\hbar}\oint_{(1-2)} \boldsymbol{A} \cdot d\boldsymbol{s}. \tag{15.33}$$

This equation tells us how the electron motion is changed by the magnetic field; with it we can find the new positions of the intensity maxima and minima at the backstop.

Before we do that, however, we want to raise the following interesting and important point. You remember that the vector potential function has some arbitrariness. Two different vector potential functions $\boldsymbol{A}$ and $\boldsymbol{A}'$ whose difference is the gradient of some scalar function $\nabla\psi$, both represent the same magnetic field, since the curl of a gradient is zero. They give, therefore, the same classical force $q\boldsymbol{v} \times \boldsymbol{B}$. If in quantum mechanics the effects depend on the vector potential, *which* of the many possible $\boldsymbol{A}$-functions is correct?

The answer is that the same arbitrariness in $\boldsymbol{A}$ continues to exist for quantum mechanics. If in Eq. (15.33) we change $\boldsymbol{A}$ to $\boldsymbol{A}' = \boldsymbol{A} + \nabla\psi$, the integral on $\boldsymbol{A}$ becomes

$$\oint_{(1-2)} \boldsymbol{A}' \cdot d\boldsymbol{s} = \oint_{(1-2)} \boldsymbol{A} \cdot d\boldsymbol{s} + \oint_{(1-2)} \nabla\psi \cdot d\boldsymbol{s}.$$

The integral of $\nabla\psi$ is around the *closed* path (1–2), but the integral of the tangential component of a gradient on a closed path is always zero, by Stokes' theorem. Therefore both $\boldsymbol{A}$ and $\boldsymbol{A}'$ give the same phase differences and the same quantum-mechanical interference effects. In both classical and quantum theory it is only the curl of $\boldsymbol{A}$ that matters; any choice of the function of $\boldsymbol{A}$ which has the correct curl gives the correct physics.

The same conclusion is evident if we use the results of Section 14–1. There we found that the line integral of $\boldsymbol{A}$ around a closed path is the flux of $\boldsymbol{B}$ through the path, which here is the flux between paths (1) and (2). Equation (15.33) can, if we wish, be written as

$$\delta = \delta(B = 0) + \frac{q}{\hbar}\,[\text{flux of } \boldsymbol{B} \text{ between (1) and (2)}], \tag{15.34}$$

where by the flux of $\boldsymbol{B}$ we mean, as usual, the surface integral of the normal component of $\boldsymbol{B}$. The result depends only on $\boldsymbol{B}$, and therefore only on the curl of $\boldsymbol{A}$.

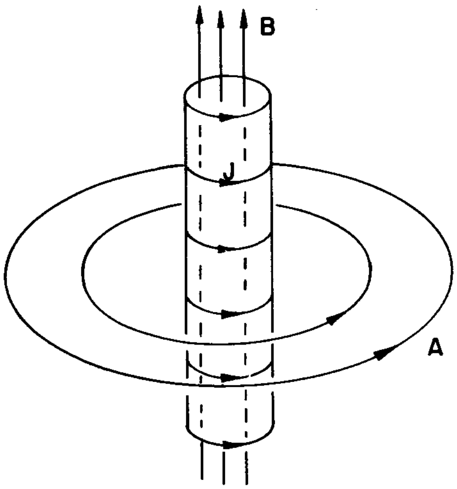


Fig. 15–6. The magnetic field and vector potential of a long solenoid.

Now because we can write the result in terms of $\boldsymbol{B}$ as well as in terms of $\boldsymbol{A}$, you might be inclined to think that the $\boldsymbol{B}$ holds its own as a "real" field and that the $\boldsymbol{A}$ can still be thought of as an artificial construction. But the definition of "real" field that we originally proposed was based on the idea that a "real" field would not act on a particle from a distance. We can, however, give an example in which $\boldsymbol{B}$ is zero—or at least arbitrarily small—at any place where there is some chance to find the particles, so that it is not possible to think of it acting *directly* on them.

You remember that for a long solenoid carrying an electric current there is a $\boldsymbol{B}$-field inside but none outside, while there is lots of $\boldsymbol{A}$ circulating around outside, as shown in Fig. 15–6. If we arrange a situation in which electrons are to be found only *outside* of the solenoid—only where there is $\boldsymbol{A}$—there will still be an influence on the motion, according to Eq. (15.33). Classically, that is impossible. Classically, the force depends only on $\boldsymbol{B}$; in order to know that the solenoid is carrying current, the particle must go through it. But quantum-mechanically you can find out that there is a magnetic field inside the solenoid by going *around* it—without ever going close to it!

Suppose that we put a very long solenoid of small diameter just behind the wall and between the two slits, as shown in Fig. 15–7. The diameter of the solenoid is to be much smaller than the distance d between the two slits. In these circumstances, the diffraction of the electrons at the slit gives no appreciable probability that the electrons will get near the solenoid. What will be the effect on our interference experiment?

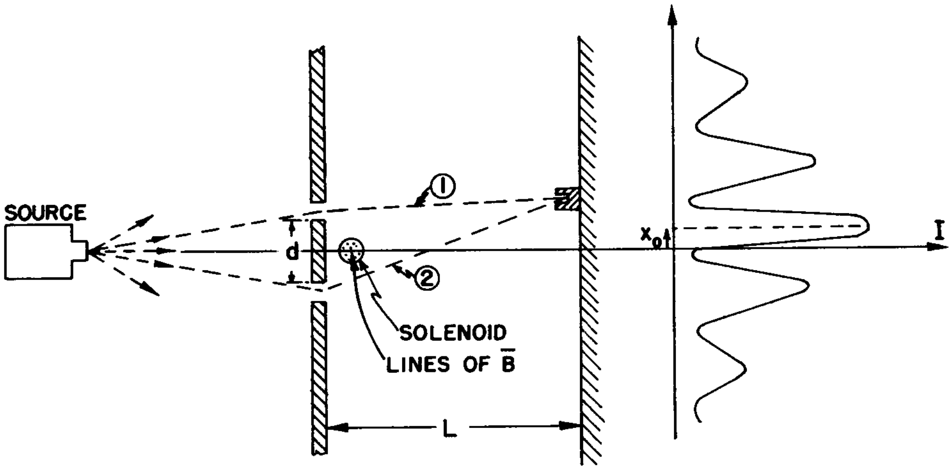


Fig. 15–7. A magnetic field can influence the motion of electrons even though it exists only in regions where there is an arbitrarily small probability of finding the electrons.

We compare the situation with and without a current through the solenoid. If we have no current, we have no $\boldsymbol{B}$ or $\boldsymbol{A}$ and we get the original pattern of electron intensity at the backstop. If we turn the current on in the solenoid and build up a magnetic field $\boldsymbol{B}$ inside, then there is an $\boldsymbol{A}$ outside. There is a shift in the phase difference proportional to the circulation of $\boldsymbol{A}$ outside the solenoid, which will mean that the pattern of maxima and minima is shifted to a new position. In fact, since the flux of $\boldsymbol{B}$ inside is a constant for any pair of paths, so also is the circulation of $\boldsymbol{A}$. For every arrival point there is the same phase change; this corresponds to shifting the entire pattern in x by a constant amount, say x_0, that we can easily calculate. The maximum intensity will occur where the phase difference between the two waves is zero. Using Eq. (15.32) or Eq. (15.33) for δ and Eq. (15.28) for $\delta(B = 0)$, we have

$$x_0 = -\frac{L}{d}\lambda\frac{q}{\hbar}\oint_{(1-2)} \boldsymbol{A}\cdot d\boldsymbol{s}, \tag{15.35}$$

or

$$x_0 = -\frac{L}{d}\lambda\frac{q}{\hbar}\,[\text{flux of } \boldsymbol{B} \text{ between (1) and (2)}]. \tag{15.36}$$

The pattern with the solenoid in place should appear* as shown in Fig. 15–7. At least, that is the prediction of quantum mechanics.

Precisely this experiment has recently been done. It is a very, very difficult experiment. Because the wavelength of the electrons is so small, the apparatus must be on a tiny scale to observe the interference. The slits must be very close together, and that means that one needs an exceedingly small solenoid. It turns out that in certain circumstances, iron crystals will grow in the form of very long, microscopically thin filaments called whiskers. When these iron whiskers are magnetized they are like a tiny solenoid, and there is no field outside except near the ends. The electron interference experiment was done with such a whisker between two slits, and the predicted displacement in the pattern of electrons was observed.

In our sense then, the **A**-field is "real." You may say: "But there *was* a magnetic field." There was, but remember our original idea—that a field is "real" if it is what must be specified *at the position* of the particle in order to get the motion. The **B**-field in the whisker acts at a distance. If we want to describe its influence not as action-at-a-distance, we must use the vector potential.

This subject has an interesting history. The theory we have described was known from the beginning of quantum mechanics in 1926. The fact that the vector potential appears in the wave equation of quantum mechanics (called the Schrödinger equation) was obvious from the day it was written. That it cannot be replaced by the magnetic field in any easy way was observed by one man after the other who tried to do so. This is also clear from our example of electrons moving in a region where there is no field and being affected nevertheless. But because in classical mechanics **A** did not appear to have any direct importance and, furthermore, because it could be changed by adding a gradient, people repeatedly said that the vector potential had no direct physical significance—that only the magnetic and electric fields are "right" even in quantum mechanics. It seems strange in retrospect that no one thought of discussing this experiment until 1956, when Bohm and Aharanov first suggested it and made the whole question crystal clear. The implication was there all the time, but no one paid attention to it. Thus many people were rather shocked when the matter was brought up. That's why someone thought it would be worth while to do the experiment to see that it really was right, even though quantum mechanics, which had been believed for so many years, gave an unequivocal answer. It is interesting that something like this can be around for thirty years but, because of certain prejudices of what is and is not significant, continues to be ignored.

Now we wish to continue in our analysis a little further. We will show the connection between the quantum-mechanical formula and the classical formula—to show why it turns out that if we look at things on a large enough scale it will look as though the particles are acted on by a force equal to $q\boldsymbol{v} \times$ the curl of $\boldsymbol{A}$. To get classical mechanics from quantum mechanics, we need to consider cases in which all the wavelengths are very small compared with distances over which external conditions, like fields, vary appreciably. We shall not prove the result in great generality, but only in a very simple example, to show how it works. Again we consider the same slit experiment. But instead of putting all the magnetic field in a very tiny region between the slits, we imagine a magnetic field that extends

* If the field **B** comes out of the plane of the figure, the flux as we have defined it is negative and x_0 is positive.

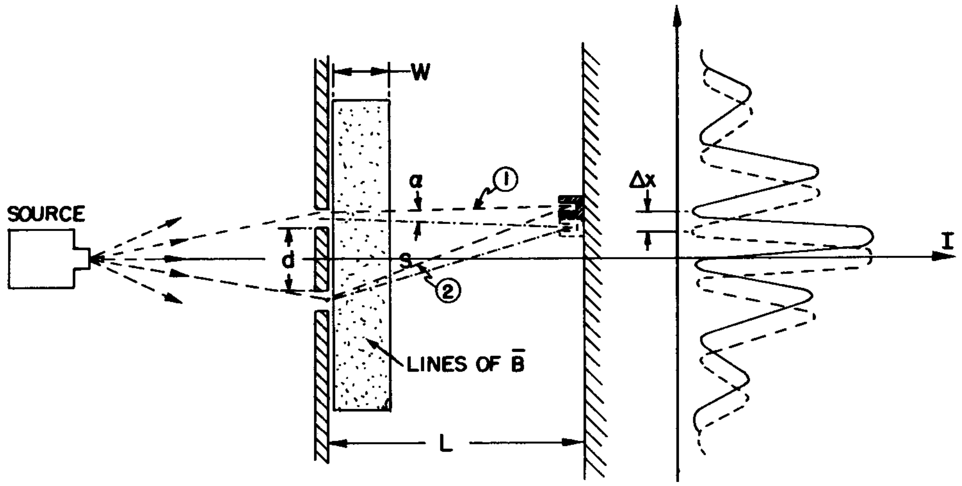


Fig. 15–8. The shift of the interference pattern due to a strip of magnetic field.

over a larger region behind the slits, as shown in Fig. 15–8. We will take the idealized case where we have a magnetic field which is uniform in a narrow strip of width w, considered small as compared with L. (That can easily be arranged; the backstop can be put as far out as we want.) In order to calculate the shift in phase, we must take the two integrals of $\boldsymbol{A}$ along the two trajectories (1) and (2). They differ, as we have seen, merely by the flux of $\boldsymbol{B}$ between the paths. To our approximation, the flux is Bwd. The phase difference for the two paths is then

$$\delta = \delta(B = 0) + \frac{q}{\hbar} Bwd. \tag{15.37}$$

We note that, to our approximation, the phase shift is independent of the angle. So again the effect will be to shift the whole pattern upward by an amount Δx. Using Eq. (15.28),

$$\Delta x = \frac{L\lambda\!\!\!^{-}}{d}\,\Delta\delta = \frac{L\lambda\!\!\!^{-}}{d}\,[\delta - \delta(B = 0)].$$

Using (15.37) for $\delta - \delta(B = 0)$,

$$\Delta x = L\lambda\!\!\!^{-}\,\frac{q}{\hbar}\,Bw. \tag{15.38}$$

Such a shift is equivalent to deflecting all the trajectories by the small angle α (see Fig. 15–8), where

$$\alpha = \frac{\Delta x}{L} = \frac{\lambda\!\!\!^{-}}{\hbar}\,qBw. \tag{15.39}$$

Now classically we would also expect a thin strip of magnetic field to deflect all trajectories through some small angle, say α', as shown in Fig. 15–9(a). As the electrons go through the magnetic field, they feel a transverse force $q\boldsymbol{v} \times \boldsymbol{B}$ which lasts for a time w/v. The change in their transverse momentum is just equal to this impulse, so

$$\Delta p_x = qwB. \tag{15.40}$$

The angular deflection [Fig. 15–9(b)] is equal to the ratio of this transverse momentum to the total momentum p. We get that

$$\alpha' = \frac{\Delta p_x}{p} = \frac{qwB}{p}. \tag{15.41}$$

We can compare this result with Eq. (15.39), which gives the same quantity computed quantum-mechanically. But the connection between classical mechanics and quantum mechanics is this: A particle of momentum p corresponds to a quan-

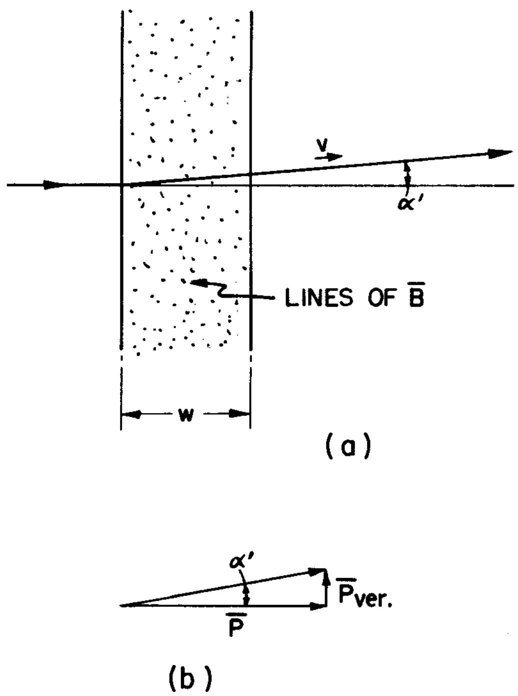


Fig. 15–9. Deflection of a particle due to passage through a strip of magnetic field.

tum amplitude varying with the wavelength $\lambdabar = \hbar/p$. With this equality, α and α' are identical; the classical and quantum calculations give the same result.

From the analysis we see how it is that the vector potential which appears in quantum mechanics in an explicit form produces a classical force which depends only on its derivatives. In quantum mechanics what matters is the interference between nearby paths; it always turns out that the effects depend only on how much the field $\boldsymbol{A}$ *changes* from point to point, and therefore only on the derivatives of $\boldsymbol{A}$ and not on the value itself. Nevertheless, the vector potential $\boldsymbol{A}$ (together with the scalar potential ϕ that goes with it) appears to give the most direct description of the physics. This becomes more and more apparent the more deeply we go into the quantum theory. In the general theory of quantum electrodynamics, one takes the vector and scalar potentials as the fundamental quantities in a set of equations that replace the Maxwell equations: $\boldsymbol{E}$ and $\boldsymbol{B}$ are slowly disappearing from the modern expression of physical laws; they are being replaced by $\boldsymbol{A}$ and ϕ.

15–6 What is true for statics is false for dynamics

We are now at the end of our exploration of the subject of static fields. Already in this chapter we have come perilously close to having to worry about what happens when fields change with time. We were barely able to avoid it in our treatment of magnetic energy by taking refuge in a relativistic argument. Even so, our treatment of the energy problem was somewhat artificial and perhaps even mysterious, because we ignored the fact that moving coils must, in fact, produce changing fields. It is now time to take up the treatment of time-varying fields—the subject of electrodynamics. We will do so in the next chapter. First, however, we would like to emphasize a few points.

Although we began this course with a presentation of the complete and correct equations of electromagnetism, we immediately began to study some incomplete pieces—because that was easier. There is a great advantage in starting with the simpler theory of static fields, and proceeding only later to the more complicated theory which includes dynamic fields. There is less new material to learn all at once, and there is time for you to develop your intellectual muscles in preparation for the bigger task.

But there is the danger in this process that before we get to see the complete story, the incomplete truths learned on the way may become ingrained and taken as the whole truth—that what is true and what is only sometimes true will become confused. So we give in Table 15–1 a summary of the important formulas we have covered, separating those which are true in general from those which are true for statics, but false for dynamics. This summary also shows, in part, where we are going, since as we treat dynamics we will be developing in detail what we must just state here without proof.

It may be useful to make a few remarks about the table. First, you should notice that the equations we started with are the *true* equations—we have not misled you there. The electromagnetic force (often called the *Lorentz force*) $\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$ is *true*. It is only Coulomb's law that is false, to be used only for statics. The four Maxwell equations for $\boldsymbol{E}$ and $\boldsymbol{B}$ are also true. The equations we took for statics are false, of course, because we left off all terms with time derivatives.

Gauss' law, $\nabla \cdot \boldsymbol{E} = \rho/\epsilon_0$, remains, but the curl of $\boldsymbol{E}$ is *not* zero in general. So $\boldsymbol{E}$ cannot always be equated to the gradient of a scalar—the electrostatic potential. We will see that a scalar potential still remains, but it is a time-varying quantity that must be used together with vector potentials for a complete description of the electric field. The equations governing this new scalar potential are, necessarily, also new.

We must also give up the idea that $\boldsymbol{E}$ is zero in conductors. When the fields are changing, the charges in conductors do not, in general, have time to rearrange themselves to make the field zero. They are set in motion, but never reach equilibrium. The only general statement is: electric fields in conductors produce cur-

Gauge transformations

Electric and magnetic fields can be written in terms of scalar and vector potentials, as follows:

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}, \tag{385}$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{386}$$

However, this prescription is not unique. There are many different potentials which can generate the same fields. We have come across this problem before. It is called *gauge invariance*. The most general transformation which leaves the $\mathbf{E}$ and $\mathbf{B}$ fields unchanged in Eqs. (385) and (386) is

$$\phi \rightarrow \phi + \frac{\partial \psi}{\partial t}, \tag{387}$$

$$\mathbf{A} \rightarrow \mathbf{A} - \nabla\psi. \tag{388}$$

This is clearly a generalization of the gauge transformation which we found earlier for static fields:

$$\phi \rightarrow \phi + c, \tag{389}$$

$$\mathbf{A} \rightarrow \mathbf{A} - \nabla\psi, \tag{390}$$

where c is a constant. In fact, if $\psi(\mathbf{r},t) \rightarrow \psi(\mathbf{r}) + c\,t$ then Eqs. (387) and (388) reduce to Eqs. (389) and (390).

Fock 1926

$$\mathbf{A} = \mathbf{A}_1 + \nabla f,$$

$$\varphi = \varphi_1 - \frac{1}{c}\frac{\partial f}{\partial t},$$

$$p = p_1 - \frac{e}{c} f,$$

and

$$\psi = \psi_0 e^{2\pi i p/h}.$$

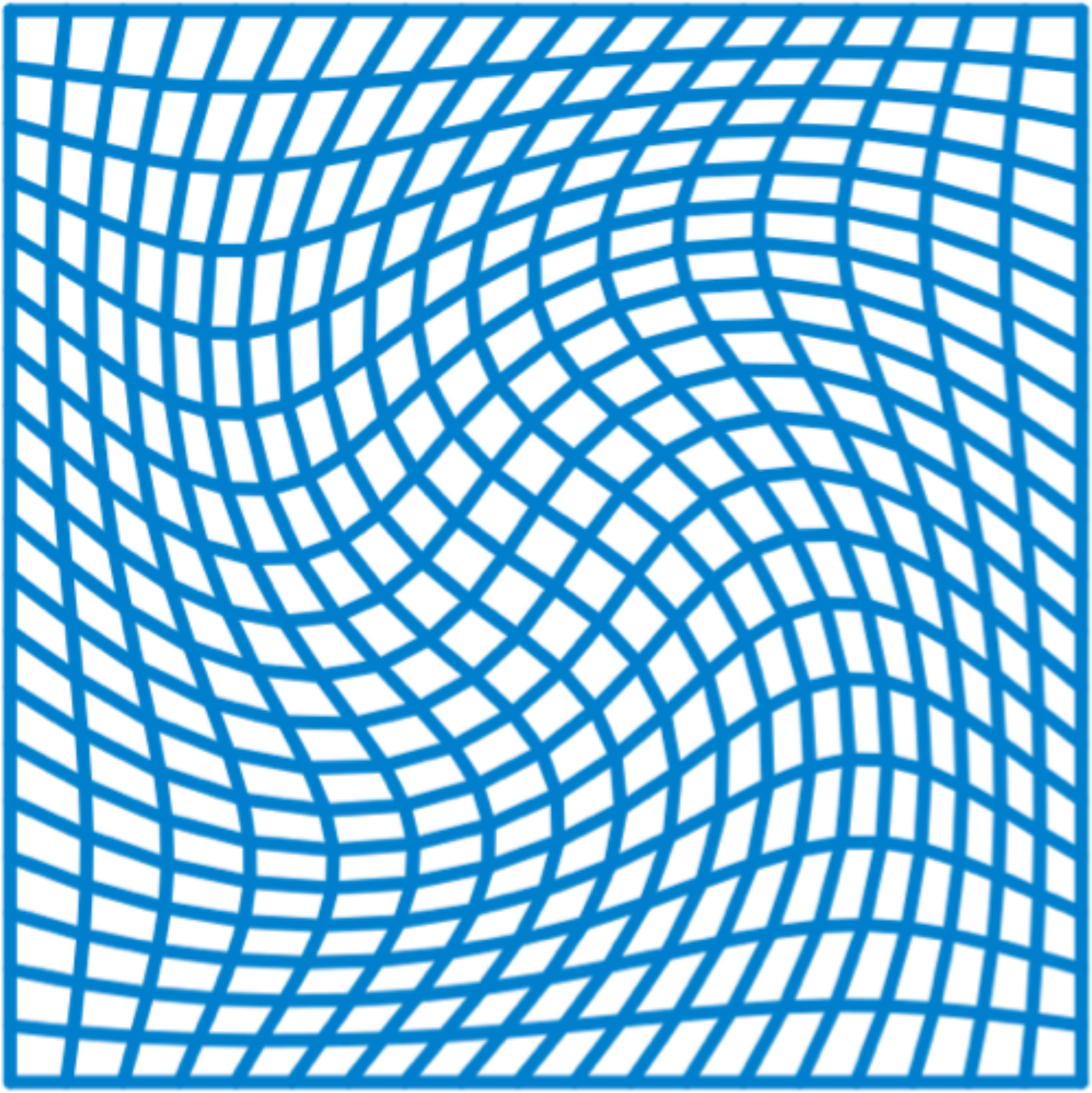
In present-day notation we write

$$\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \nabla \chi,$$

$$\Phi \to \Phi' = \Phi - \frac{1}{c}\frac{\partial \chi}{\partial t},$$

$$\psi \to \psi' = \psi \exp(ie\chi/\hbar c).$$

smooth and invertible



From Lorenz to Coulomb and other explicit gauge transformations

J. D. Jackson[a)]
Department of Physics and Lawrence Berkeley National Laboratory, University of California, Berkeley, Berkeley, California 94720

(Received 12 April 2002; accepted 10 May 2002)

The main purposes of this paper are (i) to illustrate explicitly by a number of examples the gauge functions $\chi(\mathbf{x},t)$ whose spatial and temporal derivatives transform one set of electromagnetic potentials into another equivalent set; and (ii) to show that, whatever propagation or nonpropagation characteristics are exhibited by the potentials in a particular gauge, the electric and magnetic *fields* are always the same and display the experimentally verified properties of causality and propagation at the speed of light. The example of the transformation from the Lorenz gauge (retarded solutions for both scalar and vector potential) to the Coulomb gauge (instantaneous, action-at-a-distance, scalar potential) is treated in detail. A transparent expression is obtained for the vector potential in the Coulomb gauge, with a finite nonlocality in time replacing the expected spatial nonlocality of the transverse current. A class of gauges (v-gauge) is described in which the scalar potential propagates at an arbitrary speed v relative to the speed of light. The Lorenz and Coulomb gauges are special cases of the v-gauge. The last examples of gauges and explicit gauge transformation functions are the Hamiltonian or temporal gauge, the nonrelativistic Poincaré or multipolar gauge, and the relativistic Fock–Schwinger gauge. © *2002 American Association of Physics Teachers.*
[DOI: 10.1119/1.1491265]

I. INTRODUCTION

The use of potentials in electromagnetism has a long history. The tortuous path to an understanding that the vector and scalar potentials are not unique and that different potentials describing the same physics are connected by something called a gauge transformation has been treated by Okun and me elsewhere.[1] If a given situation in electromagnetism is attributed to a scalar potential $\Phi(\mathbf{x},t)$ and a vector potential $\mathbf{A}(\mathbf{x},t)$, the physically meaningful and unique electric and magnetic fields $\mathbf{E}(\mathbf{x},t)$ and $\mathbf{B}(\mathbf{x},t)$ are determined from the potentials according to

$$\begin{aligned} \mathbf{E}(\mathbf{x},t) &= -\nabla\Phi(\mathbf{x},t) - \frac{1}{c}\frac{\partial \mathbf{A}(\mathbf{x},t)}{\partial t}, \\ \mathbf{B}(\mathbf{x},t) &= \nabla\times\mathbf{A}(\mathbf{x},t). \end{aligned} \tag{1.1}$$

Here we are using Gaussian units and considering phenomena in vacuum or as microscopic fields with localized sources. The expressions (1.1) are constituted so that they satisfy the homogeneous Maxwell equations automatically. Because the gradient of a scalar function has zero curl, it is clear that the magnetic field is unchanged if we add to $\mathbf{A}$ the gradient of a scalar function. Of course, such an addition changes the expression for the electric field. We must therefore modify the scalar potential, too. These changes are called a *gauge transformation*. Specifically, we have new potentials, $\Phi'(\mathbf{x},t)$ and $\mathbf{A}'(\mathbf{x},t)$,

$$\begin{aligned} \mathbf{A}'(\mathbf{x},t) &= \mathbf{A}(\mathbf{x},t) + \nabla\chi(\mathbf{x},t), \\ \Phi'(\mathbf{x},t) &= \Phi(\mathbf{x},t) - \frac{1}{c}\frac{\partial \chi(\mathbf{x},t)}{\partial t}, \end{aligned} \tag{1.2}$$

where the scalar function $\chi(\mathbf{x},t)$ is called the *gauge function*. The potentials $\mathbf{A}'(\mathbf{x},t)$, $\Phi'(\mathbf{x},t)$ are fully equivalent to the original set $\mathbf{A}(\mathbf{x},t)$, $\Phi(\mathbf{x},t)$, yielding the *same* electric and magnetic fields, but satisfying different dynamical equations. The chief purposes of this paper are to demonstrate some gauge transformations explicitly and to show explicitly that potentials in those different gauges, though different in detail, always yield the same electric and magnetic fields.

As is described in the textbooks cited in Refs. 2–4, common choices of gauge are $\nabla\cdot\mathbf{A}=0$, called the *Coulomb gauge*, and the relativistically covariant $\partial_\mu A^\mu = 0$ $(\nabla\cdot\mathbf{A} + (1/c)(\partial\Phi/\partial t)=0)$, called the *Lorenz gauge*.[5] There are many other gauges, but the textbooks rarely show explicitly the gauge function χ that transforms one gauge into another.

In Sec. II we review the form of the equations and the solutions for the potentials in the Lorenz gauge. We also exhibit the corresponding equations in the Coulomb gauge, focusing on the nonlocality of the source for the vector potential. The direct solution is deferred to Sec. IV. In Sec. III the gauge function $\chi(\mathbf{x},t)$ to go from the Lorenz gauge to the Coulomb gauge is constructed and used to calculate the Coulomb-gauge vector potential. The results (3.10) and (3.16) or (3.17), are surprisingly explicit and compact, with only one time integral over a finite range of the source's time t' $(t-R/c<t'<t)$, replacing the spatial nonlocality of the source with a temporal nonlocality. We return to the original equation for the Coulomb-gauge vector potential in Sec. IV and show that its straightforward solution can be transformed into that obtained in Sec. III more directly and simply through the gauge function.

In Sec. V we derive the electric and magnetic *fields* from the Coulomb-gauge potentials and show that they are the well-known expressions, causal and propagating with speed c, despite the instantaneous nature of the scalar potential. This ground has been traveled before in this journal by Brill and Goodman[6] and recently by Rohrlich.[7] There is also Problem 6.20 in my book.[2] Our discussion here is different and I think more transparent because of the form of our solution for $\mathbf{A}_C$. Some aspects of Brill and Goodman come close. In Sec. VI we discuss briefly the quasistatic limit of the vector potential in the Coulomb gauge and its use to obtain a Lagrangian for the interaction of charged particles that is cor-

 © 2002 American Association of Physics Teachers