Analytic Representation of Relativstic Wave Equations I: The Dirac Case

Tepper L. Gill[1,2] **and W. W. Zachary**[1]

[1]Department of Electrical and Computer Engineering
[2]Department of Mathematics
Howard University
Washington, D. C. 20059

Abstract

In this paper we construct an analytical separation (diagonalization) of the full (minimal coupling) Dirac equation into particle and antiparticle components. The diagonalization is analytic in that it is achieved without transforming the wave functions, as is done by the Foldy-Wouthuysen method, and reveals the nonlocal time behavior of the particle-antiparticle relationship. It is well known that the Foldy-Wouthuysen transformation leads to a diagonalization that is nonlocal in space. We interpret the zitterbewegung, and the result that a velocity measurement (of a Dirac particle) at any instant in time is $\pm c$, as reflections of the fact that the Dirac equation makes a spatially extended particle appear as a point in the present by forcing it to oscillate between the past and future at speed c. This suggests that although the Dirac Hamiltonian $\mathbf{H}_{\mathrm{D}}$ and the square-root Hamiltonian, $\mathbf{H}_{\mathrm{s}} = U_{FW}\mathbf{H}_{\mathrm{D}}U_{FW}^{-1}$, are mathematically, they are not "physically," equivalent. Furthermore, we see that although the form of the Dirac equation serves to make space and time appear on an equal footing mathematically, they are still not on an equal footing from a physical point of view. It appears that the only way to justify a physical relationship between the Dirac and the square-root equations is via their relationship to the Klein-Gordon equation.

We then show explicitly that the Pauli equation is not valid for an analysis of the Dirac hydrogen atom problem in s-states (hyperfine splitting). We conclude that there are serious physical and mathematical problems with any attempt to show that the Dirac equation is insufficient to explain the full hydrogen spectrum.

I. Introduction

A standard assertion is that quantum electrodynamics (QED) is an almost perfect theory that is in excellent agreement with experiments. Those readers with some operational contact with the subject will also point out that they are not happy with the divergences and the renormalization procedures, and that a meaningful mathematical formulation is still missing after more than fifty years. Furthermore, in a real sense, the difficult computational analysis seems a rather high price for the meager (but important) results obtained.

Most pedestrians and too many experts are unaware of, or choose to ignore, the fact that QED does not account for the complete spectrum of hydrogen. All but the Lamb shift and a minor portion of the hyperfine splitting must be computed from a combination of the Dirac equation and the corresponding Pauli approximation. These results are then given as input to the QED computation. This is, of course, irrelevant from a practical point of view, and there are other important applications. However, from a foundational (and/or theoretical) point of view, to accept this state of affairs is equivalent to changing the definition of what we mean by a physical theory.

II. Purpose

Historically, when Lamb and Retherford[1] confirmed suspicions that the $2s_{1/2}$ state hydrogen was shifted above the $2p_{1/2}$ state, the Pauli approximation to the Dirac equation was (essentially) used to decide that the Dirac equation was not sufficient. In light of the above issues and the tremendous success (historically) of eigenvalue analysis in physics and engineering, it is not inappropriate for us to reinvestigate the physical and mathematical foundations of spin 1/2 particles with an eye towards identifying the conceptual and technical limitations to our understanding of the hydrogen spectrum as an eigenvalue problem.

The first successful attempt to resolve the question of how best to handle the square-root equation:

$$i\hbar\frac{\partial\Psi}{\partial t}=\boldsymbol{\beta}\left[\sqrt{c^2\mathbf{p}^2+m^2c^4}\right]\Psi,\quad \boldsymbol{\beta}=\begin{bmatrix} I_2 & 0 \\ 0 & -I_2 \end{bmatrix}, \tag{1}$$

was made by Dirac[2] in 1926. Dirac noticed that the Pauli matrices could be used to write $c^2\mathbf{p}^2+m^2c^4$ as $\left[c\boldsymbol{\alpha}\cdot\mathbf{p}+mc^2\boldsymbol{\beta}\right]^2$. The matrix $\boldsymbol{\alpha}$ is defined (in the standard representation) by $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3)$, where

$$\alpha_i=\begin{pmatrix}\mathbf{0} & \sigma_i\\ \sigma_i & \mathbf{0}\end{pmatrix},\ \sigma_1=\begin{pmatrix}\mathbf{0} & 1\\ 1 & \mathbf{0}\end{pmatrix},\ \sigma_2=\begin{pmatrix}\mathbf{0} & -i\\ i & \mathbf{0}\end{pmatrix},\ \sigma_3=\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{-1}\end{pmatrix}.$$

This means that equation (1) can be written as:

$$i\hbar\frac{\partial\Psi}{\partial t}=\left[c\boldsymbol{\alpha}\cdot\mathbf{p}+mc^2\boldsymbol{\beta}\right]\Psi. \tag{2}$$

In this case, Ψ must be viewed as a vector valued function or spinor. To be more precise, $\Psi\in L^2\left(\mathbf{R}^3,\mathbf{C}^4\right)=L^2\left(\mathbf{R}^3\right)\otimes\mathbf{C}^4$ is a four-component column vector $\Psi=\left(\psi_1,\psi_2,\varphi_1,\varphi_2\right)^t$. In this approach, $\psi=\left(\psi_1,\psi_2\right)^t$ represents the particle (positive energy) component, and $\varphi=\left(\varphi_1,\varphi_2\right)^t$ represents the antiparticle (negative energy) component of the theory (for details, see Thaller[3]).

A fair understanding of the Dirac equation can only be claimed in recent times, and we now realize that Dirac replaced the complex number field by elements of a Clifford algebra. (See, in particular, Biedenharn[4] or deVries[5] along with the references therein. For a general reference to Clifford algebras, the work by Hestenes[6] offers a good introduction.)

Despite successes, both practical and theoretical, there still remain a number of conceptual, interpretational, and technical misunderstandings about this equation. It is generally believed that it is not possible to separate the particle and antiparticle components directly without approximations (when interactions are present). The crude, and sometimes incorrect, approximations found in the literature might have led to this belief. In addition, the algebraic approaches of Foldy-Wouthysen[7], Pauli[8], and Feynman and Gell-Mann[9] have no doubt further supported such ideas.

III. Complete Separation

It turns out that a direct analytic separation is actually quite simple and provides additional insight into the particle and antiparticle components. In order to see this, let $\mathbf{A}(\mathbf{x},t)$ and $V(\mathbf{x})$ be given vector and scalar potentials and, after adding $V(\mathbf{x})$ and making the transformation $\mathbf{p}\rightarrow\boldsymbol{\pi}=\mathbf{p}-(e/c)\mathbf{A}$, write (2) in two-component form as:

$$i\hbar\frac{\partial\psi}{\partial t}=(V+mc^2)\psi+c(\sigma\cdot\boldsymbol{\pi})\varphi, \tag{3a}$$

$$i\hbar\frac{\partial\varphi}{\partial t}=(V-mc^2)\varphi+c(\sigma\cdot\boldsymbol{\pi})\psi. \tag{3b}$$

Equation (3b) can be written in the form:

$$\left[\frac{\partial}{\partial t}+iB\right]\varphi=D\psi, \tag{4}$$

with $B = \left[(V - mc^2)/\hbar\right]$ and $D = \left[c(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]$. From an analytical point of view, we see that equation (4) is an inhomogeneous partial differential equation. This equation can be solved via the Green's function method. Thus, we then must solve

$$\left[\frac{\partial}{\partial t} + iB\right]\varphi = \delta(t). \tag{5}$$

It is easy to see that the solution to equation (5) is

$$u(t) = \theta(t)\exp\{-iBt\}, \quad \theta(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \end{cases}, \tag{6}$$

so that,

$$\varphi(t) = cu(t) * \left[(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\psi(t) = \int_{-\infty}^{\infty} c\theta(t-\tau)\exp\{-iB(t-\tau)\}\left[(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\psi(\tau)d\tau, \tag{7a}$$

$$\varphi(t) = \int_{-\infty}^{t} c\exp\{-iB(t-\tau)\}\left[(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\psi(\tau)d\tau. \tag{7b}$$

Using equation (7) in (3a), we have

$$i\hbar\frac{\partial\psi}{\partial t} = (V + mc^2)\psi + \left[c^2(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\int_{-\infty}^{t}\exp\{-iB(t-\tau)\}(\sigma \cdot \boldsymbol{\pi})\psi(\tau)d\tau. \tag{8}$$

In a similar manner, we obtain the complete equation for φ:

$$i\hbar\frac{\partial\varphi}{\partial t} = (V - mc^2)\varphi + \left[c^2(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\int_{-\infty}^{t}\exp\{-iB'(t-\tau)\}(\sigma \cdot \boldsymbol{\pi})\varphi(\tau)d\tau, \tag{9}$$

where $B' = \left[(V + mc^2)/\hbar\right]$ and $v(t) = \theta(t)\exp\{-iB't\}$, which allows us to solve for ψ:

$$\psi(t) = cv(t) * [(\sigma \cdot \boldsymbol{\pi})/i\hbar]\varphi(t) = \int_{-\infty}^{t} c\exp\{-iB'(t-\tau)\}\left[(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\varphi(\tau)d\tau. \tag{10}$$

Thus, we have decomposed $L^2\left(\mathbf{R}^3, \mathbf{C}^4\right)$ as $L^2\left(\mathbf{R}^3, \mathbf{C}^4\right) = L^2\left(\mathbf{R}^3, \mathbf{C}^2\right) \oplus L^2\left(\mathbf{R}^3, \mathbf{C}^2\right)$. The first copy of $L^2\left(\mathbf{R}^3, \mathbf{C}^2\right)$ contains the particle (positive energy) wave component $\psi = \left(\psi_1, \psi_2\right)^t$, while the second copy of $L^2\left(\mathbf{R}^3, \mathbf{C}^2\right)$ contains the antiparticle (negative energy) wave component $\varphi = \left(\varphi_1, \varphi_2\right)^t$. The only unsettled issue is the definition of the appropriate inner product for the two subspaces, which will account for the quantum constraint that the total probability integral is normalized. We can satisfy this requirement if we set $(\psi, \chi) = \psi_1\overline{\chi}_1 + \psi_2\overline{\chi}_2$, and $(\psi, \chi)_A = (A\psi, A\chi)$, $(\varphi, \eta)_{A'} = (A'\varphi, A'\eta)$, where $A\psi = cu(t) * \left[(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\psi(t)$, $A'\varphi = cv(t) * \left[(\sigma \cdot \boldsymbol{\pi})/i\hbar\right]\varphi(t)$. Define the particle and antiparticle inner products by

$$\langle \psi, \chi \rangle_p = \int_{\mathbf{R}^3} \{(\psi, \chi) + (\psi, \chi)_A\} d\mathbf{x}, \tag{11a}$$

$$\langle \varphi, \eta \rangle_{ap} = \int_{\mathbf{R}^3} \{(\varphi, \eta) + (\varphi, \eta)_{A'}\} d\mathbf{x}, \tag{12a}$$

so that:

$$\rho_\psi = |\psi|^2 + \left| \int_{-\infty}^{t} c \exp\{-iB(t-\tau)\} [(\sigma \cdot \boldsymbol{\pi})/i\hbar] \psi(\tau) d\tau \right|^2, \tag{11b}$$

$$\rho_\varphi = |\varphi|^2 + \left| \int_{-\infty}^{t} c \exp\{-iB'(t-\tau)\} [(\sigma \cdot \boldsymbol{\pi})/i\hbar] \varphi(\tau) d\tau \right|^2. \tag{12b}$$

It is clear that $\int_{\mathbf{R}^3} \rho_\psi dx = \int_{\mathbf{R}^3} \rho_\varphi dx = 1$, so we now have a complete separation of the particle and antiparticle wave functions.

In the standard representation, the charge conjugation operator is $\mathbf{C}\psi = U_C \overline{\psi}$, with $U_C = i\boldsymbol{\beta}\alpha_2$. A simple computation establishes the following theorem.

Theorem 1. *Equations (8) and (9) are mapped into each other under the charge conjugation transformation*.

Equations (8) and (9) offer an interesting alternative to the many attempts to decompose the Dirac equation into particle-antiparticle and/or parity-sensitive pairs. They also offer a different approach to the study of large Z (hydrogen-like) atoms. Although not a part of our direction, one should be able to show that (under physically) reasonable conditions, equation (8) is stable in the large Z limit for such atoms.

IV. Interpretations

Writing the Dirac equation and the direct separation in two-component matrix form, we have:

$$i\hbar \frac{\partial}{\partial t} \begin{bmatrix} \psi \\ \varphi \end{bmatrix} = \begin{bmatrix} (V + mc^2) & c(\sigma \cdot \boldsymbol{\pi}) \\ c(\sigma \cdot \boldsymbol{\pi}) & (V - mc^2) \end{bmatrix} \begin{bmatrix} \psi \\ \varphi \end{bmatrix}, \tag{13}$$

and

$$i\hbar \frac{\partial}{\partial t} \begin{bmatrix} \psi \\ \varphi \end{bmatrix} = \begin{bmatrix} (V + mc^2) & \mathbf{0} \\ +\left[c^2(\sigma \cdot \boldsymbol{\pi})/i\hbar\right][u * (\sigma \cdot \boldsymbol{\pi})] & \\ \mathbf{0} & (V - mc^2) \\ & +\left[c^2(\sigma \cdot \boldsymbol{\pi})/i\hbar\right][v * (\sigma \cdot \boldsymbol{\pi})] \end{bmatrix} \begin{bmatrix} \psi \\ \varphi \end{bmatrix}. \tag{14}$$

We call (14) the *analytic diagonalization* of the Dirac equation because the wave function has not changed.

The standard approach to the diagonalization of the Dirac equation (without an external potential V) is via the Foldy-Wouthuysen representation. Assuming that $\mathbf{A}$ does not depend on t, the following generalization can be found in deVries[5]:

$$i\hbar\frac{\partial}{\partial t}\begin{bmatrix}\Phi_1\\ \Phi_2\end{bmatrix}=\begin{bmatrix}\sqrt{c^2\boldsymbol{\pi}^2-\frac{e}{c}(\sigma\cdot\mathbf{B})+m^2c^4} & 0\\ 0 & -\sqrt{c^2\boldsymbol{\pi}^2-\frac{e}{c}(\sigma\cdot\mathbf{B})+m^2c^4}\end{bmatrix}\begin{bmatrix}\Phi_1\\ \Phi_2\end{bmatrix}. \tag{15}$$

In this case, $\begin{bmatrix}\Phi_1 & \Phi_2\end{bmatrix}^t = U_{FW}\begin{bmatrix}\psi & \varphi\end{bmatrix}^t$, and $\mathbf{H}_\mathrm{s} = U_{FW}\mathbf{H}_\mathrm{D}U_{FW}^{-1}$ (see Thaller[3] for details). Equation (15) is the object of the next paper of this series, and will be studied in detail there. However, it is known (see Gill[10]) that when $\mathbf{A}$ is zero, $\mathbf{H}_\mathrm{s}=\boldsymbol{\beta}\sqrt{c^2\mathbf{p}^2+m^2c^4}$ has the following analytic representation:

$$\mathbf{H}_\mathrm{s}f(\mathbf{x}) =$$
$$-\frac{\mu^2\hbar^2c\boldsymbol{\beta}}{\pi^2}\int_{\mathbf{R}^3}\left\{\left[\frac{\mathbf{K}_0\left[\mu\|\mathbf{x}-\mathbf{y}\|\right]}{\|\mathbf{x}-\mathbf{y}\|}+\frac{2\mathbf{K}_1\left[\mu\|\mathbf{x}-\mathbf{y}\|\right]}{\mu\|\mathbf{x}-\mathbf{y}\|^2}\right]\left[\frac{1}{\|\mathbf{x}-\mathbf{y}\|}-2\pi\delta(\mathbf{x}-\mathbf{y})\right]\right\}f(\mathbf{y})d\mathbf{y}. \tag{16}$$

Here, the $\mathbf{K}_n$ are modified Bessel functions of the third kind and $\mu = mc/\hbar$. *Equation* (16) *is the first example of a physically relevant operator, which has a "natural" representation as the confinement of a composite of three singularities, two negative and one (hard core) positive, within a Compton wavelength such that, at the point of singularity, they cancel each other providing a finite result.*

Gill first derived equation (16) using the method of fractional powers of closed operators. As will be shown in the follow-up paper, this method (simplified) has the advantage of allowing us to construct corresponding representations for a large class of vector potentials. Loss and Lieb[11] derive a version of (16) using Fourier methods. However, this approach does not generalize to nonzero $\mathbf{A}$ and/or nonconstant μ, both of which are of independent interest.

We can now interpret the zitterbewegung, and the result that a velocity measurement (of a Dirac particle) at any instant in time is $\pm c$, as reflections of the fact that the Dirac equation makes a spatially extended particle appear as a point in the present by forcing it to oscillate between the past and future at speed c.

From equation (14), we conclude that the coupling of the particle and antiparticle wave functions in the first-order form of the Dirac equation hides the second order nonlocal time nature of the equation. From (16), we see explicitly that (15) is nonlocal in space. *Thus, the implicit time nonlocality of the Dirac equation is mapped into the*

explicit spatial nonlocality of the square-root equation by the Foldy-Wouthuysen transformation. These observations imply that the Dirac Hamiltonian $\mathbf{H}_{\mathrm{D}}$ and the square-root Hamiltonian, $\mathbf{H}_{\mathrm{s}} = U_{FW}\mathbf{H}_{\mathrm{D}}U_{FW}^{-1}$, are mathematically, but not physically, equivalent. Furthermore, the only way we can justify using the square-root equation to interpret the Dirac equation is their relation to the Klein-Gordon equation. In the second paper, we show that there is an alternate relationship between the square-root and Dirac equations.

V. The Hydrogen Atom

In this section, we reconsider the standard analysis of the Dirac equation for the hydrogen atom problem from an exact point of view. We assume that $\mathbf{A} = (\boldsymbol{\mu}_I \times \mathbf{r})/r^3$, $V = -\hbar c\gamma/r$, and $\gamma = e^2/\hbar c$. Rewrite (3a) and (3b) in eigenvalue form:

$$(E - V - mc^2)\psi = c(\sigma \cdot \boldsymbol{\pi})\varphi, \tag{17a}$$

$$(E - V + mc^2)\varphi = c(\sigma \cdot \boldsymbol{\pi})\psi. \tag{17b}$$

Eliminating φ in terms of ψ and vice versa, we obtain the following equations:

$$(E - V - mc^2)\psi = \frac{c^2(\sigma \cdot \mathbf{p}V)(\sigma \cdot \boldsymbol{\pi})}{(E - V + mc^2)^2}\psi + \frac{c^2(\sigma \cdot \boldsymbol{\pi})(\sigma \cdot \boldsymbol{\pi})}{(E - V + mc^2)}\psi, \tag{18a}$$

$$(E - V + mc^2)\varphi = \frac{c^2(\sigma \cdot \mathbf{p}V)(\sigma \cdot \boldsymbol{\pi})}{(E - V - mc^2)^2}\varphi + \frac{c^2(\sigma \cdot \boldsymbol{\pi})(\sigma \cdot \boldsymbol{\pi})}{(E - V - mc^2)}\varphi. \tag{18b}$$

(We also get (18) from equations (8) and (9) via straightforward integration, using the Riemann-Lebesgue Lemma.) We call (18a) and (18b) the Slater equations since they were first used by one of his students as early as 1940[12], and appeared in his book[13], first published in 1960 (see Appendix 29). (It is difficult to understand, given the continuous historical interest in the Dirac equation, why Slater's work on this equation is not well known.) For obvious reasons, we concentrate on (18a). First note that, if we drop the middle term and replace $(E - V + mc^2)$ by $2mc^2$, we get the Pauli approximation to the Dirac equation:

$$(E - V - mc^2)\psi = -\frac{e\hbar}{2mc}(\sigma \cdot \mathbf{B})\psi + \frac{\boldsymbol{\pi}^2}{2m}\psi. \tag{19}$$

As noted earlier, the Pauli equation was used to extract the hyperfine splitting portion of the hydrogen spectrum to complete the predictions of QED. It follows that the conditions that justify the Pauli approximation and the dropping of the middle term of (18a) are both of fundamental importance for the foundations of QED.

There are a number of other equations and/or "apparent" approximations that have been given the name and/or used in lieu of the Pauli equation (see for example, Greiner[14], Mizushima[15] or Bethe and Salpeter[16]). We do not consider these equations since, although they are related to the Dirac equation, they do not give additional information and it is far from obvious that they have any mathematical or physical justification when applied to the only calculation that matters, the s-states of hydrogen.

Recall that there is a finite probability of finding the electron at the origin in s-states, but the required condition for the validity of (19) is $(E-V+mc^2) << 2mc^2$. Thus, this condition is not satisfied for any s-state calculation. *It follows that, from a foundational point of view, use of the Pauli equation to compute the hyperfine splitting of s-states is far from convincing.* On the other hand, the condition is easily seen to be satisfied for all other states. A more reasonable approximation is to use $|mc^2 - E| << mc^2$ to replace $(E-V+mc^2)$ by $2mc^2(1+r_0/r)$, where $r_0 = e^2/(E+mc^2) \cong e^2/2mc^2$. The above condition is always satisfied ($13ev$ compared to $0.5Mev$). This approach also has the additional advantage of removing the nonlinear eigenvalue problem posed by (18a) without substantially affecting the final result. In this case we have

$$(E-V-mc^2)\psi = \frac{(\sigma\cdot\mathbf{p}V)(\sigma\cdot\boldsymbol{\pi})}{4m^2c^2(1+r_0/r)^2}\psi + \frac{(\sigma\cdot\boldsymbol{\pi})(\sigma\cdot\boldsymbol{\pi})}{2m(1+r_0/r)}\psi. \tag{20}$$

Using standard computations, we get (see Slater[13], $\hbar\mathbf{L} = \mathbf{r}\times\mathbf{p}$ is the angular momentum, and $\hbar\mathbf{S}$ is the spin, $\mathbf{S} = \sigma/2$)

$$(\sigma\cdot\boldsymbol{\pi})(\sigma\cdot\boldsymbol{\pi}) = \boldsymbol{\pi}^2 - \frac{2e\hbar}{c}\left\{\frac{8\pi}{3}(\mathbf{S}\cdot\boldsymbol{\mu}_I)\delta(\mathbf{r}) + \left[\frac{3(\mathbf{S}\cdot\mathbf{r})(\boldsymbol{\mu}_I\cdot\mathbf{r})}{r^5} - \frac{(\mathbf{S}\cdot\boldsymbol{\mu}_I)}{r^3}\right]\right\}, \tag{21a}$$

$$(\sigma\cdot\mathbf{p}V)(\sigma\cdot\boldsymbol{\pi}) = \frac{2e^2\hbar}{r^2}\left\{\hbar\left[\frac{(\mathbf{S}\cdot\mathbf{L})}{r} - \frac{d}{dr}\right] + \frac{e}{c}\left[\frac{(\mathbf{S}\cdot\boldsymbol{\mu}_I)}{r^2} - \frac{(\mathbf{S}\cdot\mathbf{r})(\boldsymbol{\mu}_I\cdot\mathbf{r})}{r^4}\right]\right\}. \tag{21b}$$

Putting these expressions in (20), we have:

$$\begin{aligned}(E-V-mc^2)\psi = {} & \frac{r_0\hbar}{m(1+r_0/r)^2r^2}\left\{\hbar\left[\frac{(\mathbf{S}\cdot\mathbf{L})}{r} - \frac{d}{dr}\right] + \frac{e}{c}\left[\frac{(\mathbf{S}\cdot\boldsymbol{\mu}_I)}{r^2} - \frac{(\mathbf{S}\cdot\mathbf{r})(\boldsymbol{\mu}_I\cdot\mathbf{r})}{r^4}\right]\right\}\psi \\ & - \frac{e\hbar}{mc(1+r_0/r)}\left\{\frac{8\pi}{3}(\mathbf{S}\cdot\boldsymbol{\mu}_I)\delta(\mathbf{r}) + \left[\frac{3(\mathbf{S}\cdot\mathbf{r})(\boldsymbol{\mu}_I\cdot\mathbf{r})}{r^5} - \frac{(\mathbf{S}\cdot\boldsymbol{\mu}_I)}{r^3}\right]\right\}\psi + \frac{\boldsymbol{\pi}^2}{2m(1+r_0/r)}\psi\end{aligned}. \tag{22}$$

When $\boldsymbol{\mu}_I = 0$, (22) becomes

$$(E-V-mc^2)\psi = \frac{r_0\hbar^2}{m(1+r_0/r)^2r^2}\left[\frac{(\mathbf{S}\cdot\mathbf{L})}{r} - \frac{d}{dr}\right]\psi + \frac{\mathbf{p}^2}{2m(1+r_0/r)}\psi. \tag{23}$$

Equation (23) has (using $r_0 = e^2/(E+mc^2)$) the same eigenvalues as the unperturbed Dirac equation so that our interest centers on the following terms (op means operator)

$$-\frac{e\hbar}{mc(1+r_0/r)}\left\{\frac{8\pi}{3}(\mathbf{S}\cdot\boldsymbol{\mu}_I)\delta(\mathbf{r})+\left[\frac{3(\mathbf{S}\cdot\mathbf{r})(\boldsymbol{\mu}_I\cdot\mathbf{r})}{r^5}-\frac{(\mathbf{S}\cdot\boldsymbol{\mu}_I)}{r^3}\right]_{op}\right\}, \qquad (24a)$$

$$\frac{er_0\hbar}{mc(1+r_0/r)^2 r^2}\left[\frac{(\mathbf{S}\cdot\boldsymbol{\mu}_I)}{r^2}-\frac{(\mathbf{S}\cdot\mathbf{r})(\boldsymbol{\mu}_I\cdot\mathbf{r})}{r^4}\right]_{op}. \qquad (24b)$$

The delta term in equation (24a), except for the additional factor $(1+r_0/r)^{-1}$, would normally be used to compute the hyperfine splitting of s-states in the Pauli approximation. It is easy to see that, with this additional factor, the same calculation would give a value of zero for the splitting. In all other states this factor is small $(1>>r_0/r)$ and may be dropped.

Slater[13] used equation (24b) to compute the s-state (hyperfine) splitting and obtained the correct result. Since this term is (part of the) focus of our investigation, we repeat some of Slater's calculations. In the s-state the total angular momentum **J** is equal to **S**. Hence, following standard procedures, we replace $\left[(\mathbf{S}\cdot\boldsymbol{\mu}_I/r^2)-((\mathbf{S}\cdot\mathbf{r})(\boldsymbol{\mu}_I\cdot\mathbf{r})/r^4)\right]_{op}$ by $(\mathbf{S}\cdot\boldsymbol{\mu}_I/\mathbf{S}^2)\left[(\mathbf{S}^2/r^2)-((\mathbf{S}\cdot\mathbf{r})^2/r^4)\right]_{op}$. It is easy to see that, as far as operator averages are concerned, $\left[(\mathbf{S}\cdot\mathbf{r})^2\right]_{op}=\frac{1}{4}(r)^2_{op}$ and $\left[\mathbf{S}^2\right]_{op}=\frac{3}{4}$. The term of interest becomes

$$\frac{2er_0\hbar}{3mc(1+r_0/r)^2 r^4}(\mathbf{S}\cdot\boldsymbol{\mu}_I)_{op}. \qquad (25a)$$

The important issue is the computation of the s-state expected value of

$$\frac{r_0\lambda}{(1+r_0/r)^2 r^4}, \qquad (25b)$$

where $\lambda = 2e\hbar/3mc\left\langle(\mathbf{S}\cdot\boldsymbol{\mu}_I)_{op}\right\rangle_{ave}$. Slater[13] assumed the nonrelativistic radial wave function for s-states. (For the $2s_{1/2}$ state, $R(r)=\frac{1}{\sqrt{2}}\eta^{3/2}(1-\frac{1}{2}r\eta)\exp(-\frac{1}{2}r\eta)$ and $\eta=1/r_B$, where $r_B = 0.529178\times10^{-10}m$ is the Bohr radius.) Using the normalization $\int_0^\infty r^2R(r)^2dr=1$, this led him to the computation of

$$\lambda\int_0^\infty\frac{r_0R(r)^2r^2}{(1+r_0/r)^2r^4}dr. \qquad (26a)$$

Setting $\rho=\eta r$ and $\rho_0=\eta r_0$, we have

$$\tfrac{1}{2}\eta\rho_0\lambda\int_0^\infty \frac{(1-\tfrac{1}{2}\rho)^2\exp(-\rho)d\rho}{(1+\rho_0/\rho)^2\rho^2}. \qquad (26b)$$

By a change of variables ($u=\rho+\rho_0$) and integration by parts, it is easy to see that ρ_0 is a cutoff and that the dominant contribution is ρ_0^{-1}. We get the same result for all s-states, while it is not hard to show that equation (24b) is (almost) zero for all other states.

It would appear that the correct approach for s-state (hyperfine) splitting gives the same results as those obtained from the Pauli equation. Furthermore, equation (24b) introduces a natural cutoff, which removes the conceptual difficulty of a point magnetic dipole interaction as implied by use of the delta term in the Pauli equation. In addition, it is not hard to show that Slater's approach goes through, giving the same result, if we use (the correct) Dirac solution for the first-order calculation.

However, to provide input for the precise results of QED, we must first correct the normalization condition to

$$\int\left[|\psi|^2+|(\sigma\cdot\mathbf{p})\psi|^2\Big/4m^2c^2(1+r_0/r)^2\right]d\mathbf{x}=1. \qquad (27)$$

Clearly, the additional term will give a small correction. However, if it changes the hyperfine splitting values in the eight or ninth decimal place (in GHz), it is important. (For example, the measured values of the $2s_{1/2}$ state hyperfine splitting in hydrogen is 0.177566850(10) Ghz, see Mizushima[15].) As we shall see later, the correction is much smaller.

Problem 1. (Physical)

The first problem for the physical foundations is to provide a complete computational analysis of equation (25a) to the second order using (27) for input in QED.

Problem 1. (Mathematical)

The first problem for the mathematical foundations of QED is to prove that perturbation theory can (or cannot) be applied to equation (25a) (using (27)).

We now approach the more difficult issue facing any attempt to completely understand the Dirac problem for full coupling, namely, the $\mathbf{A}^2$ term:

$$\frac{e^2\mathbf{A}^2}{2mc^2(1+r_0/r)}=\frac{e^2\mu_I^2\sin^2\theta}{2mc^2(1+r_0/r)r^4}. \qquad (28)$$

In every treatment of the Dirac hydrogen atom problem, this term (with $\rho_0=0$) is either ignored or assumed to be small. Clearly, it cannot be ignored in any investigation of the

foundations. Furthermore, it is easy to show that this term will be small in all except s-states.

The first observation is that this term appears to be more singular than the Coulomb potential, so that perturbation analysis may not be appropriate. However, this is not completely clear since the $\sin^2\theta$ term vanishes on the spin axis and could strongly modify the singular nature of this term. This brings us to another set of important problems for the foundations of QED:

Problem 2. (Physical)

The second problem for the physical foundations of quantum electrodynamics is to construct a complete solution of the eigenvalue problem for the Dirac equation with the Coulomb and magnetic dipole interaction (or equivalently (18a)).

This problem must be solved in order to provide conclusive support for the assumption that the Dirac equation does not account for the complete spectrum of hydrogen.

Problem 2. (Mathematical)

The second problem for the mathematical foundations of quantum electrodynamics is to prove or disprove that the $\mathbf{A}^2$ term can be treated as a perturbation of the Coulomb problem.

If we take an engineering approach and assume that we can treat the $\mathbf{A}^2$ term as a perturbation, then for the 2s- state the expected value is

$$\int_0^\infty \frac{r_0\mu_I^2 R(r)^2 r^2}{(1+r_0/r)r^4}\,dr\int_0^\pi \sin^2\theta(\sin\theta)d\theta = \tfrac{1}{3}\eta^3\rho_0\mu_I^2\int_0^\infty \frac{(1-\frac{1}{2}\rho)^2\exp(-\rho)d\rho}{(1+\rho_0/\rho)\rho^2}. \qquad (29)$$

In atomic units, $\eta = 1$, $\mu_0 = (1/2)\gamma$, $r_0 = \rho_0 = (1/2)\gamma^2$, $g_N^2 = 30.9136$, $\left\langle \mathbf{I}_{op}^2 \right\rangle_{ave} = (3/4)$ and $\mu_I^2 = (1/1836)^2 g_N^2 \mu_0^2 \mathbf{I}_{op}^2$, so we can write (29) as

$$\tfrac{1}{3}\mu_I^2\int_0^\infty \left[\frac{1}{\rho}+\frac{\rho_0}{4}-\frac{(1+\rho_0+\frac{1}{4}\rho_0{}^2)}{(\rho_0+\rho)}\right]\exp(-\rho)d\rho. \qquad (30)$$

Using a table of integrals (see Gradshteyn and Ryzhik[17]) and the cutoff prescription of Bethe[16] (page 110), we have $\int_\varepsilon^\infty (1/\rho)\exp(-\rho)d\rho = -Ei(-\varepsilon)$, $(\rho_0/4)\int_0^\infty \exp(-\rho)d\rho = (\rho_0/4)$ and $-\int_0^\infty 1/(\rho_0+\rho)\exp(-\rho)d\rho = e^{\rho_0}Ei(-\rho_0)$, where $Ei(-\varepsilon) = C + \ln\varepsilon + \sum_{k=1}^{\infty}(-1)^k\left[\varepsilon^k/k(k!)\right]$ and C is Euler's constant. Using these results in (30), we get

$$\tfrac{1}{3}\mu_I^2\int_0^\infty\left[\frac{1}{\rho}+\frac{\rho_0}{4}-\frac{(1+\rho_0+\frac{1}{4}{\rho_0}^2)}{(\rho_0+\rho)}\right]\exp(-\rho)d\rho$$

$$=\tfrac{1}{3}\mu_I^2\left[-Ei(-\varepsilon)+\frac{\rho_0}{4}+e^{\rho_0}(1+\rho_0+\tfrac{1}{4}\rho_0^2)Ei(-\rho_0)\right]. \qquad (31)$$

It is clear that $-Ei(-\varepsilon)$ will diverge to ∞ like $-\ln\varepsilon$ as $\varepsilon\to 0$. If we fix ε at ρ_0, and note that $e^{\rho_0}\cong 1+\rho_0$, then

$$\tfrac{1}{3}\mu_I^2\left[-Ei(-\varepsilon)+\frac{\rho_0}{4}+e^{\rho_0}(1+\rho_0+\tfrac{1}{4}\rho_0^2)Ei(-\varepsilon)\right]$$

$$\cong\tfrac{1}{3}\mu_I^2\left\{\frac{\rho_0}{4}+(2\rho_0+\tfrac{5}{4}\rho_0^2)\left[C+\ln\rho_0+\sum_{k=1}^{\infty}\frac{(-1)^k\rho_0^k}{k(k!)}\right]\right\}\Rightarrow$$

$$\tfrac{1}{3}\rho_0\mu_I^2\int_0^\infty\frac{(1-\frac{1}{2}\rho)^2\exp(-\rho)d\rho}{(1+\rho_0/\rho)\rho^2}\cong\tfrac{1}{16}\gamma^2(\tfrac{1}{1836})^2g_N^2\left\{\tfrac{1}{8}\gamma^2+(\gamma^2+\tfrac{5}{16}\gamma^4)\left[C+\ln\tfrac{1}{2}\gamma^2-\tfrac{1}{2}\gamma^2\right]\right\}. \qquad (32)$$

If we note that $(1/1836)^2\cong(1/13)^2\gamma^2$, then this last term is of order (>) γ^7. Thus, if there is any (mathematical) justification for the calculation procedures, we can see that the $\mathbf{A}^2$ term does not make a significant contribution.

Conclusion

In this paper we have shown that the full (minimal coupling) Dirac equation can be analytically separated (diagonalized) into particle and antiparticle components without transforming the wave functions, as is done by the Foldy-Wouthuysen method. This diagonalization reveals the nonlocal time behavior of the particle-antiparticle relationship. We have shown that a more physically reasonable interpretation of the zitterbewegung, and the result that a velocity measurement (of a Dirac particle) at any instant in time is $\pm c$, are reflections of the fact that the Dirac equation makes a spatially extended particle appear as a point in the present by forcing it to oscillate between the past and future at speed c.

We have also shown that one of the most difficult issues facing any attempt to completely understand the Dirac problem for full coupling is the singular nature of the $\mathbf{A}^2$ term. This term is small in all but s- states, where it diverges when treated as a perturbation. If we introduce a cutoff, the contribution is of order γ^7, so one might be inclined to dismiss the term (as is traditionally done). However, this term appears to be more singular than the Coulomb potential, so that perturbation analysis, and indeed, the whole eigenvalue approach may be incorrect. Although this is not completely clear since the $\sin^2\theta$ term vanishes on the spin axis and could strongly modify the singular nature of this term, this problem must be solved in order to provide conclusive support for the assumption that the Dirac equation does not account for the complete spectrum of hydrogen. This is clearly an important problem for the foundations of QED.

Assuming that the problems posed in this paper can be solved in the positive, it would appear that the correct approach for s-state (hyperfine) splitting gives the same results as those obtained from using the Pauli equation. Furthermore, equation (24b) introduces a natural cutoff, which removes the conceptual difficulty of a point magnetic dipole interaction as implied by use of the delta term in the Pauli equation.

REFERENCES

1. W. E. Lamb Jr. and R. C. Retherford, *Phys. Rev.* **72** (1947) 241.
2. P. A. M. Dirac, *Proc. Roy. Soc (London)* **A117** (1928) 610, **A118** (1928) 351.
3. B. Thaller, *The Dirac Equation,* (Springer-Verlag, New York, 1992).
4. L. C. Biedenharn, *Phys. Rev.* **126** (1962) 845.
5. E. deVries, *Fortsch. d. Physik* **18** (1970) 149.
6. D. Hestenes, *Space Time Algebra,* (Gordon and Breach, New York, 1966).
7. L. L. Foldy and S. A. Wouthuysen, *Phys. Rev.* **78** (1950) 29.
8. W. Pauli, *Handbuch der Physik,* 2d ed., **24** (1933).
9. R. P. Feynman and M. Gell-Mann, Phys. Rev. **109** (1958) 193.
10. T. L. Gill, FermiLab-Pub-82/60-THY.
11. E. H. Lieb and M. Loss, *Analysis,* Grad. Studies in Math. **14**, Amer. Math. Soc., 1991 (see pages 169-70).
12. A. O. Williams Jr., *Phys. Rev.* **58** (1940) 723.
13. J. C. Slater, *Quantum Theory of Atomic Structure, Vol. II,* (McGraw-Hill, New York, 1960).
14. W. Greiner, *Relativistic Quantum Mechanics,* (Springer-Verlag, New York, 1994).
15. M. Mizushima, *Quantum Mechanics of Atomic Spectra and Atomic Structure* (Benjamin Inc. New York, 1970).
16. H. Bethe and E. E. Salpeter, *Quantum Mechanics of One-and Two-Electron Atoms,* (Springer-Verlag, New York, 1957).
17. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic Press, New York, 1980).