The zitterbewegung interpretation of quantum mechanics — **Source link**

David Hestenes

Institutions: Arizona State University

Related papers:

- Zitterbewegung and the internal geometry of the electron

- On the Zitterbewegung of the Dirac Electron

- Quantum mechanics from self-interaction

- Space-time algebra

- Classical Model of the Dirac Electron

The Zitterbewegung Interpretation of
Quantum Mechanics

David Hestenes

Abstract. The zitterbewegung is a local circulatory motion of the electron presumed to be the basis of the electron spin and magnetic moment. A reformulation of the Dirac theory shows that the zitterbewegung need not be attributed to interference between positive and negative energy states as originally proposed by Schroedinger. Rather, it provides a physical interpretation for the complex phase factor in the Dirac wave function generally. Moreover, it extends to a coherent physical interpretation of the entire Dirac theory, and it implies a zitterbewegung interpretation for the Schroedinger theory as well.

1. INTRODUCTION

The idea that the electron spin and magnetic moment are generated by a localized circulatory motion of the electron has been proposed independently by many physicists. Schroedinger's *zitterbewegung* (zbw) model for such motion is especially noteworthy, because it is grounded in an analysis of solutions to the Dirac equation.[1,2,3] Surely, if the zbw is a real physical phenomena, then it tells us something fundamental about the nature of the electron. However, the role ascribed to the zbw in standard formulations of quantum mechanics has been metaphorical at best.

My purpose in this paper is to argue that the zbw is not an esoteric or adventitious relativistic feature of the Dirac theory as commonly believed. Rather, the zbw is a ubiquitous phenomena with manifestations in every application of quantum mechanics, even in the nonrelativistic domain.

A wholesale reinterpretation of quantum mechanics is necessary to recognize this. But the reinterpretation is not arbitrarily imposed on the mathematical formalism. Rather, like Schroedinger's original ideal about the zbw, it arises from analyzing solutions of the Dirac equation to see what they tell us about the structure of the electron and its motion. In analyzing free-particle wave packet solutions of the Dirac equation, Schroedinger noted the existence of "interference" between positive and negative energy states oscillating with circular frequency

$$\omega_0 = 2mc^2/\hbar = 1.6 \times 10^{21} \text{s}^{-1} \tag{1}$$

He interpreted this as a fluctuation in the position of the electron with radius

$$\lambdabar_0 = c/\omega_0 = \hbar/2mc = 1.9 \times 10^{-13}\text{m} \tag{2}$$

For an electron moving at the speed of light about a mean position this entails an angular momentum

$$\lambdabar_0(mc) = \hbar/2 \tag{3}$$

which is naturally interpreted as the spin angular momentum of the electron.

Interpreting the oscillation between positive and negative energy states quite literally, the zbw is sometimes supposed to originate in electron-positron pair creation and annihilation.[4] That would seem to make the zbw a purely relativistic phenomena. But then how could the zbw be the origin of spin, which is surely significant in the nonrelativistic domain? We resolve this paradox by interpreting the mathematics differently.

I shall show that *the complex phase factor in the electron wave function can be associated directly with the zbw*. I call this *the zbw interpretation of quantum mechanics*. Obviously, it has far-reaching implications, since complex phase factors play a critical role in every application of quantum mechanics. In the first place, it implies that Schroedinger's original wave packet oscillation is merely an epiphenomenon revealing the zbw periodicity which was already inherent in the complex phase factors of both electron and positron plane wave states. The essential feature of the zbw idea is the association of the spin with a local circulatory motion characterized by the phase factor. Since the complex phase factor is the main feature which the Dirac wave function shares with its nonrelativistic limit, it follows that the Schroedinger equation for an electron inherits a zbw interpretation from the Dirac theory. It follows that such familiar consequences of the Schroedinger theory as barrier penetration can be interpreted as manifestations of the zbw.

To show how natural it is to impose the zbw interpretation on the Dirac theory, I employ a reformulation of the theory in terms of the real *spacetime algebra* which has been developed and discussed at length in several previous publications.[5,11] A brief review of this reformulation is given below to establish a foundation for reinterpretation. The main point of the reformulation is to reveal geometric structure in the Dirac theory which is hidden in the conventional formulation. This makes it possible to see clearly how the Dirac wave function describes the kinematics of electron motion and how this depends on definitions (i.e., interpretations) of observables in the theory.

Finally, in a study of plane wave solutions of the Dirac equation we come to the essential new idea proposed in this article: A simple change in the definition (i.e., interpretation) of the electron velocity implies that *all components of the electron wave function (including its phase) directly describe kinematical features of electron motion*. The uniquely compelling feature of this new interpretation is that it ascribes maximal geometric content to the Dirac wave function. Moreover, it entails the essential features of the zbw described above, and it generalizes automatically to the electron with arbitrary electromagnetic interactions.

2. THE GEOMETRIC CONTENT OF THE DIRAC ALGEBRA

My first contention is that the Dirac algebra contains irrelevant features which must be eliminated to reveal its true geometric content. The simplest way to carry out the elimination is to define the "geometrically purified" version of the Dirac algebra, which I call the *spacetime algebra (STA)*, and then compare it with the conventional matrix version.

We begin with the usual Minkowski model of spacetime as a four-dimensional vector space $\mathcal{M}^4$. In mathematical parlance, STA is the real Clifford algebra of the Minkowski metric on $\mathcal{M}^4$. More specifically, STA is a real associative algebra generated from $\mathcal{M}^4$ by defining an associative product on $\mathcal{M}^4$ with the special property that the square of every vector is scalar-valued. I call this product the *geometric product* to emphasize the

fact that it has a definite geometric interpretation which fully characterizes the geometrical properties of spacetime. The geometric product uv of vectors u and v can be interpreted by decomposing it into symmetric and skew-symmetric parts; thus,

$$uv = u \cdot v + u \wedge v \tag{4}$$

where two new products have been introduced and defined by

$$u \cdot v = \tfrac{1}{2}(uv + vu) \tag{5}$$

$$u \wedge v = \tfrac{1}{2}(uv - vu) = -v \wedge u \tag{6}$$

It follows from the definition of the geometric product that $u \cdot v$ is a scalar; indeed, it is the usual inner product defined on Minkowski space. The quantity $u \wedge v$ is called a *bivector*, and it represents a directed plane segment in the same way that a vector represents a directed line segment.

Note that for orthogonal vectors (as defined by $u \cdot v = 0$), Eq. (4) gives $uv = u \wedge v = -vu$. Thus, the geometric relation of orthogonality is expressed algebraically by an anticommutative geometric product. Similarly, collinearity is expressed by a commutative geometric product. For in that case $u \wedge v = 0$ and (4) gives $uv = u \cdot v = vu$. In general, (4) shows that the geometric product represents the relative direction of any two vectors by a combination of commutative and anticommutative parts.

To facilitate comparison with the Dirac matrix algebra, it is convenient to characterize the structure of STA in terms of a basis. Let $\{\gamma_\mu; \mu = 0, 1, 2, 3\}$ be a right-handed orthonormal basis for $\mathcal{M}^4$ with timelike vector γ_0 in the forward (future) light cone. In terms of this basis the spacetime metric is expressed by the equations

$$\gamma_0^2 = 1 = -\gamma_k^2 \qquad \text{for} \quad k = 1, 2, 3 \tag{7}$$

and

$$\gamma_\mu \cdot \gamma_\nu = 0 \qquad \text{for} \quad \mu \neq \nu \tag{8}$$

Other basis elements of STA, each with a definite geometric interpretation, can be generated from the γ_μ by multiplication. For example, $\gamma_2\gamma_1 = \gamma_2 \wedge \gamma_1$ is a bivector of unit magnitude, as expressed by

$$(\gamma_2\gamma_1)^2 = -1 \tag{9}$$

Geometrically, $\gamma_2\gamma_1$ can be interpreted as the unit (directed) area element for the plane containing γ_1, and γ_2. It is also the generator of rotations in that plane, for it is easily proved that multiplication by $\gamma_2\gamma_1$ rotates any vector in that plane through a quarter turn. Of great importance to us later on will be the fact that all Lorentz rotations are generated by the bivectors of STA.

The unit *pseudoscalar* for spacetime is so important that the special symbol i will be reserved to represent it. lts generation by the vector basis is expressed by

$$i = \gamma_0\gamma_1\gamma_2\gamma_3 \tag{10}$$

Geometrically, it represents the unit oriented 4-volume element for spacetime. Its algebraic properties

$$i^2 = -1 \tag{11}$$

3

$$i\gamma_\mu = -\gamma_\mu i \tag{12}$$

make it easy to manipulate. Multiplication of (10) by γ_0 yields the pseudovector

$$\gamma_1\gamma_2\gamma_3 = \gamma_0 i \tag{13}$$

Geometrically, this is the (directed) unit 3-volume element for a hyperplane with normal γ_0.

By forming all distinct products of the γ_μ, we obtain a complete basis for STA consisting of the $2^4 = 16$ linearly independent elements

$$1, \quad \gamma_\mu, \quad \gamma_\mu \wedge \gamma_\nu, \quad \gamma_\mu i, \quad i \tag{14}$$

It follows that a generic element M in STA, called a *multivector*, can be written in the *expanded form*

$$M = \alpha + a + B + bi + \beta i \,, \tag{15}$$

where α and β are scalars, a and b are vectors, and (with summation over repeated indices)

$$B = \tfrac{1}{2}B^{\mu\nu}\gamma_\mu \wedge \gamma_\nu \tag{16}$$

is a bivector.

To facilitate decompositions relative to the basis $\{\gamma_\mu\}$, it is convenient to introduce a reciprocal basis $\{\gamma^\mu\}$ defined by the conditions

$$\gamma^\mu \cdot \gamma_\nu = \delta^\mu_\nu \tag{17}$$

This implies that $\gamma_0 = \gamma^0$ and $\gamma^k = -\gamma_k$. The components of the vector a in (15) are then given by $a^\mu = a \cdot \gamma^\mu$ so $a = a^\mu\gamma_\mu$. Similarly, the bivector components $B^{\mu\nu}$ in (16) are given by

$$B^{\mu\nu} = \gamma^\mu \cdot (\gamma^\nu \cdot B) = (\gamma^\mu \wedge \gamma^\nu) \cdot B \tag{18}$$

The multivector M in (15) can be decomposed into an even part M_+ and an odd part M_-, as expressed by

$$M = M_+ + M_- \tag{19a}$$

$$M_+ = \alpha + B + \beta i \tag{19b}$$

$$M_- = a + bi \tag{19c}$$

A multivector is said to be even (odd) if its odd (even) part vanishes.

For M in the expanded form (15), the operation of reversion in STA is defined by

$$\widetilde{M} = \alpha + a - B - bi + \beta i \tag{20}$$

It follows that for any multivectors M and N,

$$(MN)^{\sim} = \widetilde{N}\widetilde{M} \tag{21}$$

Essentially, reversion amounts to reversing the order of geometric products.

The relation of STA to the Dirac algebra is now easy to state. The Dirac matrices, commonly denoted by the symbols γ_μ, can be put into one-to-one correspondence with the basis vectors denoted by the same symbols above. Then the algebra generated by the Dirac matrices *over the real* is isomorphic to STA. It follows that the geometric meaning attributed to the vectors γ_μ and their products above is inherent in the Dirac algebra, though it is scarcely recognized in the literature. This isomorphism completely defines the geometric content of the Dirac algebra with respect to spacetime. It suggests also that the representation of the γ_μ by matrices is irrelevant to their function in physical theory. This suggestion is confirmed in the next section by casting the Dirac theory in terms of STA with no reference at all to matrices.

The full Dirac algebra is generated by the γ_μ, over a complex instead of a real number field. The fact that the real field suffices to express the full geometric content of the algebra suggests that the 16 additional degrees of freedom introduced by employing a complex field instead are physically irrelevant. This suggestion is also confirmed in the next section by formulating the Dirac theory without them. Elimination of the irrelevant $\sqrt{-1}$ in the complex number field opens up the possibility of discovering a geometric meaning for the $\sqrt{-1}$ which occurs so prominently in the equations of quantum mechanics. Indeed, Eqs. (7), (9), and (11) show that STA contains *many different roots of minus one*, including three geometrically different types. Each type plays a different role in the Dirac theory.

3. THE GEOMETRIC STRUCTURE OF THE DIRAC THEORY

A review of the reformulation of the Dirac theory in terms of STA is given in Ref. 11, and many more details are supplied in Refs. 8 and 9. These references also contain proofs of mathematical equivalence to the conventional matrix formulation. Proofs will be omitted in presenting the results of interest here in order to concentrate on their geometrical and physical interpretations.

The Dirac wave function at a spacetime point x is represented in STA as an even multivector $\psi = \psi(x)$. It has the Lorentz invariant factorization

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R \tag{22}$$

where i is the unit pseudoscalar, ρ and β are scalars, and R satisfies

$$R\widetilde{R} = \widetilde{R}R = 1 \tag{23}$$

At each point x, the function $R = R(x)$ determines a Lorentz rotation (i.e., a proper, orthochronous Lorentz transformation) of a given fixed frame of vectors $\{\gamma_\mu\}$ into a frame $\{e_\mu = e_\mu(x)\}$ given by

$$e_\mu = R\gamma_\mu\widetilde{R} \tag{24}$$

In other words, R determines a unique frame field on spacetime. This frame field has a physical interpretation. First, the vector field

$$\psi\gamma_0\widetilde{\psi} = \rho e_0 = \rho v \tag{25}$$

is the *Dirac current*, which according to the Born interpretation is to be interpreted as a probabililty current. Accordingly, at each point x, the timelike vector $v = v(x) = e_0(x)$ is

5

interpreted as the probable (proper) velocity of the electron, and $\rho = \rho(x)$ is the relative probability (i.e., proper probability density) that the electron actually is at x. Second, the vector field

$$\frac{\hbar}{2}\,\psi\gamma_3\widetilde{\psi} = \rho\,\frac{\hbar}{2}\,e_3 = \rho s\,. \tag{26}$$

is the *spin* (or polarization) *vector* density. The *spin angular momentum* $S = S(x)$ is actually a bivector quantity, related to the spin vector s by

$$S = isv = \frac{\hbar}{2}\,ie_3e_0 = \frac{\hbar}{2}\,e_2e_1 = \frac{\hbar}{2}\,R\gamma_2\gamma_1\widetilde{R} \tag{27}$$

Multiplying this on the right by (22) and using (23), one easily obtains

$$S\psi = \tfrac{1}{2}\psi\gamma_2\gamma_1\hbar \tag{28}$$

which relates the spin S to the bivector $\gamma_2\gamma_1\hbar$.

In general, six parameters are needed to specify an arbitrary Lorentz rotation. Five of the parameters in the Lorentz rotation (24) are needed to specify the direction of the electron velocity v and spin s. This also determines the plane containing e_1 and e_2, as shown in (27). The remainlng parameter determines the orientation of e_1 and e_2 in the e_2e_1-plane. This parameter is the phase of the wave function, and here we have a geometrical interpretation of the phase. The vectors e_1 and e_2 are not given a physical interpretation in the conventional formulation of the Dirac theory, because the matrix formalism suppresses them completely. They will be given a kinematical interpretation in the next section when the zbw interpretation is introduced.

The factorization (22) of the wave function ψ can now be seen as a decomposition into a six-parameter *kinematical factor* R and a two-parameter statistical factor $(\rho e^{i\beta})^{\frac{1}{2}}$. The parameter ρ is clearly a probability density, but the interpretation of β is problematical. In the next section it will be seen that the value of β distinguishes electron and positron solutions of the free-particle Dirac equation, which suggests that $\cos\beta$ can be interpreted as the relative probability of observing an electron or a positron. The significance of β is discussed further in the references.[7,10]

For analytical purposes, coordinates x^μ of a spacetime point x can be defined by

$$x^\mu = \gamma^\mu \cdot x \tag{29}$$

$x = x^\mu \cdot \gamma_\mu$. The derivative with respect to x can be defined by

$$\square = \gamma^\mu \partial_\mu \tag{30}$$

where

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \gamma_\mu \cdot \square\,. \tag{31}$$

Since $\square$ is a vector differential operator, the derivative of a vector field $J = J(x)$ can be decomposed into scalar and bivector parts by using (4); thus

$$\square J = \square \cdot J + \square \wedge J \tag{32}$$

6

where $\square \cdot J = \partial_\mu J^\mu$ is the usual *divergence* and $\square \wedge J$ is the *curl* of J.

For the above representation $\psi = \psi(x)$ of the Dirac wave function, the *Dirac equation* takes the form

$$\square \psi \mathbf{i} \hbar - \frac{e}{c} A \psi = mc \psi \gamma_0 \tag{33}$$

where $A = A_\mu \gamma^\mu$ is the usual electromagnetic vector potential, and $\mathbf{i}$ is the unit bivector

$$\mathbf{i} = \gamma_2 \gamma_1 = i \gamma_3 \gamma_0 \tag{34}$$

It has been proved that (33) is completely equivalent (isomorphic) to the conventional matrix version of the Dirac equation.[2,7,8,12] Equation (33) is Lorentz invariant, despite the explicit appearance of the constants γ_0 and $\mathbf{i} = \gamma_2 \gamma_1$ in it. These constants are arbitrarily specified by writing (33). They need not be identified with the vectors of a particular coordinate system, though it is often convenient to do so. The only requirement is that γ_0 be a fixed timelike unit vector, while $\mathbf{i}$ is a spacelike unit bivector which commutes with γ_0. Of course, the γ_0 and $\mathbf{i} = \gamma_2 \gamma_1$, in (33) are the same constants that appear in the expressions (25) and (27) for the Dirac current and the spin.

The most striking thing about (33) is that the role of the unit imaginary in the matrix version of the Dirac equation has been taken over by the unit bivector $\mathbf{i}$, and this reveals that it has a geometric meaning. Indeed, Eqs. (27) and (28) show that $\mathbf{i}\hbar$ is to be identified with the spin. It enables us also to recognize that the spin was "smuggled" into the Dirac theory by the (kinetic) energy-momentum operators $\underline{p}_\mu$, defined in STA by

$$\underline{p}_\mu \psi = \partial_\mu \psi \mathbf{i} \hbar - \frac{e}{c} A_\mu \psi \tag{35}$$

These operators are given a physical meaning by using them to define the components $T_{\mu\nu}$ of the electron *energy-momentum tensor*:

$$T_{\mu\nu} = \langle \gamma_0 \widetilde{\psi} \gamma_\nu (\underline{p}_\mu \psi) \rangle \tag{36}$$

where $\langle M \rangle$ means "scalar part" of M. At each spacetime point x, the vector

$$T(n) = n^\mu T_{\mu\nu} \gamma^\nu \tag{37}$$

is the energy-momentum flux through a hypersurface with unit normal $n = n^\mu \gamma_\mu = n(x)$. The predicted *proper momentum* p for an electron at x is defined by

$$T(v) = \rho p \tag{38}$$

This is the energy-momentum flux in the direction v of electron motion; in other words, it is the energy-momentum density in the instantaneous rest system of the electron. In general, the momentum p is not collinear with the velocity v, because it includes a contribution from the spin.

From the Dirac equation it follows that[7]

$$\partial_\mu T^\mu = F \times J \tag{39}$$

7

where $T^\mu = T(\gamma^\mu)$, $F = \Box \wedge A$ is the electromagnetic field,

$$J = e\rho v = e\psi\gamma_0\widetilde{\psi} \tag{40}$$

is the electron charge current (density), and $F \times J$ is the commutator product defined by

$$F \times J = \tfrac{1}{2}(FJ - JF) \tag{41}$$

The conventional tensor component form of (39) is

$$\partial_\mu T^{\mu\nu} = F^{\mu\nu}J_\mu \tag{42}$$

Note that the right side of (39) is exactly the classical Lorentz force and does not include the "Stern-Gerlach type" force that one might expect from a medium with an intrinsic magnetic moment density. This is one clue that the electron magnetic moment arises from a circulation of electron charge.

It follows by combining the Dirac equation with (36) and (35) that[7]

$$T^\mu \wedge \gamma_\mu = T_{\mu\nu}\gamma^\nu \wedge \gamma^\mu = \mathbf{i}(\Box \wedge (\rho s)) \tag{43}$$

where ρs is the vector spin density by (26). With the help of this result and $\partial_\mu x = \gamma_\mu$, it follows easily from (39) that

$$\partial_\mu[T^\mu \wedge x + i\rho(s \wedge \gamma^\mu)] = (F \times J) \wedge x \tag{44}$$

This leads to the identification of the bivector-valued tensor

$$J(n) = T(n) \wedge x + i\rho(s \wedge n) \tag{45}$$

as the *total angular momentum* tensor, describing the angular momentum flux in the direction n at x. In particular, by virtue of (38) and (27),

$$J(v) = \rho(p \wedge x + S) \tag{46}$$

This expresses the total angular momentum of the electron as the sum of an orbital angular momentum $p \wedge x$ and a spin angular momentum S. It justifies the interpretation of the bivector $S = isv$ as spin and establishes its origin in the $\mathbf{i}\hbar$ included in the definition (35) of the energy-momentum operators.

Insight into the kinematic significance of the momentum vector can be gained by further consideration of the kinematic factor R in wave function decomposition (22). The normalization condition (23) implies that the derivative of R can be written in the form

$$\partial_\mu R = \tfrac{1}{2}\Omega_\mu R \tag{47}$$

here the Ω_μ are bivectors. Differentiation of (24) then gives

$$\partial_\mu e_\nu = \Omega_\mu \times e_\nu \tag{48}$$

Clearly Ω_μ is the *rotation rate* of the frame $\{e_\nu\}$ as it is displaced in the direction γ_ν.

As mentioned before, the Dirac equation implies the probability concervation law

$$\square \cdot (\rho v) = 0 \tag{49}$$

It follows that through each point x where $\rho(x)$ is nonvanishing there passes a unique timelike curve (here called a *streamline*) with tangent v at every point. Thus, the Dirac equation, by determining the conserved vector field ρv, also determines a bundle of nonintersecting streamlines in spacetime. In accordance with the statistical interpretation, each streamline can be regarded as a possible spacetime history of the electron. The theory predicts that the electron follows one of these streamlines, though it predicts which one only with a relative probability. The fact that attributing a continuous trajectory to the electron is fully consistent with such quantum phenomena as diffraction has been explained by Bohm and others.[13,14]

Along any streamline $x = x(\tau)$ parametrized by its proper time τ, Eq. (47) gives

$$\dot{R} = \tfrac{1}{2}\Omega R \tag{50}$$

where $\Omega = cv^\mu \Omega_\mu$ and the overdot indicates the proper time derivative

$$\dot{e}_\mu = \Omega \times e_\mu \tag{51}$$

The derivative of the frame $\{e_\mu = e_\mu(x(\tau))\}$ along the streamline is therefore given by

$$\frac{1}{c}\frac{d}{d\tau} = v \cdot \square = v^\mu \partial_\mu \tag{52}$$

This is a purely kinematical equation. The dynamics is in specifying the functional form of the bivector Ω, and that comes from the Dirac equation.

Applying the energy-momentum operator (35) to R and using (47) with (27), one obtains

$$(\underline{p}_\mu R)\widetilde{R} + \frac{e}{c}A_\mu = \Omega_\mu S = \Omega_\mu \cdot S + \Omega_\mu \wedge S + \partial_\mu S \tag{53}$$

where, in accord with (48),
$$\partial_\mu S = \Omega_\mu \times S \tag{54}$$

When (53) is used in (36) to evaluate $T_{\mu\nu}$ one finds for the proper momentum in (38) the expression

$$p_\mu = \Omega_\mu \cdot S - \frac{e}{c}A_\mu \tag{55}$$

Hence, the energy in the local rest frame of the electron is given by

$$p \cdot v = \Omega \cdot S - \frac{e}{c}A \cdot v \tag{56}$$

These equations show that the momentum p arises from the kinematics of the frame $\{e_\mu\}$. Indeed, $\Omega \cdot S$ has the familiar form of rotational kinetic energy for a body with angular momentum S and rotational velocity Ω. This raises a fundamental question of physical interpretation: Exactly what is rotating? How can the entire frame $\{e_\mu\}$ and its kinematics

9

be related to a sensible model of the electron? A general answer is proposed in the next section based on the study of a specific solution to the Dirac equation.

4. PLANE WAVES AND ZITTERBEWEGUNG

For a free particle, the plane-wave solution of the Dirac equation has the general form

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R_0 \, e^{-\mathbf{i} p \cdot x / \hbar} \tag{57}$$

where all the x-dependence is explicitly shown in the phase factor. Inserting this in the Dirac equation (33) with $A = 0$ and using $\Box \, p \cdot x = p$, one obtains

$$p\psi = mc\psi\lambda_0$$

Multiplication on the right by $\widetilde{\psi}$ yields

$$p e^{i\beta} = mcv \tag{58}$$

which can be satisfied only for $e^{i\beta} = \pm 1$; whence $p = \pm mcv$. The unphysical minus sign in this relation between p and v can be eliminated by reversing the sign in the phase of (57). Thus, the Dirac equation has two distinct types of plane wave solutions with $p = mcv$, and electron solution

$$\psi = \rho^{\frac{1}{2}} R_0 e^{\mathbf{i} p \cdot x / \hbar} \tag{59}$$

and a positron solution

$$\psi_C = \rho^{\frac{1}{2}} i R_0 e^{-\mathbf{i} p \cdot x / \hbar} \tag{60}$$

Let us examine the geometry of the electron solution (59) in detail.

For constant v, the equation

$$v \cdot x = c\tau \tag{61}$$

describes a one-parameter family of hyperplanes with unit normal v. The streamlines of ψ are straight lines with tangent v. The kinematical factor in (59) can be written in the form

$$R = e^{\frac{1}{2}\Omega\tau} R_0 \tag{62}$$

where Ω is the constant bivector

$$\Omega = mc^2 S^{-1} = \frac{2mc^2}{\hbar} e_1 e_2 \tag{63}$$

with $e_1 e_2 = R\gamma_1\gamma_2\widetilde{R} = -R_0 \mathbf{i} \widetilde{R}_0$. This R satisfies Eq. (50). Accordingly, Ω is the angular velocity of the frame $\{e_\mu = e_\mu(x(\tau))\}$ as it moves along a streamline. Both $e_0 = v$ and $e_s = \hat{s}$ are constants of the motion, but

$$
\begin{aligned}
e_1(\tau) &= e^{\Omega\tau} e_1(0) = e_1(0) \cos \omega_0\tau + e_2(0) \sin \omega_0\tau \\
e_2(\tau) &= e^{\Omega\tau} e_2(0) = e_2(0) \cos \omega_0\tau - e_1(0) \sin \omega_0\tau
\end{aligned}
\tag{64}
$$

10

where $\omega_0 = |\Omega|$ is the zbw frequency with the value given by (1). Thus e_1 and e_2 rotate with the zbw frequency in the plane of the spin S. Moreover, (63) in (56) associates an energy with this rotation:

$$p \cdot v = \Omega \cdot S = mc^2 \tag{65}$$

This strongly suggests that the "rest mass" of the electron has a kinetic origin, but that can be so only if the electron has a velocity component in the spin plane. Thus, the interpretation of v as electron velocity should be modified. Fortunately, there is a simple and elegant way to do this which actually increases the intelligibility of the Dirac theory.

Suppose that the velocity of the electron can be identified with the null vector

$$u = e_0 - e_2 \tag{66}$$

Of course, this means that the electron moves with the speed of light, as in Schroedinger's original zbw model. However, for a free particle with momentum p, this motion is completely described by the "positive energy" wave function (59) or more simply by its kinematical factor (62). The Dirac current describes the average velocity $\bar{u}$ over a zbw period, that is,

$$\bar{u} = e_0 = v \tag{67}$$

since $\dot{e}_2 = 0$. With the time dependence of e_2 given by (64) and $u = c^{-1}\dot{z}$, Eq. (66) is easily integrated to get the history $z = z(\tau)$ of the electron; thus,

$$z(\tau) = vc\tau = (e^{\Omega\tau} - 1)r_0 + z_0 \tag{68}$$

This is a parametric equation for a lightlike helix $z(\tau) = x(\tau) + r(\tau)$ centered on the streamline $x(\tau) = vc\tau + z_0 - r_0$ with radius vector

$$r(\tau) = e^{\Omega\tau}r_0 = -\frac{c}{\omega_0}e_1 = -\frac{c}{\omega_0^2}\dot{u} \tag{69}$$

The diameter of the helix is the electron Compton wavelength $2\lambda_0 = 2c/\omega_0 = \hbar/mc$, as suggested by Eq. (2).

By (63) and (69), which implies $\dot{r} = \Omega r$,

$$S = mc^2\Omega^{-1} = mr^2\Omega = mr\dot{r} \tag{70}$$

Thus, the spin angular momentum can be regarded as the angular momentum of zbw fluctuations. This conclusion from the free-particle solution can be generalized.

Suppose that the zbw average $\bar{u} = v$ applies generally even with interacting fields. Similarly, the momentum defined by (39) should be regarded as a zbw average momentum defined by

$$\bar{p} = \overline{m(v \cdot u)u} \tag{71}$$

This must be related to the "actual" momentum p by

$$p = \bar{p} + \Delta p \tag{72}$$

where $\overline{\Delta p} = 0$ by definition. The corresponding angular momentum can then be written

$$p \wedge z = (\bar{p} + \Delta p) \wedge (x + r) \tag{73}$$

11

where $x = \bar{z}$ and $r = 0$. When $\bar{p}$ and x vary slowly over a zbw period, the values of $\overline{\bar{p} \wedge r}$ and $\overline{(\Delta p) \wedge x}$ must be relatively small, so the average angular momentum is

$$\overline{\bar{p} \wedge z} = \bar{p} \wedge x + S \tag{74}$$

where

$$S = \overline{(\Delta p) \wedge r} \tag{75}$$

is identified as the spin angular momentum. This agrees with the expression (46) for the total angular momentum and suggests that observables in the Dirac theory represent averages over a zbw substructure. It thereby also points to a domain where deviations from the Dirac theory may be expected, though pair production may confound efforts to probe this domain.

With $c\tau = v \cdot z = v \cdot x$, Eq. (68) describes a spacetime-filling congruence of lightlike helixes with exactly one helix through each spacetime point. According to the statistical interpretation, the electron is a point particle which may traverse any of these helixes with equal probability. Alternatively, it may be suggested that the electron is an extended body and the helixes are world lines of its component parts. This suggestion faces difficulties which seem to rule it out. First, there is an absence of evidence for any interaction among the parts which would be needed to make the body cohere. Second, the dimensions of the body would have to be on the order of a Compton wavelength ($\sim 10^{-13}$m). But this is much too big! Scattering experiments limit the size of the electron (i.e., the size of the domain in which momentum transfer takes place) to less than 10^{-18} m.[15]

The *zbw interpretation of the Dirac Theory*, founded on the identification of the null vector $u = e_0 - e_2$ as electron velocity, provides an explanation for the electron spin and magnetic moment in the circulation of momentum and charge. It also explains the mass as the energy of this circulation. It is natural to attribute the origin of the zbw to the self-interaction of the electron with its own electromagnetic field. The relation (65) suggests that *the self-interaction is of magnetic origin*, since it has the form of a Larmor precession energy if Ω is proportional to a self-generated magnetic field. The so-called "rest mass" of the electron is therefore a kinetic energy of magnetic self-interaction. It is this that gives the electron its inertial properties. And the flywheel-like nature of this inertia may be the ultimate origin of spin dependence in electron scattering.

It has also been suggested that the electron zbw is the source of an electromagnetic field which fluctuates with the zbw frequency.[11] This frequency is too high to detect directly, but it provides a mechanism for explaining such strange features of quantum mechanics as the Pauli principle.[11]

The formulation of the zbw given here implies that the most common features of quantum mechanics are manifestations of the zbw, so we may speak of a *zbw interpretation of quantum mechanics* generally. The *key ingredients* of this interpretation are the energy momentum operators $\underline{p}_\mu$ defined by (35) and the complex phase factor in the wave function. The imaginary unit $\mathbf{i}$ in both is a bivector for a plane in space, the "spin plane" in which the zbw circulation takes place. The phase factor literally represents a physical rotation, a zbw rotation. Operating on the phase factor, the $\underline{p}_\mu$, computes the rotation rates of the phase in time and space directions, identifying them with the energy and momentum. Thus, the zbw interpretation explains why the mysterious operators $\underline{p}_\mu$ are so significant in quantum mechanics.

The key ingredients of the zbw interpretation are preserved in the approximations leading from the Dirac theory through the Pauli theory to the Schroedinger theory. The nonrelativistic approximation to the STA version of the Dirac theory has been treated in detail elsewhere.[8,9] But the main point can be seen by expressing the general Dirac wave function (22) in the form

$$\psi = \rho^{\frac{1}{2}} e^{i\beta/2} L U e^{-\mathbf{i}(mc^2/\hbar)t} \tag{76}$$

where the kinematical factor R has been broken into three parts, the L factor describing a boost of the vector γ_0 into the velocity v. In the nonrelativistic approximation, three of the factors in (76) are neglected or eliminated and ψ is reduced to the Pauli wave function

$$\psi_P = \rho^{\frac{1}{2}} U = \rho^{\frac{1}{2}} U_0 e^{-\mathbf{i}\varphi/\hbar} \tag{77}$$

where the kinematical factor U has been broken into a phase factor describing the zbw rotation and a spatial rotation factor U_0 which rotates $\mathbf{i}$ into the direction of the spin. Many aspects of spin and the zbw in the Pauli theory have been discussed in Ref. 10, though the present article brings some improvements. In the Schroedinger approximation the factor U_0 is neglected, so ψ_P reduces to the Schroedinger wave function

$$\psi_{Sh} = \rho^{\frac{1}{2}} e^{-\mathbf{i}\varphi/\hbar} \tag{78}$$

It follows from this derivation of the Schroedinger wave function that the phase $\varphi/\hbar$ describes the zbw, just as in the Dirac theory, and $\partial_\mu \varphi$ describes the zbw energy and momentum.

This being so, there should be a zbw interpretation for every feature of the Schroedinger theory. Indeed, the Heisenberg uncertainty relations can be attributed to the "zbw fact" that an electron cannot be confined to a region smaller than a Compton wavelength.[10,12] Also, the stationary states of a bound electron exhibit a resonance of the orbital frequency with harmonics of the zbw frequency, which is imposed formally in the theory by requiring single-valuedness of the wave function. Evidently such resonances, so prominent in quantum mechanics, can be interpreted as zbw resonances. And should we not be able to explain barrier penetration as due to zbw fluctuations in momentum? Or the Aharonhov-Bohm effect as a shift in zbw phase? These are among the new challenges raised by the zbw interpretation. But the greatest challenge will be to prove that the zbw is a real physical phenomena and not just a metaphorical trick. That will probably require physical detection of some zbw effect which is not already characterized by the Dirac theory or even quantum electrodynamics.

5. THE GEOMETRY OF LIGHTLIKE HISTORIES

From a theoretical perspective, the most convincing feature of the zbw interpretation is, perhaps, the fact that it maximizes the geometric meaning of the Dirac wave function by giving the complex phase factor a kinematical interpretation. To emphasize that geometrical fact, the general *Frenet equations* for an arbitrary lightlike history are developed below, and from these the precise geometrical specifications for the zbw history of a free electron

are readily seen. The compact STA formulation of the Frenet equations for timelike histories was given in Ref. 16, but a comparable treatment of lightlike histories has not been published before.

A *Frenet frame* $\{e_\mu = e_\mu(x(\tau))\}$ for a particle history $x = x(\tau)$ is an orthonormal comoving frame which characterizes the shape of the curve at every point. The parameter τ is naturally taken to be the proper time for a timelike curve, but proper time cannot be defined on a lightlike curve. However, the parameter can be defined indirectly by specifying the curvature, and this will be taken for granted in the following.

The velocity $u = dx/d\tau$ of a lightlike history is a null vector related to the Frenet frame by

$$u = \dot{x} = e_- \tag{79}$$

where $e_\pm = e_0 \pm e_2$. As noted before, the motion of the frame can be regarded as a variable Lorentz rotation

$$e_\mu = R\gamma_\mu \widetilde{R} \tag{80}$$

determined by the equation

$$\dot{R} = \tfrac{1}{2}\Omega R \tag{81}$$

so that

$$\dot{e}_\mu = \Omega \times e_\mu \tag{82}$$

The e_μ can be related to derivatives of u by the following definition of the "Darboux" bivector:

$$\begin{aligned}
\Omega &= \tfrac{1}{2}(\kappa_1 e_1 e_+ + \kappa_2 e_- e_1 + \kappa_3 e_3 e_-) \\
&= \tfrac{1}{2}(\kappa_1 - \kappa_2)e_1 e_0 + \tfrac{1}{2}(\kappa_1 + \kappa_2)e_1 e_2 + \tfrac{1}{2}\kappa_3 e_3(e_0 - e_2)
\end{aligned} \tag{83}$$

The scalar coefficient κ_i is called the ith curvature of the curve. The curvatures κ_1 and κ_2 are defined to be positive and the sign of κ_3 is determined by the assumption that $\{e_1, e_2, e_3\}$ is a right-handed set of vectors. Note that

$$e_+ e_- = 2 + 2e_2 e_0 \tag{84}$$

so $e_+ \cdot e_- = 2$ and $e_+ \wedge e_- = 2e_2 e_0$.

Insertion of (83) into (82) yields

$$\begin{aligned}
\dot{e}_- &= \dot{u} = \kappa \cdot e_1 \\
\dot{e}_1 &= \tfrac{1}{2}(\kappa_1 e_- - \kappa_2 e_-) = \tfrac{1}{2}(\kappa_1 - \kappa_2)e_0 + \tfrac{1}{2}(\kappa_1 + \kappa_2)e_2 \\
\dot{e}_+ &= -\kappa_2 e_1 + \kappa_3 e_3 \\
\dot{e}_3 &= \tfrac{1}{2}\kappa_3 e_-
\end{aligned} \tag{85}$$

and a combination of Eqs. (85) gives

$$\begin{aligned}
\dot{e}_0 &= \tfrac{1}{2}(\kappa_1 - \kappa_2)e_1 + \tfrac{1}{2}\kappa_3 e_3 \\
\dot{e}_2 &= -\tfrac{1}{2}(\kappa_1 - \kappa_2)e_1 + \tfrac{1}{2}\kappa_3 e_3
\end{aligned} \tag{86}$$

The curvatures are therefore related to the e_μ and $\dot{e}_\mu$ by

$$\kappa_1 = -e_1 \cdot \dot{e}_- = e_- \cdot \dot{e}_1$$
$$\kappa_2 = -e_+ \cdot \dot{e}_1 = e_1 \cdot \dot{e}_+ \tag{87}$$
$$\kappa_3 = -e_3 \cdot \dot{e}_+ = e_+ \cdot \dot{e}_3 = -2e_3 \cdot \dot{e}_0$$

The four equations (82) can be solved for[16]

$$\Omega = \tfrac{1}{2} \dot{e}_\mu e^\mu \tag{88}$$

and this can be used to regain (83) from (85) and (86).

Several general properties of Ω are worth noting. Equations (80) can be used to put (83) in the form

$$\Omega = R \Omega_0 \widetilde{R} \tag{89}$$

where

$$\Omega_0 = \tfrac{1}{2}(\kappa_1 \gamma_1 \gamma_+ + \kappa_2 \gamma_- \gamma_1 + \kappa_3 \gamma_3 \gamma_-) \tag{90}$$

and $\gamma_\pm = \gamma_0 \pm \gamma_2$, of course. Differentiation of (89) yields

$$\dot{\Omega} = \tfrac{1}{2}(\dot{\kappa}_1 e_1 e_+ + \dot{\kappa}_2 e_- e_1 + \dot{\kappa}_3 e_3 e_-) \tag{91}$$

Another property of Ω is the invariant

$$\Omega^2 = -\kappa_1(\kappa_2 + i\kappa_3) \tag{92}$$

Also, in general,

$$\dot{\Omega} \times \Omega = \tfrac{1}{2}\kappa_1^2 e_1 \left[e_0 \frac{d}{d\tau} \left(\frac{\kappa_2}{\kappa_1} \right) + e_3 \frac{d}{d\tau} \left(\frac{\kappa_3}{\kappa_1} \right) \right] \tag{93}$$

which vanishes when κ_2/κ_1, and κ_3/κ_1 are constant.

According to (91), when the curvatures are constant, $\dot{\Omega} = 0$, so (81) integrates immediately to

$$R = e^{\frac{1}{2}\Omega\tau} R_0 \tag{94}$$

where R_0 is a constant specifying the initial conditions. With this result in $u = R\gamma_- \widetilde{R}$, Eq. (79) can be integrated to get a parametric equation for the history. The most efficient general method for evaluating this integral involves an important "trick" given in Ref. 17.

Comparison of (83) with (63) shows at once that the *zitterbewegung* plane wave solutions of the Dirac equation are curves of constant curvature, with $\kappa_3 = 0$ and

$$\kappa_1 = \kappa_2 = \omega_0 = 2mc^2/\hbar \tag{95}$$

The most significant point to be made here is that the physical variables are hereby given a completely geometrical interpretation, so prospects are opened up for a geometrical extension of electron theory.

The identification of the free electron solution with a particular class of paths with constant curvature suggests the possibility of classifying elementary particles according

to curvature. The simplest possibility is identification of the free neutrino with the case $\kappa_1 = \kappa_2 = 0$, so (83) becomes

$$\Omega = \tfrac{1}{2}\kappa_1 e_1 e_+ \tag{96}$$

Since $\Omega^2 = 0$, Ω is the generator of lightlike Lorentz transformations, and this can be sharply distinguished from the electron case where $\Omega^2 < 0$, The corresponding wave function ψ_ν for a free neutrino with momentum $p = \tfrac{1}{2}\hbar\kappa_1 e_+$ is obtained by inserting (96) into (94) and using $\tau = x \cdot e_+$; thus,

$$\psi_\nu = \rho^{\frac{1}{2}} e^{\frac{1}{2}\Omega x \cdot e_-} R_0 = \rho^{\frac{1}{2}} R_0 e^{-\gamma_+ \gamma_1 p \cdot x/\hbar} \tag{97}$$

This satisfies the usual neutrino equation

$$\Box \psi_\nu = 0 \tag{98}$$

though it is not the conventional neutrino solution. However, the similarity between the two solutions is much greater if the first curvature in (96) is a function of, say, $\cos(p \cdot x/\hbar)$. It remains to be seen whether there is a significant physical difference between them. Supposing that the present solution is physically viable, the speculation can proceed further. Presumably κ_1, would have the same value for the electron and its neutrino and a different value for the muon and its neutrino, the values of κ_1 being quantized by some unknown mechanism of self-interaction. The values $\kappa_1 = \kappa_2$ and $\kappa_2 = 0$ evidently distinguish charged from uncharged particles. Perhaps $\kappa_3 = 0$ similarly distinguishes leptons from quarks with $\kappa_3 \neq 0$. That's enough speculation for one paper!

REFERENCES

1. E. Schroedinger, *Sitzungber. Preuss. Akad. Wiss. Phys.-Math. Kl.* **24**, 418 (1930).

2. Huang, "On the *zitterbewegung* of the electron," *Am. J. Phys.* **47**, 797 (1949).

3. A. O. Barut and A. J. Bracken, "*Zitterbewegung* and the internal geometry of the electron," *Phys. Rev. D* **23**, 2454 (1981).

4. H. Thirring, *Principles of Quantum Electrodynamics* (Academic Press, New York, 1958).

5. D. Hestenes, *Space-Time Algebra* (Gordon & Breach, London, 1966).

6. D. Hestenes, "Real spinor fields," *J. Math. Phys.* **8**, 798–808 (1967).

7. D. Hestenes, "Local observables in the Dirac theory," *J. Math. Phys.* **14**, 893–905 (1973).

8. D. Hestenes, "Observables, operators and complex numbers in the Dirac theory," *J. Math. Phys.* **16**, 556–572 (1975).

9. R. Gurtler and D. Hestenes, "Consistency in the formulation of the Dirac, Pauli, and Schroedinger theories," *J. Math. Phys.* **16**, 573–583 (1975).

10. D. Hestenes, "Spin and uncertainty in the interpretation of quantum mechanics," *Am. J. Phys.* **47**, 339–415 (1979).

11. D. Hestenes, "Clifford algebra and the interpretation of quantum mechanics," in *Clifford Algehras and Their Applications in Mathematical Physics*, J. S. R. Chisholm and A. K. Common, eds. (Reidel, Dordrecht, 1986), p. 321–346.

12. D. Hestenes, "Quantum mechanics from self-interaction," *Found. Phys.* **15**, 63–87 (1985).

13. D. Bohm and B. Hiley, "Unbroken quantum realism, from microscopic to macroscopic levels." *Phys. Ret. Lett.* **55**, 2511 (1985).

14. J.-P. Vigier, C. Dewney, P. R. Holland, and A. Kypriandis, "Causal particle trajectories and the interpretation of quantum mechanics," in *Quantum Implications*, B. J. Hiley and F. D. Peat, eds. (Routledge & Kegan Paul, London, 1987).

15. D. Bender *et al.*, "Tests of QED at 29 GeV center-of-mass energy," *Phys. Ret. D* **30**, 515 (1984).

16. D. Hestenes, "Proper particle mechanics," *J. Math. Phys.* **15**, 1768–1777 (1974).

17. D. Hestenes, "Proper dynamics of a rigid point particle," *J. Math. Phys.* **15**, 1778–1786 (1974).