The recording of this lecture is missing from the Caltech Archives.

28
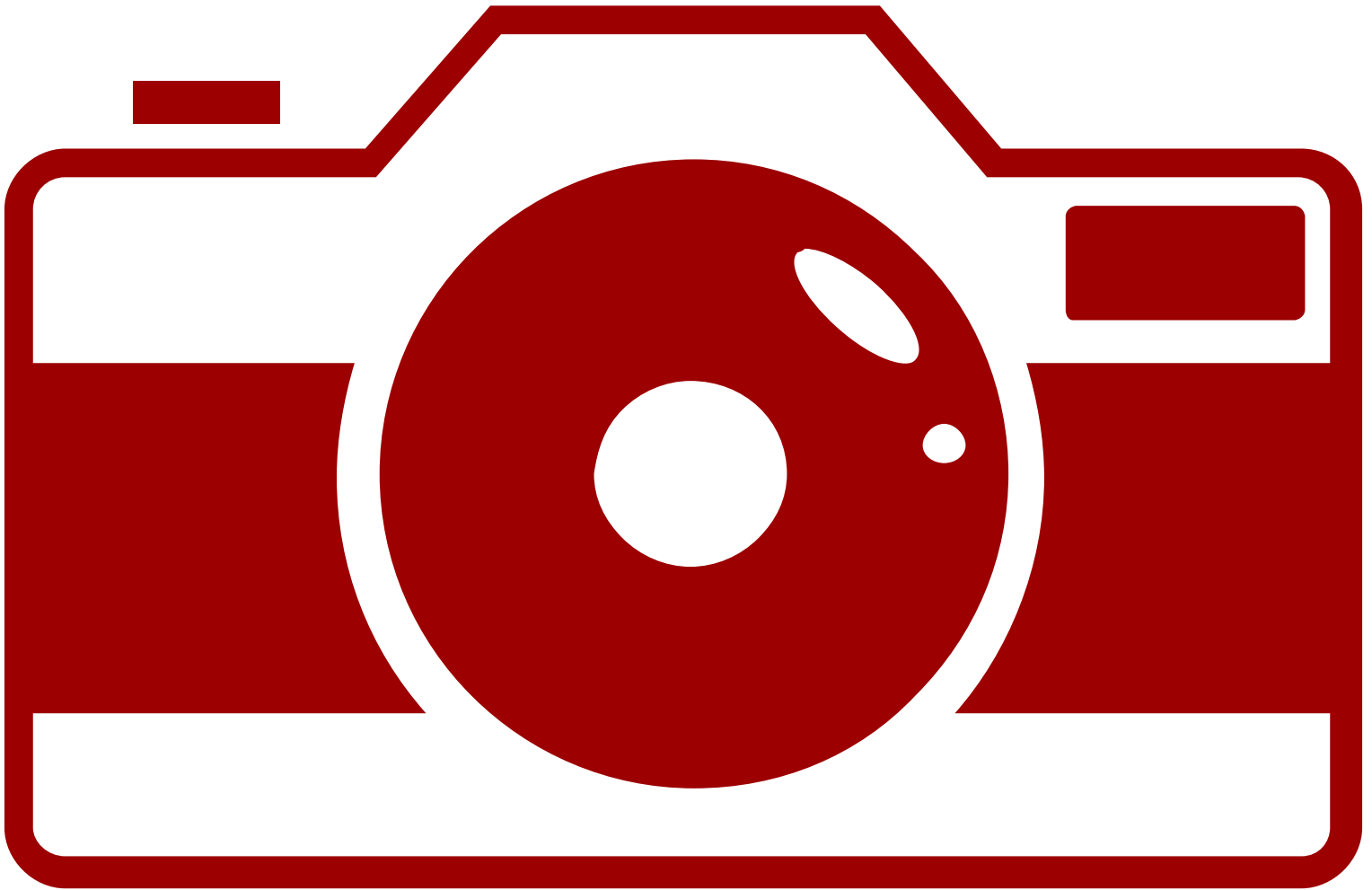
Electromagnetic Mass

Summary The field around charged particles carries energy, and, if they move, momentum so such particles have a contribution to their inertia (mass) due to electrodynamics. Classical theory gives ∞ for a point charge — quantum theory does no better — altho several experimental values are known (eg Mass$_{\pi+}$ − mass$_{\pi^0}$ = 4.6 MeV) no complete satisfactory theory for calculating them is known.

28–1 The field energy of a point charge

In bringing together relativity and Maxwell's equations, we have finished our main work on the theory of electromagnetism. There are, of course, some details we have skipped over and one large area that we will be concerned with in the future—the interaction of electromagnetic fields with matter. But we want to stop for a moment to show you that this tremendous edifice, which is such a beautiful success in explaining so many phenomena, ultimately falls on its face. When you follow any of our physics too far, you find that it always gets into some kind of trouble. Now we want to discuss a serious trouble—the failure of the classical electromagnetic theory. You can appreciate that there is a failure of all classical physics because of the quantum-mechanical effects. Classical mechanics is a mathematically consistent theory; it just doesn't agree with experience. It is interesting, though, that the classical theory of electromagnetism is an unsatisfactory theory all by itself. There are difficulties associated with the *ideas* of Maxwell's theory which are not solved by and not directly associated with quantum mechanics. You may say, "Perhaps there's no use worrying about these difficulties. Since the quantum mechanics is going to change the laws of electrodynamics, we should wait to see what difficulties there are after the modification." However, when electromagnetism is joined to quantum mechanics, the difficulties remain. So it will not be a waste of our time now to look at what these difficulties are. Also, they are of great historical importance. Furthermore, you may get some feeling of accomplishment from being able to go far enough with the theory to see everything—including all of its troubles.

The difficulty we speak of is associated with the concepts of electromagnetic momentum and energy, when applied to the electron or any charged particle. The concepts of simple charged particles and the electromagnetic field are in some way inconsistent. To describe the difficulty, we begin by doing some exercises with our energy and momentum concepts.

First, we compute the energy of a charged particle. Suppose we take a simple model of an electron in which all of its charge q is uniformly distributed on the surface of a sphere of radius a, which we may take to be zero for the special case of a point charge. Now let's calculate the energy in the electromagnetic field. If the charge is standing still, there is no magnetic field, and the energy per unit volume is proportional to the square of the electric field. The magnitude of the electric field is $q/4\pi\epsilon_0 r^2$, and the energy density is

$$u = \frac{\epsilon_0}{2} E^2 = \frac{q^2}{32\pi^2\epsilon_0 r^4}.$$

To get the total energy, we must integrate this density over all space. Using the volume element $4\pi r^2\, dr$, the total energy, which we will call U_{elec}, is

$$U_{\text{elec}} = \int \frac{q^2}{8\pi\epsilon_0 r^2}\, dr.$$

This is readily integrated. The lower limit is a, and the upper limit is ∞, so

$$U_{\text{elec}} = \frac{1}{2} \frac{q^2}{4\pi\epsilon_0} \frac{1}{a}. \tag{28.1}$$

If we use the electronic charge q_e for q and the symbol e^2 for $q_e^2/4\pi\epsilon_0$, then

$$U_{\text{elec}} = \frac{1}{2} \frac{e^2}{a}. \tag{28.2}$$

It is all fine until we set a equal to zero for a point charge—there's the great difficulty. Because the energy of the field varies inversely as the fourth power of the distance from the center, its volume integral is infinite. There is an infinite amount of energy in the field surrounding a point charge.

What's wrong with an infinite energy? If the energy can't get out, but must stay there forever, is there any real difficulty with an infinite energy? Of course, a quantity that comes out infinite may be annoying, but what really matters is only whether there are any *observable* physical effects. To answer that question, we must turn to something else besides the energy. Suppose we ask how the energy *changes* when we *move* the charge. Then, if the *changes* are infinite, we will be in trouble.
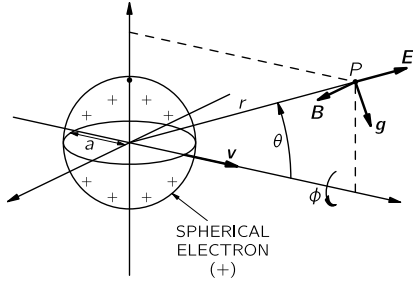
28–2The field momentum of a moving charge

Fig. 28–1.The fields $\boldsymbol{E}$ and $\boldsymbol{B}$ and the momentum density $\boldsymbol{g}$ for a positive electron. For a negative electron, $\boldsymbol{E}$ and $\boldsymbol{B}$ are reversed but $\boldsymbol{g}$ is not.

Suppose an electron is moving at a uniform velocity through space, assuming for a moment that the velocity is low compared with the speed of light. Associated with this moving electron there is a momentum—even if the electron had no mass before it was charged—because of the momentum in the electromagnetic field. We can show that the field momentum is in the direction of the velocity $\boldsymbol{v}$ of the charge and is, for small velocities, proportional to v. For a point P at the distance r from the center of the charge and at the angle θ with respect to the line of motion (see Fig. 28–1) the electric field is radial and, as we have seen, the magnetic field is $\boldsymbol{v} \times \boldsymbol{E}/c^2$. The momentum density, Eq. (27.21), is

$$\boldsymbol{g} = \epsilon_0 \boldsymbol{E} \times \boldsymbol{B}.$$

It is directed obliquely toward the line of motion, as shown in the figure, and has the magnitude

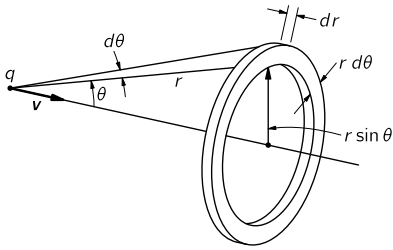$$g = \frac{\epsilon_0 v}{c^2} E^2 \sin \theta.$$

Fig. 28–2.The volume element $2\pi r^2 \sin \theta \, d\theta \, dr$ used for calculating the field momentum.

The fields are symmetric about the line of motion, so when we integrate over space, the transverse components will sum to zero, giving a resultant momentum parallel to $\boldsymbol{v}$. The component of $\boldsymbol{g}$ in this direction is $g \sin \theta$, which we must integrate over all space. We take as our volume element a ring with its plane perpendicular to $\boldsymbol{v}$, as shown in Fig. 28–2. Its volume is $2\pi r^2 \sin \theta \, d\theta \, dr$. The total momentum is then

$$\boldsymbol{p} = \int \frac{\epsilon_0 \boldsymbol{v}}{c^2} E^2 \sin^2 \theta \, 2\pi r^2 \sin \theta \, d\theta \, dr.$$

Since E is independent of θ (for $v \ll c$), we can immediately integrate over θ; the integral is

$$\int \sin^3 \theta \, d\theta = -\int (1 - \cos^2 \theta) \, d(\cos \theta) = -\cos \theta + \frac{\cos^3 \theta}{3}.$$

The limits of θ are 0 and π, so the θ-integral gives merely a factor of $4/3$, and

$$\boldsymbol{p} = \frac{8\pi}{3} \frac{\epsilon_0 \boldsymbol{v}}{c^2} \int E^2 r^2 \, dr.$$

The integral (for $v \ll c$) is the one we have just evaluated to find the energy; it is $q^2/16\pi^2 \epsilon_0^2 a$, and

$$\boldsymbol{p} = \frac{2}{3} \frac{q^2}{4\pi \epsilon_0} \frac{\boldsymbol{v}}{ac^2},$$

or

$$\boldsymbol{p} = \frac{2}{3} \frac{e^2}{ac^2} \boldsymbol{v}. \tag{28.3}$$

The momentum in the field—the electromagnetic momentum—is proportional to $\boldsymbol{v}$. It is just what we should have for a particle with the mass equal to the coefficient of $\boldsymbol{v}$. We can, therefore, call this coefficient the *electromagnetic mass*, m_{elec}, and write it as

$$m_{\text{elec}} = \frac{2}{3}\frac{e^2}{ac^2}. \tag{28.4}$$

28–3 Electromagnetic mass

Where does the mass come from? In our laws of mechanics we have supposed that every object "carries" a thing we call the mass—which also means that it "carries" a momentum proportional to its velocity. Now we discover that it is understandable that a charged particle carries a momentum proportional to its velocity. It might, in fact, be that the mass is just the effect of electrodynamics. The origin of mass has until now been unexplained. We have at last in the theory of electrodynamics a grand opportunity to understand something that we never understood before. It comes out of the blue—or rather, from Maxwell and Poynting—that any charged particle will have a momentum proportional to its velocity just from electromagnetic influences.

Let's be conservative and say, for a moment, that there are two kinds of mass—that the total momentum of an object could be the sum of a mechanical momentum and the electromagnetic momentum. The mechanical momentum is the "mechanical" mass, m_{mech}, times $\boldsymbol{v}$. In experiments where we measure the mass of a particle by seeing how much momentum it has, or how it swings around in an orbit, we are measuring the total mass. We say generally that the momentum is the total mass $(m_{\text{mech}} + m_{\text{elec}})$ times the velocity. So the observed mass can consist of two pieces (or possibly more if we include other fields): a mechanical piece plus an electromagnetic piece. We know that there is definitely an electromagnetic piece, and we have a formula for it. And there is the thrilling possibility that the mechanical piece is not there at all—that the mass is all electromagnetic.

Let's see what size the electron must have if there is to be no mechanical mass. We can find out by setting the electromagnetic mass of Eq. ([28.4](#)) equal to the observed mass m_e of an electron. We find

$$a = \frac{2}{3}\frac{e^2}{m_e c^2}. \tag{28.5}$$

The quantity

$$r_0 = \frac{e^2}{m_e c^2} \tag{28.6}$$

is called the "classical electron radius"; it has the numerical value 2.82×10^{-13} cm, about one one-hundred-thousandth of the diameter of an atom.

Why is r_0 called the electron radius, rather than our a? Because we could equally well do the same calculation with other assumed distributions of charges—the charge might be spread uniformly through the volume of a sphere or it might be smeared out like a fuzzy ball. For any particular assumption the factor $2/3$ would change to some other fraction. For instance, for a charge uniformly distributed throughout the volume of a sphere, the $2/3$ gets replaced by $4/5$. Rather than to argue over which distribution is correct, it was decided to define r_0 as the "nominal" radius. Then different theories could supply their pet coefficients.

Let's pursue our electromagnetic theory of mass. Our calculation was for $v \ll c$; what happens if we go to high velocities? Early attempts led to a certain amount of confusion, but Lorentz realized that the charged sphere would contract into a ellipsoid at high velocities and that the fields would change in accordance with the formulas ([26.6](#)) and ([26.7](#)) we derived for the relativistic case in Chapter [26](#). If you carry through the integrals for $\boldsymbol{p}$ in that case, you find that for an arbitrary velocity $\boldsymbol{v}$, the momentum is altered by the factor $1/\sqrt{1 - v^2/c^2}$:

$$\boldsymbol{p} = \frac{2}{3}\frac{e^2}{ac^2}\frac{\boldsymbol{v}}{\sqrt{1 - v^2/c^2}}. \tag{28.7}$$

In other words, the electromagnetic mass rises with velocity inversely as $\sqrt{1 - v^2/c^2}$—a discovery that was made before the theory of relativity.

Early experiments were proposed to measure the changes with velocity in the observed mass of a particle in order to determine how much of the mass was mechanical and how much was electrical. It was believed at the time that the electrical part *would* vary with velocity, whereas the mechanical part would *not*. But while the experiments were being done, the theorists were also at work. Soon the theory of relativity was developed, which proposed that no matter what the origin of the mass, it *all* should vary as $m_0/\sqrt{1 - v^2/c^2}$. Equation ([28.7](#)) was the beginning of the theory that mass depended on velocity.

Let's now go back to our calculation of the energy in the field, which led to Eq. ([28.2](#)). According to the theory of relativity, the energy U will have the mass U/c^2; Eq. ([28.2](#)) then says that the field of the electron should have the mass

$$m'_{\text{elec}} = \frac{U_{\text{elec}}}{c^2} = \frac{1}{2}\frac{e^2}{ac^2}, \tag{28.8}$$

which is not the same as the electromagnetic mass, m_{elec}, of Eq. ([28.4](#)). In fact, if we just combine Eqs. ([28.2](#)) and ([28.4](#)), we would write

$$U_{\text{elec}} = \frac{3}{4}m_{\text{elec}}c^2.$$

This formula was discovered before relativity, and when Einstein and others began to realize that it must always be that $U = mc^2$, there was great confusion.

28–4 The force of an electron on itself

The discrepancy between the two formulas for the electromagnetic mass is especially annoying, because we have carefully proved that the theory of electrodynamics is consistent with the principle of relativity. Yet the theory of relativity implies without question that the momentum must be the same as the

energy times v/c^2. So we are in some kind of trouble; we must have made a mistake. We did not make an algebraic mistake in our calculations, but we have left something out.

In deriving our equations for energy and momentum, we assumed the conservation laws. We assumed that *all* forces were taken into account and that any work done and any momentum carried by other "nonelectrical" machinery was included. Now if we have a sphere of charge, the electrical forces are all repulsive and an electron would tend to fly apart. Because the system has unbalanced forces, we can get all kinds of errors in the laws relating energy and momentum. To get a *consistent* picture, we must imagine that something holds the electron together. The charges must be *held* to the sphere by some kind of rubber bands—something that keeps the charges from flying off. It was first pointed out by Poincaré that the rubber bands—or whatever it is that holds the electron together—must be included in the energy and momentum calculations. For this reason the extra nonelectrical forces are also known by the more elegant name "the Poincaré stresses." If the extra forces are included in the calculations, the masses obtained in two ways are changed (in a way that depends on the detailed assumptions). And the results are consistent with relativity; i.e., the mass that comes out from the momentum calculation is the same as the one that comes from the energy calculation. However, both of them contain *two* contributions: an electromagnetic mass and contribution from the Poincaré stresses. Only when the two are added together do we get a consistent theory.

It is therefore impossible to get all the mass to be electromagnetic in the way we hoped. It is not a legal theory if we have nothing but electrodynamics. Something else has to be added. Whatever you call them—"rubber bands," or "Poincaré stresses," or something else—there have to be other forces in nature to make a consistent theory of this kind.

Clearly, as soon as we have to put forces on the inside of the electron, the beauty of the whole idea begins to disappear. Things get very complicated. You would want to ask: How strong are the stresses? How does the electron shake? Does it oscillate? What are all its internal properties? And so on. It might be possible that an electron does have some complicated internal properties. If we made a theory of the electron along these lines, it would predict odd properties, like modes of oscillation, which haven't apparently been observed. We say "apparently" because we observe a lot of things in nature that still do not make sense. We may someday find out that one of the things we don't understand today (for example, the muon) can, in fact, be explained as an oscillation of the Poincaré stresses. It doesn't seem likely, but no one can say for sure. There are so many things about fundamental particles that we still don't understand. Anyway, the complex structure implied by this theory is undesirable, and the attempt to explain all mass in terms of electromagnetism—at least in the way we have described—has led to a blind alley.

We would like to think a little more about why we say we have a mass when the momentum in the field is proportional to the velocity. Easy! The mass is the coefficient between momentum and velocity. But we can look at the mass in another way: a particle has mass if you have to exert a force in order to accelerate it. So it may help our understanding if we look a little more closely at where the forces come from. How do we know that there has to be a force? Because we have proved the law of the conservation of momentum for the fields. If we have a charged particle and push on it for awhile, there will be some momentum in the electromagnetic field. Momentum must have been poured into the field somehow. Therefore there must have been a force pushing on the electron in order to get it going—a force in addition to that required by its mechanical inertia, a force due to its electromagnetic interaction. And there must be a corresponding force back on the "pusher." But where does that force come from?
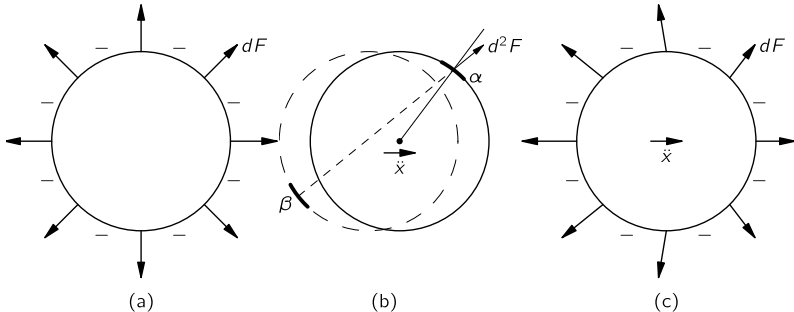

Fig. 28–3. The self-force on an accelerating electron is not zero because of the retardation. (By dF we mean the force on a surface element da; by d^2F we mean the force on the surface element da_α from the charge on the surface element da_β.)

The picture is something like this. We can think of the electron as a charged sphere. When it is at rest, each piece of charge repels electrically each other piece, but the forces all balance in pairs, so that there is no *net* force. [See Fig. 28–3(a).] However, when the electron is being accelerated, the forces will no longer be in balance because of the fact that the electromagnetic influences take time to go from one piece to another. For instance, the force on the piece α in Fig. 28–3(b) from a piece β on the opposite side depends on the position of β at an earlier time, as shown. Both the magnitude and direction of the force depend on the motion of the charge. If the charge is accelerating, the forces on various parts of the electron might be as shown in Fig. 28–3(c). When all these forces are added up, they don't cancel out. They would cancel for a uniform velocity, even though it looks at first glance as though the retardation would give an unbalanced force even for a uniform velocity. But it turns out that there is no net force unless the electron is being accelerated. With acceleration, if we look at the forces between the various parts of the electron, action and reaction are not exactly equal, and the electron exerts a force *on itself* that tries to hold back the acceleration. It holds itself back by its own bootstraps.

It is possible, but difficult, to calculate this self-reaction force; however, we don't want to go into such an elaborate calculation here. We will tell you what the result is for the special case of relatively uncomplicated motion in one dimension, say x. Then, the self-force can be written in a series. The first term in the series depends on the acceleration $\ddot{x}$, the next term is proportional to $\dddot{x}$, and so on.[1] The result is

$$F = \alpha \frac{e^2}{ac^2}\ddot{x} - \frac{2}{3}\frac{e^2}{c^3}\dddot{x} + \gamma\frac{e^2a}{c^4}\ddddot{x} + \cdots, \tag{28.9}$$

where α and γ are numerical coefficients of the order of 1. The coefficient α of the $\ddot{x}$ term depends on what charge distribution is assumed; if the charge is

distributed uniformly on a sphere, then $\alpha = 2/3$. So there is a term, proportional to the acceleration, which varies inversely as the radius a of the electron and agrees exactly with the value we got in Eq. (28.4) for m_{elec}. If the charge distribution is chosen to be different, so that α is changed, the fraction $2/3$ in Eq. (28.4) would be changed in the same way. The term in $\ddot{x}$ is *independent* of the assumed radius a, and also of the assumed distribution of the charge; its coefficient is *always* $2/3$. The next term is proportional to the radius a, and its coefficient γ depends on the charge distribution. You will notice that if we let the electron radius a go to zero, the last term (and all higher terms) will go to zero; the second term remains constant, but the first term—the electromagnetic mass—goes to infinity. And we can see that the infinity arises because of the force of one part of the electron on another—because we have allowed what is perhaps a silly thing, the possibility of the "point" electron acting on itself.

28–5Attempts to modify the Maxwell theory

We would like now to discuss how it might be possible to modify Maxwell's theory of electrodynamics so that the idea of an electron as a simple point charge could be maintained. Many attempts have been made, and some of the theories were even able to arrange things so that all the electron mass was electromagnetic. But all of these theories have died. It is still interesting to discuss some of the possibilities that have been suggested—to see the struggles of the human mind.

We started out our theory of electricity by talking about the interaction of one charge with another. Then we made up a theory of these interacting charges and ended up with a field theory. We believe it so much that we allow it to tell us about the force of one part of an electron on another. Perhaps the entire difficulty is that electrons do not act on themselves; perhaps we are making too great an extrapolation from the interaction of separate electrons to the idea that an electron interacts with itself. Therefore some theories have been proposed in which the possibility that an electron acts on itself is ruled out. Then there is no longer the infinity due to the self-action. Also, there is no longer any electromagnetic mass associated with the particle; all the mass is back to being mechanical, but there are new difficulties in the theory.

We must say immediately that such theories require a modification of the idea of the electromagnetic field. You remember we said at the start that the force on a particle at any point was determined by just two quantities—$\boldsymbol{E}$ and $\boldsymbol{B}$. If we abandon the "self-force" this can no longer be true, because if there is an electron in a certain place, the force isn't given by the total $\boldsymbol{E}$ and $\boldsymbol{B}$, but by only those parts due to *other* charges. So we have to keep track always of how much of $\boldsymbol{E}$ and $\boldsymbol{B}$ is due to the charge on which you are calculating the force and how much is due to the other charges. This makes the theory much more elaborate, but it gets rid of the difficulty of the infinity.

So we can, *if we want to*, say that there is no such thing as the electron acting upon itself, and throw away the whole set of forces in Eq. (28.9). However, we have then thrown away the baby with the bath! Because the second term in Eq. (28.9), the term in $\ddot{x}$, is needed. That force does something very definite. If you throw it away, you're in trouble again. When we accelerate a charge, it radiates electromagnetic waves, so it loses energy. Therefore, to accelerate a charge, we must require more force than is required to accelerate a neutral object of the same mass; otherwise energy wouldn't be conserved. The rate at which we do work on an accelerating charge must be equal to the rate of loss of energy by radiation. We have talked about this effect before—it is called the radiation resistance. We still have to answer the question: Where does the extra force, against which we must do this work, come from? When a big antenna is radiating, the forces come from the influence of one part of the antenna current on another. For a single accelerating electron radiating into otherwise empty space, there would seem to be only one place the force could come from—the action of one part of the electron on another part.

We found back in Chapter 32 of Vol. I that an oscillating charge radiates energy at the rate

$$\frac{dW}{dt} = \frac{2}{3}\frac{e^2(\ddot{x})^2}{c^3}. \tag{28.10}$$

Let's see what we get for the rate of doing work *on* an electron against the bootstrap force of Eq. (28.9). The rate of work is the force times the velocity, or $F\dot{x}$:

$$\frac{dW}{dt} = \alpha\,\frac{e^2}{ac^2}\,\ddot{x}\dot{x} - \frac{2}{3}\frac{e^2}{c^3}\,\dddot{x}\,\dot{x} + \cdots \tag{28.11}$$

The first term is proportional to $d\dot{x}^2/dt$, and therefore just corresponds to the rate of change of the kinetic energy $\frac{1}{2}mv^2$ associated with the electromagnetic mass. The second term should correspond to the radiated power in Eq. (28.10). But it is different. The discrepancy comes from the fact that the term in Eq. (28.11) is generally true, whereas Eq. (28.10) is right only for an *oscillating* charge. We can show that the two are equivalent if the motion of the charge is periodic. To do that, we rewrite the second term of Eq. (28.11) as

$$-\frac{2}{3}\frac{e^2}{c^3}\frac{d}{dt}(\dot{x}\ddot{x}) + \frac{2}{3}\frac{e^2}{c^3}(\ddot{x})^2,$$

which is just an algebraic transformation. If the motion of the electron is periodic, the quantity $\dot{x}\ddot{x}$ returns periodically to the same value, so that if we take the *average* of its time derivative, we get zero. The second term, however, is always positive (it's a square), so its average is also positive. This term gives the net work done and is just equal to Eq. (28.10).

The term in $\ddot{x}$ of the bootstrap force is required in order to have energy conservation in radiating systems, and we can't throw it away. It was, in fact, one of the triumphs of Lorentz to show that there is such a force and that it comes from the action of the electron on itself. We must believe in the idea of the action of the electron on itself, and we *need* the term in $\ddot{x}$. The problem is how we can get that term without getting the first term in Eq. (28.9), which gives all the trouble. We don't know how. You see that the classical electron theory has pushed itself into a tight corner.

There have been several other attempts to modify the laws in order to straighten the thing out. One way, proposed by Born and Infeld, is to change the Maxwell equations in a complicated way so that they are no longer linear. Then the electromagnetic energy and momentum can be made to come out finite. But the laws they suggest predict phenomena which have never been observed. Their theory also suffers from another difficulty we will come to later, which

is common to all the attempts to avoid the troubles we have described.

The following peculiar possibility was suggested by Dirac. He said: Let's admit that an electron acts on itself through the *second* term in Eq. (28.9) but not through the first. He then had an ingenious idea for getting rid of one but not the other. Look, he said, we made a special assumption when we took only the *retarded* wave solutions of Maxwell's equations; if we were to take the *advanced* waves instead, we would get something different. The formula for the self-force would be

$$F = \alpha \frac{e^2}{ac^2} \ddot{x} + \frac{2}{3} \frac{e^2}{c^3} \dddot{x} + \gamma \frac{e^2 a}{c^4} \ddddot{x} + \cdots \qquad (28.12)$$

This equation is just like Eq. (28.9) except for the sign of the second term—and some higher terms—of the series. [Changing from retarded to advanced waves is just changing the *sign* of the delay which, it is not hard to see, is equivalent to changing the sign of t everywhere. The only effect on Eq. (28.9) is to change the sign of all the odd time derivatives.] So, Dirac said, let's make the new rule that an electron acts on itself by one-half the *difference* of the retarded and advanced fields which it produces. The difference of Eqs. (28.9) and (28.12), divided by two, is then

$$F = -\frac{2}{3} \frac{e^2}{c^3} \dddot{x} + \text{higher terms.}$$

In all the higher terms, the radius a appears to some positive power in the numerator. Therefore, when we go to the limit of a point charge, we get only the one term—just what is needed. In this way, Dirac got the radiation resistance force and none of the inertial forces. There is no electromagnetic mass, and the classical theory is saved—but at the expense of an arbitrary assumption about the self-force.

The arbitrariness of the extra assumption of Dirac was removed, to some extent at least, by Wheeler and Feynman, who proposed a still stranger theory. They suggest that point charges interact *only* with other charges, but that the interaction is half through the advanced and half through the retarded waves. It turns out, most surprisingly, that in most situations you won't see any effects of the advanced waves, but they do have the effect of producing just the radiation reaction force. The radiation resistance is *not* due to the electron acting on itself, but from the following peculiar effect. When an electron is accelerated at the time t, it shakes all the other charges in the world at a *later* time $t' = t + r/c$ (where r is the distance to the other charge), because of the *retarded* waves. But then these other charges react back on the original electron through their *advanced* waves, which will arrive at the time t'', equal to t' *minus* r/c, which is, of course, just t. (They also react back with their retarded waves too, but that just corresponds to the normal "reflected" waves.) The combination of the advanced and retarded waves means that at the instant it is accelerated an oscillating charge feels a force from all the charges that are "going to" absorb its radiated waves. You see what tight knots people have gotten into in trying to get a theory of the electron!

We'll describe now still another kind of theory, to show the kind of things that people think of when they are stuck. This is another modification of the laws of electrodynamics, proposed by Bopp. You realize that once you decide to change the equations of electromagnetism you can start anywhere you want. You can change the force law for an electron, or you can change the Maxwell equations (as we saw in the examples we have described), or you can make a change somewhere else. One possibility is to change the formulas that give the potentials in terms of the charges and currents. One of our formulas has been that the potentials at some point are given by the current density (or charge) at each other point at an earlier time. Using our four-vector notation for the potentials, we write

$$A_\mu(1, t) = \frac{1}{4\pi\epsilon_0 c^2} \int \frac{j_\mu(2, t - r_{12}/c)}{r_{12}} dV_2. \qquad (28.13)$$

Bopp's beautifully simple idea is that: Maybe the trouble is in the $1/r$ factor in the integral. Suppose we were to start out by assuming only that the potential at one point depends on the charge density at any other point as *some* function of the distance between the points, say as $f(r_{12})$. The total potential at point (1) will then be given by the integral of j_μ times this function over all space:

$$A_\mu(1, t) = \int j_\mu(2, t - r_{12}/c) f(r_{12}) dV_2.$$

That's all. No differential equation, nothing else. Well, one more thing. We also ask that the result should be relativistically invariant. So by "distance" we should take the invariant "distance" between two points in space-time. This distance squared (within a sign which doesn't matter) is

$$s_{12}^2 = c^2(t_1 - t_2)^2 - r_{12}^2$$
$$= c^2(t_1 - t_2)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2 - (z_1 - z_2)^2. \qquad (28.14)$$

So, for a relativistically invariant theory, we should take some function of the magnitude of s_{12}, or what is the same thing, some function of s_{12}^2. So Bopp's theory is that

$$A_\mu(1, t_1) = \int j_\mu(2, t_2) F(s_{12}^2) dV_2 \, cdt_2. \qquad (28.15)$$

(The integral must, of course, be over the four-dimensional volume $cdt_2 \, dx_2 \, dy_2 \, dz_2$.)

All that remains is to choose a suitable function for F. We assume only one thing about F—that it is very small except when its argument is near zero—so that a graph of F would be a curve like the one in Fig. 28–4. It is a narrow spike with a finite area centered at $s^2 = 0$, and with a width which we can say is roughly a^2. We can say, crudely, that when we calculate the potential at point (1), only those points (2) produce any appreciable effect if $s_{12}^2 = c^2(t_1 - t_2)^2 - r_{12}^2$ is within $\pm a^2$ of zero. We can indicate this by saying that F is important only for

$$s_{12}^2 = c^2(t_1 - t_2)^2 - r_{12}^2 \approx \pm a^2. \qquad (28.16)$$

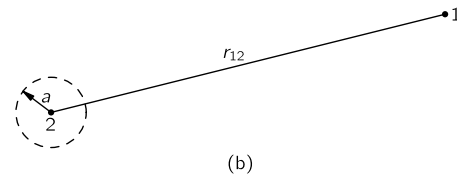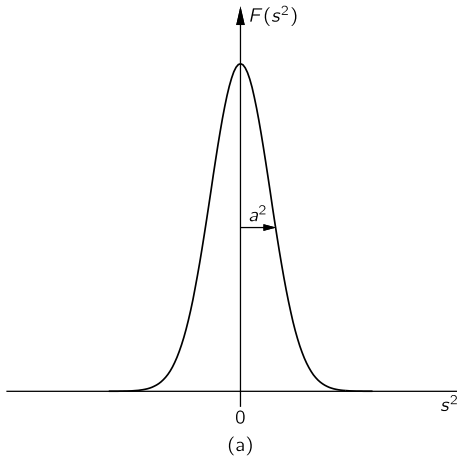You can make it more mathematical if you want to, but that's the idea.

Fig. 28–4.The function $F(s^2)$ used in the nonlocal theory of Bopp.

Now suppose that a is very small in comparison with the size of ordinary objects like motors, generators, and the like so that for normal problems $r_{12} \gg a$. Then Eq. (28.16) says that charges contribute to the integral of Eq. (28.15) only when $t_1 - t_2$ is in the small range

$$c(t_1 - t_2) \approx \sqrt{r_{12}^2 \pm a^2} = r_{12}\sqrt{1 \pm \frac{a^2}{r_{12}^2}}.$$

Since $a^2/r_{12}^2 \ll 1$, the square root can be approximated by $1 \pm a^2/2r_{12}^2$, so

$$t_1 - t_2 = \frac{r_{12}}{c}\left(1 \pm \frac{a^2}{2r_{12}^2}\right) = \frac{r_{12}}{c} \pm \frac{a^2}{2r_{12}c}.$$

What is the significance? This result says that the only *times* t_2 that are important in the integral of A_μ are those which differ from the time t_1, at which we want the potential, by the delay r_{12}/c—with a negligible correction so long as $r_{12} \gg a$. In other words, this theory of Bopp approaches the Maxwell theory—so long as we are far away from any particular charge—in the sense that it gives the retarded wave effects.

We can, in fact, see approximately what the integral of Eq. (28.15) is going to give. If we integrate first over t_2 from $-\infty$ to $+\infty$—keeping r_{12} fixed—then s_{12}^2 is also going to go from $-\infty$ to $+\infty$. The integral will all come from t_2's in a small interval of width $\Delta t_2 = 2 \times a^2/2r_{12}c$, centered at $t_1 - r_{12}/c$. Say that the function $F(s^2)$ has the value K at $s^2 = 0$; then the integral over t_2 gives approximately $Kj_\mu\Delta t_2$, or

$$\frac{Ka^2}{c} \frac{j_\mu}{r_{12}}.$$

We should, of course, take the value of j_μ at $t_2 = t_1 - r_{12}/c$, so that Eq. (28.15) becomes

$$A_\mu(1, t_1) = \frac{Ka^2}{c} \int \frac{j_\mu(2, t_1 - r_{12}/c)}{r_{12}} \, dV_2.$$

If we pick $K = 1/4\pi\epsilon_0 ca^2$, we are right back to the retarded potential solution of Maxwell's equations—including automatically the $1/r$ dependence! And it all came out of the simple proposition that the potential at one point in space-time depends on the current density at all other points in space-time, but with a weighting factor that is some narrow function of the four-dimensional distance between the two points. This theory again predicts a finite electromagnetic mass for the electron, and the energy and mass have the right relation for the relativity theory. They must, because the theory is relativistically invariant from the start, and everything seems to be all right.

There is, however, one fundamental objection to this theory and to all the other theories we have described. All particles we know obey the laws of quantum mechanics, so a quantum-mechanical modification of electrodynamics has to be made. Light behaves like photons. It isn't 100 percent like the Maxwell theory. So the electrodynamic theory has to be changed. We have already mentioned that it might be a waste of time to work so hard to straighten out the classical theory, because it could turn out that in quantum electrodynamics the difficulties will disappear or may be resolved in some other fashion. But the difficulties do not disappear in quantum electrodynamics. That is one of the reasons that people have spent so much effort trying to straighten out the classical difficulties, hoping that if they *could* straighten out the classical difficulty and *then* make the quantum modifications, everything would be straightened out.

The Maxwell theory still has the difficulties after the quantum mechanics modifications are made.

The quantum effects do make some changes—the formula for the mass is modified, and Planck's constant $\hbar$ appears—but the answer still comes out infinite unless you cut off an integration somehow—just as we had to stop the classical integrals at $r = a$. And the answers depend on how you stop the integrals. We cannot, unfortunately, demonstrate for you here that the difficulties are really basically the same, because we have developed so little of the theory of quantum mechanics and even less of quantum electrodynamics. So you must just take our word that the quantized theory of Maxwell's electrodynamics gives an infinite mass for a point electron.

It turns out, however, that nobody has ever succeeded in making a *self-consistent* quantum theory out of *any* of the modified theories. Born and Infeld's ideas have never been satisfactorily made into a quantum theory. The theories with the advanced and retarded waves of Dirac, or of Wheeler and Feynman, have never been made into a satisfactory quantum theory. The theory of Bopp has never been made into a satisfactory quantum theory. So today, there is no known solution to this problem. We do not know how to make a consistent theory—including the quantum mechanics—which does not produce an infinity for the self-energy of an electron, or any point charge. And at the same time, there is no satisfactory theory that describes a non-point charge. It's an unsolved problem.

In case you are deciding to rush off to make a theory in which the action of an electron on itself is completely removed, so that electromagnetic mass is no longer meaningful, and then to make a quantum theory of it, you should be warned that you are certain to be in trouble. There is definite experimental evidence of the existence of electromagnetic inertia—there is evidence that some of the mass of charged particles is electromagnetic in origin.

It used to be said in the older books that since Nature will obviously not present us with two particles—one neutral and the other charged, but otherwise the same—we will never be able to tell how much of the mass is electromagnetic and how much is mechanical. But it turns out that Nature *has* been kind enough to present us with just such objects, so that by comparing the observed mass of the charged one with the observed mass of the neutral one, we can tell whether there is any electromagnetic mass. For example, there are the neutrons and protons. They interact with tremendous forces—the nuclear forces—whose origin is unknown. However, as we have already described, the nuclear forces have one remarkable property. So far as they are concerned, the neutron and proton are exactly the same. The *nuclear* forces between neutron and neutron, neutron and proton, and proton and proton are all identical as far as we can tell. Only the little electromagnetic forces are different; electrically the proton and neutron are as different as night and day. This is just what we wanted. There are two particles, identical from the point of view of the strong interactions, but different electrically. And they have a small difference in mass. The mass difference between the proton and the neutron—expressed as the difference in the rest-energy mc^2 in units of MeV—is about 1.3 MeV, which is about 2.6 times the electron mass. The classical theory would then predict a radius of about $\frac{1}{3}$ to $\frac{1}{2}$ the classical electron radius, or about 10^{-13} cm. Of course, one should really use the quantum theory, but by some strange accident, all the constants—2π's and $\hbar$'s, etc.—come out so that the quantum theory gives roughly the same radius as the classical theory. The only trouble is that the *sign* is wrong! The neutron is *heavier* than the proton.

Nature has also given us several other pairs—or triplets—of particles which appear to be exactly the same except for their electrical charge. They interact with protons and neutrons, through the so-called "strong" interactions of the nuclear forces. In such interactions, the particles of a given kind—say the π-mesons—behave in every way like one object *except* for their electrical charge. In Table 28–1 we give a list of such particles, together with their measured masses. The charged π-mesons—positive or negative—have a mass of 139.6 MeV, but the neutral π-meson is 4.6 MeV lighter. We believe that this mass difference is electromagnetic; it would correspond to a particle radius of 3 to 4×10^{-14} cm. You will see from the table that the mass differences of the other particles are usually of the same general size.

Table 28–1 Particle Masses

Particle	Charge (electronic)	Mass (MeV)	Δm [1] (MeV)
n (neutron)	0	939.5	
p (proton)	+1	938.2	−1.3
π (π-meson)	0	135.0	
	±1	139.6	+4.6
K (K-meson)	0	497.8	
	±1	493.9	−3.9
Σ (sigma)	0	1191.5	
	+1	1189.4	−2.1
	−1	1196.0	+4.5

[1] $\Delta m = $ (mass of charged) − (mass of neutral).

Now the size of these particles can be determined by other methods, for instance by the diameters they appear to have in high-energy collisions. So the electromagnetic mass seems to be in general agreement with electromagnetic theory, if we stop our integrals of the field energy at the same radius obtained by these other methods. That's why we believe that the differences do represent electromagnetic mass.

You are no doubt worried about the different signs of the mass differences in the table. It is easy to see why the charged ones should be heavier than the neutral ones. But what about those pairs like the proton and the neutron, where the measured mass comes out the other way? Well, it turns out that these particles are complicated, and the computation of the electromagnetic mass must be more elaborate for them. For instance, although the neutron has no *net*

charge, it *does* have a charge distribution inside it—it is only the *net* charge that is zero. In fact, we believe that the neutron looks—at least sometimes—like a proton with a negative π-meson in a "cloud" around it, as shown in Fig. 28–5. Although the neutron is "neutral," because its total charge is zero, there are still electromagnetic energies (for example, it has a magnetic moment), so it's not easy to tell the sign of the electromagnetic mass difference without a detailed theory of the internal structure.
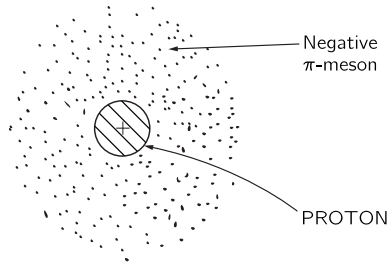

Fig. 28–5.A neutron may exist, at times, as a proton surrounded by a negative π-meson.

We only wish to emphasize here the following points: (1) the electromagnetic theory predicts the existence of an electromagnetic mass, but it also falls on its face in doing so, because it does not produce a consistent theory—and the same is true with the quantum modifications; (2) there is experimental evidence for the existence of electromagnetic mass; and (3) all these masses are roughly the same as the mass of an electron. So we come back again to the original idea of Lorentz—maybe all the mass of an electron is purely electromagnetic, maybe the whole 0.511 MeV is due to electrodynamics. Is it or isn't it? We haven't got a theory, so we cannot say.

We must mention one more piece of information, which is the most annoying. There is another particle in the world called a *muon* which, so far as we can tell, differs in no way whatsoever from an electron except for its mass. It acts in every way like an electron: it interacts with neutrinos and with the electromagnetic field, and it has no nuclear forces. It does nothing different from what an electron does—at least, nothing which cannot be understood as merely a consequence of its higher mass (206.77 times the electron mass). Therefore, whenever someone finally gets the explanation of the mass of an electron, he will then have the puzzle of where a muon gets its mass. Why? Because whatever the electron does, the muon does the same—so the mass ought to come out the same. There are those who believe faithfully in the idea that the muon and the electron are the same particle and that, in the final theory of the mass, the formula for the mass will be a quadratic equation with two roots—one for each particle. There are also those who propose it will be a transcendental equation with an infinite number of roots, and who are engaged in guessing what the masses of the other particles in the series must be, and why these particles haven't been discovered yet.

28–6The nuclear force field

We would like to make some further remarks about the part of the mass of nuclear particles that is not electromagnetic. Where does this other large fraction come from? There are other forces besides electrodynamics—like nuclear forces—that have their own field theories, although no one knows whether the current theories are right. These theories also predict a field energy which gives the nuclear particles a mass term analogous to electromagnetic mass; we could call it the "π-mesic-field-mass." It is presumably very large, because the forces are great, and it is the possible origin of the mass of the heavy particles. But the meson field theories are still in a most rudimentary state. Even with the well-developed theory of electromagnetism, we found it impossible to get beyond first base in explaining the electron mass. With the theory of the mesons, we strike out.

We may take a moment to outline the theory of the mesons, because of its interesting connection with electrodynamics. In electrodynamics, the field can be described in terms of a four-potential that satisfies the equation

$$\Box^2 A_\mu = \text{sources}.$$

Now we have seen that pieces of the field can be radiated away so that they exist separated from the sources. These are the photons of light, and they are described by a differential equation without sources:

$$\Box^2 A_\mu = 0.$$

People have argued that the field of nuclear forces ought also to have its own "photons"—they would presumably be the π-mesons—and that they should be described by an analogous differential equation. (Because of the weakness of the human brain, we can't think of something really new; so we argue by analogy with what we know.) So the meson equation might be

$$\Box^2 \phi = 0,$$

where ϕ could be a different four-vector or perhaps a scalar. It turns out that the pion has no polarization, so ϕ should be a scalar. With the simple equation $\Box^2 \phi = 0$, the meson field would vary with distance from a source as $1/r^2$, just as the electric field does. But we know that nuclear forces have much shorter distances of action, so the simple equation won't work. There is one way we can change things without disrupting the relativistic invariance: we can add or subtract from the d'Alembertian a constant, times ϕ. So Yukawa suggested that the free quanta of the nuclear force field might obey the equation

$$-\Box^2 \phi - \mu^2 \phi = 0, \tag{28.17}$$

where μ^2 is a constant—that is, an invariant scalar. (Since $\Box^2$ is a scalar differential operator in four dimensions, its invariance is unchanged if we add another scalar to it.)

Let's see what Eq. (28.17) gives for the nuclear force when things are not changing with time. We want a spherically symmetric solution of

$$\nabla^2\phi - \mu^2\phi = 0$$

around some point source at, say, the origin. If ϕ depends only on r, we know that

$$\nabla^2\phi = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\phi).$$

So we have the equation

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}(r\phi) - \mu^2\phi = 0$$

or

$$\frac{\partial^2}{\partial r^2}(r\phi) = \mu^2(r\phi).$$

Thinking of $(r\phi)$ as our dependent variable, this is an equation we have seen many times. Its solution is

$$r\phi = Ke^{\pm\mu r}.$$

Clearly, ϕ cannot become infinite for large r, so the $+$ sign in the exponent is ruled out. The solution is

$$\phi = K\frac{e^{-\mu r}}{r}. \tag{28.18}$$

This function is called the *Yukawa potential*. For an attractive force, K is a negative number whose magnitude must be adjusted to fit the experimentally observed strength of the forces.

The Yukawa potential of the nuclear forces dies off more rapidly than $1/r$ by the exponential factor. The potential—and therefore the force—falls to zero much more rapidly than $1/r$ for distances beyond $1/\mu$, as shown in Fig. 28–6. The "range" of nuclear forces is much less than the "range" of electrostatic forces. It is found experimentally that the nuclear forces do not extend beyond about 10^{-13} cm, so $\mu \approx 10^{15}$ m^{-1}.
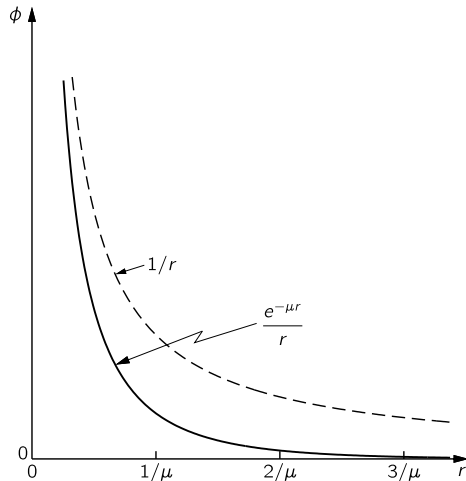

Fig. 28–6.The Yukawa potential $e^{-\mu r}/r$, compared with the Coulomb potential $1/r$.

Finally, let's look at the free-wave solution of Eq. (28.17). If we substitute

$$\phi = \phi_0 e^{i(\omega t - kz)}$$

into Eq. (28.17), we get that

$$\frac{\omega^2}{c^2} - k^2 - \mu^2 = 0.$$

Relating frequency to energy and wave number to momentum, as we did at the end of Chapter 34 of Vol. I, we get that

$$\frac{E^2}{c^2} - p^2 = \mu^2\hbar^2,$$

which says that the Yukawa "photon" has a mass equal to $\mu\hbar/c$. If we use for μ the estimate 10^{15} m^{-1}, which gives the observed range of the nuclear forces, the mass comes out to 3×10^{-25} g, or 170 MeV, which is roughly the observed mass of the π-meson. So, by an analogy with electrodynamics, we would say that the π-meson is the "photon" of the nuclear force field. But now we have pushed the ideas of electrodynamics into regions where they may not really be valid—we have gone beyond electrodynamics to the problem of the nuclear forces.

1.   We are using the notation: $\dot{x} = dx/dt$, $\ddot{x} = d^2x/dt^2$, $\dddot{x} = d^3x/dt^3$, etc.