IOP Publishing

Modern Quantum Mechanics and Quantum Information

J S Faulkner

Chapter 3

Relativistic quantum mechanics

3.1 The necessity for relativistic quantum mechanics

Not many years ago it was thought that even a graduate text on quantum mechanics could focus entirely on the Schrödinger equation. The assumption was that it would be nice to include a section on the Dirac theory for the sake of completeness, but it was not a necessity. Today, quantum mechanics is widely used to calculate the properties of condensed matter and also the chemistry of aggregates of atoms and molecules. These calculations are extremely complex, but they have become practical due to the development of cluster workstations and massively parallel supercomputers. Condensed matter physicists, chemists, and materials scientists rely on these calculations to provide insights into the wide variety of phenomena that they must understand in order to develop new and useful materials, compounds, devices, and structures. The earliest computer codes contained many approximations, one of which was ignoring relativistic effects. The reason for this was that the computers that were available in the 1960s and 1970s were relatively slow. The speed and number of parallel processors of computers continues to increase at a great rate, so calculations are expected to include fewer approximations. Condensed matter physicists now base their codes on Dirac theory [1] because it must be used for all but the lightest atoms. Modern theories of magnetism make no sense if the calculations are not relativistic [2]. Quantum chemists have likewise found that the errors in their calculations caused by ignoring relativistic effects are unacceptable [3].

In the later chapters of this book, most discussion will be based on the Schrödinger formulation because it contains the essential physics and the Dirac equation requires the inclusion of many subscripts and superscripts that are not necessary for the discussion. It is assumed that the relativistic effects will be added later as needed.

3.2 Klein–Gordon equation

In many ways, the Klein–Gordon (K-G) equation is easier to derive than the Schrödinger equation. As was pointed out in the preceding chapter, Schrödinger

doi:10.1088/978-0-7503-2167-9ch3

derived a relativistic equation that was essentially the K-G equation, but it gave the wrong results for the hydrogen atom. The first version of this equation is obtained by writing the relativistic equation for the energy of a free particle

$$E^2 = p^2c^2 + m^2c^4. \tag{3.1}$$

If the momentum is set equal to zero, this equation reduces to the famous mass–energy relation

$$E = mc^2. \tag{3.2}$$

It is hoped that these equations of special relativity will be familiar to the reader. Courses in classical mechanics or modern physics normally contain a discussion of that theory. There are useful discussions on the web, and books on special relativity are very cheap.

The experience of Schrödinger is now invoked to replace E and $\vec{p}$ with the operators,

$$(E^2 - c^2\mathbf{p}^2 - m^2c^4)|\,\psi(t)\rangle \tag{3.3}$$

and then transform into the position representation

$$\langle \vec{r}\,|\,E\,|\,\psi(t)\rangle \to i\hbar\frac{\partial\psi(\vec{r},\,t)}{\partial t}\langle \vec{r}\,|\,\vec{\mathbf{p}}\,|\,\psi(t)\rangle = -i\hbar\nabla\psi(\vec{r},\,t), \tag{3.4}$$

to obtain the differential equation

$$-\hbar^2\frac{\partial^2\psi(\vec{r},\,t)}{\partial t^2} + c^2\hbar^2\nabla^2\psi(\vec{r},\,t) - m^2c^4\psi(\vec{r},\,t) = 0. \tag{3.5}$$

Dividing this equation by $c^2\hbar^2$ yields the Klein–Gordon equation for a free particle

$$-\frac{1}{c^2}\frac{\partial^2\psi(\vec{r},\,t)}{\partial t^2} + \nabla^2\psi(\vec{r},\,t) - \frac{m^2c^2}{\hbar^2}\psi(\vec{r},\,t) = 0. \tag{3.6}$$

The position vector in four-dimensional space is

$$x^0 = ct,\; x^1 = x,\; x^2 = y,\; x^3 = z. \tag{3.7}$$

The length of a vector in that space is found with the use of a metric

$$l = \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} x^\mu g_{\mu\nu} x^\nu, \tag{3.8}$$

where

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{3.9}$$

Einstein's second postulate of relativity is that the length of a vector in the four-dimensional spacetime is invariant under a Lorentz transformation

$$l = c^2t'^2 - r'^2 = c^2t^2 - r^2. \tag{3.10}$$

The four-vector momentum is

$$p^0 = \frac{E}{c}, p^1 = p_x, p^2 = p_y, p^3 = p_z, \tag{3.11}$$

and its length is

$$m^2c^2 = \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} p^\mu g_{\mu\mu} p^\nu. \tag{3.12}$$

This is identical with equation (3.1) and since the length of the four momentum is invariant under a Lorentz transformation, it follows that the Klein–Gordon is invariant and is a proper relativistic equation.

The next step is to consider a particle in a force field. In special relativity, this is achieved in the same way that it is in classical Hamiltonian theory. The total energy is modified to include a scalar potential $\phi(\vec{r})$ and a vector potential $\vec{A}(\vec{r})$ as shown in this equation

$$(E - e\phi(\vec{r}))^2 = c^2(\vec{p} - e\vec{A}(\vec{r}))^2 + m^2c^4. \tag{3.13}$$

Converting E, $\vec{p}$, and $\vec{r}$ into operators and transforming into the position representation leads to

$$\left(ih\frac{\partial}{\partial t} - e\phi(\vec{r})\right)^2 \psi(\vec{r}, t) - c^2(-i\hbar\nabla - e\vec{A}(\vec{r}))^2\psi(\vec{r}, t) = m^2c^4\psi(\vec{r}, t). \tag{3.14}$$

It simplifies the analysis to set the vector potential equal to zero and call $e\phi(\mathbf{r}) = V(\mathbf{r})$ and this leads to

$$\left[-h^2\frac{\partial^2}{\partial t^2} - ih2V(\vec{r})\frac{\partial}{\partial t} + V(\vec{r})^2\right]\psi(\vec{r}, t) + c^2\hbar^2\nabla^2\psi(\vec{r}, t) - m^2c^4\psi(\vec{r}, t) = 0. \tag{3.15}$$

This equation takes an interesting form in the non-relativistic limit. The first step is to eliminate the rest mass energy, equation (3.2), because that constant is not normally included in non-relativistic equations. This is achieved by replacing the wave function in equation (3.15) with

$$\psi(r, t) = \hat{\psi}(r, t)e^{-imc^2t/\hbar}. \tag{3.16}$$

Dividing the resulting expression by $2mc^2$ and carrying out some manipulations gives

$$\begin{aligned} -\frac{\hbar^2}{2m}\nabla^2\hat{\psi}(\vec{r}, t) + V(\vec{r})\hat{\psi}(\vec{r}, t) &= i\hbar\frac{\partial\hat{\psi}}{\partial t} \\ -\frac{\hbar^2}{2mc^2}\frac{\partial^2\hat{\psi}}{\partial t^2} - i\hbar\frac{V(\vec{r})}{mc^2}\frac{\partial\hat{\psi}}{\partial t} + \frac{V(\vec{r})^2}{2mc^2}\hat{\psi}(\vec{r}, t) & \end{aligned}. \tag{3.17}$$

In the non-relativistic limit, all of the terms in the second line of this equation are ignored, and the result is the non-relativistic Schrödinger equation that revolutionized quantum mechanics

$$-\frac{\hbar^2}{2m}\nabla^2\hat{\psi}(\vec{r},\,t)+V(\vec{r})\hat{\psi}(\vec{r},\,t)=i\hbar\frac{\partial\hat{\psi}}{\partial t}. \tag{3.18}$$

It is interesting that the second time derivative in the Klein–Gordon equation disappears, leaving only the first derivative.

3.3 Problems with the Klein–Gordon equation

It was discovered almost immediately by the theorists who derived a Klein–Gordon equation, Klein [4], Gordon [5], Schrödinger *et al*, that the eigenfunctions and eigenvalues produced by the equation showed unphysical features. The first point has to do with Born's interpretation of the absolute square of a time independent wave function $\psi(\vec{r})$ as the probability for finding a particle at the point $\vec{r}$. If the wave function depends on time $\psi(\vec{r},\,t)$, then the conservation of probability is analogous to the conservation of particles in an incompressible liquid. If the particle density is $\rho(\vec{r},\,t)$ and the flux of particles through a small surface is $\vec{j}\,(\vec{r},\,t)$, then the conservation of particles requires that

$$\frac{\partial\rho}{\partial t}=\nabla\cdot\vec{j}\,. \tag{3.19}$$

Premultiplying the Schrödinger equation, equation (3.18), by $\psi*(\vec{r},\,t)$ and premultiplying the complex conjugate of equation (3.18) by $\psi(\vec{r},\,t)$ produces two new equations. Subtracting the second from the first and dividing by $-i\hbar$ leads to an equation like equation (3.19) with

$$\rho=\psi*\psi, \tag{3.20}$$

and

$$\vec{j}\,=\frac{i\hbar}{2m}[\psi*\nabla\psi-\psi\nabla\psi*]=\frac{\hbar}{m}\mathrm{Im}\psi*\nabla\psi. \tag{3.21}$$

These equations fit nicely with Born's interpretation of the wave function.

Premultiplying the Klein–Gordon equation, equation (3.15), by $\psi*(\vec{r},\,t)$ and premultiplying the complex conjugate of equation (3.15) by $\psi(\vec{r},\,t)$ and then subtracting the second equation from the first and dividing by $-i\hbar 2mc^2$ leads to an equation like equation (3.19) with the same probability flux as is shown in equation (3.21). However, the density $\rho(\vec{r},\,t)$ is

$$\rho(\vec{r},\,t)=-\frac{i\hbar}{2mc^2}\left[\psi*\frac{\partial\psi}{\partial t}-\psi\frac{\partial\psi*}{\partial t}\right]+\frac{V(\vec{r})}{mc^2}[\psi*\psi]. \tag{3.22}$$

The strange looking time dependence of this density can be interpreted as a relativistic distortion of the infinitesimal cube that is used in its definition.

However the modification of the non-relativistic density by a factor proportional to $V(\vec{r})$ is really disturbing. The potential function is negative more often than not, and there is nothing to prevent the second term from being larger than the first making $\rho(\vec{r}, t)$ negative. A negative probability density has no physical meaning. It can be argued that this is not a conclusive demonstration of the short-comings of the Klein–Gordon equation because there may be other ways to define the probability density and flux.

A more definitive way to illustrate the problems with the Klein–Gordon equation is to apply it to an elementary scattering problem. Consider a one-dimensional system with a potential that is zero for x less than zero and has a value V_0 when x is greater than zero

$$V(x) = 0x < 0V(x) = V_0x \geqslant 0. \tag{3.23}$$

In classical mechanics, a particle with energy less than V_0 would bounce back from the step at $x = 0$, and if the energy is greater than V_0 the particle would proceed with no effect from the potential. If the non-relativistic Schrödinger equation, equation (3.18), is used to solve this problem, the reflection of the particle for $0 \leqslant E \leqslant V_0$ is not 100% because the particle has a non-zero probability of tunneling into the range where the potential is V_0. The wave function for $x \geqslant 0$ is

$$\psi(x, t) = Te^{-\kappa x}e^{-iEt/\hbar}, \tag{3.24}$$

where

$$\kappa = \frac{\sqrt{2m(V_0 - E)}}{\hbar}. \tag{3.25}$$

It can be seen that κ is real throughout the energy range. When $E \geqslant V_0$, κ becomes imaginary, which means the the particle is propagating. However, even when $E \geqslant V_0$ the reflection coefficient is not zero. It is analogous to light striking a block of glass. Some of the light is transmitted and some is reflected. The classical and Schrödinger results are both reasonable in their way, but the Klein–Gordon equation gives a different answer.

In one dimension, equation (3.14) is

$$\left[ih\frac{\partial}{\partial t} - V(x)\right]^2\psi(x, t) + c^2\hbar^2\frac{\partial^2\psi(x, t)}{\partial x^2} - m^2c^4\psi(x, t) = 0. \tag{3.26}$$

For x less than zero, the equation is

$$-h^2\frac{\partial^2\psi(x, t)}{\partial t^2} + c^2\hbar^2\frac{\partial^2\psi(x, t)}{\partial x^2} - m^2c^4\psi(x, t) = 0, \tag{3.27}$$

and the solution is propagating

$$\psi(x, t) = (e^{ikx} + Re^{-ikx})e^{-iEt/\hbar}. \tag{3.28}$$

where

$$p = \hbar k \tag{3.29}$$

and

$$k = \frac{\sqrt{E^2 - m^2c^4}}{\hbar c}. \tag{3.30}$$

For x greater than zero, the solution is as shown in equation (3.24), a function that does not propagate. The difference between the Klein–Gordon and Schrödinger solutions is that for the former, κ is a solution of

$$(E - V_0)^2 = -\hbar^2\kappa^2c^2 + m^2c^4, \tag{3.31}$$

or

$$\kappa = \frac{[m^2c^4 - (E - V_0)^2]^{1/2}}{\hbar c}. \tag{3.32}$$

It is clear from this equation that κ is real when E is such that

$$-mc^2 \leqslant E - V_0 \leqslant mc^2, \tag{3.33}$$

and pure imaginary otherwise.

The real part of κ is plotted in figure 3.1. This figure makes the predictions of the Klein–Gordon theory very clear. For $E \leqslant V_0 - mc^2$, κ is imaginary and the particle will propagate in the $x \geqslant 0$ region. This not only contradicts the Schrödinger result described above, but also it is hard to believe on physical grounds. The Schrödinger equation gives the physically reasonable result that, when E is greater than the height of the step, the wave will propagate. The Klein–Gordon equation, however, predicts that for energies in the range $V_0 \leqslant E \leqslant V_0 + mc^2$, κ is real and the solution is not propagating. Such a result violates physical expectations.

It is very difficult to solve the hydrogen atom problem with the Klein–Gordon equation, but it has been done. The eigenvalues are considerably different from the ones predicted by the Schrödinger equation. The Schrödinger eigenvalues agree very

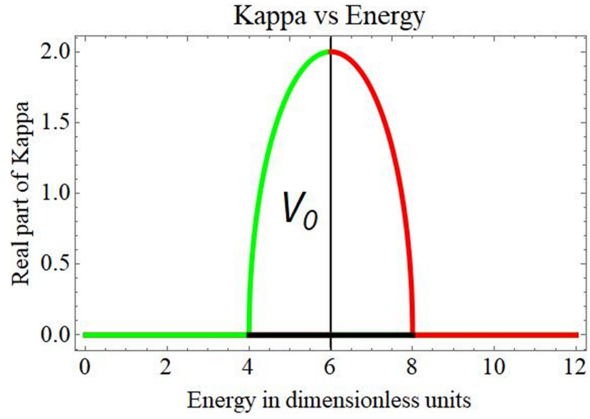

Figure 3.1. The quantity κ from equation (3.32) as a function of energy in dimensionless units. The perpendicular line shows the height of the potential for $x \geqslant 0$. Only the real part of κ is shown.

well with experiments, which means that the Klein–Gordon eigenvalues do not. In addition, heavier atoms could not be treated with the Klein–Gordon equation because the equation for the innermost shell of electrons has no solution for an atomic number Z greater than 69.

For the reasons outlined in this section, most physicists have given up on the Klein–Gordon equation as a wave equation. However, the fact that it is Lorentz invariant means that it serves a useful purpose in field theory.

3.4 Dirac theory

Physicists have often discovered new mathematical principles in the process of developing a theory of some phenomenon. This is particularly true of P A M Dirac. In chapter 1 of this book, the theory of distributions was referenced in order to put into a mathematical context the delta-function that Dirac introduced. Mathematicians have developed new areas of algebra and group theory in order to elucidate the manipulations Dirac performed in his derivation of relativistic quantum theory, in particular, the matrices that he used. His equation led to a number of new concepts in physics, such as spin and antimatter. The theoretical physicist Victor F Weisskopf spent much of his career working with Dirac's equation. He said the following, 'Today it is hard to realize the excitement, the skepticism, and the enthusiasm aroused in the early years by the development of all the new insights that emerged from the Dirac equation. A great deal more was hidden in the Dirac equation than the author had expected when he wrote it down in 1928'. Because the Dirac equation provides the foundation for quantum electrodynamics and field theory, many sophisticated analyses of that equation are found in books on those subjects. Dirac himself remarked in one of his talks that his equation was more intelligent than its author. It should be added, however, that it was Dirac who found many of the additional insights.

Dirac [6] realized that the problem with the Klein–Gordon approach is the second derivative of $\psi(\vec{r}, t)$ with respect to time. The experience with the non-relativistic Schrödinger equation indicated that it would be good to eliminate that feature, but it also made clear that, in order to achieve Lorentz invariance, the derivative with respect to position must also be of first order. He therefore decided to take the square root of both sides of equation (3.1). He was aware that the square root of the quadratic form $c^2(p_x^2 + p_y^2 + p_z^2) + m^2c^4$ doesn't exist within the field of real numbers, so he wrote it as

$$E = c\vec{\alpha} \cdot \vec{p} + \beta mc^2 \tag{3.34}$$

where the elements of $\vec{\alpha}$ and also β have to be matrices. Multiplying the right side of equation (3.34) by itself leads to the quadratic form only if these matrices are such that

$$\alpha_x^2 = \alpha_y^2 = \alpha_z^2 = \beta^2. \tag{3.35}$$

Also

$$\vec{\alpha}\beta + \beta\vec{\alpha} = 0, \tag{3.36}$$

as well as

$$\{\alpha_x, \alpha_y\} = \{\alpha_y, \alpha_z\} = \{\alpha_z, \alpha_x\} = 0, \tag{3.37}$$

where $\{\alpha_a, \alpha_b\}$ is the anticommutator, $\{\alpha_a, \alpha_b\} = \alpha_a\alpha_b + \alpha_b\alpha_a$. These relations are easy to find by simply multiplying the right side of equation (3.34) by itself.

Dirac found a set of matrices that satisfy these conditions. They are 4×4 matrices, but it is more convenient to write them in block matrix form using 2×2 matrices. For example, the 2×2 unit matrix is

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.38}$$

and the matrix β can be written

$$\beta = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{3.39}$$

The components of $\vec{\alpha}$ are

$$\alpha_x = \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, \ \alpha_y = \begin{pmatrix} 0 & \sigma_y \\ \sigma_y & 0 \end{pmatrix}, \ \alpha_z = \begin{pmatrix} 0 & \sigma_z \\ \sigma_z & 0 \end{pmatrix}, \tag{3.40}$$

where the sigmas are the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.41}$$

They are Hermitean, which means that the α and β matrices are also Hermitean.

The square of a Pauli matrix is a unit matrix

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = -i\sigma_x\sigma_y\sigma_z = I_2, \tag{3.42}$$

and the anticommutators are all zero

$$\{\sigma_x, \sigma_y\} = \{\sigma_y, \sigma_z\} = \{\sigma_z, \sigma_x\} = 0. \tag{3.43}$$

These two properties can be written

$$\{\sigma_\alpha, \sigma_\beta\} = 2\delta_{\alpha\beta}I_2. \tag{3.44}$$

Later, the commutators of these matrices will be needed

$$[\sigma_x, \sigma_y] = 2i\sigma_z, \ [\sigma_y, \sigma_z] = 2i\sigma_x, [\sigma_z, \sigma_x] = 2i\sigma_y. \tag{3.45}$$

Defining the vector

$$\vec{\sigma} = \sigma_x \hat{x} + \sigma_y \hat{y} + \sigma_z \hat{z}, \tag{3.46}$$

and using the properties outlined above it can be shown that

$$e^{i\theta \vec{n}\cdot\vec{\sigma}} = \cos\theta + i \sin\theta \vec{n} \cdot \vec{\sigma}. \tag{3.47}$$

The commutators in equation (3.45) are obtained by matrix multiplication. It will be seen that multiplying the Pauli matrices by $\hbar/2$ gives operators that have the same commutation rules as the spin operators introduced in chapter 1

$$S_i = \frac{1}{2}\hbar\sigma_i. \tag{3.48}$$

This is proved by comparing equation (3.45) with equation (1.87).

Following Schrödinger, Dirac considered E and $\vec{p}$ as operators and then replaced them with

$$E \Rightarrow i\hbar\frac{\partial}{\partial t}, \vec{p} \Rightarrow -i\hbar\nabla, \tag{3.49}$$

so that equation (3.34) becomes

$$-i\hbar c\vec{\alpha} \cdot \nabla\psi + \beta mc^2\psi = i\hbar\frac{\partial\psi}{\partial t}. \tag{3.50}$$

Since this is a 4×4 matrix equation, the wave function is a column vector

$$\psi(\vec{r}, t) = \begin{pmatrix} \psi^1(\vec{r}, t) \\ \psi^2(\vec{r}, t) \\ \psi^3(\vec{r}, t) \\ \psi^4(\vec{r}, t) \end{pmatrix}. \tag{3.51}$$

It will be explained later why this wave function is more properly called a bispinor. Equation (3.50) is the original, and still a very useful, form of the Dirac equation for a free particle. As described in connection with equation (3.13), the equation can be generalized to treat a particle in a force field by including a scalar potential $\phi(\vec{r})$ and a vector potential $\vec{A}(\vec{r})$ to obtain

$$c\vec{\alpha} \cdot \left(-i\hbar\nabla - \frac{e}{c}\vec{A}(\vec{r})\right)\psi + e\phi(\vec{r})\psi + \beta mc^2\psi = i\hbar\frac{\partial\psi}{\partial t}. \tag{3.52}$$

3.5 Proof of the Lorentz covariance of the Dirac equation

The invariance of the Klein–Gordon equation under a Lorentz transformation was shown in the first section of this chapter. The behavior of the Dirac equation under a Lorentz transformation is not as obvious. In order to investigate this question, it is useful to put the Dirac equation into a different form.

The position and also time was written as a four-vector in equation (3.7). It is useful to first consider a common example of the Lorentz transformation of that vector. According to the rules of special relativity the position of a particle in a coordinate system in which the origin is moving to the right along the x-axis with velocity v is

$$x' = \frac{x - vt}{\sqrt{1 - \beta^2}} = \gamma x - \beta\gamma ct, \tag{3.53}$$

where

$$\beta = \frac{v}{c}, \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}. \tag{3.54}$$

Obviously, the inverse of this transformation would be

$$x = \frac{x' + vt'}{\sqrt{1 - \beta^2}} = \gamma x' + \beta\gamma ct'. \tag{3.55}$$

The time coordinate also changes in the moving coordinate system

$$ct' = \frac{ct - \beta x}{\sqrt{1 - \beta^2}} = \gamma ct - \beta\gamma x, \tag{3.56}$$

and the inverse is

$$ct = \frac{ct' - \beta x'}{\sqrt{1 - \beta^2}} = \gamma ct' + \beta\gamma x'. \tag{3.57}$$

The Lorentz transformation of the four-vector may thus be written in matrix form

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}, \tag{3.58}$$

or

$$x'^{\mu} = \sum_{\nu=0}^{3} \Lambda^{\mu}_{\nu} x^{\nu}. \tag{3.59}$$

Using the fact that

$$\gamma^2(1 - \beta^2) = 1, \tag{3.60}$$

it can be seen algebraically that the inverse of the above equation is

$$\begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix}, \tag{3.61}$$

which is also obvious from the physical arguments above. The compact way to write these matrix equations is

$$x' = \Lambda x. \tag{3.62}$$

There are other four vectors

$$a_4 = \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix}, \tag{3.63}$$

that transform under the Lorentz transformation like

$$\begin{pmatrix} a'^0 \\ a'^1 \\ a'^2 \\ a'^3 \end{pmatrix} = \begin{pmatrix} \Lambda_0^0 & \Lambda_1^0 & \Lambda_2^0 & \Lambda_3^0 \\ \Lambda_0^1 & \Lambda_1^1 & \Lambda_2^1 & \Lambda_3^1 \\ \Lambda_0^2 & \Lambda_1^2 & \Lambda_2^2 & \Lambda_3^2 \\ \Lambda_0^3 & \Lambda_1^3 & \Lambda_2^3 & \Lambda_3^3 \end{pmatrix} \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix}, \tag{3.64}$$

or

$$a' = \Lambda a, \tag{3.65}$$

in a moving reference frame. The length of the vector requires the use of the metric defined in equation (3.9)

$$\tilde{a} g a = l, \tag{3.66}$$

where $\tilde{a}$ is the transpose

$$\tilde{a} = (a^0 a^1 a^2 a^3). \tag{3.67}$$

It follows from the invariance of this length that

$$\tilde{a}' g a' = \tilde{a} \tilde{\Lambda} g \Lambda a = \tilde{a} g a = l, \tag{3.68}$$

and

$$\tilde{\Lambda} g \Lambda = g. \tag{3.69}$$

Algebraically, the definition of a proper Lorentz transformation matrix Λ is a real four-by-four matrix that satisfies the preceding equation and has a determinant equal to one.

From the definition in equation (3.7) the components x^1, x^2, and x^3 are the Cartesian coordinates in ordinary three-dimensional space. A rotation by the angle θ around the z axis in that space is also described by a Lorentz transformation

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.70}$$

This relation is familiar from undergraduate physics classes. Obviously, any combination of rotations with translations in time are described by a Lorentz transformation Λ. A transformation in space and time, like the one in equation (3.58), is called a boost.

Defining the matrices

$$\gamma^0 = \beta,\ \gamma^1 = \beta\alpha_x,\ \gamma^2 = \beta\alpha_y,\ \gamma^3 = \beta\alpha_z, \tag{3.71}$$

and multiplying equation by β leads to a new form for the Dirac equation

$$\sum_{\mu=0}^{4} \gamma^\mu \left(i\hbar \frac{\partial}{\partial x^\mu} - eA_\mu \right) \psi - mc\psi = 0 \tag{3.72}$$

where the four-vector potential A is defined by

$$A = \left[\frac{\phi(\vec{r})}{c}, -\vec{A}(\vec{r}) \right]. \tag{3.73}$$

The gamma matrices defined above above and equation (3.72) are similar to the form of his equation carved in Dirac's memorial stone in Westminster Abbey.

Writing the matrices out in detail

$$\begin{gathered} \gamma^0 = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \\ \gamma^1 = \begin{pmatrix} 0 & \sigma_x \\ -\sigma_x & 0 \end{pmatrix},\ \gamma^2 = \begin{pmatrix} 0 & \sigma_y \\ -\sigma_y & 0 \end{pmatrix},\ \gamma^3 = \begin{pmatrix} 0 & \sigma_z \\ -\sigma_z & 0 \end{pmatrix}, \end{gathered} \tag{3.74}$$

it can be seen that γ^0 is Hermitean and the others are anti-Hermitean. The conditions previously worked out for the alphas and beta, the anticommutators of the gamma matrices can be summed up

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g_{\mu\nu} I_4, \tag{3.75}$$

where the $g_{\mu\nu}$ are the elements of the metric matrix in equation (3.9).

Using the gamma matrices and the four-vector form of the momentum from equation (3.11), equation (3.34) can be written

$$\sum_{\mu=0}^{4}\sum_{\nu=0}^{4} \gamma^\mu g_{\mu\nu} p^\nu = mc. \tag{3.76}$$

It might appear that it would be useful to define gamma matrices in a moving reference frame as

$$\gamma'^{\mu} = \sum_{\nu=0}^{3} \Lambda_{\nu}^{\mu} \gamma^{\nu}, \tag{3.77}$$

and the four-vector momentum in the usual way as

$$p'^{\mu} = \sum_{\nu=0}^{3} \Lambda_{\nu}^{\mu} p^{\nu}. \tag{3.78}$$

Then the primed version of equation (3.76) would transform into the unprimed form easily using the condition in equation (3.69). This would lead to time and velocity dependent gamma matrices, and is not the definition of covariance that is desired.

The Lorentz transformed version of the Dirac equation for a free particle is

$$\sum_{\nu=0}^{4} \gamma^{\nu} \left(i\hbar \frac{\partial}{\partial x'^{\nu}} \right) \psi'(x') - mc\psi'(x') = 0, \tag{3.79}$$

where $x' = \Lambda x$ are the coordinates in the moving reference frame. It is assumed that the four-dimensional wave vector must transform according to the equation

$$\psi'(x') = \psi'(\Lambda x) = S\psi(x), \tag{3.80}$$

with S being a four-by-four matrix that clearly depends on Λ.

It is at this point that it is seen that $\psi(x)$ is not just a 4×1 vector. A function that transforms according to equation (3.80) is called a bispinor. In later chapters, spinors and bispinors will be discussed in more detail.

Using the inverse of equation (3.59)

$$x^{\mu} = \sum_{\nu=0}^{3} (\Lambda^{-1})_{\nu}^{\mu} x'^{\nu}, \tag{3.81}$$

and the rule from differential calculus, the derivatives become

$$\frac{\partial}{\partial x'^{\nu}} = \sum_{\mu=0}^{4} \left[(\Lambda^{-1})_{\nu}^{\mu} \frac{\partial}{\partial x^{\mu}} \right]. \tag{3.82}$$

Multiplying equation (3.79) on the left by S^{-1} and assuming that the differential operator has no effect on S, the equation takes the intermediate form

$$\sum_{\nu=0}^{4} S^{-1} \gamma^{\nu} S \left(i\hbar \sum_{\mu=0}^{4} \left[(\Lambda^{-1})_{\nu}^{\mu} \frac{\partial}{\partial x^{\mu}} \right] \right) S^{-1} \psi'(x') - mcS^{-1} \psi'(x') = 0. \tag{3.83}$$

This is the same as the free particle equation in the unprimed variables

$$\sum_{\mu=0}^{4} \gamma^{\mu} \left(i\hbar \frac{\partial}{\partial x^{\mu}} \right) \psi(x) - mc\psi(x) = 0, \tag{3.84}$$

identifying

$$\gamma^{\mu} = \sum_{\nu=0}^{4} S^{-1}\gamma^{\nu}S(\Lambda^{-1})^{\mu}_{\nu}, \tag{3.85}$$

or

$$\sum_{\mu=0}^{4} \gamma^{\mu}\Lambda^{\kappa}_{\mu} = S^{-1}\gamma^{\kappa}S. \tag{3.86}$$

All of the matrices in this equation (3.86) are known with the exception of S. The question is if it exists and can be found. It was proved by Pauli [7] that if there is another set of four four-by-four matrices that satisfy equation (3.75)

$$\{\gamma'^{\mu}, \gamma'^{\nu}\} = \gamma'^{\mu}\gamma'^{\nu} + \gamma'^{\nu}\gamma'^{\mu} = 2g_{\mu\nu}I_4, \tag{3.87}$$

then the two sets of matrices are connected by a similarity transformation

$$\gamma'^{\kappa} = S^{-1}\gamma^{\kappa}S. \tag{3.88}$$

He called this the fundamental theorem on Dirac matrices. It follows from it that the matrices defined in equation (3.86)

$$\gamma'^{\kappa} = \sum_{\mu=0}^{4} \gamma^{\mu}\Lambda^{\kappa}_{\mu}, \tag{3.89}$$

are such a set of matrices and therefore S must exist. Since the gamma matrices are the same for all Dirac equations, S depends only on the Lorentz transformation, Λ.

The algebra required to solve equation (3.85) for S is lengthy and not very illuminating. It has been explained in a number of publications, so the final results will just be quoted. For the case that Λ describes a simple rotation by an angle θ about the z axis, shown in equation (3.70), the matrix S is

$$S = \begin{pmatrix} \cos\frac{\theta}{2}I_2 + i\sigma_3 \sin\frac{\theta}{2} & 0 \\ 0 & \cos\frac{\theta}{2}I_2 + i\sigma_3 \sin\frac{\theta}{2} \end{pmatrix}. \tag{3.90}$$

The Λ for a boost in the x direction is shown in equation (3.58). The corresponding S matrix for this case is

$$S = \begin{pmatrix} I_2 \cosh\frac{\eta}{2} & -\sigma_x \sinh\frac{\eta}{2} \\ -\sigma_x \sinh\frac{\eta}{2} & I_2 \cosh\frac{\eta}{2} \end{pmatrix}, \tag{3.91}$$

where

$$\tanh \eta = \frac{v}{c}. \tag{3.92}$$

It turns out that the four-vector potential in equation (3.72) transforms like

$$A'_{\nu} = \sum_{\mu=0}^{4} \left[(\Lambda^{-1})^{\mu}_{\nu} A_{\mu} \right], \tag{3.93}$$

so the argument for the covariance of the Dirac equation is unchanged for particles moving in a potential. Of course, for that case S will depend on the potential as well as the motion of the coordinate system.

3.6 The fifth gamma matrix

In the preceding section it was shown that the gamma matrices represent a fixed basis of unit vectors in spacetime. The unit volume element on spacetime is, in terms of the gammas,

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3, \tag{3.94}$$

or

$$\gamma^5 = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix}. \tag{3.95}$$

This is not one of the gamma matrices, but it is frequently used in connection with them. It anticommutes with the basic four gamma matrices

$$\{\gamma^{\mu}, \gamma^5\} = 0. \tag{3.96}$$

It takes a leading role when questions of parity arise because it changes sign under a spacetime reflection.

3.7 Free particle solution of the Dirac equation

The Dirac equation for a free particle at rest is

$$\beta mc^2 \psi = i\hbar \frac{\partial \psi}{\partial t} \tag{3.97}$$

with $E = \pm mc^2$. Taking β from equation (3.39), leads to

$$\begin{matrix} \left[\dfrac{\partial}{\partial t} + \dfrac{imc^2}{\hbar} \right] \psi_i = 0 i = 1, 2 \\ \left[\dfrac{\partial}{\partial t} - \dfrac{imc^2}{\hbar} \right] \psi_i = 0 i = 3, 4 \end{matrix}, \tag{3.98}$$

so that

$$\psi_1 = \frac{1}{\sqrt{V}}\begin{pmatrix}1\\0\\0\\0\end{pmatrix}e^{-\frac{imc^2}{\hbar}t}\ \psi_2 = \frac{1}{\sqrt{V}}\begin{pmatrix}0\\1\\0\\0\end{pmatrix}e^{-\frac{imc^2}{\hbar}t}, \tag{3.99}$$

for positive energies and

$$\psi_3 = \frac{1}{\sqrt{V}}\begin{pmatrix}0\\0\\1\\0\end{pmatrix}e^{\frac{imc^2}{\hbar}t}\ \psi_4 = \frac{1}{\sqrt{V}}\begin{pmatrix}0\\0\\0\\1\end{pmatrix}e^{\frac{imc^2}{\hbar}t}, \tag{3.100}$$

for negative energies. Obviously,

$$\int_V \psi_i^\dagger \psi_i d\vec{r} = 1 \tag{3.101}$$

when working in a finite space V.
Solutions of the Dirac equation could be written in many different ways, but the theory derived in the preceding sections requires that they remain bispinors. Thus, it must be possible to obtain the free particle solution for a particle with finite momentum $\vec{p}$ by applying a boost to the preceding equations. Generalizing equation (3.91) for the case of a particle moving in the direction $\hat{n}$, $\vec{p} = p\hat{n}$, is equivalent to giving the origin a boost in the opposite direction

$$S = \begin{pmatrix} I_2 \cosh\frac{\eta}{2} & \vec{\sigma}\cdot\vec{n}\sinh\frac{\eta}{2} \\ \vec{\sigma}\cdot\vec{n}\sinh\frac{\eta}{2} & I_2\cosh\frac{\eta}{2}\end{pmatrix}. \tag{3.102}$$

The definition of η in equation (3.92) is not useful for quantum mechanics. It is better to define

$$\tanh\eta = \frac{cp}{E}, \tag{3.103}$$

which leads to

$$\cosh\eta = \frac{E}{mc^2}. \tag{3.104}$$

Trigonometric identities lead to

$$\cosh\frac{\eta}{2} = \sqrt{\frac{1+\cosh\eta}{2}} = \sqrt{\frac{E+mc^2}{2mc^2}}, \tag{3.105}$$

and

$$\sinh\frac{\eta}{2} = \sqrt{\frac{\cosh\eta - 1}{2}} = \sqrt{\frac{E-mc^2}{2mc^2}}. \tag{3.106}$$

These relations make it possible to write the boost matrix S in terms of the energy, and applying that matrix to ψ_1 leads to

$$\psi_1(x,\,p) = S\psi_1(x) = \frac{1}{\sqrt{V}}\cosh\frac{\eta}{2}\begin{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\\ \tanh\frac{\eta}{2}\vec{\sigma}\cdot\vec{n}\begin{pmatrix}1\\0\end{pmatrix}\end{pmatrix}e^{-i\frac{mc^2}{\hbar}t}. \tag{3.107}$$

The four momentum is $\mathbf{p} = \left(\frac{E}{c},\, p_x,\, p_y,\, p_z\right)$ and the four-vector position is $\mathbf{x} = (ct,\, x,\, y,\, z)$. Since both of these vectors transform according the the same Lorentz transformation, Λ, it follows that their inner product is invariant. Thus,

$$(\mathbf{p},\,\mathbf{x}) = Et - \vec{p}\cdot\vec{r} = mc^2t, \tag{3.108}$$

the last form being obtained by evaluating the inner product at $t = 0$. It is also useful to write the hyperbolic tangent in a different way

$$\tanh\frac{\eta}{2} = \sqrt{\frac{E-mc^2}{E+mc^2}} = \frac{\sqrt{E^2-m^2c^4}}{E+mc^2} = \frac{cp}{E+mc^2}. \tag{3.109}$$

Using these relations, the final form of $\psi_1(x,\,p)$ is

$$\psi_1(x,\,p) = \frac{1}{\sqrt{V}}\sqrt{\frac{E+mc^2}{2mc^2}}\begin{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\\ \dfrac{c\vec{\sigma}\cdot\vec{p}}{E+mc^2}\begin{pmatrix}1\\0\end{pmatrix}\end{pmatrix}e^{i\frac{(\vec{p}\cdot\vec{r}-Et)}{\hbar}}. \tag{3.110}$$

Similar manipulations lead to

$$\psi_2(x,\,p) = \frac{1}{\sqrt{V}}\sqrt{\frac{E+mc^2}{2mc^2}}\begin{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\\ \dfrac{c\vec{\sigma}\cdot\vec{p}}{E+mc^2}\begin{pmatrix}0\\1\end{pmatrix}\end{pmatrix}e^{i\frac{(\vec{p}\cdot\vec{r}-Et)}{\hbar}}. \tag{3.111}$$

These are the positive energy states, $E > mc^2$. The top two components make up the large part of the wave function and the bottom two are generally smaller because of $1/\sqrt{E+mc^2}$.

$$\psi_3(x,\,p) = \frac{1}{\sqrt{V}}\sqrt{\frac{E+mc^2}{2mc^2}}\begin{pmatrix}-\dfrac{c\vec{\sigma}\cdot\vec{p}}{E+mc^2}\begin{pmatrix}1\\0\end{pmatrix}\\ \begin{pmatrix}1\\0\end{pmatrix}\end{pmatrix}e^{i\frac{(\vec{p}\cdot\vec{r}+Et)}{\hbar}}, \tag{3.112}$$

and

$$\psi_3(x,\,p) = \frac{1}{\sqrt{V}}\sqrt{\frac{E+mc^2}{2mc^2}}\begin{pmatrix}-\dfrac{c\vec{\sigma}\cdot\vec{p}}{E+mc^2}\begin{pmatrix}0\\1\end{pmatrix}\\ \begin{pmatrix}0\\1\end{pmatrix}\end{pmatrix}e^{i\frac{(\vec{p}\cdot\vec{r}+Et)}{\hbar}}. \tag{3.113}$$

These are the negative energy states, $E < -mc^2$. The bottom two components make up the large part of the wave function and the top two are generally smaller. Naming the parts of the wave function as to large and small parts and positive and negative energy is brought up frequently in discussions of the solutions of the Dirac equation.

A rotation is normally represented mathematically by a three-by-three matrix operating on a three-dimensional vector. It can also be represented by a two-by-two matrix operating on a two-dimensional object called a spinor. For the case of a rotation through the angle θ about an axis that is pointing in the direction of a unit vector $\vec{n}$, the two-by-two matrix is

$$s(\theta) = e^{i\vec{n}\cdot\vec{\sigma}\frac{\theta}{2}}, \tag{3.114}$$

where $\vec{\sigma}$ is the matrix defined in equation (3.46). The simplest way to write the complete set of spinors is

$$\chi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \chi_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.115}$$

These objects will appear often in succeeding chapters. Comparing them with the solutions of the Dirac equation in equations (3.110) through (3.113) makes it obvious why these solutions are called bispinors.

The free particle solutions found by the boost method are exactly the same as those obtained from a conventional solution of the Dirac equation

$$(c\vec{\alpha}\cdot\vec{p} + \beta mc^2)\psi = i\hbar\frac{\partial\psi}{\partial t}. \tag{3.116}$$

There are two advantages to the boost method. First, it is easier. Second, it demonstrates more clearly the theoretical point that the bispinors transform as expected under the Lorentz transformation.

3.8 Angular momentum and spin

Using equation (2.148) the rate of change of the angular momentum $\vec{L} = \vec{r} \times \vec{p}$ can be found for the case of a free particle by calculating the commutator with the Hamiltonian

$$\frac{d\vec{L}}{dt} = \frac{i}{\hbar}[H, \vec{L}]. \tag{3.117}$$

The math is easier when the focus is on one component

$$L_x = yp_z - zp_y, \tag{3.118}$$

and

$$H = c\alpha_x p_x + c\alpha_y p_y + c\alpha_z p_z + \beta mc^2, \tag{3.119}$$

so

$$[H, L_x] = c\alpha_y[p_y, y]p_z - c\alpha_z[p_z, z]p_y = -i\hbar c(\alpha_y p_z - \alpha_z p_y), \tag{3.120}$$

using the standard position-momentum commutation rule. With a few more manipulations, the result is

$$\frac{d\vec{L}}{dt} = c(\vec{\alpha} \times \vec{p}). \tag{3.121}$$

This result shows that angular momentum is not conserved for a Dirac particle, while it is conserved for a particle described by the Schrödinger equation.

The spin of a Dirac particle is represented as a four-by-four matrix by

$$\vec{S} = \frac{\hbar}{2}\vec{\Sigma}, \tag{3.122}$$

where

$$\vec{\Sigma} = \hat{\mathbf{x}}\Sigma_x + \hat{\mathbf{y}}\Sigma_y + \hat{\mathbf{z}}\Sigma_z, \tag{3.123}$$

and

$$\Sigma_i = \begin{pmatrix} \sigma_i & 0 \\ 0 & \sigma_i \end{pmatrix}. \tag{3.124}$$

The time derivative of Σ_z is given by

$$\frac{d\Sigma_z}{dt} = \frac{i}{\hbar}[H, \Sigma_z] \tag{3.125}$$

where the useful form of H is in equation (3.119). Using the commutation rules in equation (3.45),

$$\frac{i}{\hbar}[H, \Sigma_z] = \frac{c}{\hbar}[\alpha_x, \Sigma_z]p_x + \frac{c}{\hbar}[\alpha_y, \Sigma_z]p_y, \tag{3.126}$$

and hence

$$\frac{d\Sigma_z}{dt} = -\frac{2c}{\hbar}(\vec{\alpha} \times \vec{p})_z, \tag{3.127}$$

and finally

$$\frac{d\vec{S}}{dt} = -c(\vec{\alpha} \times \vec{p}), \tag{3.128}$$

which means that the spin of the Dirac particle is not conserved.

It can be seen from the preceding equations that neither $\vec{L}$ nor $\vec{S}$ is conserved for a Dirac particle, but the total angular momentum $\vec{L} + \vec{S}$ is conserved

$$\frac{d(\vec{L} + \vec{S})}{dt} = 0. \tag{3.129}$$

The spin of an electron was used before Dirac published his equation. It appeared in the Pauli equation, and was used in the explanation of atomic spectra. However, the appearance of the spin in those equations was completely ad hoc. The first equation in which the spin and its behavior appeared naturally is the Dirac equation. This first-principles theory of spin is one of Dirac's great achievements.

3.9 The magnetic moment of the electron

The Dirac equation for a system with non-zero electric and magnetic fields is

$$\left(c\vec{\alpha}\cdot\left(\vec{p}-\frac{e}{c}\vec{A}\right)+\beta mc^2\right)\psi=(E-e\phi)\psi, \tag{3.130}$$

where $\vec{A}$ and ϕ are the scalar and vector potentials. This is obtained in the usual way by replacing

$$\vec{p}\rightarrow\vec{p}-\frac{e}{c}\vec{A},\ E\rightarrow E-E\phi. \tag{3.131}$$

Defining

$$\vec{\pi}=\vec{p}-\frac{e}{c}\vec{A}, \tag{3.132}$$

and writing the wave function in terms of two component spinors

$$\psi=\begin{pmatrix}\psi_A\\ \psi_B\end{pmatrix} \tag{3.133}$$

the Dirac equation can be written

$$\begin{aligned}c\vec{\sigma}\cdot\vec{\pi}\psi_B&=(E-mc^2-e\phi)\psi_A\\ c\vec{\sigma}\cdot\vec{\pi}\psi_A&=(E+mc^2-e\phi)\psi_B\end{aligned}. \tag{3.134}$$

Solving the bottom equation for ψ_B in terms of ψ_A and inserting that into the top equation leads to

$$\frac{-i\hbar ec^2(\vec{\sigma}\cdot\nabla\phi)\vec{\sigma}\cdot\vec{\pi}}{(2mc^2+W-e\phi)^2}\psi_A+\frac{c^2(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi})}{2mc^2+W-e\phi}\psi_A=(W-e\phi)\psi_A, \tag{3.135}$$

where the energy that appears in classical mechanics has been introduced

$$W=E-mc^2. \tag{3.136}$$

The first term in equation (3.135) is small relative to the others and will be ignored for the rest of this section. It will play a large role in the following section.

In the low-energy limit

$$W-E\phi\ll mc^2, \tag{3.137}$$

so

$$\frac{1}{2m}(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi})\psi_A = (W - e\phi)\psi_A. \tag{3.138}$$

Using the properties of the Pauli matrices described previously, the identity

$$(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi}) = \vec{\pi}\cdot\vec{\pi} + i\sigma\cdot(\vec{\pi}\times\vec{\pi}) \tag{3.139}$$

can be proved. Pursuing the argument further, it is shown that

$$\vec{\pi}\times\vec{\pi} = \left(-i\hbar\nabla - \frac{e}{c}\vec{A}\right)\times\left(-i\hbar\nabla - \frac{e}{c}\vec{A}\right) = i\frac{e\hbar}{c}\vec{B} \tag{3.140}$$

so the low-energy limit of the Dirac equation is

$$\frac{1}{2m}\left(-i\hbar\nabla - \frac{e}{c}\vec{A}\right)^2\psi_A - \frac{e\hbar}{2mc}\vec{\sigma}\cdot\vec{B}\psi_A + e\phi\psi_A = W\psi_A. \tag{3.141}$$

The preceding equation is usually called the Pauli equation because it was written down by him before Dirac's equation was available. The difference is that Pauli's version of the spin magnetic moment was ad hoc. Dirac's equation gives a first principles derivation of it

$$\vec{\mu} = \frac{e\hbar}{2mc}\vec{\sigma} = 2\times\frac{e}{2mc}\vec{S}. \tag{3.142}$$

The reason for writing this moment as the above is that a classical derivation for an angular momentum $\vec{L}$ would give

$$\vec{\mu}_L = \frac{e}{2mc}\vec{L}. \tag{3.143}$$

The factor of 2 in equation (3.142) is called a g-factor. It was understood from experiments on atomic spectra that it had to have this value, but it arises naturally from Dirac's equation.

Feynman was extremely proud that his quantum electrodynamics (QED) predicts the radiative corrections to the g-factor

$$\frac{g-2}{2} = 0.0011596521564, \tag{3.144}$$

which agrees with experiment to the first eleven decimal places. He liked to point out that this was the most accurate theoretical prediction of a physical constant ever.

3.10 Scalar relativistic approximation

Starting in the 1960s, quantum mechanical calculations of the electronic structure of atoms, molecules, and solids became increasingly common. These calculations allowed the fields of atomic physics, condensed matter physics and quantum chemistry to become analytical. The rapid increase in the understanding of materials

that arose from the ability to calculate properties from first principles is the basis for the technological advances in communications, computers, and consumer products that are the basis for the present economic development.

The earliest calculations were based on the Schrödinger equation, but it was soon realized that relativistic corrections are significant. With the computers that were available at the time, it was thought that solving the Dirac equation for large systems would be too difficult. Therefore, approximations were made, the most common being the scalar relativistic approximation [8].

Let us assume that there is no magnetic field. The Dirac equation is written as in equation (3.134), but with $E\phi = V$

$$\begin{aligned} c\vec{\sigma}\cdot\vec{p}\,\psi_B &= (W - V)\psi_A \\ c\vec{\sigma}\cdot\vec{p}\,\psi_A &= (W + 2mc^2 - V)\psi_B \end{aligned}. \tag{3.145}$$

The spinor ψ_B from the lower equation is inserted into the upper to obtain

$$\begin{aligned} &-i\hbar c^2(W + 2mc^2 - V)^{-2}\vec{\sigma}\cdot\nabla V\vec{\sigma}\cdot\vec{p}\,\psi_A + (W + 2mc^2 - V)^{-1}c^2(\vec{\sigma}\cdot\vec{p})^2\psi_A \\ &= (W - V)\psi_A \end{aligned}. \tag{3.146}$$

Manipulating this equation leads to

$$\begin{aligned} &\frac{\vec{p}^2\psi_A}{2m} + V\psi_A \\ &-\frac{(W - V)^2}{2mc^2}\psi_A + \frac{i\hbar}{2m(W + 2mc^2 - V)}[\nabla V\cdot\vec{p} + i\vec{\sigma}\cdot(\nabla V\times\vec{p})]\psi_A, \\ &= W\psi_A \end{aligned} \tag{3.147}$$

which looks like the Schrödinger equation but with a more complicated potential function. The first addition to the potential,

$$-\frac{(W - V)^2}{2mc^2}, \tag{3.148}$$

is called the mass–velocity correction. The second,

$$\frac{i\hbar\nabla V\cdot\vec{p}}{2m(W + 2mc^2 - V)}, \tag{3.149}$$

is called the Darwin term. The third,

$$\frac{-\hbar\vec{\sigma}\cdot(\nabla V\times\vec{p})}{2m(W + 2mc^2 - V)}, \tag{3.150}$$

is the spin–orbit correction. In scalar relativistic calculations, only the mass–velocity and Darwin terms are used. If it is desirable to see the effect of the spin–orbit interactions, they can be calculated later using perturbation theory.

The scalar relativistic approximation reproduces relativistic effects reasonably well, and is still used in calculations where great accuracy is not required. However, the gold standard for calculations of atoms, molecules, and solids is to use the full Dirac theory.

3.11 The Dirac theory of the hydrogen atom

The measure of success for a quantum theory in 1928 was how well it treated the hydrogen atom. Schrödinger abandoned his first effort at a quantum equation, which was relativistic, because it gave a very bad result for hydrogen. His non-relativistic equation successfully explained the main features of the hydrogen spectrum, The Dirac equation does the same, but it then goes on to predict fine structure that Schrödinger's equation can not explain.

Since 1928 when the Dirac equation for the hydrogen atom was first solved correctly, other approaches have been used that emphasize different aspects of the solutions. The one shown here was used by Martin and Glauber in a study of orbital electron capture [9, 10]. It requires that the Dirac equation is written in a form that, like the Klein–Gordon equation, is second order in time and space derivatives. Starting with equation (3.52) and defining a Hamiltonian operator leads to

$$c\vec{\alpha}\cdot\left(-i\hbar\nabla-\frac{e}{c}\vec{A}(\vec{r})\right)\psi+e\phi(\vec{r})\psi+\beta mc^2\psi=i\hbar\frac{\partial\psi}{\partial t}=H\psi. \tag{3.151}$$

An operator P is defined by

$$P\beta=\frac{1}{2mc^2}\left[\beta\left(i\hbar\frac{\partial}{\partial t}-H\right)+2mc^2\right]\beta. \tag{3.152}$$

Applying this operator to equation (3.151)

$$P\beta\left(i\hbar\frac{\partial}{\partial t}-H\right)=0 \tag{3.153}$$

leads, after a lot of manipulations, to

$$\left[\frac{1}{c^2}\left(i\hbar\frac{\partial\psi}{\partial t}-e\phi\right)^2-\left(-i\hbar\nabla-\frac{e}{c}\vec{A}\right)^2-m^2c^2+\frac{e\hbar}{c}(\vec{\sigma}\cdot\vec{H}-i\vec{\alpha}\cdot\vec{E})\right]\psi=0. \tag{3.154}$$

This is the Klein–Gordon equation (3.14) with the addition of the term $(\vec{\sigma}\cdot\vec{H}-i\alpha\cdot\vec{E})$ that couples the electron to the magnetic field $\vec{H}$ and the electric field $\vec{E}$.

A solution of this second order equation doesn't have to be a solution of the Dirac equation, although a solution of the Dirac equation is always a solution of the second order equation. The operator $\beta(i\hbar\frac{\partial}{\partial t}-H)$ commutes with the P operator. It follows that a spinor obtained by operating P on any solution of the second order equation will be such that

$$P\beta\left(i\hbar\frac{\partial}{\partial t} - H\right)\psi = \beta\left(i\hbar\frac{\partial}{\partial t} - H\right)P\psi = 0. \tag{3.155}$$

Therefore, if ψ is any solution of the second order equation, $P\psi$ is a solution to the Dirac equation. If the original ψ is already a solution of the Dirac equation, then $P\psi = \psi$.

For the hydrogen atom or a nucleus with charge Ze and one electron,

$$\vec{A} = 0 \quad \vec{H} = 0 \quad e\phi = -\frac{Ze^2}{r}. \tag{3.156}$$

For a stationary state with energy E, equation (3.154) becomes

$$\left[\frac{1}{c^2}\left(E + \frac{Ze^2}{r}\right)^2 + \hbar^2\nabla^2 - m^2c^2 + \frac{i\hbar}{c}\frac{Ze^2}{r^2}\vec{\alpha}\cdot\hat{r}\right]\psi = 0 \tag{3.157}$$

and $\hat{r}$ is the unit vector in the direction of $\vec{r}$. It was observed in equation (2.100) that

$$-\hbar^2\nabla^2 = -\hbar^2\frac{1}{r^2}\frac{\partial^2}{\partial r^2}r^2 + \frac{\mathbf{L}^2}{r^2}. \tag{3.158}$$

Equation (3.157) then becomes

$$\left[\frac{E^2 - m^2c^4}{c^2} + \frac{2EZe^2}{c^2r} + \hbar^2\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{\mathbf{L}^2 - (Ze^2/c)^2 - (i\hbar Ze^2/c)\vec{\alpha}\cdot\hat{r}}{r^2}\right]\psi = 0. \tag{3.159}$$

Let us now define the scalar operator

$$K = \beta\left(1 + \frac{1}{\hbar}\vec{\sigma}\cdot\vec{L}\right). \tag{3.160}$$

This operator commutes with

$$\vec{\alpha}\cdot\vec{p}\,\vec{\alpha}\cdot\vec{r}r^2\vec{J} = \vec{L} + \vec{S}. \tag{3.161}$$

From the equations in the preceding sections, it can be seen that these relations imply that K commutes with the usual Dirac equation for the hydrogen atom

$$H = c\vec{\alpha}\cdot\vec{p} + \beta mc^2 - \frac{Ze^2}{r}, \tag{3.162}$$

and is a constant of the motion for that system. Since J^2 and J_z are also constants of motion, the energy levels of the relativistic hydrogen atom can be labeled with the eigenvalues of K, J^2, and J_z.

The eigenvalues of J^2, and J_z are $\hbar^2 j(j+1)$ and $\hbar j$ where $j = 1/2, 3/2, 5/2, \ldots$. The square of K is

$$K^2 = \frac{1}{\hbar^2}J^2 + \frac{1}{4}, \tag{3.163}$$

and the eigenvalues of K^2 are thus $(j + 1/2)^2$. It can be shown that K anticommutes with the fifth gamma matrix discussed in section 3.6 of this chapter

$$\{K, \gamma^5\} = 0. \tag{3.164}$$

It follows that, if there is an eigenstate of K corresponding to an eigenvalue k,

$$K \mid k\rangle = k \mid k\rangle, \tag{3.165}$$

then

$$K\gamma_5 \mid k\rangle = -\gamma_5 K \mid k\rangle = -k\gamma_5 \mid k\rangle. \tag{3.166}$$

This means that, for every positive eigenvalue and eigenvector, there is a negative eigenvalue and an eigenvector that has the opposite parity. From the preceding analysis it is concluded that the eigenvalues of K are

$$k = \pm 1, \pm 2, \pm 3, \ldots. \tag{3.167}$$

If an operator Λ can be found such that

$$\hbar^2 \Lambda(\Lambda + 1) = \mathbf{L}^2 - (Ze^2/c)^2 - (i\hbar Ze^2/c)\vec{\alpha} \cdot \hat{r}, \tag{3.168}$$

then equation (3.159) can be written

$$\left\{-\frac{\hbar^2}{2m}\left[\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}\right) + \frac{\Lambda(\Lambda + 1)}{r^2}\right] - \frac{Z'e^2}{r}\right\}R_{nl}(r) = E'R_{nl}(r). \tag{3.169}$$

This equation looks just like the radial part of the non-relativistic Schrödinger equation for the hydrogen atom, equation (2.102), except that

$$Z' = \frac{EZ}{mc^2}, \tag{3.170}$$

and

$$E' = \frac{E^2 - m^2c^4}{2mc^2}. \tag{3.171}$$

Of course the angular momentum quantum number l has been replaced by the operator Λ. The advantage of this manipulation is that, assuming the solution of the relativistic equation is also an eigenstate of Λ, $\Lambda\psi = l'\psi$, then it follows immediately that the solutions of the equation are the generalized Laguerre polynomials and the energy eigenvalue is

$$E' = -\frac{mZ'^2e^4}{2\hbar^2 n'^2}, \tag{3.172}$$

where n' is to be determined. This analogy also defines the relation between the eigenvalues of Λ, l', and n'. It turns out that the operator

$$\Lambda = -\beta K - (iZe^2/\hbar c)\vec{\alpha} \cdot \hat{r}, \tag{3.173}$$

satisfies all the requirements outlined above. It can also be shown that

$$n' = n - j - 1/2 + \sqrt{(j+1/2)^2 - \alpha^2 Z^2}, \tag{3.174}$$

where n is the non-relativistic principle quantum number. The constant α is the well known fine structure constant

$$\alpha = \frac{e^2}{\hbar c}, \tag{3.175}$$

which has the numerical value

$$\alpha^{-1} = 137.035999084. \tag{3.176}$$

It should not be confused with the matrices in $\vec{\alpha}$.

Eliminating the primed quantities from equations (3.170), (3.171) and (3.172) leads to the expression for the energy eigenvalue of the relativistic equation

$$E = \frac{mc^2}{\left[1 + \dfrac{\alpha^2 Z^2}{n'^2}\right]^{1/2}}. \tag{3.177}$$

This is perhaps the easiest way to find the energy eigenvalues of the Dirac equation. It also has the advantage that it is easy to relate these eigenvalues to the ones for the non-relativistic Schrödinger equation. For example, Taylor's expansion of the energy in equation (3.177) in powers of $(\alpha Z)^2$ is

$$E = mc^2 - 1/2mc^2\left(\frac{\alpha Z}{n'}\right)^2 + 3/8mc^2\left(\frac{\alpha Z}{n'}\right)^4 + \ldots. \tag{3.178}$$

The expansion of n' in powers of $(\alpha Z)^2$ is

$$n' = n - \frac{(\alpha Z)^2}{2j+1} + \ldots. \tag{3.179}$$

Ignoring the second term in this equation, the second term in equation (3.178) is just the non-relativistic Schrödinger energy. The terms of order $(\alpha Z)^4$ and higher are the relativistic corrections. They clearly depend on j as well as n.

One of the first successes of the Dirac equation was the calculation of the splitting of the energy levels of the hydrogen atom. The Schrödinger equation predicts that the 2s and 2p levels are degenerate. They are split by the spin–orbit effects in the Dirac equation. The splitting of the Lyman-alpha spectral lines that is the consequence of this effect was seen experimentally years before the Dirac equation was proposed.

3.12 Advantages and disadvantages

The energy levels and total energy of atoms with the atomic number $Z>1$ can be calculated with relativistic Hartree–Fock codes based on the Dirac equation. This technique will be discussed in more detail later. The results of these calculations

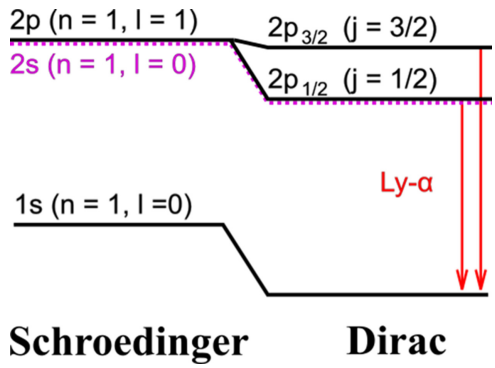

Figure 3.2. The lowest few energy levels of the hydrogen atom calculated with the Schrödinger and the Dirac equations.

shown in figure 3.2 agree very well with experiment. The difference between them and non-relativistic calculations increases with increasing Z. However, the solution of the Dirac equation for the 1S electron is oscillatory and unbounded for $Z \geqslant 137$. This can be seen from equation (3.174) because the square root becomes imaginary for such values of $j + 1/2 = 1$ and $Z \geqslant 137$. It could be argued that there is no atom on the periodic table with a Z that large, but this still indicates that Dirac equation is fundamentally limited as a theory of electrons.

Another feature of the Dirac equation that has led to great debate over the years is the occurrence of negative energy solutions with $E < -mc^2$. Dirac suggested that, in the equilibrium state, these negative energy levels are filled with two electrons each. In an excited state, an electron goes into a positive energy state, leaving a hole in the negative energy sea. This hole behaves like a particle of positive charge. The only positive charged particle known at the time was a proton, so Dirac gave the 1930 paper in which he put forward this theory 'A Theory of Electrons and Protons'. It was shown theoretically that Dirac's theory could not work unless the positively charged particle had the same mass as an electron. Dirac's prediction seemed to be vindicated in 1932 when Carl D Anderson discovered the positron in a cloud chamber. This discovery added to the enthusiasm for Dirac's theory, and he shared the 1933 Nobel prize with Schrödinger. Dirac's idea of an infinite sea of electrons with negative energies is not only illogical, but efforts to use it to make predictions of specific phenomena were not successful.

It is now generally agreed that the explanation for the difficulties in Dirac's theory is that his theory starts with ordinary mechanics in which potential functions are used to describe interactions between particles. With the development of quantum electrodynamics and field theory, this picture has been replaced with one in which the force between particles is caused by the exchange of virtual particles. For example, the Coulomb force between electrons is mediated by the exchange of photons. The forces between protons and neutrons are mediated by pi and rho mesons. From this point of view, the surprise is that Dirac's theory works so well for atoms with reasonable atomic numbers.

The pair production of a positron and an electron, or the reverse interaction in which the positron and electron are annihilated with the resulting energy being

carried away by photons is also best explained by field theory. In 1973 Julian Schwinger remarked, 'The picture of an infinite sea of negative energy electrons is now best regarded as a historical curiosity, and forgotten'.

Schwinger's remark does not mean the Dirac equation should be forgotten. The field theory for spin one half particles is based on the Dirac equation. Also, as mentioned before, practical calculations of the energetics, chemical behavior, cohesion, and magnetic properties of atoms, molecules, and solids require the solution of the Dirac equation with the help of high speed supercomputers.

Problems

P3.1 Find in the literature the relative size of relativistic corrections on the total energy of atoms as a function of atomic number.

P3.2 Derive equation (3.22) from equation (3.15) putting in all the steps.

P3.3 What is the difference between the ground state of the hydrogen atom as obtained from the Schröedinger equation and the Klein–Gordon equation?

P3.4 Use the properties of the Pauli spin matrices to derive equations (3.35), (3.36), and (3.37).

P3.5 Derive an equation similar to equations (3.21) and (3.22) for the Dirac equation.

P3.6 What is the difference between a 4×1 matrix and a bispinor?

P3.7 A transformation in space and time, like the one in equation (3.58), is called a boost. Use such a transformation to find the wave function for a particle moving in the y direction.

P3.8 Extend the preceding proofs by calculating $[H, L_y]$,$[H, L_z]$,$[H, \Sigma_x]$, and $[H, \Sigma_y]$.

P3.9 Use Maxwell's equations to show that the classical magnetic moment is proportional to the angular momentum $\vec{m} = \frac{e}{2m}\vec{L}$.

P3.10 Derive an expression for the zitterbewegung of a moving electron.

P3.11 Use the expansion of n' to get the correct expression for the lowest relativistic corrections to E from equation (3.178).

References

[1] Strange P 1998 *Relativistic Quantum Mechanics: With Applications in Condensed Matter and Atomic Physics* (Cambridge: Cambridge University Press)

[2] Mohn P 2003 *Magnetism in the Solid State: An Introduction* (*Springer Series in Solid-State Sciences*) (Berlin: Springer)

[3] Reiher M and Wolf A 2009 *Relativistic Quantum Chemistry* (New York: Wiley)

[4] Klein O 1926 *Z Phys.* **37** 895

[5] Gordon W 1926 *Z Phys.* **40** 117

[6] Dirac P A M 1928 *Proc. R. Soc. Lond.* **A117** 610 ibid **A118** 351

[7] Pauli W 1936 *Ann. Inst. Henri Poincare* **6** 109

[8] Koelling D D and Harmon B N 2001 *J. Phys.* C **10** 3107

[9] Martin P C and Glauber R J 1958 *Phys. Rev.* **109** 1307

[10] Baym G 1969 Lectures on quantum mechanics *Lecture Notes and Supplements in Physics* (New York: The Benjamin/Cummings Publishing Company)