Electron Paths, Tunnelling and Diffraction in the Spacetime Algebra

AUTHORS
Stephen Gull
Anthony Lasenby
Chris Doran

Found. Phys. **23**(10), 1329-1356 (1993)

Abstract

This paper employs the ideas of geometric algebra to investigate the physical content of Dirac's electron theory. The basis is Hestenes' discovery of the geometric significance of the Dirac spinor, which now represents a Lorentz transformation in spacetime. This transformation specifies a definite velocity, which might be interpreted as that of a real electron. Taken literally, this velocity yields predictions of tunnelling times through potential barriers, and defines streamlines in spacetime that would correspond to electron paths. We also present a general, first-order diffraction theory for electromagnetic and Dirac waves. We conclude with a critical appraisal of the Dirac theory.

1 Introduction

In this, the last of a 4-paper series [1, 2, 3], we are concerned with one of the main areas of David Hestenes' work — the Dirac equation. His thesis has always been that the theory of the electron is central to quantum mechanics and that the electron wavefunction, or Dirac spinor, contains important geometric information [4, 5, 6, 7]. This information is usually hidden by the conventional matrix notation, but it can be revealed by systematic use of a better mathematical language — the geometric (Clifford) algebra of spacetime, or *spacetime algebra* (STA) [8]. This algebra provides a powerful coordinate-free language for dealing with all aspects of relativistic physics — not just relativistic quantum mechanics [9]. Indeed, it is a formalism that makes the conventional 4-vector/tensor approach look decidedly primitive.

In earlier papers [1, 2] we saw that the 8 real components of a Dirac spinor encode a 6-component Lorentz transformation and a 2-component scalar + pseudoscalar factor. Hestenes interprets this as a decomposition into dynamical and statistical terms respectively [10]. But what, in that case, is the physical meaning of the Dirac wavefunction $\psi(x)$, which is a Dirac spinor defined at every point x in spacetime? Hestenes' innovation is that the Dirac wavefunction $\psi(x)$ contains the spacetime rotation necessary to define the velocity of an electron, the direction of its spin axis, and the phase angle. The wavefunction varies with spacetime position because it is a *summary of the properties the electron would have if it happened to be at point* x. In this interpretation the electron is a point particle, moving along a streamline in spacetime in the direction determined by the wavefunction (the direction of the Dirac current). The amplitude, $|\psi|^2$, then gives the probability density that the electron will be found on the streamlines through a particular region. In recent versions of this interpretation (the zitterbewegung or ZBW model) the electron executes a light-like helical motion about this average streamline at a frequency of $2mc^2/h = 2.5 \times 10^{20}$ Hz.

We emphasise that there are two distinct issues here. The first is that the STA is the best available way of dealing with relativistic physics. This is a matter of formalism and language, which we strongly believe to be correct. The second is the contention that the velocity Hestenes identifies in the Dirac theory makes useful predictions about the motion of an electron. This is a matter of physics, which we examine here.

In Section 2 we review the STA formulation of Dirac theory, emphasising how the wavefunction contains a spacetime transformation, including a Lorentz boost.

The non-relativistic approximation is set out in Section 3, and yields the Pauli equation. By a careful study of the limiting process by which the Pauli theory is related to the Dirac theory, we are able to retain the velocity information, which is encoded as the gradient of the Pauli 2-spinor. Section 4 concerns the tunnelling of Dirac waves through a potential barrier, a topic of practical interest for the design of semiconductor devices. We derive formulae for transit times similar to those of Bohm's approach [11, 12, 13], although the full relativistic treatment, including spin, is given here for the first time.

Diffraction theory is considered in Section 5, and powerful formulae for the propagation of both electromagnetic and Dirac waves are given. In the STA, the Helmholtz equation is first-order (like the Dirac equation), and Huygens' principle of re-radiation of wavelets is directly applicable. This has clear advantages over the conventional, second-order treatment. We use these results in Section 6, plotting electron streamlines for a double pinhole arrangement. Again, the results are similar to those obtained from Bohm's approach [14], but the relativistic calculations include some surprises. In particular, the streamlines from a point source show a circulating current in the plane defined by the spin. This circulation takes place over a characteristic radius defined by the electron's 3-momentum. We finish on a note of caution with a critical discussion of the Dirac theory. The predictions made in this paper are only as good as the Dirac equation itself, and, since it predicts antiparticles with negative mass instead of the opposite charge that is observed, the first-quantised Dirac equation is fatally flawed.

2 Dirac Theory and Spacetime Algebra

The spacetime algebra (STA) is a real, geometric (Clifford) algebra developed on a 4-dimensional flat spacetime with a standard Minkowski metric. The basic ingredients of this algebra are an orthonormal frame of vectors $\{\gamma_\mu\}$, $\mu = 0, 1, 2, 3$ where

$$\gamma_0^2 = -\gamma_k^2 = 1 \quad (k = 1, 2, 3). \tag{2.1}$$

The time-like vector γ_0 defines a Lorentz frame, which we can think of as representing the *laboratory frame*. From this basis set of vectors we build up the 16 geometric elements of the STA:

$$\begin{array}{ccccc} 1 & \{\gamma_\mu\} & \{\sigma_k, i\sigma_k\} & \{i\gamma_\mu\} & i \\ \text{1 scalar} & \text{4 vectors} & \text{6 bivectors} & \text{4 pseudovectors} & \text{1 pseudoscalar.} \end{array} \tag{2.2}$$

The time-like bivectors $\sigma_k \equiv \gamma_k\gamma_0$ obey the same algebraic relations as the Pauli spin matrices, but in the STA they represent an orthonormal frame of vectors in space *relative* to the laboratory time vector γ_0. The unit pseudoscalar of spacetime is defined as

$$i \equiv \gamma_0\gamma_1\gamma_2\gamma_3 = \sigma_1\sigma_2\sigma_3. \tag{2.3}$$

Two useful operations are reversion, $\psi \mapsto \tilde{\psi}$, which reverses the order of spacetime vectors in a relativistic expression, and Hermitian conjugation, $\psi \mapsto \psi^\dagger$, which reverses the order of relative 3-vectors in a non-relativistic expression. These operations are related by

$$\psi^\dagger = \gamma_0\tilde{\psi}\gamma_0. \tag{2.4}$$

Angled brackets $\langle\psi\rangle_k$ are used to project out the grade-k components of ψ, with the scalar component written as $\langle\psi\rangle$. Further details of our notation and conventions are given in the earlier papers of this series [1, 2, 3].

In this paper we concentrate on testable aspects of the Dirac theory. We will therefore often need a 'space-time split' [8] of relativistic observables into the laboratory frame. For vectors this is achieved as follows:

$$\begin{aligned} x\gamma_0 &= x\cdot\gamma_0 + x\wedge\gamma_0 \\ &= t + \mathbf{x}, \end{aligned} \tag{2.5}$$

where $\mathbf{x} = x^i\sigma_i$. We distinguish relative 3-vectors (which are spacetime bivectors) from spacetime 4-vectors by writing the former in bold type. The sole exceptions to this rule are the Pauli basis vectors $\{\sigma_k\}$, which are unambiguously spacetime bivectors. We extend the usual convention for tensor indices and use Greek indices for spacetime frames and 4-vectors, reserving Roman indices for 3-space. These are our conventions for the space-time splits of particular quantities:

$$\begin{array}{ccc} \text{position vector} & x & x\gamma_0 = t + \mathbf{x}, \\ \text{vector derivative} & \nabla & \gamma_0\nabla = \partial_t + \boldsymbol{\nabla}, \\ \text{4-momentum} & p & p\gamma_0 = E + \mathbf{p}, \\ \text{4-potential} & A & A\gamma_0 = V + \mathbf{A}, \\ \text{electromagnetic field} & F & F = \mathbf{E} + i\mathbf{B}, \end{array} \tag{2.6}$$

where

$$\begin{aligned} \mathbf{E} &= \tfrac{1}{2}(F - \gamma_0 F\gamma_0) \\ i\mathbf{B} &= \tfrac{1}{2}(F + \gamma_0 F\gamma_0). \end{aligned} \tag{2.7}$$

We use 'natural' units ($\hbar = c = 1$) except when expressing our final results.

The even subalgebra of the STA, formed by the scalar, bivectors and pseudoscalar, comprises a representation of the spinors of spacetime [2]. Standard conventions notwithstanding, we will call these Dirac spinors. A Dirac spinor ψ defines a spacetime transformation through the relation

$$\rho e_\mu = \psi \gamma_\mu \tilde{\psi}, \tag{2.8}$$

where the $\{e_\mu\}$ form a second orthonormal basis, and ρ is a scale factor. We can write [1, 5]

$$\psi = \left(\rho e^{i\beta}\right)^{\frac{1}{2}} R, \tag{2.9}$$

where R is a spacetime rotation (rotor) satisfying $R\tilde{R} = 1$. We can decompose R into parts representing a spatial rotation (leaving γ_0 fixed) and a Lorentz boost:

$$R = L\Phi, \tag{2.10}$$

where

$$\begin{array}{rcl} \Phi\gamma_0 & = & \gamma_0\Phi, \\ L\gamma_0 & = & \gamma_0\tilde{L}, \end{array} \tag{2.11}$$

and $\Phi\tilde{\Phi} = L\tilde{L} = 1$. Explicit forms for L and Φ are

$$\begin{array}{rcl} L(\mathbf{u}) & = & e^{\mathbf{u}/2}, \\ \Phi(\mathbf{a}) & = & e^{i\mathbf{a}/2}, \end{array} \tag{2.12}$$

where L represents a boost to a velocity $v = e^{\mathbf{u}}\gamma_0$, and Φ is a rotation through $|\mathbf{a}|$ about the 3-space axis $\mathbf{a}$ [1].

The systematic use of rotors simplifies relativistic particle mechanics enormously [9]. A particle's 4-velocity v is written in terms of the fixed axis γ_0 and the rotor R as

$$\dot{x} \equiv v = e_0 = R\gamma_0\tilde{R}, \tag{2.13}$$

and the dynamical equations for the rotor R are generally much simpler than those for v itself. Furthermore, the rotor carries more information than v, since it also predicts the direction of the $\{e_1, e_2, e_3\}$ axes: a useful fact when considering the motion of rigid bodies [9].

A crucial aspect of Hestenes' interpretation of the Dirac wavefunction is the identification of the transformed time-like axis e_0 as the *velocity* of an electron, and of the e_3 axis as the direction of its *spin*. The remaining axes e_1 and e_2 rotate

rapidly in their plane, in a way that defines the *phase* of the electron.

Not all parts of this interpretation are controversial. The velocity e_0 (2.8) is, after all, in the direction of the Dirac current $J = \rho v = \psi\gamma_0\tilde{\psi}$, which is conventionally taken as the source term for the electromagnetic field in quantum electrodynamics. What is certainly controversial, though, is to assign meaning to the relation $v = \dot{x}$, implying that the electron is, in some sense, a point particle moving in the direction of v.

The variation of the electron wavefunction in space is described by the Dirac equation [15], which in the STA is written as [2, 4]

$$\nabla\psi i\sigma_3 - eA\psi = m\psi\gamma_0. \tag{2.14}$$

This admits free-space plane-wave solutions for wave-vector $k = me^{\mathbf{u}}\gamma_0$:

$$\begin{aligned} \psi_+ &= L(\mathbf{u})\Phi e^{-i\sigma_3 k\cdot x}, \\ \psi_- &= L(\mathbf{u})\Phi i\ e^{+i\sigma_3 k\cdot x}, \end{aligned} \tag{2.15}$$

which are the waves of positive and negative frequency. These two solutions differ in their values of β: ψ_+ has $\beta = 0$, whereas the negative-frequency wave ψ_- has $\beta = \pi$, so that the product $\psi_-\tilde{\psi}_-$ is real and negative.

An important quantity in the Dirac theory is the stress-energy tensor, which has been derived and discussed in a companion paper [3]. The stress-energy tensor $T(n)$ defines the energy-momentum flux in the direction of the time-like unit vector n, and for the Dirac theory $T(n)$ is given by

$$T(n) = \overset{*}{\nabla}\ \langle \overset{*}{\psi}\ i\gamma_3\tilde{\psi}n\rangle - eA\langle\psi\gamma_0\tilde{\psi}n\rangle, \tag{2.16}$$

where the overstar indicates the operand of ∇. We have included the interaction term, $-eAn\cdot J$, in order to preserve gauge invariance.

The Dirac equation singles out a preferred timelike axis γ_0, which we interpret as specifying the laboratory frame. (This does not in fact violate Lorentz invariance: see the earlier papers in this series [1, 2].) For this frame we write the momentum density as $p = T(\gamma_0)$. The total 4-momentum in the laboratory frame is then given by

$$P = \int |d^3x|\, T(\gamma_0) \tag{2.17}$$

and, in the absence of an A field, P is conserved with respect to laboratory time. For the two plane-wave solutions (2.15) we find that $p = \pm k\cosh|\mathbf{u}|$, so that ψ_+

and ψ_- have opposite signs for their 4-momentum densities.

It is interesting to consider the flux of energy-momentum along streamlines — the integral curves of $v(x)$. We find that

$$T(v) = \overset{*}{\nabla} \langle \overset{*}{\psi} i\gamma_3 \tilde{\psi} v \rangle - eA\langle \psi\gamma_0\tilde{\psi}v\rangle \tag{2.18}$$

$$= \rho \overset{*}{\nabla} \langle \overset{*}{R}\, i\sigma_3 \tilde{R}\rangle - e\rho A, \tag{2.19}$$

so that the flux is determined by the derivatives of the rotor R. This supports the identification of R as the dynamical element of the full spinor ψ. The free-particle solutions (2.15) have $T(v) = \pm k = \pm mv$, so that for ψ_- the 4-momentum points in the opposite direction to the 4-velocity; the ψ_- solution behaves as if it has negative mass. The classical relation $T(v) = mv$ holds only in special cases, of which the plane wave ψ_+ is an example.

3 Relation to Pauli and Schrödinger Theories

The Dirac theory of the electron is able to predict a well-defined velocity v at any point of spacetime because the Dirac wavefunction characterises a general Lorentz transformation. In non-relativistic approximations this information is hidden. The result is that discussions of currents and velocities within non-relativistic theories are often confused. In this section we indicate how conceptual problems are easily overcome through a clear picture of how the Schrödinger and Pauli theories emerge from the Dirac theory. Despite the simplicity of the idea, the results may be surprising to those schooled in conventional non-relativistic quantum mechanics.

We begin by examining the traditional derivation of the probability current from the time-dependent Schrödinger equation,

$$j\partial_t \Psi = -\frac{1}{2m}\boldsymbol{\nabla}^2\Psi + eV\Psi, \tag{3.1}$$

where we denote the imaginary unit by j to distinguish it from the STA pseudoscalar i, and Ψ is a conventional, complex wavefunction. The first step is to show that the 'probability density' $\Psi^*\Psi = |\Psi|^2$ satisfies a continuity equation:

$$\partial_t |\Psi|^2 = \boldsymbol{\nabla}\cdot\left(\frac{j}{2m}\left(\Psi^*\boldsymbol{\nabla}\Psi - \Psi\boldsymbol{\nabla}\Psi^*\right)\right), \tag{3.2}$$

where Ψ^* is the complex conjugate of Ψ. It is then argued that the quantity

$$\mathbf{J} \equiv -\frac{j}{2m}\left(\Psi^*\boldsymbol{\nabla}\Psi - \Psi\boldsymbol{\nabla}\Psi^*\right) \tag{3.3}$$

should be interpreted as the probability current. The probability current is not defined uniquely by this method because the curl of any vector field can be added to $\mathbf{J}$ without changing the continuity equation. Equation (3.3) *does* indicate that, for the second-order Schrödinger theory, the velocity information is encoded in the gradients of the wavefunction.

Examination of the non-relativistic approximation to the Dirac equation shows how this comes about. This analysis is particularly simple using the STA. We begin by performing a space-time split of the Dirac equation into the laboratory frame, multiplying first on the left, then on the right, by γ_0. This yields the pair of equations

$$\begin{aligned}
\partial_t\psi i\sigma_3 + \boldsymbol{\nabla}\psi i\sigma_3 - eV\psi + e\mathbf{A}\psi &= m\bar{\psi}, &(3.4)\\
\partial_t\bar{\psi} i\sigma_3 - \boldsymbol{\nabla}\bar{\psi} i\sigma_3 - eV\bar{\psi} - e\mathbf{A}\bar{\psi} &= m\psi, &(3.5)
\end{aligned}$$

where we have used the definition $\bar{\psi} \equiv \gamma_0\psi\gamma_0$. We define two new spinors $\psi_\pm \equiv \frac{1}{2}(\psi \pm \bar{\psi})$ and rewrite the equations as

$$\begin{aligned}
\partial_t\psi_+ i\sigma_3 + \boldsymbol{\nabla}\psi_- i\sigma_3 - eV\psi_+ + e\mathbf{A}\psi_- &= m\psi_+, &(3.6)\\
\partial_t\psi_- i\sigma_3 + \boldsymbol{\nabla}\psi_+ i\sigma_3 - eV\psi_- + e\mathbf{A}\psi_+ &= -m\psi_-. &(3.7)
\end{aligned}$$

These spinors $\psi_\pm$ satisfy the commutation relations

$$\psi_\pm\gamma_0 = \pm\gamma_0\psi_\pm. \tag{3.8}$$

We now remove the rapidly-varying part of ψ by writing

$$\psi(x) = \psi'(t,\mathbf{x})e^{-i\sigma_3 mt}. \tag{3.9}$$

This ensures that we are dealing with 'positive mass' solutions. Substituting into equation (3.7), we find that

$$\partial_t\psi'_- i\sigma_3 - eV\psi'_- + 2m\psi'_- = -\boldsymbol{\nabla}\psi'_+ i\sigma_3 - e\mathbf{A}\psi'_+, \tag{3.10}$$

and $\partial_t\psi'_- i\sigma_3 - eV\psi'_-$ will be of the same order as the non-relativistic energy. We

now suppose that the non-relativistic energy scale is much smaller than the mass, so that the mass term dominates the left-hand side of (3.10). A good approximation for the 'small' part ψ'_- of the Dirac spinor ψ' is therefore given by:

$$-2m\psi'_- \approx \boldsymbol{\nabla}\psi'_+ i\sigma_3 + e\mathbf{A}\psi'_+. \tag{3.11}$$

For the rest of this section we will simplify the notation by defining $\phi \equiv \psi'_+$ for the 'large' part of the Dirac spinor. ϕ is Pauli-even ($\phi = \bar{\phi}$), and can be identified with a non-relativistic 2-component Pauli spinor [2]. From (3.6) this approximation yields the STA form of the Pauli equation:

$$\begin{aligned} \partial_t \phi i\sigma_3 &= \frac{1}{2m}\left(\boldsymbol{\nabla}(\boldsymbol{\nabla}\phi i\sigma_3 + e\mathbf{A}\phi)i\sigma_3 + e\mathbf{A}(\boldsymbol{\nabla}\phi i\sigma_3 + e\mathbf{A}\phi)\right) + eV\phi \\ &= \frac{1}{2m}\left(-\boldsymbol{\nabla}^2\phi + 2e\mathbf{A}\cdot\boldsymbol{\nabla}\phi i\sigma_3 + e^2\mathbf{A}^2\phi\right) + eV\phi + \frac{e}{2m}\boldsymbol{\nabla}\mathbf{A}\phi i\sigma_3 \end{aligned} \tag{3.12}$$

In the Coulomb gauge ($\boldsymbol{\nabla}\cdot\mathbf{A} = 0$), the last term in (3.12) becomes $-\frac{e}{2m}\mathbf{B}\phi\sigma_3$, which represents the coupling of the magnetic moment of the spin to the magnetic field.

Our derivation of the Pauli equation shows clearly how the 'small' part of the Dirac spinor (which contains information about the velocity) becomes encoded as the gradient of the 'large' part ϕ. It also shows how the first-order Dirac equation has been complicated by this non-relativistic approximation, which introduces *second* order derivatives.

We now examine the non-relativistic approximation to the observables of the Dirac theory. We will ignore terms quadratic in ψ'_-, which are small because of the assumptions made in deriving (3.11) from (3.7). First, the density $\psi\tilde{\psi} = \rho e^{i\beta}$ yields the approximations:

$$\begin{aligned} \rho\cos\beta &\approx \phi\phi^\dagger, \\ i\rho\sin\beta &\approx -\frac{1}{2m}\left(\boldsymbol{\nabla}\phi i\sigma_3\phi^\dagger + \phi i\sigma_3\phi^\dagger\overleftarrow{\boldsymbol{\nabla}}\right), \end{aligned} \tag{3.13}$$

so that consistency of our non-relativistic approximation requires that $\rho\sin\beta$ be small. Hestenes [5] has argued from this that β must also be small, but an examination of hydrogen atom solutions shows that this is not necessarily true. From (3.13) the non-relativistic approximation to β is determined by

$$\phi\phi^\dagger\tan\beta \approx -\frac{1}{2m}\boldsymbol{\nabla}\cdot(\phi\sigma_3\phi^\dagger), \tag{3.14}$$

which shows that β is determined by the spin. The non-relativistic approximation to the Dirac current $\psi\gamma_0\tilde{\psi}$ gives the 3-current

$$\mathbf{J} \approx -\frac{1}{2m}\left(\boldsymbol{\nabla}\phi i\sigma_3\phi^\dagger - \phi i\sigma_3\phi^\dagger\overleftarrow{\boldsymbol{\nabla}}\right) - \frac{e}{m}\mathbf{A}\phi\phi^\dagger \tag{3.15}$$

and, since the 4-velocity v is J/J_0, we define the 3-velocity $\mathbf{v}$ as $\mathbf{J}/(\phi\phi^\dagger)$.

We now relinquish the link with the relativistic theory, and define the observables for the Pauli theory as follows:

$$\begin{aligned} \rho &= \phi\phi^\dagger, \\ m\rho\mathbf{v} &= -\langle\boldsymbol{\nabla}\phi i\sigma_3\phi^\dagger\rangle_v - e\mathbf{A}\phi\phi^\dagger, \\ \rho\mathbf{s} &= \phi\sigma_3\phi^\dagger, \\ -2m\rho\tan\beta &= \boldsymbol{\nabla}\cdot(\rho\mathbf{s}), \end{aligned} \tag{3.16}$$

where $\langle\ \rangle_v$ denotes the 3-vector part of the expression in brackets. The definitions (3.16) are summarised neatly by the relation

$$-\frac{1}{m}\left(\boldsymbol{\nabla}\phi i\sigma_3 + e\mathbf{A}\phi\right)\phi^\dagger = \rho(\mathbf{v} + i\tan\beta). \tag{3.17}$$

The conventional 'equation of continuity' argument leads to the components

$$\rho u_k = -\frac{1}{2m}\left(\partial_k\phi i\sigma_3\phi^\dagger - \phi i\sigma_3\partial_k\phi^\dagger\right) - \frac{eA_k}{m}\phi\phi^\dagger, \tag{3.18}$$

from which we construct the current

$$\rho\mathbf{u} = -\frac{1}{m}\left(\overset{*}{\boldsymbol{\nabla}}\ \langle\ \overset{*}{\phi}\ i\sigma_3\phi^\dagger\ \rangle + e\mathbf{A}\phi\phi^\dagger\right). \tag{3.19}$$

This can now be compared with the form of $\mathbf{J}$, which was derived from the relativistic charge current. Looking again at our approximation to the Dirac current, we recognise that

$$\begin{aligned} m\rho\mathbf{v} &= -\langle\ \boldsymbol{\nabla}\phi i\sigma_3\phi^\dagger\ \rangle_1 - e\mathbf{A}\phi\phi^\dagger \\ &= -\overset{*}{\boldsymbol{\nabla}}\ \langle\ \overset{*}{\phi}\ i\sigma_3\phi^\dagger\ \rangle - e\mathbf{A}\phi\phi^\dagger - \overset{*}{\boldsymbol{\nabla}}\cdot\langle\ \overset{*}{\phi}\ i\sigma_3\phi^\dagger\ \rangle_2 \\ &= m\rho\mathbf{u} + \tfrac{1}{2}\boldsymbol{\nabla}\times(\rho\mathbf{s}), \end{aligned} \tag{3.20}$$

where $\times$ is the conventional vector cross product. The correct charge current in the Pauli theory therefore differs from that found by continuity arguments by a

term that is determined by the curl of the spin. This fact is lost in discussions based solely on non-relativistic equations, and explains why, in the Pauli theory, it is conventional to supplement the Schrödinger current with a magnetisation term. The term $m\rho\mathbf{u}$ does, however, have a physical meaning — it is the *momentum*. The non-relativistic approximation to the 4-momentum $p = T(\gamma_0)$ yields the formulae

$$E = \langle \partial_t \phi i\sigma_3 \phi^\dagger \rangle + (m - eV)\phi\phi^\dagger \tag{3.21}$$

$$\mathbf{p} = -\overset{*}{\boldsymbol{\nabla}} \langle \overset{*}{\phi}\, i\sigma_3 \phi^\dagger \rangle - e\mathbf{A}\phi\phi^\dagger, \tag{3.22}$$

so that $\mathbf{p} = m\rho\mathbf{u}$. As a result of choosing the stress-energy tensor in the form of (2.16) (which was motivated by gauge invariance) equation (3.21) gives the kinetic rather than the total energy.

From (3.22) the current and momentum in the Pauli theory are simply related by

$$\mathbf{p} = m\rho\mathbf{v} - \tfrac{1}{2}\boldsymbol{\nabla}\times(\rho\mathbf{s}). \tag{3.23}$$

Unlike the Dirac equation, the Pauli equation defines streamlines for both velocity *and* momentum. Consistency with the Dirac equation, however, demands that $\mathbf{v}$ be taken as the velocity in the Pauli theory, and not $\mathbf{p}$/m. The incorrect choice $\mathbf{p}/m$ was made by Bohm *et al.* [16] when applying the causal interpretation of quantum mechanics to the Pauli theory. This choice has since been repeated in modern extensions of Bohm's ideas [17].

This inconsistency between the currents identified in the Pauli and Dirac theories has been repeatedly advertised by David Hestenes [9, 18], but the results are so surprising to most conventionally-trained physicists that it is useful to illustrate them further. We do so by considering the spin 'up' solutions of the Pauli equation for the hydrogen atom. For these we find:

$$\begin{aligned} \mathbf{p} &= \mu\rho\sigma_\phi, \\ m\rho\mathbf{v} &= \mu\rho\sigma_\phi + \tfrac{1}{2}(\boldsymbol{\nabla}\rho)\times\sigma_3, \\ -2m\tan\beta &= \partial_z(\ln\rho), \end{aligned} \tag{3.24}$$

where μ is the azimuthal quantum number and σ_ϕ is the unit vector in the azimuthal direction.

The ground state of the hydrogen atom, with spin 'up', is given by

$$\phi(r,t) = \left(\frac{1}{\pi a^3}\right)^{\frac{1}{2}} e^{-r/a} e^{-Ei\sigma_3 t}, \tag{3.25}$$

where a is the Bohr radius $(m\alpha)^{-1}$, $E = -\frac{m}{2}\alpha^2$ and α is the fine-structure constant. Applying (3.24) we find that

$$\begin{aligned} \mathbf{p} &= 0, \\ \mathbf{v} &= \alpha \sin\theta \sigma_\phi \\ \tan\beta &= \alpha\cos\theta, \end{aligned} \tag{3.26}$$

which predicts an azimuthal current corresponding to a velocity of $v = \alpha \sin\theta$. Accordingly, there is a circulation in the ground state of the hydrogen atom: a fact that is ignored in most conventional treatments. A further curious feature is the presence of a nonzero β-factor: $\beta \approx \alpha\cos\theta$. The physical significance of the β-factor is not clear, but studying its behaviour for higher-energy states reveals some very peculiar properties. For example the $n = 2$, $l = 1$, $\mu = 0$ state has a discontinuity in β in the azimuthal plane, in which the density ρ vanishes. The role of the β-parameter thus remains obscure, despite the vigour with which Hestenes has drawn attention to the problem [5, 7, 19].

4 Tunnelling times in Quantum Mechanics

When an electron wave of relativistic energy E and momentum p is incident upon an electric potential step of height eV, where $eV > E - m$, an evanescent wave is set up in the classically forbidden region. If the barrier has a finite width there is a transmitted wave. Potential wells of this type are of considerable interest for the design of semiconductor devices, and there has been a great deal of controversy over the 'correct' answer to the question: 'How long does the electron spend in the classically forbidden region?' (A comprehensive review covering the period up to 1989 is given by Hauge and Støvneng [20].) Within the Copenhagen approach to quantum mechanics it is not even clear that the question makes any sense, since time is not a Hermitian observable, and a corresponding measurement process is therefore not defined. Also, the Copenhagen interpretation rejects any notion of a particle following a trajectory, which would enable the time spent within a given region to be calculated. In the Dirac theory, however, this question has a simple answer, because given the Dirac wavefunction we can associate an unambiguous 4-velocity v with any point in space. Whether this velocity actually represents the real velocity of an electron is unclear, but it is valuable to compute the consequences of this assumption in the hope that experimental evidence will soon settle the matter.

Consider first the steady-state (continuous wave) situation in which an electron wave is incident normally upon a step of width Δ, so that

$$A = \begin{cases} V\gamma_0 & \text{if} -\Delta < z < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Using the STA, we write the transmitted wave as

$$\psi_t = e^{u\sigma_3/2}\phi e^{i\sigma_3(pz-Et)}, \tag{4.1}$$

where $E = m\cosh u$, $p = m\sinh u$ and $\nu = p/E = \tanh u$. The quantity ϕ is a normalised Pauli spinor describing the spin state, with $\phi = 1$ corresponding to longitudinal spin 'up', and $\phi = -i\sigma_2$ to spin 'down'.

If the step has potential $eV > E - m$ the wavefunction inside the barrier is evanescent. When working with the STA, however, we cannot simply let the wavenumber become imaginary. Defining a new energy $E' \equiv E - eV$, and a decay parameter κ via the relation $\kappa^2 + (E')^2 = m^2$, we write the wavefunction for longitudinal spin up as follows:

$$\psi_b \propto \left(e^{i\beta/2}e^{-\kappa z}e^{-i\sigma_3\delta/2} + e^{-i\beta/2}e^{\kappa z}e^{i\sigma_3\delta/2}\right)e^{-i\sigma_3 Et}, \tag{4.2}$$

where $E - eV = m\cos\beta$, $\kappa = m\sin\beta$ and δ is a phase factor to be determined. The matching condition at $z = 0$ gives

$$\tan(\delta/2)\tan(\beta/2) = \tanh(u/2), \tag{4.3}$$

or

$$\tan\delta = \frac{\sinh u \sin\beta}{1 - \cosh u\cos\beta}. \tag{4.4}$$

The Dirac current can now be evaluated as $\rho v = \psi\gamma_0\tilde{\psi}$, and the velocity inside the barrier is

$$\frac{dz}{dt} = \frac{\sin\beta\sin\delta}{\cos\beta\cos\delta + \cosh(2\kappa z)}. \tag{4.5}$$

By integrating this we find the tunnelling time

$$t_b = \frac{\sinh(2\kappa\Delta) + 2\kappa\Delta\cos\beta\cos\delta}{2\kappa\sin\beta\sin\delta}. \tag{4.6}$$

On translating the answer into physical units and using a non-relativistic

approximation, we find, as usual,

$$(\hbar\kappa)^2 \approx 2m(E_b - E_t); \quad p^2 \approx 2mE_t, \tag{4.7}$$

where E_t is the kinetic energy of the transmitted electrons and $E_b \equiv eV$ is the height of the barrier. Corresponding approximations are

$$\sin\beta \approx \left(\frac{2(E_b - E_t)}{mc^2}\right)^{\frac{1}{2}}, \tag{4.8}$$

$$\sin\delta \approx \frac{2(E_t(E_b - E_t))^{\frac{1}{2}}}{E_b}; \quad \cos\delta \approx \frac{E_b - 2E_t}{E_b}. \tag{4.9}$$

Our final formula is

$$t_b \approx \frac{\hbar}{8}\frac{E_b \sinh(2\kappa\Delta) + 2\kappa\Delta(E_b - 2E_t)}{(E_t(E_b - E_t)^3)^{\frac{1}{2}}}. \tag{4.10}$$

Numerical results can be obtained according to the formulae

$$1/\kappa = 0.194(E_b - E_t)^{-\frac{1}{2}} \text{ nm}, \tag{4.11}$$

$$t_b = 8.20 \times 10^{-18} \frac{E_b \sinh(2\kappa\Delta) + 2\kappa\Delta(E_b - 2E_t)}{(E_t(E_b - E_t)^3)^{\frac{1}{2}}} \text{ sec}, \tag{4.12}$$

where the energies are measured in electron-volts. As an example, for $E_b = 2$ eV, $E_t = 1$ eV and $\Delta = 1$ nm, we find $t_b = 2.4 \times 10^{-12}$ sec.

This answer is much too long compared to the (indirect) experimental evidence available so far, which favours times closer to 10^{-15} secs for energies of this magnitude. The difficulty arises from the use of a monochromatic incident wave, implying that the electron is not localized, so that we cannot expect the times found in this way to be realistic. To see why these times might be too long, consider the following simple physical picture of the tunnelling process in the monochromatic case. At the right-hand edge of the barrier ($z = 0$) the evanescent wave matches to the free-space solution, and must have a velocity $\nu = p/E$ equal to that of the escaping electrons. The density ρ rises exponentially back into the barrier (as z decreases), but since the current $\rho\nu$ must stay constant in the steady state, the velocity falls, and is predicted to be very low when the barrier is sufficiently wide. Thus the electrons, on first arriving inside the barrier at $z = -\Delta$, spend a long time at low velocity before finally accelerating to their velocity at escape, leading

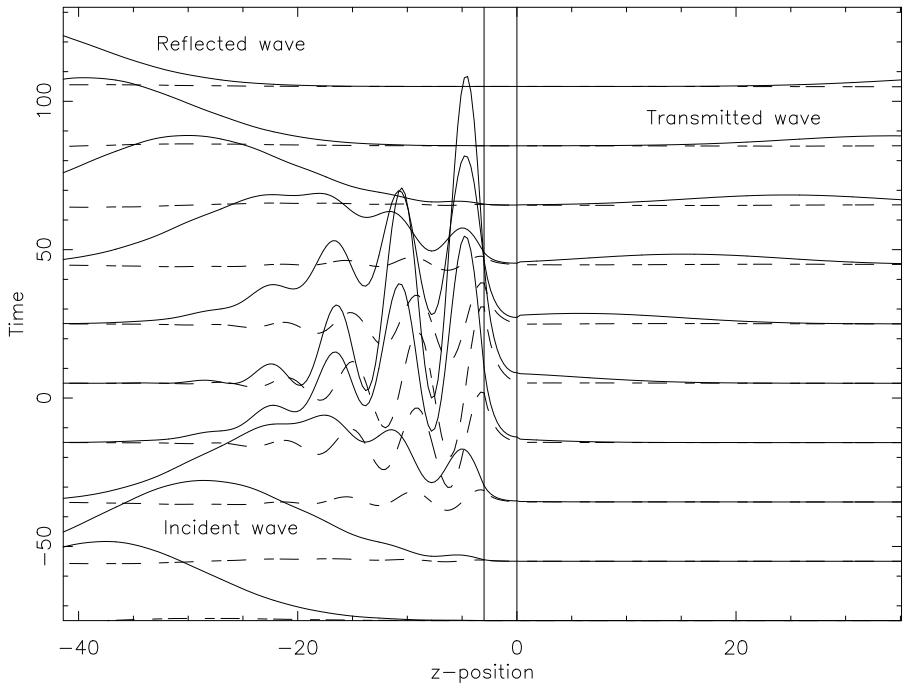

Figure 1: The evolution of a Dirac wavepacket of energy $m\cosh(0.5)$ incident upon a potential barrier of height $eV = 0.3m$. The spatial units are (Compton wavelength)$/2\pi$. The solid and dashed lines represent the scalar and pseudoscalar parts of $\psi\tilde{\psi}$, respectively.

to long tunnelling times. This simple explanation for the variation of velocity with density has a non-relativistic counterpart in the hydrodynamical formulation of the Schrödinger theory [21], to which our relativistic approach reduces when the results of Section 3 are applied. The Bohmian approach to non-relativistic quantum mechanics can also treat this problem, and for a monochromatic wave [22] yields precisely equation (4.12).

Because the tunnelling times predicted for monochromatic waves are unpromising, we now consider the behaviour of a finite-width incident wavepacket. We may calculate the transit time numerically for any given case of interest, but it is useful to have a physical picture of how the tunnelling of a wavepacket differs from the continuous wave case. We illustrate this process in Figure 1, which shows the real and imaginary parts of $\psi\tilde{\psi}$ as a function of time for a Gaussian Dirac wavepacket incident on a barrier. The packet moves through the standing wave set up as the incident and reflected waves interfere. Turning our attention to the region inside

the barrier (between the vertical lines) we see that, as the wavepacket approaches, the amplitude at the barrier at first rises, then falls as the packet moves away. The evanescent wave follows this amplitude closely, with an approximate exponential fall-off into the barrier. Since an exponential curve is self-similar, we can either imagine the exponential as rising and falling in amplitude, or as first protruding into, then retreating from the barrier. The velocity ν_b of this incoming (or outgoing) 'tide' of evanescent wave is given by

$$\kappa\nu_b \approx \frac{d}{dt}\log\rho_b, \tag{4.13}$$

where ρ_b is the density at $z = -\Delta$. For the packet motion we can replace d/dt by $\nu_g(d/dz)$, where $\nu_g = dE/dp$ is the group velocity, which we set equal to the monochromatic velocity ν considered above. The largest tidal velocity therefore occurs at the points of inflection of the packet envelope and, for a Gaussian packet of width Δx, speeds are achieved of order $\nu/(\kappa\Delta x)$, dominating the flow through the first (slow) part of the barrier. Calculations based on this 'tidal' approximation will be reported elsewhere, but the predicted tunnelling times are the right order of magnitude. We have also carried out numerical calculations of the tunnelling times for a packet incident on a barrier, by direct integration of the 4-velocity v along the streamlines starting at different points within the initial wavepacket. The choice of starting point has a profound effect, since streamlines near the front of the packet are 'helped' on their way through the barrier by the first influx of evanescent wave, while ones closer to the rear are swept out of the barrier by the withdrawing tide, and are reflected. This process yields a predicted *distribution* of particle tunnelling times, and it will be of great interest to compare such histograms with subsequent experiments. Leavens [12, 13] has carried out calculations based on Bohm's formulation of the Schrödinger equation, and found results that are very similar to those of our tidal approximation and numerical calculations. We feel that our approach has many advantages, however, in that we need not adopt the ideas of a 'quantum force' or 'quantum potential'. Furthermore, our results automatically include the effects of relativity and spin, which should facilitate more detailed comparisons with experiment.

5 Diffraction Theory

We turn now to the diffraction of electron waves. Suppose that a source produces monochromatic waves incident on an aperture in a surface S. What is the wavefunction in the interior volume V? In order to determine this we need the Dirac equivalent of Kirchoff's diffraction theory. This general theory is not well known, so we outline it here. Again, we find that first-order equations have a considerable advantage over their second-order counterparts.

5.1 The Boundary-Value Problem in Electrodynamics

We begin with a simpler case, relevant to the propagation of electromagnetic waves from a surface. The electromagnetic field is a bivector $F = \mathbf{E} + i\mathbf{B}$, but the analysis is simplified by generalising to a full Dirac spinor ψ. We express the time-dependence of the spinor as

$$\psi = \psi(\mathbf{x})e^{-i\omega t}, \tag{5.1}$$

where i is the pseudoscalar of space, which commutes with all Dirac spinors. The spacetime equation $\nabla\psi = 0$ then becomes

$$\boldsymbol{\nabla}\psi - i\omega\psi = 0, \tag{5.2}$$

which is our first-order equivalent of the Helmholtz equation.

The central theorem we exploit (Green's theorem) relates volume and surface integrals:

$$-\oint |dS(\mathbf{x}')|\tilde{G}(\mathbf{x};\mathbf{x}')\mathbf{n}(\mathbf{x}')\psi(\mathbf{x}') = \int dV(\mathbf{x}')\tilde{G}(\mathbf{x};\mathbf{x}') \overleftrightarrow{\boldsymbol{\nabla}}_{\mathbf{x}'} \psi(\mathbf{x}') = -\psi(\mathbf{x}), \tag{5.3}$$

where $\mathbf{n}$ is the normal to dS pointing *into* the volume V, and $G(\mathbf{x};\mathbf{x}')$ is a Green's function, which satisfies the equation

$$\boldsymbol{\nabla}_{\mathbf{x}'}G(\mathbf{x};\mathbf{x}') - i\omega G(\mathbf{x};\mathbf{x}') = \delta(\mathbf{x}-\mathbf{x}'). \tag{5.4}$$

The solution is easily found by standard methods: let

$$G \equiv \boldsymbol{\nabla}\phi + i\omega\phi, \tag{5.5}$$

so that

$$\boldsymbol{\nabla}^2\phi + \omega^2\phi = \delta. \tag{5.6}$$

This is a scalar operator equation, and the Green's function representing outgoing waves is given by

$$\phi = -\frac{1}{4\pi}\frac{e^{i\omega r}}{r}\ , \tag{5.7}$$

where $r \equiv |\mathbf{r}|$ and $\mathbf{r} \equiv \mathbf{x} - \mathbf{x}'$. Hence, we find that

$$G(\mathbf{x};\mathbf{x}') = -\frac{1}{4\pi}\left(i\omega\frac{e^{i\omega r}}{r} - \sigma_r\frac{d}{dr}\left(\frac{e^{i\omega r}}{r}\right)\right) \tag{5.8}$$

where σ_r is the unit vector in the direction of $\mathbf{r}$. We use this result in (5.3) to give an explicit form for $\psi(\mathbf{x})$:

$$\psi(\mathbf{x}) = -\frac{1}{4\pi}\oint |dS(\mathbf{x}')|\left(i\omega\frac{e^{i\omega r}}{r} + \sigma_r\frac{d}{dr}\left(\frac{e^{i\omega r}}{r}\right)\right)\mathbf{n}(\mathbf{x}')\psi(\mathbf{x}'). \tag{5.9}$$

An alternative expression is obtained by writing ψ in terms of a spinor potential ϕ:

$$\psi \equiv \ \boldsymbol{\nabla}\phi + i\omega\phi, \tag{5.10}$$

where

$$\phi(\mathbf{x}) = -\frac{1}{4\pi}\oint |dS(\mathbf{x}')|\mathbf{n}(\mathbf{x}')\psi(\mathbf{x}')\frac{e^{i\omega r}}{r}. \tag{5.11}$$

Formula (5.9) contains all the necessary polarisation and obliquity factors, and is equivalent to results derived at considerable length in standard optics texts [23, 24]. A great advantage of the STA approach is that *first-order equations satisfy Huygens' principle*. This is evident from (5.3) — the function ψ is propagated into the interior simply by multiplying it by a Green's function. This accords exactly with Huygen's original idea of re-radiation of wavelets from any given wavefront. The spinor potential form (5.11) makes this particularly clear.

Second-order equations, by contrast, are more complicated. The equivalent formula for the propagation of a function satisfying the scalar Helmholtz equation

$$\boldsymbol{\nabla}^2\phi + \omega^2\phi = 0 \tag{5.12}$$

is

$$\phi(\mathbf{x}) = \oint |dS(\mathbf{x}')|\left(\phi\mathbf{n}\cdot\boldsymbol{\nabla}G - G\mathbf{n}\cdot\boldsymbol{\nabla}\phi\right). \tag{5.13}$$

Now we are in a quandary: do we use the Dirichlet Green's function ($G = 0$ on S) and take the first term? Or do we set $\mathbf{n}\cdot\boldsymbol{\nabla}G = 0$ (Neumann conditions) and use the second term? The two approaches yield different results: the first term has a $\cos\theta$ obliquity factor; the second has none. Actually (as the standard texts state) one should take equal amounts of each term, so as to obtain the correct obliquity factor $\frac{1}{2}(1 + \cos\theta)$. There is no such ambiguity for the first-order equation, because the the Green's function *automatically* picks out the appropriate part of ψ, propagating *into* V rather than away from it.

We illustrate this by propagating a constant spinor $\psi(\mathbf{x}') = \psi_0$ away from the plane $z = 0$. We find that

$$\psi(\mathbf{x}) = \frac{1}{2}(1 + \sigma_3)\psi_0 e^{i\omega z} \quad (z > 0), \tag{5.14}$$

$$\psi(\mathbf{x}) = \frac{1}{2}(1 - \sigma_3)\psi_0 e^{-i\omega z} \quad (z < 0). \tag{5.15}$$

The function $\psi(\mathbf{x}') = \psi_0$ has been decomposed into two parts, each propagating separately, in different directions, away from the plane $z = 0$. This behaviour also explains why we can ignore any contribution from the closing surface at $z \to \infty$: there are only *outgoing* waves on that surface.

In order to apply these results to electromagnetic waves, we need a further condition that restricts ψ to be a pure bivector. The formulae given in standard texts [23] are simply the bivector part of our result above, but, since the gradient operator $\boldsymbol{\nabla}$ is grade-mixing, there are implied restrictions on the form of ψ on S. Not only must ψ be a bivector on S (which reduces ψ from eight independent components to six); it must also satisfy a *transversality* condition so that $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$ and $\boldsymbol{\nabla}\cdot\mathbf{B} = 0$ throughout V. Any Fourier component $\psi_{\mathbf{k}}$ with a longitudinal component generates a scalar or pseudoscalar as it propagates off the surface. Specifically, if ψ is given on $z = 0$, and we consider Fourier components of the form

$$\psi(x, y, z = 0) \propto \exp i(k_x x + k_y y), \tag{5.16}$$

then the z-component of the k-vector propagating into $z > 0$ is

$$k_z = |\omega^2 - k_x^2 - k_y^2|. \tag{5.17}$$

The transversality condition is then $\mathbf{k}\psi_{\mathbf{k}} + \psi_{\mathbf{k}}\mathbf{k} = 0$ on S, which limits ψ to four degrees of freedom per Fourier component, as required.

5.2 Diffraction of Dirac Waves

The theory developed above employed the pseudoscalar as the unit imaginary. For Dirac wavefunctions a monochromatic wave (constant energy E) has a time-dependence

$$\psi = \psi(\mathbf{x}) \exp\left(-i\sigma_3 E t\right). \tag{5.18}$$

We handle this case by splitting the wave function into two parts (ideals) defined by the idempotents $\frac{1}{2}(1 \pm \sigma_3)$:

$$\psi = \psi_+ \frac{1}{2}(1 + \sigma_3) + \psi_- \frac{1}{2}(1 - \sigma_3), \tag{5.19}$$

where $\psi_\pm \equiv \psi \frac{1}{2}(1 \pm \sigma_3)$. In conventional language, these are chirality eigenstates. The idempotents satisfy the relations

$$(1 \pm \sigma_3) i\sigma_3 = \pm(1 \pm \sigma_3) i. \tag{5.20}$$

Before examining electron waves, we treat the case of massless neutrinos, which satisfy the equation

$$\boldsymbol{\nabla}\psi - E\psi i\sigma_3 = 0, \tag{5.21}$$

so that

$$\boldsymbol{\nabla}\psi_\pm \mp iE\psi_\pm = 0. \tag{5.22}$$

The separate spinors $\{\psi_\pm\}$ can be propagated exactly as before, using two Green's functions containing the commutative pseudoscalars $\pm i$, and reassembled afterwards. The result is

$$\psi(\mathbf{x}) = -\frac{1}{4\pi} \oint |dS'| \left(\mathbf{n}'\psi(\mathbf{x}') i\sigma_3 E \frac{e^{i\sigma_3 E r}}{r} + \sigma_r \mathbf{n}'\psi(\mathbf{x}') \frac{d}{dr}\left(\frac{e^{i\sigma_3 E r}}{r}\right)\right), \tag{5.23}$$

where $S' = S(\mathbf{x}')$ and $\mathbf{n}' = \mathbf{n}(\mathbf{x}')$. Alternatively, we can define a spinor potential function:

$$\phi(\mathbf{x}) \equiv -\frac{1}{4\pi} \oint |dS'| \mathbf{n}'\psi(\mathbf{x}') \frac{e^{i\sigma_3 E r}}{r}, \tag{5.24}$$

and generate $\psi(\mathbf{x})$ by using the relation

$$\psi = \boldsymbol{\nabla}\phi + E\phi i\sigma_3. \tag{5.25}$$

We are now in a position to deal with the propagator for the Dirac equation

$$\nabla\psi i\sigma_3 = m\psi\gamma_0. \tag{5.26}$$

For a time-dependence of the form $\exp(-i\sigma_3 Et)$ we obtain

$$\boldsymbol{\nabla}\psi i\sigma_3 + E\psi - m\bar{\psi} = 0, \tag{5.27}$$

where $\bar{\psi} \equiv \gamma_0\psi\gamma_0$. We now define

$$\psi_\pm \equiv m\psi - (E \mp p)\bar{\psi}, \tag{5.28}$$

satisfying

$$\begin{array}{rcl} \boldsymbol{\nabla}\psi_+ - p\psi_+ i\sigma_3 & = & 0 \\ \boldsymbol{\nabla}\psi_- + p\psi_- i\sigma_3 & = & 0, \end{array} \tag{5.29}$$

where $p^2 = E^2 - m^2$. According to the above results we find that

$$\psi(\mathbf{x}) = -\frac{1}{4\pi}\oint |dS'| \left(\mathbf{n}'(E\psi(\mathbf{x}') - m\bar{\psi}(\mathbf{x}'))i\sigma_3 \frac{e^{i\sigma_3 pr}}{r} + \sigma_r \mathbf{n}'\psi(\mathbf{x}')\frac{d}{dr}\left(\frac{e^{i\sigma_3 pr}}{r}\right)\right), \tag{5.30}$$

Once again, there is a spinor potential for this problem:

$$\psi = \boldsymbol{\nabla}\phi + (E\phi + m\bar{\phi})i\sigma_3, \tag{5.31}$$

where

$$\phi(\mathbf{x}) = -\frac{1}{4\pi}\oint |dS'|\mathbf{n}'\psi(\mathbf{x}')\frac{e^{i\sigma_3 pr}}{r}. \tag{5.32}$$

6 Electron Paths in Diffracting Systems

Our theory of electron diffraction can be applied to find the wave field from any illuminated aperture. To illustrate this we consider diffraction from two pinholes, of unequal transmittivity. The pinholes lie in the plane $z = 0$, and are illuminated by electron waves with longitudinal spin. We plot streamlines for the Dirac current by starting at a distant point and following numerically the direction of the Dirac velocity, using $\dot{x} = v$, until we reach one of the pinholes. The results are dramatic. In Figure 2(a,b,c) we show the streamlines when the amplitudes are in the ratio $(1:0,\ 1:1/\sqrt{2},\ 1:1)$ respectively. As expected, these show interference effects and, because of the process by which they were generated, the streamlines concentrate

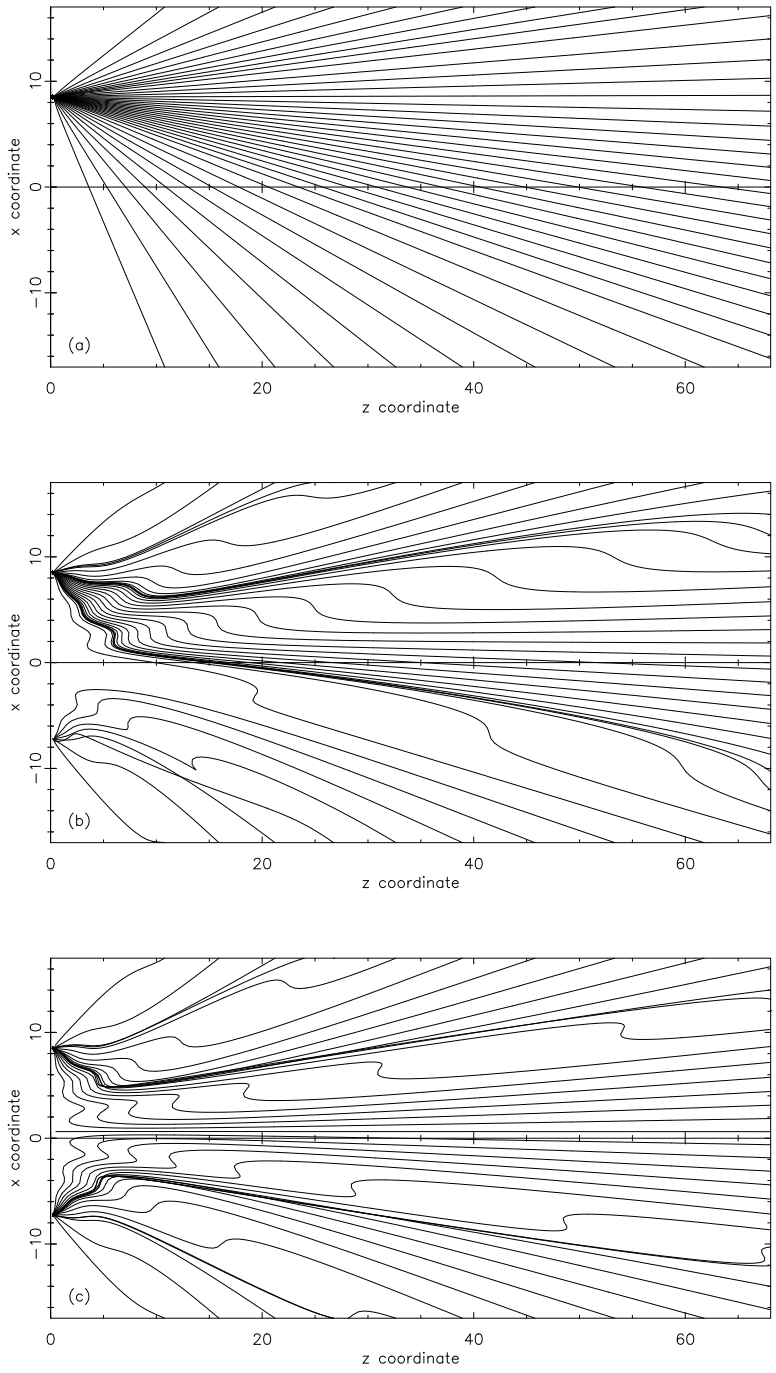

Figure 2: Streamlines of the Dirac curent for a two-pinhole configuration, illuminated by an incident wave of energy $m\cosh(1)$. The amplitudes are in the ratios (a) $1:0$; (b) $1:1/\sqrt{2}$; (c) $1:1$. The spatial units are (Compton wavelength)$/2\pi$.

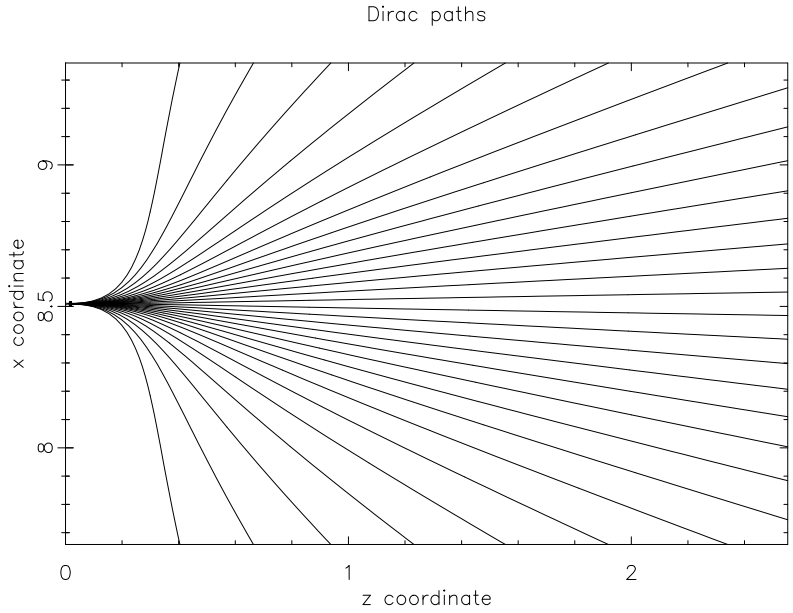

Figure 3: A detailed view of Figure 2(a), showing the behaviour of streamlines near the pinhole.

into the nodes of the interference pattern as they approach the pinholes. Since the Dirac current is strictly conserved, the streamlines never cross and there is a clear-cut boundary between the streamlines going into each pinhole. This accords with predictions made using the Bohmian approach [14]. In our case, the regions 'belonging' to each pinhole have a more complicated shape, because of the effects of electron spin.

Figure 3 is a magnified version of Figure 2a; it shows that the streamlines do not continue into the pinhole on radial trajectories, but line up into the direction of the incident wave as they approach the pinhole. The incident wave momentum dominates the effects of the spin for the majority of these streamlines. The off-axis streamlines close to the pinhole have a complicated spiral pattern, caused by the variation in amplitude of the wave. This is an example of a more general effect which we discuss later.

Streamlines from a point source of Dirac waves are shown in Figure 4. These plots do not correspond to any physically realisable situation, but illustrate some properties of the stationary Dirac Green's function. Figure 4a shows the plane perpendicular to the spin (taken to be the z-axis) and Figure 4b is a polar view. Magnified versions of this diagram are presented as Figures 4c and 4d. Direct

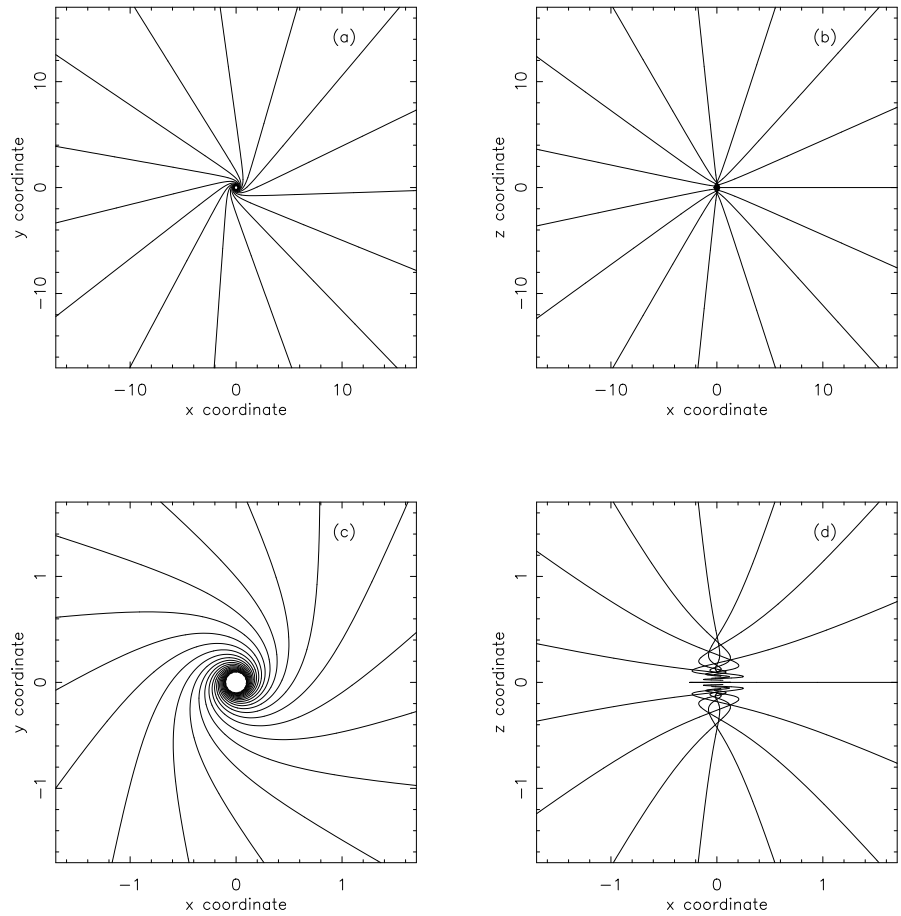

Figure 4: Streamlines of the Dirac current near a point source of electrons, with the spin oriented in the z-direction. The planes shown are (a) $z = 0$ and (b) $y = 0$. The behaviour of the streamlines close to the source (4(c) and 4(d)) reveals a circulating current.

calculation shows how these plots result from competing radial and azimuthal contributions to the current, with the azimuthal term dominating at small distances and the radial term dominating at large distances. The radius over which the circulation term dominates is characterised by the 3-momentum, so that this effect cannot be interpreted as a demonstration of electron 'zitterbewegung', which would take place over distances closer to the Compton wavelength. These diagrams should nevertheless convince sceptics that real-space circulation is inherent in the Dirac theory. Remember that our computers know nothing of the theoretical arguments that the Dirac imaginary operator can be identified with the bivector generator of rotations about the spin axis. The computer seems to have worked that out for itself, just by evaluating the Dirac current in its conventional form.

7 A Critical Look at Dirac Theory

Some further properties of the Dirac equation suggest that it cannot be correct as it stands, and that a deeper structure must be involved. The first problem arises when the free-particle solutions of Section 2 are modified by an electromagnetic field that induces a position dependence in the density, whilst leaving the wave-vector unchanged. In this case we find from the Dirac equation that

$$m\rho v = \rho(k - eA) - \tfrac{1}{2}\nabla\rho\cdot(e_1 e_2), \tag{7.1}$$

so that the velocity/current now contains a component transverse to the momentum. The magnitude of this transverse velocity is $\approx |\boldsymbol{\nabla} \log \rho|/2m$. A related phenomenon is the appearance of a non-zero β-factor, which is determined by [3]

$$-2m\rho \sin\beta = \nabla\cdot(\rho s). \tag{7.2}$$

If s is a constant, β is determined by the component of the spin parallel to the density gradient, as was found for the evanescent waves set up inside a potential barrier (Section 4).

The simplest place to look for transverse currents is in the evanescent wave at a potential step. This situation has been treated by one of us elsewhere [25] and we quote the necessary result here. Following the conventions of Section 4, with a plane wave incident on a potential step at $z = 0$ (with height $eV > E - m$), the

transmitted wavefunction for an electron beam polarised in the x-direction is

$$\psi_b \propto e^{-\kappa z}\frac{1}{\sqrt{2}}\left(e^{i\beta/2} - i\sigma_2 e^{-i\beta/2}\right)e^{-i\sigma_3 Et}. \tag{7.3}$$

This solution produces a velocity in the y-direction, with a magnitude of κ/m. It is tempting to attribute this current, which is perpendicular to the spin, to the non-cancellation of the electron circulation, caused in turn by the variation of amplitude. The longitudinal behaviour, however, has no satisfactory explanation in terms of this simple model.

Another place where these strange effects can be studied is in the ground state of Hydrogen-like atoms. In accordance with our non-relativistic approximation, the exact Dirac theory predicts a circulation in the $1s$ state, with an azimuthal velocity $v = Z\alpha\sin\theta$ where Z is the nuclear charge, α is the fine-structure constant and θ is the polar angle about the direction of electron spin. Again, the position-dependent density gives rise to a nonzero β-factor $\sin\beta = Z\alpha\cos\theta$. A peculiarity of this solution is that the velocity *does not depend on radius*. This suggests that the velocity calculated from the Dirac current does not, in fact, represent the actual velocity of an individual electron, but, rather, an *average* flow.

A further complication (one that is related to the Klein paradox) occurs when $Z\alpha > 1$, in which case the solution for the ground state takes the form:

$$\psi \propto e^{\Gamma t}e^{-Z\alpha r}\frac{1}{r}\left(1 + \sigma_r e^{i\sigma_3\delta}\right)e^{i\sigma_3\Gamma\ln r} \tag{7.4}$$

where

$$\Gamma = (Z^2\alpha^2 - 1)^{1/2}, \quad \cos\delta = \frac{\Gamma}{Z\alpha}. \tag{7.5}$$

The amplitude of this solution increases exponentially with time, with a component of the flow radially outward. The solution has $\beta = \pi/2$, supporting the contention that the β-factor represents the balance between electrons and positrons. We must stress, though, that no satisfactory interpretation of this solution has yet been given.

The final place where difficulties with the Dirac equation can be found is related to the Klein paradox. This indicates [25] that the negative-frequency states behave as though they have negative mass, *not* opposite charge, when compared to positive-frequency ones. This is very inconvenient, since it means that the negative-energy states cannot represent positrons, and it is grimly amusing to witness the contortions that physicists have undergone in order to avoid this problem. These

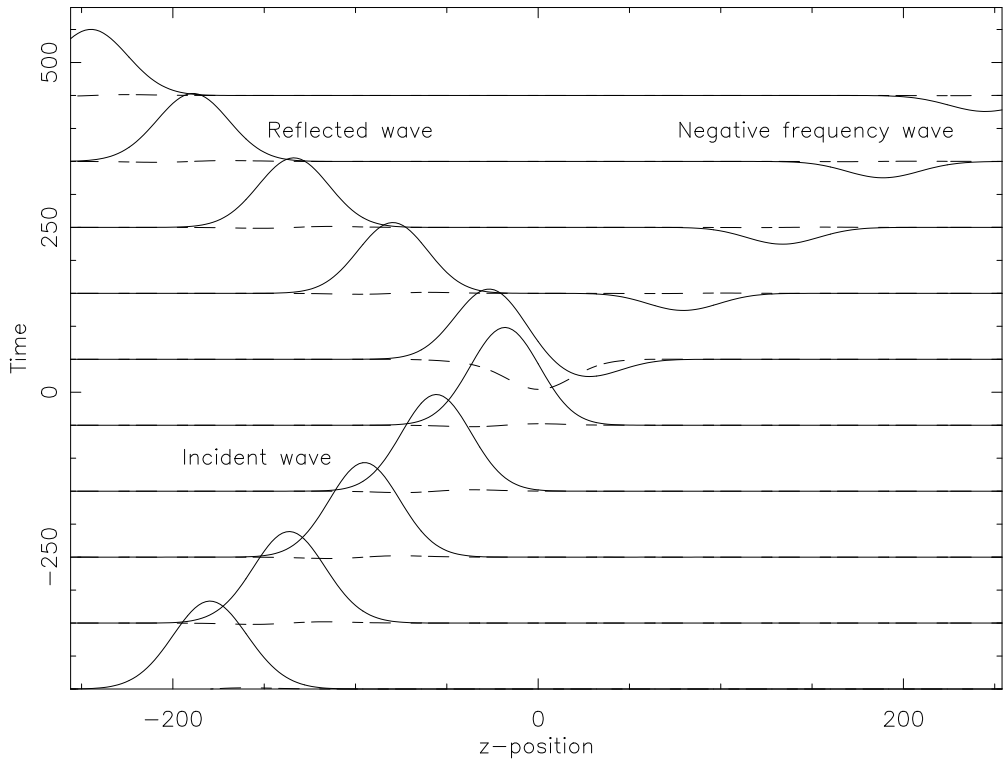

Figure 5: The evolution of a Dirac wavepacket, showing the effect of an electric shock. See text for details of the energies involved. The solid and dashed lines represent the scalar and pseudoscalar parts of $\psi\tilde{\psi}$, respectively. Note that the transmitted (antiparticle) wave has a negative real part.

range from Dirac's original 'holes' [15] and elementary errors (such as matching onto solutions with inappropriate group velocity [26]) through to the more formal method of second quantisation.

Rather than illustrate the failure of the single-particle Dirac equation by means of the conventional Klein paradox [27], we give a further graphic example of wavepacket motion. Figure 5 shows a Dirac wavepacket propagating to the right (z increasing), with an average momentum of $p = \sinh(0.5)$. At time $t = 0$ the wavepacket suffers a violent electric shock of the form

$$eV = \Delta p z \delta(t), \tag{7.6}$$

where $\Delta p = \sinh(1)$. The effect of this shock on any plane-wave component at

$t = 0$ is to multiply it by a phase factor:

$$R(\mathbf{p})e^{i\sigma_3 pz} \longrightarrow R(\mathbf{p})e^{i\sigma_3(p-\Delta p)z}. \tag{7.7}$$

The wavefunction now has a spatial phase dependence appropriate to a different momentum $p' = p - \Delta p$, and its future evolution can accordingly be analysed as a combination of the positive- and negative-frequency solutions

$$\psi_+ = R(\mathbf{p}')e^{-i\sigma_3(E't-p'z)} \quad \text{and} \quad \psi_- = iR(-\mathbf{p}')e^{i\sigma_3(E't+p'z)}. \tag{7.8}$$

After the shock, the positive and negative components separate, because they have opposite group velocities. The negative-energy wave in the future would be forbidden if Feynman boundary conditions had been chosen for this problem, but here we are concerned with analyzing the case where only a single real electron is input from the past. (Adopting Feynman boundary conditions would force us to use a combination of positive- and negative-energy waves *in the past*, in order to supress the outward going negative-energy wave in the future.) It would be nice to interpret the appearance of the negative-energy wave in Fig. 5 as being due to pair production, but there is a major difficulty, which is clearly visible in the diagram. *The amplitude of the reflected wave is less than that of the incident wave.* This demonstrates that the charge current $\rho v = \psi\gamma_0\tilde{\psi}$ has been shared between the two resultant waves: the negative-energy waves have the same sign of charge as their positive-energy counterparts. Since the behaviour of these waves in electromagnetic fields shows that their charge/mass ratio is reversed, they must have *negative mass.* The inescapable conclusion is that the first-quantised Dirac theory cannot account for positrons. There must be something wrong with with the theory of the coupling of the electron wavefunction to the electromagnetic field A. This problem is circumvented by the formal method of second quantisation, in which the Dirac equation is replaced by the Feynman rules of QED. Remarkably, this enables the electromagnetic coupling to be dismantled and reassembled in a manner that produces useful and accurately verified results. Yet QED is still based on the same coupling to the electromagnetic field that was used in the Dirac theory. At a deeper level, something must still be wrong.

The figures presented in this paper will provoke an obvious question: what do these streamlines mean? Frankly, we do not know; we have simply investigated their behaviour. In order to provide such an answer, it is necessary to adopt a definite interpretation of quantum mechanics, and it seems to us that no currently available viewpoint is remotely satisfactory. David Hestenes has commented on these issues

before [19, 18, 6] and has usually favoured the viewpoint given by Ballentine [28]. In order to account for the non-local multi-particle mysteries, however, some other, atemporal, ingredient seems to be necessary, after the manner of the theory of electromagnetism developed by Wheeler and Feynman [29, 30]. Such a viewpoint is provided by Cramer's 'transactional' interpretation of quantum mechanics [31], but this is based entirely upon a discussion of the Schrödinger equation, and Cramer's view that the conjugate wavefunction Ψ^* represents a time-advanced wave clearly fails in the case of the Dirac equation. Indeed, in the light of the spacetime algebra analysis, any discussion of quantum mechanics based solely on non-relativistic equations seems to be utterly doomed!

For these reasons, we are currently members of the agnostic, or 'none of the above' school of quantum mechanics, preferring to reserve judgement until further information is available. We present these plots and predictions only in the hope of provoking discussion and experiment. For example, from the streamlines in spacetime, we can produce a histogram of transit times through a barrier. If experiments verify this distribution of times, then it is possible that a single electron does, in some sense, follow such a trajectory. Of course, if experiments did not find this distribution, it would be very strong evidence to the contrary.

By airing these reservations, we hope that we have not detracted too much from David Hestenes' central themes: that the spacetime algebra is the best available tool for studying the Dirac equation, and that the single-particle Dirac wavefunction represents a spacetime rotation determining a Lorentz boost to a definite velocity. This part of his thesis is proved beyond doubt. What is needed now is an extension of his insights to provide a consistent interpretation of the multiparticle theory.

Acknowledgements

The four papers in this series are gratefully dedicated to David Hestenes, with whom we have had many beneficial and thought-provoking discussions. The authors thank Anton Garrett for his careful checking of these manuscripts, and his numerous suggestions for improvements.

References

[1] S.F. Gull, A.N. Lasenby, and C.J.L. Doran. Imaginary numbers are not real — the geometric algebra of spacetime. *Found. Phys.*, 23(9):1175, 1993.

[2] C.J.L. Doran, A.N. Lasenby, and S.F. Gull. States and operators in the spacetime algebra. *Found. Phys.*, 23(9):1239, 1993.

[3] A.N. Lasenby, C.J.L. Doran, and S.F. Gull. A multivector derivative approach to Lagrangian field theory. *Found. Phys.*, 23(10):1295, 1993.

[4] D. Hestenes. Vectors, spinors, and complex numbers in classical and quantum physics. *Am. J. Phys.*, 39:1013, 1971.

[5] D. Hestenes. Observables, operators, and complex numbers in the Dirac theory. *J. Math. Phys.*, 16(3):556, 1975.

[6] D. Hestenes. Clifford algebra and the interpretation of quantum mechanics. In J.S.R. Chisholm and A.K. Common, editors, *Clifford Algebras and their Applications in Mathematical Physics (1985)*, page 321. Reidel, Dordrecht, 1986.

[7] D. Hestenes. Real Dirac theory. In Preparation, 1994.

[8] D. Hestenes. *Space–Time Algebra*. Gordon and Breach, New York, 1966.

[9] D. Hestenes. Proper dynamics of a rigid point particle. *J. Math. Phys.*, 15(10):1778, 1974.

[10] D. Hestenes. On decoupling probability from kinematics in quantum mechanics. In P. Fougère, editor, *Maximum Entropy and Bayesian Methods*, page 161. Kluwer Academic, Dordrecht, 1990.

[11] D. Bohm and B.J. Hiley. Unbroken quantum realism, from microscopic to macroscopic levels. *Phys. Rev. Lett.*, 55:2511, 1985.

[12] C.R. Leavens. Transmission, reflection and dwell times within Bohm's causal interpretation of quantum mechanics. *Solid State Comm.*, 74(9):923, 1990.

[13] C.R. Leavens. Traversal times for rectangular barriers within Bohm's causal interpretation of quantum mechanics. *Solid State Comm.*, 76(3):253, 1990.

[14] J.P. Vigier, C. Dewdney, P.R. Holland, and A. Kyprianidis. Causal particle trajectories and the interpretation of quantum mechanics. In B.J. Hiley and F.D. Peat, editors, *Quantum Implications*, page 169. Routledge, London, 1987.

[15] P.A.M. Dirac. A theory of electrons and protons. *Proc. Roy. Soc. (London) A*, 126:360, 1930.

[16] D. Bohm, R. Schiller, and J. Tiomno. A causal interpretation of the Pauli equation. *Nuovo Cim. Suppl.*, 1:48, 1955.

[17] C. Dewdney, P.R. Holland, A. Kyprianidis, and J.P. Vigier. Spin and non-locality in quantum mechanics. *Nature*, 336:536, 1988.

[18] D. Hestenes. Spin and uncertainty in the interpretation of quantum mechanics. *Am. J. Phys.*, 47(5):399, 1979.

[19] D. Hestenes. Local observables in the Dirac theory. *J. Math. Phys.*, 14(7):893, 1973.

[20] E.H. Hauge and J.A. Støvneng. Tunnelling times: a critical review. *Rev. Mod. Phys.*, 61(4):917, 1989.

[21] J.O. Hirschfelder, A.C. Christoph, and W.E. Palke. Quantum mechanical streamlines. I. Square potential barrier. *J. Chem. Phys.*, 61(12):5435, 1974.

[22] T.P Spillar, T.D. Clark, R.J. Prance, and H. Prance. Barrier traversal time in the quantum potential picture. *Europhys. Lett.*, 12(1):1, 1990.

[23] J.D. Jackson. *Classical Electrodynamics*. Wiley, New York, third edition, 1999.

[24] M. Born and E. Wolf. *Principles of Optics*. London, 1959.

[25] S.F. Gull. Charged particles at potential steps. In A. Weingartshofer and D. Hestenes, editors, *The Electron*, page 37. Kluwer Academic, Dordrecht, 1991.

[26] J.D. Bjorken and S.D. Drell. *Relativistic Quantum Mechanics, vol 1*. McGraw-Hill, New York, 1964.

[27] O. Klein. Die reflexion von elektronen an einem potentialsprung nach der relativistischen dynamik von Dirac. *Z. Physik*, 53:157, 1929.

[28] L.E. Ballentine. The statistical interpretation of quantum mechanics. *Rev. Mod. Phys.*, 42(4):358, 1970.

[29] J.A. Wheeler and R.P. Feynman. Classical electrodynamics in terms of direct interparticle action. *Rev. Mod. Phys.*, 21(3):425, 1949.

[30] J.A. Wheeler and R.P. Feynman. Interaction with the absorber as the mechanism of radiation. *Rev. Mod. Phys.*, 17:157, 1945.

[31] J.G. Cramer. The transactional interpretation of quantum mechanics. *Rev. Mod. Phys.*, 58(3):647, 1986.