**Physics 221B**
**Academic Year 2021–22**
**Notes 51**
**Hole Theory and Second Quantization of the Dirac Equation†**

## 1. Introduction

In these notes we provide the ultimate resolution of the difficulty of the negative energy solutions, which appear both in the Klein-Gordon equation and in the Dirac equation. The results are dramatic on several accounts. First, they lead to the physical prediction of the existence of antimatter, something that was not suspected when Dirac began his work on relativistic wave equations. This prediction came only a short time before positrons were found experimentally, in one of the most brilliant theoretical successes in the history of physics. Second, the ultimate resolution of the difficulties of the negative energy solutions shows that in relativistic quantum mechanics, it is impossible to speak of a system consisting of a single particle. Instead, particles and antiparticles are present everywhere and in all circumstances, at least virtually, with observable consequences. And finally, the proper framework for understanding relativistic quantum mechanics appears, and it is not some Schrödinger-like equation for one or some fixed number of particles, rather it is quantum field theory.

We summarize where we are in our exploration of the Dirac equation. Although we have as yet no interpretation for the negative energy solutions, and although we do not see them physically, we have seen that we cannot simply declare them to be nonphysical. Two reasons were given in Sec. 50.11: First, the negative energy solutions of bound state problems do not span the same subspace of the Hilbert space as negative energy solutions of the free particle, so there is no consistent way to define a nonphysical subspace; and second, the negative energy solutions seem to be necessary to get the Zitterbewegung, which explains the Darwin term in atoms.

Here is another reason, which involves the fact that the positive energy solutions by themselves do not form a complete set. In second order perturbation theory, it is necessary to sum over a set of intermediate states, as shown by Eq. (43.17). This sum comes from a resolution of the identity, inserted into the terms of the Dyson series, so it must be taken over a complete set of states. But in the case of the Dirac equation, should the sum include the negative energy solutions? If they are nonphysical, it seems we should not.

Let us take an example. If we use the nonrelativistic theory to calculate the cross section for the scattering of a photon by a free electron, we obtain the Thomson formula, Eq. (43.57). See also

---

Prob. 43.2. The Thomson formula comes from the seagull diagram only, since the other diagrams are negligible in comparison. The seagull diagram in turn comes from the second order Hamiltonian $K_2$ (see Notes 43 for the notation), taken in first order perturbation theory, so there is no sum over intermediate states.

Now if we do the same calculation using the Dirac equation, there is no second order Hamiltonian, and the first nonvanishing contributions to the scattering amplitude come from the first order Hamiltonian, taken in second order perturbation theory. Thus there is a sum over intermediate states. But if we exclude the negative energy solutions in this sum, it turns out that we throw away precisely the terms that in the nonrelativistic limit give the seagull graph. Thus, the cross section calculated in this limit does not agree with the Thomson formula, which we know is correct. We know this because it is verified experimentally, and, in any case, it can be derived from purely classical electromagnetic theory. We see that the negative energy solutions are needed in sums over intermediate states in order to get physically correct results.

## 2. Hole Theory

So we cannot declare the negative energy states to be nonphysical. But there are also problems if we assume that they are real. For example, a real electron interacts with the electromagnetic field, and this is an interaction that we cannot turn off. Consequently an electron in a higher energy state can emit a photon and drop into a lower energy state. This cannot happen with a free electron, because it is impossible to satisfy energy and momentum conservation when emitting a photon and dropping from one free particle state to another. But it can happen in bound systems such as the hydrogen atom, in which the nucleus can absorb any extra momentum needed to satisfy overall energy and momentum conservation. For example, the lowest positive energy eigenstate of the Dirac hydrogen atom is the usual ground state, approximately 13.6 eV below the rest-mass-energy of the electron, $mc^2 \approx 511$ KeV. But the Dirac hydrogen atom also has a continuous spectrum of negative energy states, lying in the range $E \leq -mc^2$. Why, then, cannot a hydrogen atom in the usual ground state emit a photon, and drop into one of the negative energy states? This photon would have an energy close to $2mc^2$ or higher. Moreover, once one negative energy state has been reached, the system could emit another photon, dropping into an even more negative energy state. This process, it would seem, would continue forever, as the energy of the electron went to $-\infty$, and an infinite amount of energy in the form of photons was released. One can calculate the life time of the (usual) ground state of hydrogen according to this mechanism, and it turns out to be very short. Obviously, we do not see hydrogen atoms self-destructing in this manner and emitting an infinite amount of energy in the form of photons.

In 1930, Dirac suggested that the reason we do not see such transitions is that the negative energy states are already filled. Since electrons are fermions no more than one can occupy a given state, so transitions to negative energy states would be forbidden by the Pauli exclusion principle. These negative energy electrons constitute what is called the "Dirac sea." According to this hypothesis,

space is filled with the sea of negative energy electrons, which produce a nominally infinite density of both energy and negative charge. So we have to imagine some mechanism that would prevent the sea from having observable effects. Perhaps the infinite electric field created would cancel out from symmetry, since there would be no preferred direction for it to point in. And never mind the gravitational effects of the infinite mass density.
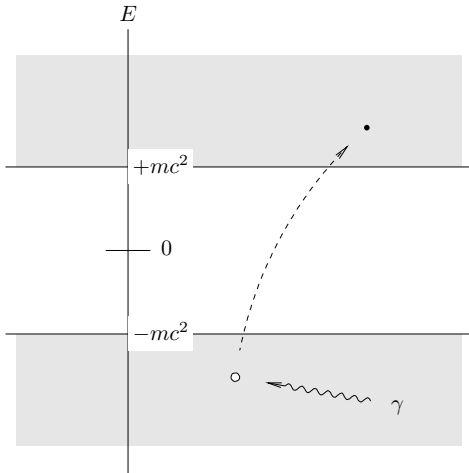


**Fig. 1.** A diagram suggestive of the physics of pair creation, according to hole theory. A photon is absorbed by a negative energy electron, which is promoted into a positive energy state, leaving behind a hole.
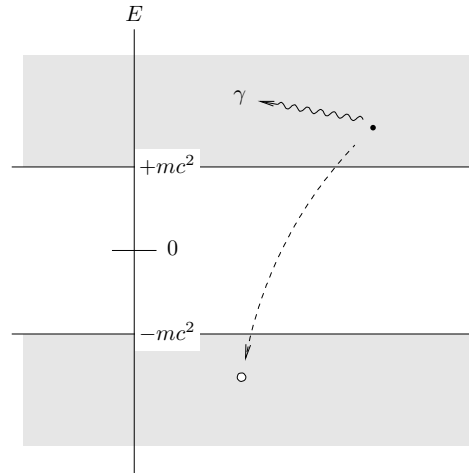
**Fig. 2.** Pair annihilation according to hole theory. A positive energy electron makes a transition to a negative-energy state, filling up the hole and emitting a photon in the process.

Pushing these problems aside, Dirac noted that his hypothesis leads to predictions of new physics. For example, although a positive energy electron could not make a transition to one of the (already filled) negative energy states, the negative energy electrons in the sea also interact with the electromagnetic field, so it is possible for one of them to absorb a photon and get lifted into a positive energy state. The photon would have to have an energy $> 2mc^2$ for this to happen. But if it did, we would see the disappearance of the photon, and the appearance of a (regular) positive energy electron that was not there before. We would see something else, too, a "hole" in the negative energy sea, that is, the absence of a negative energy electron. If the electric field of the (normally filled) negative energy sea somehow cancels, then, when we remove a negative energy electron, its electric field no longer contributes to the sum, and now the sum of the field from the negative energy sea would be the negative of the field contributed by the negative energy electron that was removed. That is, we would see the electric field of a particle of positive charge. In fact, the absence of the negative energy electron would behave overall as a particle with the opposite charge, energy, momentum and spin of the negative energy electron, that is, it would have positive energy as well as positive charge. Thus Dirac arrived at the prediction that a photon can disappear and be replaced by an ordinary electron, plus a new particle with the same mass as the electron but

opposite charge. This process is illustrated in Fig. 1.

Since such a particle had never been observed at the time of Dirac's 1930 paper, he tried to imagine that it was the proton. Unfortunately, the proton does not have the same mass as the electron, so the whole idea seemed shaky. Pauli was especially skeptical, and criticized Dirac's theory. In 1932, however, Anderson discovered the positron in cosmic ray tracks, a particle of the same mass as the electron but with the opposite charge, thereby vindicating Dirac. The process described is now called "pair creation" (the creation of a positron-electron pair). A photon cannot create an electron-positron pair in the vacuum, because it is impossible to conserve energy and momentum in the reaction $\gamma \to e^+ + e^-$. But if a nucleus is nearby to absorb the extra momentum, then it can happen. In Anderson's experiment, layers of lead (a high $Z$ material) were separated by gaps. A high energy photon, created by the interaction of a cosmic ray muon in one of the lead plates, passed down to a lower plate where it produced an electron-positron pair, which emerged into the next lower gap.

Dirac's "hole theory" also makes other predictions. If a hole has been created in the negative energy sea, then it is no longer impossible for a (regular) positive energy electron to make a transition into that particular negative energy state, emitting a photon in the process. Thus we have the prediction that an ordinary electron and a hole, which is manifested as a positron, may simultaneously disappear, with the appearance of a photon. That is, Dirac's hypothesis of the negative energy sea leads to the prediction of the reaction, $e^+ + e^- \to \gamma$. This reaction cannot occur in vacuum because of energy and momentum conservation (just look at the reaction in the rest frame of the $e^+$-$e^-$ system and you will see), but an annihilation into two photons, $e^+ + e^- \to \gamma + \gamma$, is possible, and is observed to happen. Thus, Dirac's hole theory predicts that a hole (otherwise known as a positron) and an electron can annihilate one another, leaving behind only photons. It is a direct conversion of matter into energy, as promised by the Einstein relation $E = mc^2$. This process is illustrated in Fig. 2.

### 3. Successes and Shortcomings of Hole Theory

The Dirac equation, combined with the hypothesis of the negative energy sea, constitutes "hole theory." It not only solves the problem of the negative energy solutions of the Dirac equation, but also forms the basis of a theory that can be used for many sophisticated calculations in quantum electrodynamics. Indeed, it was so used for many years, although it is definitely out of fashion nowadays, and in these notes we shall give it no more than a brief treatment. For us its main importance is as a crucial stepping stone to the modern point of view, which is based on quantum field theory.

In spite of its successes, hole theory also has a number of shortcomings. First, although it describes very well many processes in quantum electrodynamics involving the creation and annihilation of electrons and positrons, it cannot easily describe other processes in which electrons or positrons

are created and destroyed, such as beta decay. Consider, for example, the beta decay of a neutron,

$$n \to p^+ + e^- + \bar{\nu}, \tag{1}$$

where $p^+$ is a proton, $e^-$ an electron, and $\bar{\nu}$ an antineutrino. This reaction is a manifestation of the weak interactions, not the electromagnetic. If the electron that appears in beta decay is to be interpreted as one that has been promoted out of the negative energy sea, then we would expect to see a positron left behind. Instead, the positive charge is carried by the proton.

In addition, there are a number of theoretical, interpretational and esthetic difficulties with hole theory. For example, the hypothesis of the negative energy sea makes crucial use of the fact that electrons are fermions, so that the exclusion principle can prevent transitions from positive energy states to negative energy states, at least under normal circumstances in which the negative energy sea is filled. Although this resolves the difficulty of the negative energy solutions for the Dirac equation, it does not help with the negative energy solutions of the Klein-Gordon equation, which describes bosons. (The Klein-Gordon wave equation must describe bosons because the wave function is a scalar, which must represent a particle of spin 0.)

There is also the obvious problem of how to understand why the negative energy sea has no observable consequences, under the usual conditions in which it is filled. The problem of the infinite electric field has already been alluded to, and that of the gravitational effects of the infinite mass density has been dismissed without real justification. (Of course there are similar problems with the zero point energy of quantum fields.)

Finally, there is the fact that the Dirac equation possesses a symmetry between positive and negative energy solutions, or, when coupled with hole theory, between electrons and positrons. This symmetry is called *charge conjugation*, and it is one of the fundamental symmetries of nature. We have not considered charge conjugation so far in these notes because it is best formulated in terms of quantum field theory. But while the Dirac equation itself is invariant under charge conjugation, hole theory introduces an asymmetry between positrons and electrons by postulating a negative energy sea of electrons. One can argue that this is justified by the fact that nature evidently has a preference for electrons, since they are everywhere while positrons are few and far between. Thus the discussion turns to the question of the asymmetry in the universe between matter and antimatter, which was touched upon in Notes 22. Rather than go in this direction, let us just say that to many physicists hole theory seems unsymmetrical and unappealing on esthetic grounds.

For us probably the most important lesson of hole theory is that with it the Dirac equation becomes secretly a many-particle theory. Although we started out thinking that we were describing a single electron, with the addition of the negative energy sea we always have an infinite number of particles.

## 4. Second Quantization of the Dirac Equation

There is one formalism that we have seen so far that allows us to deal with many-particle systems

in which particles can be created or destroyed. That is the quantum theory of the electromagnetic field, described in Notes 41, in which the particles in question are photons. That theory was obtained by quantizing the classical electromagnetic field, that is, reinterpreting the $c$-number fields $\mathbf{E}$, $\mathbf{B}$, $\mathbf{A}$, etc, as fields of operators. This suggests that we might be able to deal with the many-particle aspects of Dirac's hole theory, including processes of creation and annihilation of electrons and positrons, by reinterpreting the Dirac wave function $\psi$, which up to this point has been a $c$-number field, as a quantum field. The process of passing from a $c$-number field to a quantum field is called "quantization," but since the Dirac wave function $\psi$ already describes the quantum mechanics of a single particle, when this process is applied to the Dirac wave function, it is called "second quantization." In this process the Dirac equation, as we have dealt with it up to this point, may be called the "first quantized" version of the theory, that is, the quantum theory of a single particle. In the process of second quantizing the Dirac equation, the first quantized Dirac theory will be treated as a "classical" theory, insofar as the process of quantization is concerned. That is, the "classical" theory concerns a $c$-number field that satisfies certain equations of motion. When used in this context, we will put the word "classical" in quotes, to acknowledge that the Dirac equation actually describes a single-particle quantum system.

The idea of second quantizing the Dirac wave field did not come out of the blue. In 1928 Wigner and Jordan explored the question of what would happen if the usual Schrödinger wave function were reinterpreted as a quantum field. This was in the days in which the interpretation of the wave function was not settled, and different possibilities were being tried. Wigner and Jordan were slightly disappointed in their results, because they found that second quantizing the Schrödinger theory led to a theory physically equivalent to the first quantized version, but one in which the symmetry or antisymmetry of multiparticle wave functions of identical particles was handled in a natural manner. To achieve this result in the case of fermions, it turned out to be necessary to modify the process of Dirac quantization, which we described in Notes 41 in application to the electromagnetic field. Nowadays the second quantized approach of Wigner and Jordan is the preferred method for handling many body problems in atomic, nuclear and condensed matter physics, often for nonrelativistic processes in which massive particles are not actually created or destroyed. This is mainly because of the convenience with which the second quantized formalism handles the requirements of the symmetrization postulate, and because of the modes of thinking and physical insight that follow from the use of quantum field theory.

To second quantize the Dirac equation, we will follow the same general procedure that we used in quantizing the electromagnetic field. First we seek a classical Hamiltonian and a definition of $q$'s and $p$'s such that Hamilton's classical equations of motion are equivalent to the known equations of motion for the classical field. In the case of the electromagnetic field, the classical equations of motion are Maxwell's equations, while for the Dirac equation, the "classical" equations of motion are the Dirac equation itself. Next we reinterpret the $q$'s and $p$'s as operators satisfying the canonical commutation relations, and acting on a certain ket space. In the case of the Dirac equation, we

will see that this step requires modification, in comparison to what we did with the electromagnetic field, in order to satisfy the Fermi-Dirac statistics that electrons must satisfy. Finally, the ket space is interpreted as a space of multiparticle states, with a variable number of particles, and operators emerge that create and destroy these particles, thereby changing the particle number.

We now proceed to the second quantization of the Dirac field in detail.

## 5. The "Classical" Hamiltonian for the Dirac Field

In the remainder of these notes we use natural units, in which $\hbar = c = 1$. In these units, the unit of electric charge $e$ is not 1, rather $e^2 = \alpha \approx 1/137$.

In Notes 40 we guessed that the Hamiltonian for the classical electromagnetic field was the energy of the system. In the case of the free field, this energy is

$$H = \frac{1}{8\pi} \int d^3\mathbf{x}\,(E^2 + B^2). \tag{2}$$

The Hamiltonian is interpreted as a functional of the fields $\mathbf{E}$ and $\mathbf{B}$, or, equivalently, of the potentials $\Phi$ and $\mathbf{A}$ and their space-time derivatives.

For the first quantized Dirac equation, the "classical" Hamiltonian must be a functional of the field $\psi(\mathbf{x})$, that has an interpretation as the energy of the field. For simplicity we deal only with the case of a free particle. The most obvious candidate for such a functional is

$$H[\psi(\mathbf{x})] = \int d^3\mathbf{x}\, \psi^\dagger(\mathbf{x})(-i\boldsymbol{\alpha} \cdot \nabla + m\beta)\psi(\mathbf{x}), \tag{3}$$

where the operator sandwiched between $\psi^\dagger$ and $\psi$ is the free particle Hamiltonian operator of the first quantized theory. Notice that there are two Hamiltonians here, the Hamiltonian operator of the first quantized theory, which acts on wave functions of a single electron, and the "classical" field Hamiltonian, denoted by $H$ in Eq. (3).

To distinguish these, let us write $h$ for the free particle Hamiltonian operator of the first quantized theory,

$$h = -i\boldsymbol{\alpha} \cdot \nabla + m\beta = \boldsymbol{\alpha} \cdot \mathbf{p} + m\beta, \tag{4}$$

so that

$$H[\psi(\mathbf{x})] = \langle \psi|h|\psi \rangle. \tag{5}$$

The "classical" Hamiltonian of the Dirac field is the expectation value of the energy in the first quantized system with respect to the wave function $\psi(\mathbf{x})$, when the latter is normalized. However, the "classical" field Hamiltonian $H$ is defined for any wave function, not just normalized ones.

## 6. Classical Lagrangians in Mechanics and Field Theory

Here as in Notes 40 we have simply guessed the classical Hamiltonian, but in both cases it can be derived in a systematic manner from the classical Lagrangian. We will now summarize how this

is done. We begin by reviewing some aspects of Lagrangian mechanics. A more complete treatment is given in Appendix B.

For a mechanical system with coordinates $q_i$, $i = 1, 2, \ldots$, the *Lagrangian* is a function of the $q$'s and their time derivatives,

$$L = L(q, \dot{q}), \tag{6}$$

where the bare symbol $q$ stands for all the coordinates $q_i$, $i = 1, 2, \ldots$. Given the Lagrangian, the *action* is the integral of the Lagrangian over time,

$$S[q(t)] = \int dt\, L(q, \dot{q}), \tag{7}$$

which, as indicated, is considered a functional of the path $q(t)$. The path $q(t)$ used in the action functional $S[q(t)]$ need not be physical, but according to *Hamilton's principle*, the path is physical if and only if $\delta S = 0$, where $\delta S$ is the variation in the action caused by a variation $\delta q(t)$ in the path. The condition $\delta S = 0$ is equivalent to the *Euler-Lagrange equations*,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i}, \tag{8}$$

as explained in detail in Appendix B.

Consider now a classical field $\phi_a(\mathbf{x})$, with components indicated by $a = 1, 2, \ldots$. The *Lagrangian density* is a function of the fields and their space-time derivatives,

$$\mathcal{L} = \mathcal{L}(\phi, \partial_\mu \phi), \tag{9}$$

where the bare symbol $\phi$ stands for all the components, $\phi_a$, $a = 1, 2, \ldots$. Then the Lagrangian is defined as the spatial integral of the Lagrangian density,

$$L = \int d^3\mathbf{x}\, \mathcal{L}(\phi, \partial_\mu \phi), \tag{10}$$

while the action (as in mechanics) is defined as the integral of the Lagrangian over time,

$$S[\phi(x)] = \int dt\, L = \int d^4x\, \mathcal{L}(\phi, \partial_\mu \phi). \tag{11}$$

In field theory the action is an integral of the Lagrangian density over a 4-dimensional region of space-time. As indicated, it is considered a functional of the field $\phi(x)$, where $x$ stands for $(\mathbf{x}, t)$. The field $\phi(x)$ used in the action need not be physical, but according to Hamilton's principle it is physical if and only if $\delta S = 0$, where $\delta S$ is the change in the action engendered by a change $\delta\phi$ in the field. The condition $\delta S = 0$ is equivalent to the Euler-Lagrange equations in field theory,

$$\frac{\partial}{\partial x^\mu}\left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)}\right) = \frac{\partial \mathcal{L}}{\partial \phi_a}, \qquad a = 1, 2, \ldots. \tag{12}$$

In relativistic theories we require equations of motion that are covariant, that is, they must have the same form in all Lorentz frames. This goal is achieved if the action is a Lorentz scalar.

Since the 4-dimensional volume element $d^4 x$ is a Lorentz scalar, the action will be a scalar if the Lagrangian density is a scalar, so normally we require this. The requirement that a physically acceptable Lagrangian density be a Lorentz scalar imposes severe restrictions on the possible forms of allowed Lagrangian densities.

## 7. The Electromagnetic Lagrangian

For example, in the case of the free electromagnetic field, the simplest Lorentz scalars that can be constructed out of the 4-vector potential $A^\mu$ and its first derivatives are $A^\mu A_\mu$ and $F^{\mu\nu} F_{\nu\mu}$, where only the latter is gauge-invariant. See Sec. E.22 for a summary of the covariant formulation of electromagnetism. It turns out that the Lagrangian density $F^{\mu\nu} F_{\nu\mu}$ alone gives Maxwell's equations in vacuum, while the term $A^\mu A_\mu$, if included, gives the photon a finite mass (and it breaks gauge invariance). We normally assume that photons are massless, so we keep only the term $F^{\mu\nu} F_{\nu\mu}$ for the free field. The experimental upper bound on the photon mass is very small, currently about $1 \times 10^{-18}$ eV.

With a conventional normalization and choice of sign, the Lagrangian density for the free electromagnetic field is

$$\mathcal{L}_{\text{em}} = \frac{1}{16\pi} F^{\mu\nu} F_{\nu\mu} = \frac{1}{8\pi} (E^2 - B^2), \tag{13}$$

where we write the the Lagrangian density both in covariant form and in $3 + 1$-form, the latter following from Eqs. (E.90) and (E.92).

We now show that the Lagrangian density (13) gives Maxwell's equations in vacuum, working with the $3+1$-form. It is easier to use the condition $\delta S = 0$ directly, rather than the Euler-Lagrange equations (12). It is understood that **E** and **B** are given in terms of the potentials $\Phi$ and **A** by the usual expressions (here with $c = 1$),

$$\mathbf{E} = -\nabla\Phi - \frac{\partial \mathbf{A}}{\partial t}, \qquad \mathbf{B} = \nabla\times\mathbf{A}, \tag{14}$$

and that $\mathcal{L}_{\text{em}}$ is a function of $\Phi$ and **A** and their space-time derivatives. Thus, $\delta S$ is the variation in the action engendered by variations $\delta\Phi$ and $\delta\mathbf{A}$ in the potentials. To calculate $\delta S$ we write

$$\delta S = \frac{1}{8\pi} \int d^3\mathbf{x}\, dt\, \delta(E^2 - B^2) = \frac{1}{4\pi} \int d^3\mathbf{x}\, dt\, \left(\mathbf{E} \cdot \delta\mathbf{E} - \mathbf{B} \cdot \delta\mathbf{B}\right)$$
$$= \frac{1}{4\pi} \int d^3\mathbf{x}\, dt \left[\mathbf{E} \cdot \left(-\nabla\delta\Phi - \frac{\partial\delta\mathbf{A}}{\partial t}\right) - \mathbf{B} \cdot (\nabla\times\delta\mathbf{A})\right]. \tag{15}$$

Next we integrate by parts, either in space or in time, in order to remove the derivative operators acting on $\delta\Phi$ and $\delta\mathbf{A}$. We use the identities

$$\mathbf{E} \cdot \nabla\delta\Phi = \nabla \cdot (\mathbf{E}\delta\Phi) - (\nabla \cdot \mathbf{E})\delta\Phi,$$
$$\mathbf{E} \cdot \frac{\partial\delta\mathbf{A}}{\partial t} = \frac{\partial}{\partial t}(\mathbf{E} \cdot \delta\mathbf{A}) - \frac{\partial\mathbf{E}}{\partial t} \cdot \delta\mathbf{A}, \tag{16}$$
$$\mathbf{B} \cdot (\nabla\times\delta\mathbf{A}) = \nabla \cdot (\mathbf{B}\times\delta\mathbf{A}) + (\nabla\times\mathbf{B}) \cdot \delta\mathbf{A}.$$

We substitute these into Eq. (15), whereupon the exact time derivatives and exact divergences can be integrated, giving boundary terms which we drop. See Sec. B.7 for the vanishing of boundary terms in classical mechanics; the logic used in classical field theory is similar. Altogether we obtain

$$\delta S = \frac{1}{4\pi} \int d^3\mathbf{x}\, dt \left[ (\nabla \cdot \mathbf{E})\delta\Phi + \left( \frac{\partial \mathbf{E}}{\partial t} - \nabla\times\mathbf{B} \right) \cdot \delta\mathbf{A} \right]. \tag{17}$$

Demanding that $\delta S = 0$ for all variations $\delta\Phi$ and $\delta\mathbf{A}$ gives Maxwell's equations in vacuum,

$$\nabla \cdot \mathbf{E} = 0, \qquad \nabla\times\mathbf{B} = \frac{\partial \mathbf{E}}{\partial t}. \tag{18}$$

To incorporate interactions between the electromagnetic field and matter, described by the 4-current $J^\mu$, we must add an interaction Lagrangian. This must be a Lorentz scalar composed of the fields $A^\mu$ and $F^{\mu\nu}$ and the current $J^\mu$. The simplest such scalar is $J^\mu A_\mu$, and the interaction Lagrangian itself is

$$\mathcal{L}_{\text{int}} = -J^\mu A_\mu = -\rho\Phi + \mathbf{A} \cdot \mathbf{J}, \tag{19}$$

which we write in covariant form and $3+1$-form. Using this as an interaction Lagrangian amounts to the minimal coupling prescription, which is discussed in a different context in Sec. B.13.

The sign on $\mathcal{L}_{\text{int}}$ was chosen so that $\mathcal{L}_{\text{em}} + \mathcal{L}_{\text{int}}$ will give Maxwell's equations including the sources contained in $J^\mu$. This follows easily by adding the term

$$\delta \int d^3\mathbf{x}\, dt\, \mathcal{L}_{\text{int}} = \int d^3\mathbf{x}\, dt\, (-\rho\, \delta\Phi + \mathbf{J} \cdot \delta\mathbf{A}) \tag{20}$$

to $\delta S$, given in Eq. (17). We do not vary $\rho$ or $\mathbf{J}$ in this equation because they are fixed sources. The equations of motion become

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad \nabla\times\mathbf{B} = 4\pi\mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}, \tag{21}$$

Maxwell's equations with sources.

### 8. The Dirac Lagrangian

Similarly, to find the Lagrangian density for the first quantized, "classical" Dirac equation, we seek Lorentz scalars that can be constructed out of the field $\psi$ and its first space-time derivatives. For simplicity we work with the case of the free particle. The simplest such scalars are $\bar{\psi}\psi$ and $\bar{\psi}\gamma^\mu\partial_\mu\psi$ (see Sec. 48.16 for a discussion of the bilinear covariants of the Dirac field). The scalar $(\partial_\mu\bar{\psi})\gamma^\mu\psi$ is not independent, since it can be obtained from $\bar{\psi}\gamma^\mu\partial_\mu\psi$ by integration by parts, that is, with the help of the identity,

$$(\partial_\mu\bar{\psi})\gamma^\mu\psi = \partial_\mu(\bar{\psi}\gamma^\mu\psi) - \bar{\psi}\gamma^\mu\partial_\mu\psi. \tag{22}$$

The exact divergence (the first term on the right) produces only boundary terms when integrated over the space-time region, which we discard. By trying a Lagrangian density for the Dirac field

that is a linear combination of $\bar{\psi}\psi$ and $\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi$ and adjusting the coefficients to make the equations of motion come out right, we find the Dirac Lagrangian density,

$$\mathcal{L}_D = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi = \bar{\psi}(\not{p} - m)\psi, \tag{23}$$

with a conventional normalization and sign.

We now check that this Lagrangian density gives the right equations of motion. This time we use the Euler-Lagrange equations (12). Since the field $\psi$ is complex, we must vary both its real and imaginary parts. Equivalently, we can take $\psi$ and $\psi^{\dagger}$, or $\psi$ and $\bar{\psi}$, as independent, complex fields. These fields (or rather their components) are identified with the fields $\phi_a$ that appear in Eq. (12). First we apply Eq. (12), with $\phi_a$ identified with the components of $\bar{\psi}$. We find

$$\frac{\partial \mathcal{L}_D}{\partial \bar{\psi}} = (i\gamma^{\mu}\partial_{\mu} - m)\psi, \qquad \frac{\partial \mathcal{L}_D}{\partial(\partial_{\mu}\bar{\psi})} = 0. \tag{24}$$

Thus the equations of motion are

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = (\not{p} - m)\psi = 0, \tag{25}$$

which is the free particle Dirac equation. Now varying with respect to $\psi$, we find

$$\frac{\partial \mathcal{L}_D}{\partial \psi} = -m\bar{\psi}, \qquad \frac{\partial \mathcal{L}_D}{\partial(\partial_{\mu}\psi)} = i\bar{\psi}\gamma^{\mu}, \tag{26}$$

so that the Euler-Lagrange equations are

$$i(\partial_{\mu}\bar{\psi})\gamma^{\mu} = -m\bar{\psi}. \tag{27}$$

This is also the Dirac equation, in slightly disguised form. To show this, we take the Hermitian conjugate, multiply by $\gamma^0$ and use Eq. (48.79). Both the Euler-Lagrange equation in $\psi$ and that in $\bar{\psi}$ give the Dirac equation, confirming that $\mathcal{L}_D$ is the correct Lagrangian density.

### 9. The Classical Field Hamiltonian

We now explain the Hamiltonian in classical field theory, continuing with the general notation introduced in Sec. 6. We begin by reviewing Hamiltonians in classical mechanics.

In classical mechanics, the momentum $p_i$ conjugate to $q_i$ is defined by

$$p_i = \frac{\partial L(q, \dot{q})}{\partial \dot{q}_i}, \tag{28}$$

and the Hamiltonian is defined by

$$H = \sum_i p_i \dot{q}_i - L. \tag{29}$$

See Appendix B for a more detailed discussion of these subjects.

Now let $\mathcal{L}(\phi, \partial_\mu \phi)$ be a Lagrangian density in classical field theory. Then the momentum conjugate to $\phi_a$ is defined by

$$\pi_a = \frac{\partial \mathcal{L}}{\partial \dot{\phi}_a}, \tag{30}$$

where $\dot{\phi}_a$ means $\partial \phi_a / \partial t$. Notice that $\pi_a$ is a field itself, just like $\phi_a$. Then the *Hamiltonian density* is defined by

$$\mathcal{H} = \sum_a \pi_a \dot{\phi}_a - \mathcal{L}. \tag{31}$$

Finally, the Hamiltonian is defined as the spatial integral of the Hamiltonian density,

$$H = \int d^3\mathbf{x} \, \mathcal{H}. \tag{32}$$

In physical applications, $H$ is usually the energy of the system.

## 10. The "Classical" Hamiltonian for the Dirac Field, Revisited

In Sec. 5 we guessed the "classical" Hamiltonian for the first quantized Dirac equation. We will now derive it from the Lagrangian density $\mathcal{L}_D$, using the formalism of Sec. 9.

Since the Dirac Lagrangian density (23) depends on both $\psi$ and $\bar{\psi}$, which we are regarding as independent fields, there are two momenta, which we denote by $\pi$ and $\bar{\pi}$. To compute these we write $\mathcal{L}_D$ in $3 + 1$-form, in order to bring out the time derivatives:

$$\mathcal{L}_D = \bar{\psi}\Big(i\gamma^0 \frac{\partial}{\partial t} + i\boldsymbol{\gamma} \cdot \nabla - m\Big)\psi, \tag{33}$$

where $\boldsymbol{\gamma}$ is the 3-vector with components $\gamma^i$. Then we define the momenta conjugate to $\psi$ and $\bar{\psi}$ by

$$\pi = \frac{\partial \mathcal{L}_D}{\partial \dot{\psi}} = \bar{\psi}(i\gamma^0),$$
$$\bar{\pi} = \frac{\partial \mathcal{L}_D}{\partial \dot{\bar{\psi}}} = 0. \tag{34}$$

Note that by these definitions, $\pi$ is a row spinor and $\bar{\pi}$ (which vanishes) is a column spinor. Then the Dirac Hamiltonian density is

$$\mathcal{H}_D = \pi \dot{\psi} + \dot{\bar{\psi}} \, \bar{\pi} - \mathcal{L}_D, \tag{35}$$

or,

$$\mathcal{H}_D = \bar{\psi}(-i\boldsymbol{\gamma} \cdot \nabla + m)\psi = \psi^\dagger(-i\boldsymbol{\alpha} \cdot \nabla + m\beta)\psi = \psi^\dagger h \psi. \tag{36}$$

The "classical" field Hamiltonian is the spatial integral of this, which confirms the guess made in Eq. (3).

## 11. The Mode Expansion, and Classical $q$'s and $p$'s

We have now guessed the "classical" field Hamiltonian of the first quantized theory, and checked that it is the same as the Hamiltonian obtained from the Lagrangian. Next we need the classical $q$'s and $p$'s, which, when used in Hamilton's equations, give equations of motion equivalent to the Dirac equation.

In Notes 40 we guessed that the $q$'s and $p$'s of the electromagnetic field were the real and imaginary parts of the mode amplitudes, because the time evolution of the (complex) mode amplitudes in the complex plane looks like the evolution of a harmonic oscillator in phase space. A similar procedure works for the "classical" Dirac field.

The modes of the electromagnetic field are basically the normal modes of the vacuum Maxwell equations, that is, they are solutions that evolve with a definite frequency. Similarly, the modes of the free Dirac field are the free particle solutions, which we investigated in Notes 50. Here we use box normalization in a box of volume $V$. This means that the momentum that appears in a plane wave $e^{\pm i \mathbf{p} \cdot \mathbf{x}}$ takes on discrete values,

$$\mathbf{p} = \frac{2\pi}{L}\mathbf{n}, \tag{37}$$

where $V = L^3$ and where $\mathbf{n}$ is a vector of integers, positive, negative and zero. The allowed plane waves form a lattice in $\mathbf{p}$-space. As in Notes 50, free particle solutions are parameterized by the 4-vectors $p^\mu$ and $s^\mu$, but $p^\mu = (E, \mathbf{p})$ is actually a function of the 3-momentum $\mathbf{p}$, since $E$ is understood to be $\sqrt{\mathbf{p}^2 + m^2}$ (the positive square root). Sometimes we will write $E_p$ to emphasize that the energy depends on $p^\mu$ (which in turn depends on $\mathbf{p}$). Likewise, the 4-vector $s^\mu$ is a unit vector in the rest frame of the electron, which must run over two opposite choices in the rest frame in order to create a complete set of wave functions.

Then the normalized, positive and negative energy free particle solutions are

$$\psi_{ps+}(\mathbf{x}) = \sqrt{\frac{m}{VE}}\, u_{ps}\, e^{i\mathbf{p} \cdot \mathbf{x}},$$

$$\psi_{ps-}(\mathbf{x}) = \sqrt{\frac{m}{VE}}\, v_{ps}\, e^{-i\mathbf{p} \cdot \mathbf{x}}. \tag{38}$$

We will sometimes write these in ket language, using $|ps+\rangle$ to stand for $\psi_{ps+}(\mathbf{x})$ and $|ps-\rangle$ to stand for $\psi_{ps-}(\mathbf{x})$. The state $|ps+\rangle$ is an eigenstate of energy and momentum with eigenvalues $E$ and $\mathbf{p}$, while $|ps-\rangle$ has eigenvalues $-E$ and $-\mathbf{p}$. In particular, we have

$$h|ps+\rangle = E_p|ps+\rangle, \qquad h|ps-\rangle = -E_p|ps-\rangle. \tag{39}$$

These states satisfy the orthonormality relations,

$$\langle ps+|p's'+\rangle = \delta_{pp'}\,\delta_{ss'}, \qquad \langle ps+|p's'-\rangle = 0, \qquad \langle ps-|p's'-\rangle = \delta_{pp'}\,\delta_{ss'}, \tag{40}$$

where $\delta_{pp'}$ really means $\delta_{\mathbf{p},\mathbf{p}'}$ in which $\mathbf{p}$, $\mathbf{p}'$ take on the discrete values on the lattice in momentum

space. The explicit calculation of these proceeds as follows. For the $uu$-relation we have

$$
\begin{aligned}
\langle ps+|p's'+\rangle &= \sqrt{\frac{m^2}{V^2 EE'}} \int d^3\mathbf{x} \, (u_{ps}^\dagger u_{p's'}) \, e^{i(\mathbf{p}-\mathbf{p}')\cdot\mathbf{x}} \\
&= \sqrt{\frac{m^2}{EE'}} \, (u_{ps}^\dagger u_{p's'}) \, \delta_{pp'} = \frac{m}{E} \, (u_{ps}^\dagger u_{ps'}) \, \delta_{pp'} = \delta_{pp'} \, \delta_{ss'},
\end{aligned}
\tag{41}
$$

where the integral is taken over the volume of the box and where we have used Eq. (50.46$a$). Similarly, the calculation of the cross terms is

$$
\begin{aligned}
\langle ps+|p's'-\rangle &= \sqrt{\frac{m^2}{V^2 EE'}} \int d^3\mathbf{x} \, (u_{ps}^\dagger v_{p's'}) \, e^{i(\mathbf{p}+\mathbf{p}')\cdot\mathbf{x}} \\
&= \sqrt{\frac{m^2}{EE'}} \, (u_{ps}^\dagger v_{p's'}) \, \delta_{\tilde{p}p'} = \frac{m}{E} \, (u_{ps}^\dagger v_{\tilde{p}s'}) \, \delta_{\tilde{p}p'} = 0,
\end{aligned}
\tag{42}
$$

where $\tilde{p} = (E, -\mathbf{p})$ and where we have used Eq. (50.46$b$). The $vv$-scalar product is proved similarly. These orthogonality relations are really just special cases of the theorem that eigenstates of a collection of commuting operators with distinct eigenvalues are orthogonal, where in this case the operators are energy and momentum. See Theorem 1.2. The normalization follows from relations (50.46).

Since the free particle solutions form a complete set, we can expand an arbitrary Dirac wave function $\psi(\mathbf{x})$ as a linear combination of them,

$$
\psi(\mathbf{x}) = \sqrt{\frac{1}{V}} \sum_{ps} \sqrt{\frac{m}{E}} (b_{ps} \, u_{ps} \, e^{i\mathbf{p}\cdot\mathbf{x}} + c_{ps} \, v_{ps} \, e^{-i\mathbf{p}\cdot\mathbf{x}}),
\tag{43}
$$

where $b_{ps}$ and $c_{ps}$ are the expansion coefficients, or mode amplitudes, of the expansion into positive and negative energy free particle states, respectively. In ket language this is simply

$$
|\psi\rangle = \sum_{ps} (b_{ps}|ps+\rangle + c_{ps}|ps-\rangle).
\tag{44}
$$

Equation (43) or (44) is the mode expansion of the "classical" Dirac field.

If we allow $\psi$ to have a time dependence, then the amplitudes $b_{ps}$ and $c_{ps}$ also have one. If $\psi(\mathbf{x}, t)$ satisfies the free particle Dirac equation,

$$
i\frac{\partial}{\partial t}|\psi\rangle = h|\psi\rangle,
\tag{45}
$$

then the mode amplitudes satisfy

$$
i\dot{b}_{ps} = E_p \, b_{ps}, \qquad i\dot{c}_{ps} = -E_p \, c_{ps}.
\tag{46}
$$

These have the solutions,

$$
b_{ps}(t) = b_{ps}(0) \, e^{-iEt}, \qquad c_{ps}(t) = c_{ps}(0) \, e^{+iEt},
\tag{47}
$$

so that

$$\psi(\mathbf{x}, t) = \sqrt{\frac{1}{V}} \sum_{ps} \sqrt{\frac{m}{E}} \big[ b_{ps}(0)\, u_{ps}\, e^{i(\mathbf{p}\cdot\mathbf{x} - Et)} + c_{ps}(0)\, v_{ps}\, e^{-i(\mathbf{p}\cdot\mathbf{x} - Et)} \big]. \tag{48}$$

Notice that the exponents can be written in covariant form, $\mathbf{p} \cdot \mathbf{x} - Et = -p_\mu x^\mu = -(p \cdot x)$. The motion of the complex mode amplitudes $b_{ps}$ and $c_{ps}$ is a circle in the complex plane, clockwise for $b_{ps}$ and counterclockwise for $c_{ps}$. This reminds us of the motion of a harmonic oscillator in phase space, so we guess that the $q$'s and $p$'s of the "classical" Dirac field are the real and imaginary parts of the mode amplitudes.

To bring this out we express the "classical" field Hamiltonian in terms of the mode amplitudes. This is fairly easy in ket language. We write

$$\begin{aligned} H = \langle \psi | h | \psi \rangle &= \sum_{\substack{ps \\ p's'}} \big( b_{ps}^* \langle ps+ | + c_{ps}^* \langle ps- | \big) h \big( b_{p's'} | p's'+ \rangle + c_{p's'} | p's'- \rangle \big) \\ &= \sum_{ps} E_p (|b_{ps}|^2 - |c_{ps}|^2), \end{aligned} \tag{49}$$

where we use Eqs. (39) and the orthonormality relations (40). The final expression has a simple interpretation in the first quantized theory, since (for a normalized $\psi$) the quantities $|b_{ps}|^2$ and $|c_{ps}|^2$ are the probabilities of finding the state $\psi$ in various energy eigenstates, and the sum is just the average value of the energy, taken over these probabilities.

If we now write each mode amplitude $b_{ps}$ or $c_{ps}$ in the form $(Q + iP)/\sqrt{2}$, then the Hamiltonian $H$ becomes a sum of harmonic oscillators of the form

$$\pm \frac{E}{2}(Q^2 + P^2), \tag{50}$$

where the sign indicates positive or negative energy modes. There is one $(Q, P)$ pair for each mode, of positive or negative energy. Then it is easily verified that Hamilton's equations, with these $Q$'s and $P$'s, reproduce the time evolution (47), and hence that of the free particle wave function $\psi(\mathbf{x})$. That is, Hamilton's equations are equivalent to the free particle Dirac equation.

In addition to the Hamiltonian, there are two other observables of the "classical" field that are worth mentioning. One is the momentum of the field, which is logically defined as

$$\mathbf{P} = \int d^3\mathbf{x}\, \psi^\dagger(\mathbf{x})(-i\nabla)\psi(\mathbf{x}) = \langle \psi | \mathbf{p} | \psi \rangle, \tag{51}$$

where $\mathbf{p}$ is the momentum operator of the first quantized theory [it is $-i\nabla$ when acting on wave functions $\psi(\mathbf{x})$]. The other is the normalization integral,

$$N = \int d^3\mathbf{x}\, \psi^\dagger(\mathbf{x})\psi(\mathbf{x}) = \langle \psi | \psi \rangle, \tag{52}$$

so that $N = 1$ for a normalized wave function.

## 12. Second Quantizing the Dirac Field

To second quantize the Dirac field we attempt to follow the same steps we used in the quantization of the electromagnetic field. We interpret the $Q$'s and $P$'s as operators, with the usual, canonical commutation relations [see Eq. (41.6)]. Then the mode amplitudes $b_{ps}$ and $c_{ps}$ become operators, satisfying commutation relations just like Eqs. (41.6). These are raising and lowering operators for the harmonic oscillators that make up the Hamiltonian.

This procedure, however, runs into a serious problem, since commutation relations like Eq. (41.6) imply that the particles in question are bosons. This is the correct conclusion in the case of photons, as explained in Sec. 41.18, but electrons are fermions and the wave function of multielectron states must be antisymmetric under exchange.

Therefore we postulate that the creation and annihilation operators for electrons, both of positive and of negative energy, satisfy *anticommutation* relations, that is,

$$\{b_{ps}, b^\dagger_{p's'}\} = \delta_{pp'}\,\delta_{ss'}, \qquad \{b_{ps}, b_{p's'}\} = \{b^\dagger_{ps}, b^\dagger_{p's'}\} = 0,$$

$$\{b_{ps}, c_{p's'}\} = \{b_{ps}, c^\dagger_{p's'}\} = \{b^\dagger_{ps}, c_{p's'}\} = \{b^\dagger_{ps}, c^\dagger_{p's'}\} = 0, \tag{53}$$

$$\{c_{ps}, c^\dagger_{p's'}\} = \delta_{pp'}\,\delta_{ss'}, \qquad \{c_{ps}, c_{p's'}\} = \{c^\dagger_{ps}, c^\dagger_{p's'}\} = 0.$$

These are just like the commutation relations that would follow from the quantization of harmonic oscillators except for the replacement of commutators by anticommutators. Note that all anticommutators except $\{b_{ps}, b^\dagger_{p's'}\}$ and $\{c_{ps}, c^\dagger_{p's'}\}$ vanish.

This is a major departure from the procedure followed in the quantization of the electromagnetic field, and it detaches the definitions of the operators $b_{ps}$, $b^\dagger_{ps}$, $c_{ps}$, $c^\dagger_{ps}$ from the $Q$'s and $P$'s of the classical fields. Normally in calculations involving the second quantized Dirac field (or other fermion fields) we work with the creation and annihilation operators, not some version of $Q$'s and $P$'s.

The anticommutators (53) imply that multielectron states are antisymmetric under exchange. For example, the state with two electrons in modes $(ps)$ and $(p's')$ is given by

$$b^\dagger_{ps} b^\dagger_{p's'} |0\rangle = -b^\dagger_{p's'} b^\dagger_{ps} |0\rangle. \tag{54}$$

As in the case of the electromagnetic field, we borrow the mode expansion (43) of the classical field, and reinterpret it as an expansion of the quantum field $\psi(\mathbf{x})$ in terms of creation and annihilation operators. Under this reinterpretation (as in the case of the electromagnetic field), the argument $\mathbf{x}$ of the field is not an operator, rather $\psi(\mathbf{x})$ is the operator, and $\mathbf{x}$ is just a label. Similarly, on the right hand side, the operators are $b_{ps}$ and $c_{ps}$, while the spinors $u_{ps}$ and $v_{ps}$ have their same meaning as in the first quantized theory, that is, they are the 4-component spinors of complex numbers defined in Notes 50.

### 13. The Field Hamiltonian

We also borrow definitions of interesting observables from the "classical" theory. For example, we take the Hamiltonian of the quantized theory to be a quantized version of the "classical" Hamiltonian in Eq. (3), that is,

$$H = \int d^3\mathbf{x} \; :\psi^\dagger(\mathbf{x})(-i\boldsymbol{\alpha} \cdot \nabla + m\beta)\psi(\mathbf{x}): , \tag{55}$$

which is the same as the "classical" formula, apart from the normal ordering that has been introduced to make vacuum expectation values vanish.

However, fermion normal ordering is slightly different than in the case of bosons. If $Q$ is a monomial in fermion creation and annihilation operators, we define $:Q:$ as the monomial obtained by migrating all creation operators to the left, using the anticommutation relations (53) *and retaining the sign changes*, but dropping the anticommutators themselves. For example, we have

$$:b_{ps}b^\dagger_{p's'}: = -b^\dagger_{p's'}b_{ps}. \tag{56}$$

With this understanding, the Hamiltonian operator becomes

$$H = \frac{1}{V} \int d^3\mathbf{x} \sum_{\substack{ps \\ p's'}} \sqrt{\frac{m^2}{EE'}} \; :\!\left(b^\dagger_{ps}u^\dagger_{ps}e^{-i\mathbf{p}\cdot\mathbf{x}} + c^\dagger_{ps}v^\dagger_{ps}e^{+i\mathbf{p}\cdot\mathbf{x}}\right)$$

$$\times (-i\alpha \cdot \nabla + m\beta)\left(b_{p's'}u_{p's'}e^{i\mathbf{p}'\cdot\mathbf{x}} + c_{p's'}v_{p's'}e^{-i\mathbf{p}'\cdot\mathbf{x}}\right): , \tag{57}$$

where the operator in the middle is the same free-particle Hamiltonian operator $h$ of the first quantized theory which is seen in Eq. (49). The expression (57) can be reduced exactly as in Eq. (49), with the help of the orthonormality relations among the free-particle wave functions of the first quantized theory, except that the mode amplitudes $b_{ps}$, $c_{ps}$ are interpreted as operators whose ordering must be respected. The normal ordering does nothing since all creation operators are already to the left of all annihilation operators. Thus we obtain the field Hamiltonian of the second quantized theory,

$$H = \int d^3\mathbf{x} \; :\psi^\dagger(\mathbf{x})(-i\boldsymbol{\alpha} \cdot \nabla + m\beta)\psi(\mathbf{x}): = \sum_{ps} E_p(b^\dagger_{ps}b_{ps} - c^\dagger_{ps}c_{ps}). \tag{58}$$

Compare this with the analogous expression for the electromagnetic field,

$$H = \frac{1}{8\pi} \int d^3\mathbf{x} \; :(E^2 + B^2): = \sum_{\mathbf{k}\mu} \omega_{\mathbf{k}} a^\dagger_{\mathbf{k}\mu} a_{\mathbf{k}\mu}. \tag{59}$$

Note that the photon momentum $\mathbf{k}$ is like the electron momentum $\mathbf{p}$, and that the polarization index $\mu$ is like the spin index $s$.

## 14. Algebra of Fermion Creation and Annihilation Operators

Since the operators $b_{ps}$, $c_{ps}$, etc obey anticommutation relations instead of commutation relations, their algebraic relations are somewhat different than those of the operators of harmonic oscillator theory, which are useful for boson fields. To begin, we notice that since all fermion creation and annihilation operators anticommute with themselves, we have

$$b_{ps}^2 = b_{ps}^{\dagger 2} = c_{ps}^2 = c_{ps}^{\dagger 2} = 0. \tag{60}$$

Next we explore the properties of the number operators. The calculations are the same for any mode we pick, so to be definite we choose a positive energy mode and temporarily drop the $ps$ subscripts. Thus there are two operators under discussion, $b$ and $b^\dagger$. We define the Hermitian number operator by

$$N = b^\dagger b, \tag{61}$$

exactly as in the case of bosons. Now using the anticommutation relations $\{b, b^\dagger\} = 1$, $\{b, b\} = \{b^\dagger, b^\dagger\} = 0$, we find

$$N^2 = b^\dagger b b^\dagger b = -b^\dagger b^2 b^\dagger + b^\dagger b = N, \tag{62}$$

where we use $b^2 = 0$. Thus

$$N(N-1) = 0, \tag{63}$$

so the only eigenvalues of $N$ are 0 and 1. Recall that in the case of bosons or ordinary harmonic oscillators, the spectrum of $N$ is all nonnegative integers, $0, 1, 2, \ldots$. The fact that it is restricted to 0 and 1 for fermions is an expression of the Pauli exclusion principle, that a given state can have no more than one electron in it.

We assume that the eigenstates $|0\rangle$ and $|1\rangle$ of $N$ are nondegenerate. This is the simplest assumption we can make about the ket space upon which the operators $b$ and $b^\dagger$ act. As in the case of harmonic oscillators, it turns out that $b|n\rangle$ and $b^\dagger|n\rangle$ are eigenstates of the number operator with eigenvalues $n-1$ and $n+1$, respectively, except that here $n$ can take on only the values 0 and 1. Also, if $n \pm 1$ steps outside the range $(0, 1)$, then the result is zero. In summary, $b$ and $b^\dagger$ are annihilation and creation operators.

To show this in detail, we first consider $Nb|0\rangle = b^\dagger b^2|0\rangle = 0$, since $b^2 = 0$. Thus $b|0\rangle$ is an eigenket of $N$ with eigenvalue 0, so we must have $b|0\rangle = c_0|0\rangle$ for some complex number $c_0$. But squaring both sides, we obtain $\langle 0|b^\dagger b|0\rangle = \langle 0|N|0\rangle = 0 = |c_0|^2$, so $c_0 = 0$ and we have $b|0\rangle = 0$. Similarly, $Nb|1\rangle = b^\dagger b^2|1\rangle = 0$, so $b|1\rangle = c_1|0\rangle$, for some complex number $c_1$. Squaring both sides, we obtain $\langle 1|N|1\rangle = \langle 1|1\rangle = 1 = |c_1|^2$, or $c_1 = e^{i\alpha}$, a phase factor. Thus $b|1\rangle = e^{i\alpha}|0\rangle$. By redefining the phase convention for $|1\rangle$ the phase can be absorbed, and we have $b|1\rangle = |0\rangle$. Now consider $Nb^\dagger|0\rangle = b^\dagger b b^\dagger|0\rangle = -b b^{\dagger 2}|0\rangle + b^\dagger|0\rangle = b^\dagger|0\rangle$. Thus $b^\dagger|0\rangle$ is an eigenket of $N$ with eigenvalue 1, so it must be proportional to $|1\rangle$, $b^\dagger|0\rangle = c_0'|1\rangle$. But multiplying this by $b$, we have $bb^\dagger|0\rangle = -b^\dagger b|0\rangle + |0\rangle = c_0'b|1\rangle = c_0'|0\rangle$, or $c_0' = 1$. Thus we find $b^\dagger|0\rangle = |1\rangle$. Similarly, we find

$b^\dagger|1\rangle = b^{\dagger 2}|0\rangle = 0$. This derivation may be compared to what we did with the harmonic oscillator raising and lowering operators in Sec. 8.4. Altogether, we obtain

$$b|0\rangle = 0, \qquad b|1\rangle = |0\rangle, \qquad b^\dagger|0\rangle = |1\rangle, \qquad b^\dagger|1\rangle = 0. \tag{64}$$

This same analysis applies to each mode. Including now the mode indices, we can define number operators for positive and negative energy electrons,

$$N^+_{ps} = b^\dagger_{ps}b_{ps}, \qquad N^-_{ps} = c^\dagger_{ps}c_{ps}. \tag{65}$$

Now we define the vacuum to be the state with no electrons, either of positive or negative energy,

$$N^+_{ps}|0\rangle = N^-_{ps}|0\rangle = 0, \tag{66}$$

and we define occupation number basis states by applying creation operators $b^\dagger_{ps}$ or $c^\dagger_{ps}$ to the vacuum.

This is just like the occupation number basis in the case of photons, except that the occupation numbers $n^\pm_{ps}$ can only take on the values 0 or 1. Another difference is that an occupation number basis state is not uniquely determined by the list of occupation numbers, since the overall sign is determined by the order in which the creation operators are applied. See Eq. (54). But with the understanding that the notation is ambiguous with respect to an overall sign, we can write such a basis state as $|\ldots n^+_{ps}\ldots n^-_{ps}\ldots\rangle$. These basis states are eigenstates of all the number operators $N^\pm_{ps}$ with eigenvalues $n^\pm_{ps}$.

## 15. Field Energy, Momentum and Charge

The occupation number basis states are eigenstates of the field Hamiltonian (49) with eigenvalues given by

$$H|\ldots n^+_{ps}\ldots n^-_{ps}\ldots\rangle = \left[\sum_{ps} E_p\left(n^+_{ps} - n^-_{ps}\right)\right]|\ldots n^+_{ps}\ldots n^-_{ps}\ldots\rangle. \tag{67}$$

This is what we would expect for a system with $n^+_{ps}$ electrons in positive energy mode $(ps)$, and $n^-_{ps}$ electrons in negative energy mode $(ps)$.

The occupation number basis states are also eigenstates of momentum. The momentum operator of the field is the quantized version of the "classical" momentum $\mathbf{P}$, given in Eq. (51). Borrowing the definition of "classical" momentum and reinterpreting it as a field operator, we define

$$\mathbf{P} = \int d^3\mathbf{x} \; :\psi^\dagger(\mathbf{x})(-i\nabla)\psi(\mathbf{x})\!: \; = \sum_{ps}\mathbf{p}\left(b^\dagger_{ps}b_{ps} - c^\dagger_{ps}c_{ps}\right), \tag{68}$$

where the final expression follows by substituting the mode expansion of the field (43) (with $b_{ps}$ and $c_{ps}$ interpreted as operators) and simplifying. Then we have

$$\mathbf{P}|\ldots n^+_{ps}\ldots n^-_{ps}\ldots\rangle = \left[\sum_{ps}\mathbf{p}\left(n^+_{ps} - n^-_{ps}\right)\right]|\ldots n^+_{ps}\ldots n^-_{ps}\ldots\rangle, \tag{69}$$

where $\mathbf{p}$ under the sum refers to the momentum implicit in the $p$ index of the sum. This is just what we would expect for a collection of positive and negative energy electrons; as in Notes 50, the momentum labels of the negative energy electrons are the negatives of the eigenvalues.

The "classical" observable $N$, defined in the first quantized theory as the normalization integral (52) of the wave function $\psi(\mathbf{x})$, can also be converted in a field operator. It becomes

$$N = \int d^3\mathbf{x} \; :\psi^\dagger(\mathbf{x})\psi(\mathbf{x}): = \sum_{ps}\left(b^\dagger_{ps}b_{ps} + c^\dagger_{ps}c_{ps}\right), \tag{70}$$

which has the interpretation as the total number of electrons, of both positive and negative energy. If we multiply by $-e$, we obtain a charge operator,

$$Q = -e\int d^3\mathbf{x} \; :\psi^\dagger(\mathbf{x})\psi(\mathbf{x}): = -e\sum_{ps}\left(b^\dagger_{ps}b_{ps} + c^\dagger_{ps}c_{ps}\right), \tag{71}$$

The occupation number eigenstates are also eigenstates of $Q$,

$$Q|\ldots n^+_{ps}\ldots n^-_{ps}\ldots\rangle = -e\left[\sum_{ps}\left(n^+_{ps} + n^-_{ps}\right)\right]|\ldots n^+_{ps}\ldots n^-_{ps}\ldots\rangle. \tag{72}$$

It is the operator that gives the total charge of positive and negative energy electrons.

## 16. Heisenberg Equations of Motion

Now we examine the Heisenberg equations of motion of the Dirac field $\psi(\mathbf{x})$, since it is of interest to see whether and how the quantization by means of anticommutators affects these. The Heisenberg equations of motion are expressed in terms of the usual commutators (see Sec. 5.5); the quantization by means of anticommutation relations on the creation and annihilation operators does not change the rest of quantum mechanics.

The Heisenberg equation of motion for the operator $b_{ps}$ is

$$\dot{b}_{ps} = -i[b_{ps}, H] = -i\sum_{p's'}E_{p'}\left([b_{ps}, b^\dagger_{p's'}b_{p's'}] - [b_{ps}, c^\dagger_{p's'}c_{p's'}]\right). \tag{73}$$

The second commutator vanishes,

$$[b_{ps}, c^\dagger_{p's'}c_{p's'}] = b_{ps}c^\dagger_{p's'}c_{p's'} - c^\dagger_{p's'}c_{p's'}b_{ps} = 0, \tag{74}$$

since in the second term we can migrate $b_{ps}$ to the left past the two $c$-operators, incurring two minus signs. As for the first commutator, it is

$$[b_{ps}, b^\dagger_{p's'}b_{p's'}] = b_{ps}b^\dagger_{p's'}b_{p's'} - b^\dagger_{p's'}b_{p's'}b_{ps}, \tag{75}$$

in which the second term is

$$-b^\dagger_{p's'}b_{p's'}b_{ps} = b^\dagger_{p's'}b_{ps}b_{p's'} = -b_{ps}b^\dagger_{p's'}b_{p's'} + \delta_{pp'}\,\delta_{ss'}\,b_{p's'}. \tag{76}$$

Putting this back into Eq. (73) and similarly evaluating the Heisenberg equation of motion for $c_{ps}$, we find

$$\dot{b}_{ps} = -iE\,b_{ps}, \qquad \dot{c}_{ps} = +iE\,c_{ps}, \tag{77}$$

which are exactly the "classical" equations of motion (46).

This implies that the second quantized Dirac field $\psi(\mathbf{x})$ obeys Heisenberg equations of motion that are precisely of the same form as the first quantized Dirac equation. Recall that Maxwell's equations are valid in quantum electrodynamics, it is just that they are reinterpreted as the Heisenberg equations of motion for the quantum fields (see Sec. 41.10).

## 17. Positrons, not Holes

So far we have created a nice field theory of positive and negative energy electrons. It gives the correct Fermi-Dirac statistics for multiparticle problems but otherwise its physical content does not go beyond what we had in the first quantized theory. In particular, it does not address the interpretational difficulties of the negative energy solutions.

To fix this up, we borrow ideas from hole theory. If all of the negative energy states are filled, as Dirac supposed, then when we excite an electron out of a negative energy state, we create a hole. Thus the destruction of a negative energy electron is equivalent to the creation of a hole, that is, a positron. Therefore let us define

$$d_{ps}^{\dagger} = c_{ps}, \tag{78}$$

where $d_{ps}^{\dagger}$ creates a positron of charge, energy, momentum and spin $+e$, $+E$, $+\mathbf{p}$ and $+s$. Likewise, if an electron makes a transition to an unoccupied negative energy state (a hole), it creates a negative energy electron or destroys a hole. So let us write

$$d_{ps} = c_{ps}^{\dagger}. \tag{79}$$

Note that with these definitions, the operators $d_{ps}$ and $d_{ps}^{\dagger}$ satisfy anticommutation relations exactly like those of the $b$'s and $c$'s, that is,

$$\{d_{ps}, d_{p's'}^{\dagger}\} = \delta_{pp'}\,\delta_{ss'}, \tag{80}$$

with all other anticommutators being zero.

Now the quantum field (43) becomes

$$\psi(\mathbf{x}) = \sqrt{\frac{1}{V}} \sum_{ps} \sqrt{\frac{m}{E}} (b_{ps}\,u_{ps}\,e^{i\mathbf{p}\cdot\mathbf{x}} + d_{ps}^{\dagger}\,v_{ps}\,e^{-i\mathbf{p}\cdot\mathbf{x}}), \tag{81}$$

which is also worth recording in its adjoint version,

$$\bar{\psi}(\mathbf{x}) = \sqrt{\frac{1}{V}} \sum_{ps} \sqrt{\frac{m}{E}} (b_{ps}^{\dagger}\,\bar{u}_{ps}\,e^{-i\mathbf{p}\cdot\mathbf{x}} + d_{ps}\,\bar{v}_{ps}\,e^{i\mathbf{p}\cdot\mathbf{x}}). \tag{82}$$

Also, the Hamiltonian (58) becomes

$$H = \sum_{ps} E_p (b_{ps}^\dagger b_{ps} - d_{ps} d_{ps}^\dagger). \tag{83}$$

Notice the order of the operators in the last term. If we anticommute them, we obtain

$$H = \sum_{ps} E_p (b_{ps}^\dagger b_{ps} + d_{ps}^\dagger d_{ps}) - \sum_{ps} E_p. \tag{84}$$

The final term is infinite. One interpretation is that it is the infinite (negative) energy of the electrons in the filled Dirac sea. Another interpretation is that it is an infinite term resulting from the ordering ambiguities inherent in passing from a "classical" expression to a quantum operator. That is, it is a zero-point term.

We throw away zero-point terms in order to make vacuum expectation values vanish. But is the vacuum the state in which there are no electrons, either of positive or negative energy, or is it the state in which the Dirac sea is filled, that is, a state with no electrons and no positrons? The latter is more physical, and if we want $\langle 0|H|0\rangle = 0$ for this vacuum then we must throw away the final, infinite term in Eq. (84).

This leads to a new interpretation of normal ordering: We migrate all $b_{ps}^\dagger$ and all $d_{ps}^\dagger$ to the left, keeping sign changes but throwing away anticommutators. This differs from what we did above insofar as the operators $d_{ps}$, $d_{ps}^\dagger$ are concerned. Thus, the energy, momentum and charge of the field become

$$H = \int d^3\mathbf{x} \; {:}\psi^\dagger(\mathbf{x})(-i\boldsymbol{\alpha}\cdot\nabla + m\beta)\psi(\mathbf{x}){:} = \sum_{ps} E_p (b_{ps}^\dagger b_{ps} + d_{ps}^\dagger d_{ps}), \tag{85a}$$

$$\mathbf{P} = \int d^3\mathbf{x} \; {:}\psi^\dagger(\mathbf{x})(-i\nabla)\psi(\mathbf{x}){:} = \sum_{ps} \mathbf{p}(b_{ps}^\dagger b_{ps} + d_{ps}^\dagger d_{ps}), \tag{85b}$$

$$Q = -e \int d^3\mathbf{x} \; {:}\psi^\dagger(\mathbf{x})\psi(\mathbf{x}){:} = -e \sum_{ps} (b_{ps}^\dagger b_{ps} - d_{ps}^\dagger d_{ps}). \tag{85c}$$

Note especially the sign changes. Now positrons have positive charge and energy (and momentum and spin), but the charge operator is no longer $(-e)$ times a positive definite operator. The quantity Dirac worked so hard to make positive definite in the first quantized theory is now replaced by an operator of either sign in the second quantized theory. Energies are strictly positive, and we deal strictly with observable objects (electrons and positrons). Dirac's original goal of curing the problems with the Klein-Gordon equation now appears as irrelevant, although there was no way to see this within the framework of the single-particle (first quantized) theory. The real conceptual breakthrough came with hole theory.

**18. The Klein-Gordon Equation Redux**

These successes with the Dirac equation also help us to fix the problems with the Klein-Gordon equation. This equation should be regarded as the Heisenberg equations of motion for a quantum field, in which the annihilation operators for negative energy states are reinterpreted as creation operators for antiparticles. Of course these operators satisfy boson commutation relations, not fermion anticommutation relations. This makes the energy of the particles and antiparticles positive definite. As for the negative probability density that occurs in the Klein-Gordon theory, the entire probability current 4-vector, when multiplied by the unit of charge, becomes a charge current for particles and antiparticles. It is not positive definite, but that is no problem since charges of both signs appear in the theory. This is the "rehabilitation" of the Klein-Gordon equation that was referred to earlier. For lack of time and space we will not go further into the Klein-Gordon theory.