In J. Keller and Z. Oziewicz (Eds.), *The Theory of the Electron*,
UNAM, Facultad de Estudios Superiores, Cuautitlan, Mexico (1996), p. 1–50.

REAL DIRAC THEORY

David Hestenes
Department of Physics & Astronomy
Arizona State University
Tempe, AZ 85287

Abstract

The Dirac theory is completely reformulated in terms of *Spacetime Algebra*, a real Clifford Algebra characterizing the geometrical properties of spacetime. This eliminates redundancy in the conventional matrix formulation and reveals a hidden geometric structure in the theory. Among other things, it reveals that complex numbers in the Dirac equation have a kinematical interpretation, with the unit imaginary identified as the generator of rotations in a spacelike plane representing the direction of electron spin. Thus, spin and complex numbers are shown to be inextricably related in the Dirac Theory. This leads to a new version of the zitterbewegung, wherein local circular motion of the electron is directly associated with the phase factor of the wave function. In consequence, the electron spin and magnetic moment can be attributed to the zitterbewegung, and many other features of quantum mechanics can be explained as zitterbewegung resonances.

INTRODUCTION.

This paper reviews and consolidates results from a line of research[1–14] aimed at clarifying the Dirac electron theory and simplifying its mathematical formulation. The central result is a representation of the Dirac wave function which reveals geometric structure in the Dirac theory — structure which is not at all apparent in the conventional formulation. Besides computational benefits, this result has many implications for the interpretation of quantum mechanics, ranging from the classical limit to possibilities of a substructure in the Dirac theory. Before we can address these issues, however, an analysis and revision of the underlying mathematical formalism is needed.

Most physicists regard the Dirac algebra as the algebra of a relativistic spin-$\frac{1}{2}$ particle. However, there is a broader interpretation with far-reaching consequences. Dirac's gamma matrices can be regarded as representations of spacetime vectors. These vectors generate a *real* Clifford algebra, which has been dubbed the *Spacetime Algebra* (STA), because it provides a complete, nonredundant characterization of the metrical and directional properties of spacetime. All the elements of the STA have straightforward geometric interpretations. This brings to light a geometric significance of the Dirac algebra which was implicit in Dirac's original formulation but only dimly recognized since. Since STA is a universal algebra of spacetime properties, it is more than a special tool for characterizing spin-$\frac{1}{2}$ particles. It is equally applicable to every domain of classical and quantum physics.

Reformulation of the Dirac Theory in terms of STA reveals that the conventional matrix formulation contains superfluous degrees of freedom in its use of complex numbers. This

redundancy is eliminated in the STA formulation wherein the base number field is real rather than complex. Its most immediate and striking consequence is a revelation of the geometrical role played by complex numbers in the Dirac Theory. It reveals that the unit imaginary appearing in the Dirac equation and the energy-momentum operators represents the bivector generator of rotations in a spacelike plane corresponding to the direction of electron spin. In other words, the *factor $i\hbar$ represents spin in the Dirac theory*—the i represents the spin direction while $\hbar/2$ is the magnitude of the spin. It must be emphasized that this is not an adventitious interpretation of $\hbar$ arbitrarily imposed on the Dirac theory. It has been implicit in the theory from the beginning. The STA formulation only makes it explicit.

In as much as complex wave functions clearly play a crucial role in quantum mechanics, the discovery that the ubiquitous factor $i\hbar$ is inextricably associated with spin in the Dirac theory must surely be a critical clue to a deeper interpretation of quantum mechanics. Its implications are not immediately obvious, however, and they evidently cannot be understood apart from a thorough analysis of the Dirac theory. The approach reviewed in this paper is to *systematically study the geometric structure* of the Dirac equation and its solutions as revealed by the STA formulation. *The geometric structure of the Dirac theory is taken to be the most reliable guide to its physical interpretation* as well as to feasible modifications and extensions.

The central result is an invariant decomposition of the Dirac wave function into a 2-parameter statistical factor and a 6-parameter kinematical factor. Among its implications are the following: The statistical factor characterizes the admixture of electron-positron states without employing Fourier analysis. The kinematical factor admits a kinematical interpretation of the complex phase factor and relates it to the spin. The fullest interpretation is achieved by a simple change in identification of the electron velocity which reveals a *zitterbewegung* substructure inherent in the Dirac theory. This opens up possibilities for probing the substructure experimentally. Other possibilities arise from noting that the invariance group of the Dirac current has the structure of the electroweak gauge group.

In two appendixes the coupled Dirac-Maxwell field equations are derived from a Lagrangian and the associated conservation laws are derived from Poincaré invariance.

This article was originally prepared in 1990 and circulated as a preprint. The whole approach has been further elaborated along several lines since.[37–46] References 45 and 46 are especially noteworthy because they present a number of physical applications, including tunneling, electron diffraction and Stern-Gerlach splitting, with new insights into each of them.

1. SPACETIME ALGEBRA AND CALCULUS.

We shall employ a flat space model of the physical spacetime manifold in this article, so each point event can be represented by a unique element x in a real 4-dimensional vector space $\mathcal{M}^4$. The properties of scalar multiplication and vector addition in $\mathcal{M}^4$ provide only a partial characterization of spacetime geometry. To complete the characterization we introduce an associative *geometric product* among vectors with the property that the square x^2 of any vector x is a (real) scalar. As usual, we say that the vector x is *timelike*, *lightlike* or *spacelike* if $x^2 > 0,\ x^2 = 0,\ x^2 < 0$ respectively.

The vector space $\mathcal{M}^4$ is not closed under the geometric product. Rather, by multiplication and addition it generates a real associative (but noncommutative) algebra of dimension $2^4 = 16$, commonly called the *Geometric Algebra* (or Clifford Algebra) of $\mathcal{M}^4$. We call it the *spacetime algebra* (STA), because all its elements and operations represent geometric elements and relations, and it suffices for the representation of any geometric structure on spacetime.

To facilitate applications of STA to physics it is necessary to build a system of definitions and theorems. From the *geometric product* uv of two vectors it is convenient to define two other products. The *inner product* $u \cdot v$ is defined by

$$u \cdot v = \tfrac{1}{2}(uv + vu) = v \cdot u \,, \tag{1.1}$$

while the *outer product* $u \wedge v$ is defined by

$$u \wedge v = \tfrac{1}{2}(uv - vu) = v \wedge u \,. \tag{1.2}$$

The three products are therefore related by

$$uv = u \cdot v + u \wedge v \,, \tag{1.3}$$

and this can be regarded as a decomposition of the product uv into symmetric and skewsymmetric parts.

The inner and outer products can be generalized. We define the outer product along with the notion of *k-vector* iteratively as follows: Scalars are defined to be 0-vectors, vectors are 1-vectors, and bivectors, such as $u \wedge v$, are 2-vectors. For a given k-vector K, the integer k is called the *step* (or grade) of K. For $k \geq 1$, the outer product of a vector v with a k-vector K is a $(k+1)$-vector defined in terms of the geometric product by

$$v \wedge K = \tfrac{1}{2}(vK + (-1)^k Kv) = (-1)^k K \wedge v \,. \tag{1.4}$$

The corresponding inner product is defined by

$$v \cdot K = \tfrac{1}{2}(vK + (-1)^{k+1} Kv) = (-1)^{k+1} K \cdot v \,, \tag{1.5}$$

and it can be proved that the result is a $(k-1)$-vector. Adding (1.4) and (1.5) we obtain

$$vK = v \cdot K + v \wedge K \,, \tag{1.6}$$

which obviously generalize (1.3). The important thing about (1.6), is that it decomposes vK into $(k-1)$-vector and $(k+1)$-vector parts.

By continuing as above, the STA as been developed into a complete coordinate-free calculus for spacetime physics. However, to hasten comparison with standard Dirac theory, we interrupt that process to introduce coordinates and a basis for the algebra. Let $\{\gamma_\mu;\ 0, 1, 2, 3\}$ be a *right-handed orthonormal frame* of vectors with γ_0 in the forward light cone. In accordance with (1.1), the components $\mathrm{g}_{\mu\nu}$ of the metric tensor for this frame are given by

$$\mathrm{g}_{\mu\nu} = \gamma_\mu \cdot \gamma_\nu = \tfrac{1}{2}(\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu) \,. \tag{1.7}$$

The unit pseudoscalar i for spacetime is related to such a frame by the equation

$$i = \gamma_0\gamma_1\gamma_2\gamma_3 = \gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3 \,. \tag{1.8}$$

It is readily verified from (1.8) that $i^2 = -1$, and the geometric product of i with any vector is anticommutative. Hereafter it will be convenient to refer to $\{\gamma_\mu\}$ as a *standard frame*.

For manipulating coordinates it is convenient to introduce the *reciprocal frame* $\{\gamma^\mu\}$ defined by the equations

$$\gamma_\mu = \mathrm{g}_{\mu\nu}\gamma^\nu \qquad \text{or} \qquad \gamma_\mu \cdot \gamma^\nu = \delta_\mu^\nu \,. \tag{1.9}$$

(Summation convention in force!) The relation of scalar coordinates x^μ to the spacetime point x they designate is then given by

$$x^\mu = \gamma^\mu \cdot x \quad \text{and} \quad x = x^\mu\gamma_\mu \,. \tag{1.10}$$

By multiplication the γ_μ generate a complete basis of k-vectors for STA, consisting of the $2^4 = 16$ linearly independent elements

$$1, \quad \gamma_\mu, \quad \gamma_\mu \wedge \gamma_\nu, \quad \gamma_\mu i, \quad i \,. \tag{1.11}$$

Any multivector can be expressed as a linear combination of these elements. For example, a bivector F has the expansion

$$F = \tfrac{1}{2}F^{\mu\nu}\gamma_\mu \wedge \gamma_\nu \,, \tag{1.12a}$$

with its "scalar components" $F^{\mu\nu}$ given by

$$F^{\mu\nu} = \gamma^\mu \cdot F \cdot \gamma^\nu = \gamma^\mu \cdot (\gamma^\nu \cdot F) = (\gamma^\mu \wedge \gamma^\nu) \cdot F \,. \tag{1.12b}$$

Note that the two inner products in the middle term can be performed in either order, so a parenthesis is not needed.

The entire spacetime algebra is obtained by taking linear combinations of basis k-vectors in (1.11) obtained by outer multiplication of vectors in $\mathcal{M}^4$. A generic element M of the STA, called a *multivector*, can thus be written in the *expanded form*

$$M = \alpha + a + F + bi + \beta i \,, \tag{1.13}$$

where α and β are scalars, a and b are vectors, and F is a bivector. This is a decomposition of M into its k-vector parts, with $k = 0, 1, 2, 3, 4$, as is expressed more explicitly by putting it in the form

$$M = \sum_{k=0}^{4} M_{(k)} \tag{1.13$'$}$$

where the subscript (k) means "*k-vector part*." Of course, $M_{(0)} = \alpha$, $M_{(1)} = a$, $M_{(2)} = F$, $M_{(3)} = bi$, $M_{(4)} = \beta i$

Computations are facilitated by the operation of *reversion*. For M in the expanded form (1.13) the *reverse* $\widetilde{M}$ can be defined by

$$\widetilde{M} = \alpha + a - F - bi + \beta i \,. \tag{1.14}$$

Note, in particular, the effect of reversion on the various k-vector parts.

$$\widetilde{\alpha} = \alpha, \quad \widetilde{a} = a, \quad \widetilde{F} = -F, \quad \widetilde{i} = i\,.$$

It is not difficult to prove that

$$(MN)^{\sim} = \widetilde{N}\widetilde{M}\,, \tag{1.15}$$

for arbitrary multivectors M and N.

Any multivector M can be decomposed into the sum of an *even* part M_+ and an *odd* part M_- defined in terms of the expanded form (1.13) by

$$M_+ = \alpha + F + \beta i\,, \tag{1.16a}$$

$$M_- = a + bi\,, \tag{1.16b}$$

or, equivalently, by

$$M_\pm = \tfrac{1}{2}(M \mp iMi)\,. \tag{1.16c}$$

The set $\{M_+\}$ of all even multivectors forms an important subalgebra of STA called the *even subalgebra*.

We are now in position to describe the relation of STA to the Dirac algebra. The *Dirac matrices* are representations of the vectors γ_μ by 4×4 matrices, and we emphasize this correspondence by denoting the vectors with the same symbols γ_μ ordinarily used to represent the Dirac matrices. In view of what we know about STA, this correspondence reveals the physical significance of the Dirac matrices, appearing so mysteriously in relativistic quantum mechanics: *The Dirac matrices are no more and no less than matrix representations of an orthonormal frame of spacetime vectors* and thereby they characterize spacetime geometry. But how can this be? Dirac never said any such thing! And physicists today regard the set $\{\gamma_\mu\}$ as a single vector with matrices for components. Nevertheless, their practice shows that the "frame interpretation" is the correct one, though we shall see later that the "component interpretation" is actually equivalent to it in certain circumstances. The correct interpretation was actually inherent in Dirac's argument to derive the matrices in the first place: First he put the γ_μ in one-to-one correspondence with orthogonal directions in spacetime by indexing them. Second, he related the γ_μ to the metric tensor by imposing the "peculiar condition" (1.7) on the matrices for formal algebraic reasons. But we have seen that this condition has a clear geometric meaning in STA, a meaning which demands the "frame interpretation" though it is compatible with the "component interpretation." Finally, Dirac introduced associativity automatically by employing matrix algebra, without realizing that it has a geometric meaning. The geometric meaning of associativity in geometric algebra is discussed at length in Chap. 1 of Ref. 16.

If indeed the physical significance of the Dirac matrices derives entirely from their interpretation as a frame of vectors, then their specific matrix properties must be irrelevant to physics. We shall prove this by dispensing with matrices altogether and formulating the Dirac theory entirely in terms of STA. A step in this direction has already been taken in conventional approaches by proving that physical predictions are invariant under a change of matrix representation. Though the particular representation is irrelevant, it is required that γ_0 be hermitian while the γ_k are antihermitian, as expressed by

$${\gamma_0}^\dagger = \gamma_0\,, \qquad {\gamma_k}^\dagger = -\gamma_k\,. \tag{1.17}$$

Though hermitian conjugation is a matrix operation, it has a representation in STA that gives it a physical meaning. For any multivector M, we define the hermitian conjugate by

$$M^\dagger = \gamma_0 \widetilde{M} \gamma_0 \,. \tag{1.18}$$

Clearly, this is not an invariant operation because it depends on the selection of a particular timelike direction. Its physical significance will be made clear in the next Section.

Finally, we note that the *Dirac algebra* is the matrix algebra generated by the Dirac matrices over the base field of the complex numbers. Let $i' = \sqrt{-1}$ denote the unit imaginary in the base number field. In contrast to the γ_μ, i' does not represent any property of spacetime. Furthermore, STA already contains a number of "geometrical roots" of minus one, including

$$\gamma_k^2 = -1, \qquad (\gamma_2\gamma_1)^2 = -1, \qquad i^2 = (\gamma_0\gamma_1\gamma_2\gamma_3) = -1 \,.$$

We shall see that these roots can take over the role of i' in the Dirac theory, thereby revealing its geometrical meaning. Note that the complex base field gives the Dirac algebra twice as many degrees of freedom as STA. Evidently these 16 additional degrees of freedom are devoid of geometrical or physical significance, serving only to obscure the geometric content of the algebra. For these reasons we stick with STA and eschew the Dirac algebra until we need it for translating the matrix version of the Dirac theory into STA.

With STA we can describe physical processes by equations which are invariant in the sense that they are not referred to any inertial system. However, observations and measurements are usually expressed in terms of variables tied to a particular inertial system, so we need to know how to reformulate invariant equations in terms of those variables. STA provides a very simple way to do that called a *space-time split*.

In STA a given inertial system is completely characterized by a single future-pointing, timelike unit vector. Refer to the inertial system characterized by the vector γ_0 as the γ_0-*system*. The vector γ_0 is tangent to the world line of an observer at rest in the γ_0-system, so it is convenient to use γ_0 as *a name for the observer*. The observer γ_0 is represented algebraically in STA in the same way as any other physical system, and the spacetime split amounts to no more than comparing the motion of a given system (the observer) to other physical systems.

An *inertial observer* γ_0 determines a unique mapping of spacetime into the even subalgebra of STA. For each spacetime point (or *event*) x the mapping is specified by

$$x\gamma_0 = t + \mathbf{x} \,, \tag{1.19a}$$

where

$$t = x \cdot \gamma_0 \tag{1.19b}$$

and

$$\mathbf{x} = x \wedge \gamma_0 \,. \tag{1.19c}$$

This defines the γ_0-*split* of spacetime. In "relativistic units" where the speed of light $c = 1$, t is the *time parameter* for the γ_0-system. Equation (1.19b) assigns a unique time t to every event x; indeed, (1.19b) is the equation for a one parameter family of spacelike hyperplanes with normal γ_0.

The set of all position vectors (1.19c) is the 3-dimensional *position space of the observer* γ_0, which we designate by $\mathcal{P}^3 = \mathcal{P}^3(\gamma_0) = \{\mathbf{x} = x \wedge \gamma_0\}$. Note that $\mathcal{P}^3$ consists of all bivectors in STA with γ_0 as a common factor. In agreement with common parlance, we refer to the elements of $\mathcal{P}^3$ as vectors. Thus, we have two kinds of vectors, those in $\mathcal{M}^4$ and those in $\mathcal{P}^3$. To distinguish between them, we may refer to elements of $\mathcal{M}^4$ as *proper vectors* and to elements of $\mathcal{P}^3$ as relative vectors (relative to γ_0, of course!). Also, relative vectors will be designated in boldface.

By the geometric product and sum the vectors in $\mathcal{P}^3$ generate the entire even subalgebra of STA as the geometric algebra of $\mathcal{P}^3$. This is made obvious by constructing a basis. Corresponding to a standard basis $\{\gamma_\mu\}$ for $\mathcal{M}^4$, we have a standard basis $\{\boldsymbol{\sigma}_k; k = 1, 2, 3\}$ for $\mathcal{P}^3$, where

$$\boldsymbol{\sigma}_k = \gamma_k \wedge \gamma_0 = \gamma_k\gamma_0 \,. \tag{1.20a}$$

These generate a basis for the relative bivectors:

$$\boldsymbol{\sigma}_i \wedge \boldsymbol{\sigma}_j = \boldsymbol{\sigma}_i\boldsymbol{\sigma}_j = i\boldsymbol{\sigma}_k = \gamma_j\gamma_i \,, \tag{1.20b}$$

where the allowed values of the indices $\{i, j, k\}$ are cyclic permutations of 1,2,3. The right sides of (1.20) and (2.5) show how the bivectors for spacetime are split into vectors and bivectors for $\mathcal{P}^3$. Remarkably, the right-handed pseudoscalar for $\mathcal{P}^3$ is identical to that for $\mathcal{M}^4$; thus,

$$\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2\boldsymbol{\sigma}_3 = i = \gamma_0\gamma_1\gamma_2\gamma_3 \,. \tag{1.20c}$$

It will be noted that the geometric algebra of $\mathcal{P}^3$ is isomorphic to the Pauli algebra, with the $\boldsymbol{\sigma}_k$ corresponding to the 2×2 *Pauli matrices*. However, if the γ_μ are regarded as Dirac matrices, then (1.20) defines 4×4 matrix representations of the $\boldsymbol{\sigma}_k$; these are the α_k matrices of Dirac.[7] This awkward distinction between 2×2 and 4×4 matrix representations is another unnecessary complication in the conventional formulation of quantum mechanics. It is entirely irrelevant to the simple relation between relative and proper basis vectors defined by (1.20a).

To complete the correspondence with matrix representations, consider hermitian conjugation again. According to the definition (1.18), the $\boldsymbol{\sigma}_k$ are hermitian, that is,

$$\boldsymbol{\sigma}_k^\dagger = \boldsymbol{\sigma}_k \qquad \text{and} \qquad (\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j)^\dagger = \boldsymbol{\sigma}_j\boldsymbol{\sigma}_i \,. \tag{1.21}$$

But this defines the geometrical operation of *reversion* in the algebra of $\mathcal{P}^3$. Thus, hermitian conjugation has the geometrical meaning of reversion "in" an inertial system.

Let $x = x(\tau)$ be the history of a particle with *proper time* τ and *proper velocity* $v = dx/dt$. The space-time split of v is obtained by differentiating (1.19a); whence

$$v\gamma_0 = v_0(1 + \mathbf{v}) \,, \tag{1.22a}$$

where

$$v_0 = v \cdot \gamma_0 = \frac{dt}{d\tau} = \left(1 - \mathbf{v}^2\right)^{-\frac{1}{2}} \tag{1.22b}$$

is the "time dilation" factor, and

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} = \frac{d\tau}{dt}\frac{d\mathbf{x}}{d\tau} = \frac{v \wedge \gamma_0}{v \cdot \gamma_0} \tag{1.22c}$$

is the *relative velocity* in the γ_0-system. The last equality in (1.22b) was obtained from

$$1 = v^2 = (v\gamma_0)(\gamma_0 v) = v_0(1+\mathbf{v})v_0(1-\mathbf{v}) = v_0^2(1-\mathbf{v}^2)\,.$$

An electromagnetic field is a bivector-valued function $F = F(x)$ on spacetime. An observer γ_0 splits it into an electric (relative vector) part $\mathbf{E}$ and, a magnetic (relative bivector) part $i\mathbf{B}$; thus

$$F = \mathbf{E} + i\mathbf{B}\,, \tag{1.23a}$$

where

$$\mathbf{E} = (F\cdot\gamma_0)\gamma_0 = \tfrac{1}{2}(F + F^\dagger)\,, \tag{1.23b}$$

$$i\mathbf{B} = (F\wedge\gamma_0)\gamma_0 = \tfrac{1}{2}(F - F^\dagger)\,, \tag{1.23c}$$

and, in accordance with (1.18), $F^\dagger = \mathbf{E} - i\mathbf{B}$. Equation (1.23a) represents the field formally as a complex (relative) vector; but it must be remembered that the imaginary i here is the unit pseudoscalar and so has a definite geometric meaning. Indeed, (1.23a) shows that the magnetic field is actually a bivector quantity $i\mathbf{B}$, and its conventional representation as a vector $\mathbf{B}$ is a historical accident in which the duality is hidden in the notion of "axial vector."[18,16]

At this point it is worth noting that the geometric product of relative vectors $\mathbf{E}$ and $\mathbf{B}$ can be decomposed into symmetric and antisymmetric parts in the same way that we decomposed the product of proper vectors. Thus, we obtain

$$\mathbf{EB} = \mathbf{E}\cdot\mathbf{B} = i(\mathbf{E}\times\mathbf{B})\,, \tag{1.24a}$$

where

$$\mathbf{E}\cdot\mathbf{B} = \tfrac{1}{2}(\mathbf{EB} + \mathbf{BE}) \tag{1.24b}$$

is the usual dot product for Euclidean 3-space, and

$$\mathbf{E}\times\mathbf{B} = \frac{1}{2i}\,(\mathbf{EB} - \mathbf{BE}) = i^{-1}(\mathbf{E}\wedge\mathbf{B}) \tag{1.24c}$$

is usual *cross product* of Gibbs. Thus, the standard vector algebra of Gibbs is smoothly imbedded in STA and simply related to invariant spacetime relations by a spacetime split. Consequently, translations from STA to vector algebra are effortless. Moreover, the combination (1.24) of the dot and cross products into the single geometric product simplifies many aspects of classical nonrelativistic physics, as demonstrated at length in Ref. 16.

In terms of spacetime coordinates defined by (1.10), an operator ∇ interpreted as the derivative with respect to a spacetime point x can be defined by

$$\nabla = \gamma^\mu\partial_\mu \tag{1.25}$$

where

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \gamma_\mu\cdot\nabla\,. \tag{1.26}$$

The square of is the d'Alembertian

$$\nabla^2 = \mathrm{g}^{\mu\nu}\partial_\mu\partial_\nu\,. \tag{1.27}$$

The matrix representation of (1.25) will be recognized as the "Dirac Operator," originally discovered by Dirac by seeking a "square root" of ∇^2 in order to find a first order relativistically invariant wave equation, the famous "Dirac equation" discussed in the next Section. However, in STA where the γ_μ are vectors, it is clear that ∇ is a vector operator, because it is the derivative with respect to a vector representing a spacetime point. Contrary to the impression given by conventional accounts of relativistic quantum theory, ∇ is not an operator specially adapted to spin-1/2 wave equations. Here we show that it is equally apt for electromagnetic field equations.

In STA an electromagnetic field is represented by a bivector-valued function $F = F(x)$ on spacetime. The field produced by a source with spacetime current density $J = J(x)$ is determined by *Maxwell's Equation*

$$\nabla F = J\,. \tag{1.28}$$

Using (1.11) again to write

$$\nabla F = \nabla \cdot F + \nabla \wedge F\,, \tag{1.29}$$

(1.28) can be separated into a vector part

$$\nabla \cdot F = J \tag{1.30a}$$

and a trivector part

$$\nabla \wedge F = 0\,. \tag{1.30b}$$

Expressed in terms of a basis by using (1.25) and (1.12a), these two equations are seen to as equivalent to the usual tensor form of Maxwell's equations.

Sometimes the source current J can be decomposed into a *conduction current* J^C and a magnetization current $\nabla \cdot M$, where the generalized *magnetization* $M = M(x)$ is a bivector field; thus

$$J = J^C + \nabla \cdot M\,. \tag{1.31}$$

The Gordon decomposition of the Dirac current is of this ilk. Because of the mathematical identity $\nabla \cdot (\nabla \cdot M) = (\nabla \wedge \nabla) \cdot M = 0$, the conservation law $\nabla \cdot J = 0$ implies also that $\nabla \cdot J^C = 0$. Using (1.31), equation (1.30a) can be put in the form

$$\nabla \cdot G = J^C \tag{1.32}$$

where we have defined a new field

$$G = F - M\,. \tag{1.33}$$

A disadvantage of this approach is that it mixes up physically different kinds of entities, an E-M field F and a matter field M. However, in most materials M is a function of the field F, so when a "constitutive equation" $M = M(F)$ is known (1.32) becomes a well defined equation for F.

STA enables us to write the usual Maxwell energy-momentum tensor $T(n) = T(n(x), x)$ for the electromagnetic field in the compact form

$$T(n) = \tfrac{1}{2} F n \widetilde{F} = -\tfrac{1}{2} F n F\,. \tag{1.34}$$

Recall that the tensor field $T(n)$ is a vector-valued linear function on the tangent space at each spacetime point x describing the flow of energy-momentum through a surface with normal $n = n(x)$, By linearity $T(n) = n_\mu T^\mu$, where $n_\mu = n \cdot \gamma_\mu$ and

$$T^\mu \equiv T(\gamma^\mu) = \tfrac{1}{2} F \gamma^\mu \widetilde{F} \,. \tag{1.35}$$

The divergence of $T(n)$ can be evaluated by using Maxwell's equation (1.28), with the result

$$\partial_\mu T^\mu = T(\nabla) = J \cdot F \,. \tag{1.36}$$

Its value is the negative of the *Lorentz Force* $F \cdot J$, which is the rate of energy-momentum transfer from the source J to the field F.

2. THE REAL DIRAC EQUATION

To find a representation of the Dirac theory in terms of the STA, we begin with a Dirac spinor Ψ, a column matrix of 4 complex numbers. Let u be a fixed spinor with the properties

$$u^\dagger u = 1 \,, \tag{2.1a}$$

$$\gamma_0 u = u \,, \tag{2.1b}$$

$$\gamma_2 \gamma_1 u = i' u \,. \tag{2.1c}$$

In writing this we regard the γ_μ, for the time being, as 4×4 Dirac matrices, and i' as the unit imaginary in the complex number field of the Dirac algebra. Now, we can write any Dirac spinor

$$\Psi = \psi u \,, \tag{2.2}$$

where Ψ is a matrix which can be expressed as a polynomial in the γ_μ. The coefficients in this polynomial can be taken as real, for if there is a term with an imaginary coefficient, then (2.1c) enables us to make it real without altering (2.2) by replacing i' in the term by $\gamma_2 \gamma_1$ on the right of the term. Furthermore, the polynomial can be taken to be an even multivector, for if any term is odd, then (2.1b) allows us to make it even by multiplying on the right by γ_0. Thus, in (2.2) we may assume that ψ *is a real even multivector*. Now we may reinterpret the γ_μ in ψ as vectors in the STA instead of matrices. Thus, we have established a correspondence between Dirac spinors and even multivectors in the STA. The correspondence must be one-to-one, because the space of even multivectors (like the space of Dirac spinors) is exactly 8-dimensional, with 1 scalar, 1 pseudoscalar and 6 bivector dimensions.

There are other ways to represent a Dirac spinor in the STA,[12] but all representations are, of course, mathematically equivalent. The representation chosen here has the advantages of simplicity and, as we shall see, ease of interpretation.

To distinguish a spinor ψ in the STA from its matrix representation Ψ in the Dirac algebra, let us call it a *real spinor* to emphasize the elimination of the ungeometrical imaginary i'. Alternatively, we might refer to ψ as the *operator representation* of a Dirac spinor, because, as shown below, it plays the role of an operator generating observables in the theory.

In terms of the real wave function ψ, the Dirac equation for an electron can be written in the form

$$\gamma^\mu(\partial_\mu\psi\gamma_2\gamma_1\hbar - eA_\mu\psi) = m\psi\gamma_0\,, \tag{2.3}$$

where m is the mass and $e = -|\,e\,|$ is the charge of the electron, while the $A_\mu = A\cdot\gamma_\mu$ are components of the electromagnetic vector potential. To prove that this is equivalent to the standard matrix form of the Dirac equation,[21] we simply interpret the γ_μ as matrices, multiply by u on the right and use (2.1a, b, c) and (2.2) to get the standard form

$$\gamma^\mu(i'\hbar\partial_\mu - eA_\mu)\Psi = m\Psi\,. \tag{2.4}$$

This completes the proof. Alternative proofs are given elsewhere.[4,7] The original converse derivation of (2.3) from (2.4) was much more difficult.[2]

Henceforth, we can work with the *real Dirac equation* (2.3) without reference to its matrix representation (2.4). We know that computations in STA can be carried out without introducing a basis, so let us use (1.25) to write the real Dirac equation in the coordinate-free form

$$\nabla\psi\mathbf{i}\hbar - eA\psi = m\psi\gamma_0\,, \tag{2.5}$$

where $A = A_\mu\gamma^\mu$ is the electromagnetic vector potential, and the notation

$$\mathbf{i} \equiv \gamma_2\gamma_1 = i\gamma_3\gamma_0 = i\boldsymbol{\sigma}_3 \tag{2.6}$$

emphasizes that this bivector plays the role of the imaginary i' that appears explicitly in the matrix form (2.4) of the Dirac equation. To interpret the theory, it is crucial to note that the bivector $\mathbf{i}$ has a definite geometrical interpretation while i' does not.

Equation (2.5) is Lorentz invariant, despite the explicit appearance of the constants γ_0 and $\mathbf{i} = \gamma_2\gamma_1$ in it. These constants need not be associated with vectors in a particular reference frame, though it is often convenient to do so. It is only required that γ_0 be a fixed, future-pointing, timelike unit vector while $\mathbf{i}$ is a spacelike unit bivector which commutes with γ_0. The constants can be changed by a Lorentz rotation

$$\gamma_\mu \to \gamma'_\mu = U\gamma_\mu\widetilde{U}\,, \tag{2.7}$$

where U is a constant rotor, so $U\widetilde{U} = 1$,

$$\gamma'_0 = U\gamma_0\widetilde{U} \qquad \text{and} \qquad \mathbf{i}' = U\mathbf{i}\widetilde{U}\,. \tag{2.8}$$

A corresponding change in the wave function,

$$\psi \to \psi' = \psi\widetilde{U}\,, \tag{2.9}$$

induces a mapping of the Dirac equation (2.5) into an equation of the same form:

$$\nabla\psi\mathbf{i}'\hbar - eA\psi' = m\psi'\gamma'_0\,. \tag{2.10}$$

This transformation is no more than a change of constants in the Dirac equation. It need not be coupled to a change in reference frame. Indeed, in the matrix formulation it can be

interpreted as a mere change in matrix representation, that is, in the particular matrices selected to be associated with the vectors γ_μ, for (2.2) gives

$$\Psi = \psi u = \psi' u' \,, \tag{2.11}$$

where $u' = Uu$.

For the special case

$$U = e^{i\varphi_0} \,, \tag{2.12}$$

where φ_0 is a scalar constant, (2.8) gives $\gamma_0' = \gamma_0$ and $\mathbf{i}' = \mathbf{i}$, so ψ and

$$\psi' = \psi e^{i\varphi_0} \tag{2.13}$$

are solutions of the same equation. In other words, the Dirac equation does not distinguish solutions differing by a constant phase factor.

Note that $\boldsymbol{\sigma}_2 = \gamma_2\gamma_0$ anticommutes with both γ_0 and $\mathbf{i} = i\boldsymbol{\sigma}_3$, so multiplication of the Dirac equation (2.5) on the right by $\boldsymbol{\sigma}_2$ yields

$$\nabla\psi^C \mathbf{i}\hbar - eA\psi^C = m\psi^C\gamma_0 \,, \tag{2.14}$$

where

$$\psi^C = \psi\boldsymbol{\sigma}_2 \,. \tag{2.15}$$

The net effect is to change the sign of the charge in the Dirac equation, therefore, the transformation $\psi \to \psi^C$ can be interpreted as *charge conjugation*. Of course, the definition of charge conjugate is arbitrary up to a constant phase factor such as in (2.13). The main thing to notice here is that in (2.15) charge conjugation, like parity conjugation, is formulated as a completely geometrical transformation, without any reference to a complex conjugation operation of obscure physical meaning. Its geometrical meaning is determined by what it does to the "frame of observables" identified below.

For any even multivector ψ, $\psi\widetilde{\psi}$ is also even but, according to (1.14), its bivector part must vanish because $(\psi\widetilde{\psi})\tilde{} = \psi\widetilde{\psi}$. Therefore, we can write

$$\psi\widetilde{\psi} = \rho e^{i\beta} \,, \tag{2.16a}$$

where ρ and β are scalars. If $\rho \neq 0$ we can derive from ψ an even multivector $R = \psi(\psi\widetilde{\psi})^{-\frac{1}{2}}$ satisfying

$$R\widetilde{R} = \widetilde{R}R = 1 \,. \tag{2.16b}$$

Hence ψ has the Lorentz *invariant decomposition*

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R \,. \tag{2.17}$$

This decomposition applies to the real Dirac wave function $\psi = \psi(x)$, because it is an even multivector. At each spacetime point x, the rotor $R = R(x)$ determines a Lorentz rotation of a given fixed frame of vectors $\{\gamma_\mu\}$ into a frame $\{e_\mu = e_\mu(x)\}$ given by

$$e_\mu = R\gamma_\mu\widetilde{R} \,. \tag{2.18}$$

In other words, R determines a unique frame field on spacetime.

We shall see that that the physical interpretation given to the frame field $\{e_\mu\}$ is a key to the interpretation of the entire Dirac theory. Specifically, it will be seen that the e_μ can be interpreted directly as descriptors of the kinematics of electron motion. It follows from (2.18), therefore, that the rotor field $R = R(x)$ is a descriptor of electron kinematics. The factor $(\rho e^{i\beta})^{\frac{1}{2}}$ will be given a statistical interpretation. Thus, the canonical form (2.17) is an invariant decomposition of the Dirac wave function into a 2-parameter *statistical factor* $(\rho e^{i\beta})^{\frac{1}{2}}$ and a 6-parameter *kinematical factor* R.

From (2.17) and (2.18) we find that

$$\psi\gamma_\mu\widetilde{\psi} = \psi'\gamma_\mu\widetilde{\psi}' = \rho e_\mu \,. \tag{2.19}$$

Note that that we have here a set of four linearly independent vector fields which are invariant under the transformation specified by (2.7) and (2.8). Thus these fields do not depend on any coordinate system, despite the appearance of γ_μ on the left side of (2.19). Note also that the factor $e^{i\beta/2}$ in (2.17) does not contribute to (2.19), because the pseudoscalar i anticommutes with the γ_μ.

Two of the vector fields in (2.19) are given physical interpretations in the standard Dirac theory. First, the vector field

$$\psi\gamma_0\widetilde{\psi} = \rho e_0 = \rho v \tag{2.20}$$

is the *Dirac current*, which, in accord with the standard *Born interpretation*, we interpret as a *probability current*. Thus, at each spacetime point x the timelike vector $v = v(x) = e_0(x)$ is 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. The correspondence of (2.20) to the conventional definition of the Dirac current is displayed in Table I.

The second vector field

$$\tfrac{1}{2}\hbar\psi\gamma_3\widetilde{\psi} = \rho\tfrac{1}{2}\hbar e_3 = \rho s \tag{2.21}$$

will be interpreted as the *spin vector* density. Justification for this interpretation comes from angular momentum conservation treated in the next Section. Note in Table I that this vector quantity is represented as a pseudovector (or axial vector) quantity in the conventional matrix formulation. The spin pseudovector is correctly identified as is, as shown below.

Angular momentum is actually a bivector quantity. The *spin angular momentum* $S = S(x)$ is a bivector field related to the spin vector field $s = s(x)$ by

$$S = isv = \tfrac{1}{2}\hbar ie_3e_0 = \tfrac{1}{2}\hbar R\gamma_2\gamma_1\widetilde{R} = \tfrac{1}{2}R(\mathbf{i}\hbar)\widetilde{R}. \tag{2.22}$$

The right side of this chain of equivalent representations shows the relation of the spin to the unit imaginary $\mathbf{i}$ appearing in the Dirac equation (2.5). Indeed, it shows that *the bivector* $\frac{1}{2}\mathbf{i}\hbar$ is a *reference representation* of the spin which is rotated by the kinematical factor R into the local spin direction at each spacetime point. This establishes an explicit connection between spin and imaginary numbers which is inherent in the Dirac theory but hidden in the conventional formulation, a connection, moreover, which remains even in the Schroedinger approximation, as seen in a later Section.

TABLE I: BILINEAR COVARIANTS

Scalar	$\widetilde{\Psi}\Psi = \Psi^\dagger\gamma_0\Psi = (\psi\widetilde{\psi})_{(0)} = \rho\cos\beta$
Vector	$\widetilde{\Psi}\gamma_\mu\Psi = \Psi^\dagger\gamma_0\gamma_\mu\Psi = (\psi\gamma_0\widetilde{\psi}\gamma_\mu)_{(0)} = (\psi^\dagger\gamma_0\gamma_\mu\psi)_{(0)}$ $= (\psi\gamma_0\widetilde{\psi})\cdot\gamma_\mu = (\rho v)\cdot\gamma_\mu = \rho v_\mu$
Bivector	$\frac{e}{m}\frac{i'\hbar}{2}\widetilde{\Psi}\frac{1}{2}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)\Psi = \frac{e\hbar}{2m}\left(\gamma_\mu\gamma_\nu\psi\gamma_2\gamma_1\widetilde{\psi}\right)_{(0)}$ $= (\gamma_\mu\wedge\gamma_\nu)\cdot(M) = M_{\mu\nu} = \frac{e}{m}\rho(ie^{i\beta}sv)\cdot(\gamma_\mu\wedge\gamma_\nu)$
Pseudovector*	$\frac{1}{2}i'\hbar\widetilde{\Psi}\gamma_\mu\gamma_5\Psi = \frac{1}{2}\hbar(\gamma_\mu\psi\gamma_3\widetilde{\psi})_{(0)} = \gamma_\mu\cdot(\rho s) = \rho s_\mu$
Pseudoscalar*	$\widetilde{\Psi}\gamma_5\Psi = (i\psi\widetilde{\psi})_{(0)} = -\rho\sin\beta$

*Here we use the more conventional symbol $\gamma_5{=}\gamma_0\gamma_1\gamma_2\gamma_3$ for the matrix representation of the unit pseudoscalar i.

The hidden relation of spin to the imaginary i' in the Dirac theory can be made manifest in another way. Multiplying (2.21) on the right by ψ and using (2.17) we obtain

$$S\psi = \tfrac{1}{2}\psi\mathbf{i}\hbar\,. \tag{2.23}$$

Then using (2.1c) and (2.2) to translate this into the matrix formalism, we obtain

$$S\Psi = \tfrac{1}{2}\mathbf{i}\hbar\Psi\,. \tag{2.24}$$

Thus, $\frac{1}{2}i'\hbar$ is the eigenvalue of the invariant "spin operator"

$$S = \tfrac{1}{2}S^{\alpha\beta}\gamma_\alpha\gamma_\beta\,. \tag{2.25}$$

Otherwise said, the factor $i'\hbar$ in the Dirac theory is a *representation* of the spin bivector by its eigenvalue. The eigenvalue is imaginary because the "spin tensor" $S^{\alpha\beta}$ is skewsymmetric. The fact that $S = S(x)$ specifies a definite spacelike tangent plane at each point x is completely suppressed in the $i'\hbar$ representation. It should be noted also that (2.24) is completely general, applying to any Dirac wave function whatsoever.

The identification of $S^{\alpha\beta}$ in (2.25) as spin tensor is not made in standard accounts of the Dirac theory, though of course it must be implicated. Standard accounts (e.g. p. 59 of Ref. 22) either explicitly or implicitly introduce the spin (density) tensor

$$\rho S^{\nu\alpha\beta} = \frac{i'\hbar}{2}\widetilde{\Psi}\gamma^\nu\wedge\gamma^\alpha\wedge\gamma^\beta\Psi = \frac{i'\hbar}{2}\widetilde{\Psi}\gamma_5\gamma_\mu\Psi\epsilon^{\mu\nu\alpha\beta} = \rho s_\mu\epsilon^{\mu\nu\alpha\beta}\,, \tag{2.26}$$

where use has been made of the identity

$$\gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta = \gamma_5 \gamma_\mu \epsilon^{\mu\nu\alpha\beta} \tag{2.27a}$$

and the expression for s_μ in Table I. Note that the "alternating tensor" $\epsilon^{\mu\nu\alpha\beta}$ can be *defined* simply as the product of two pseudoscalars, thus

$$\begin{aligned} \epsilon^{\mu\nu\alpha\beta} &= i(\gamma^\mu \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta) = (i\gamma^\mu\gamma^\nu\gamma^\alpha\gamma^\beta)_{(0)} \\ &= (\gamma_3 \wedge \gamma_2 \wedge \gamma_1 \wedge \gamma_0) \cdot (\gamma^\mu \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta)\,. \end{aligned} \tag{2.27b}$$

Alternatively,

$$\gamma^\mu \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta = -i\epsilon^{\mu\nu\alpha\beta}\,. \tag{2.27c}$$

From (2.26) and (2.27b) we find

$$S^{\nu\alpha\beta} = s_\mu \epsilon^{\mu\nu\alpha\beta} = i(s \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta) = (is) \cdot (\gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta)\,. \tag{2.28}$$

The last expression shows that the $S^{\nu\alpha\beta}$ are simply tensor components of the pseudovector is. Contraction of (2.28) with $v_\nu = v \cdot \gamma_\nu$ and use of duality (1.16b) gives the desired relation between $S^{\nu\alpha\beta}$ and $S^{\alpha\beta}$:

$$v_\nu S^{\nu\alpha\beta} = i(s \wedge v \wedge \gamma^\alpha \wedge \gamma^\beta) = [\,i(s \wedge v)\,] \cdot (\gamma^\alpha \wedge \gamma^\beta) = S^{\alpha\beta}\,. \tag{2.29}$$

Its significance will be made clear in the discussion of angular momentum conservation.

Note that the spin bivector and its relation to the unit imaginary is invisible in the standard version of the bilinear covariants in Table I. The spin S is buried there in the magnetization (tensor or bivector). The magnetization M can be defined and related to the spin by

$$M = \frac{e\hbar}{2m}\,\psi\gamma_2\gamma_1\widetilde{\psi} = \frac{e\hbar}{2m}\,\rho e^{i\beta} e_2 e_1 = \frac{e}{2m}\,\rho S e^{i\beta}\,. \tag{2.30}$$

The interpretation of M as magnetization comes from the Gordon decomposition considered in the next Section. Equation (2.30) reveals that in the Dirac theory the magnetic moment is not simply proportional to the spin as often asserted; the two are related by a duality rotation represented by the factor $e^{i\beta}$. It may be appreciated that this relation of M to S is much simpler than any relation of $M^{\alpha\beta}$ to $S^{\nu\alpha\beta}$ in the literature, another indication that S is the most appropriate representation for spin. By the way, note that (2.30) provides some justification for referring to β henceforth as the *duality parameter*. The name is noncommittal to the physical interpretation of β, a debatable issue discussed later.

We are now better able to assess the content of Table I. There are $1+4+6+4+1=16$ distinct bilinear covariants but only 8 parameters in the wave function, so the various covariants are not mutually independent. Their interdependence has been expressed in the literature by a system of algebraic relations known as "Fierz Identities" (e.g., see Ref. 23). However, the invariant decomposition of the wave function (2.17) reduces the relations to their simplest common terms. Table I shows exactly how the covariants are related by expressing them in terms of ρ, β, v_μ, s_μ, which constitutes a set of 7 independent parameters, since the velocity and spin vectors are constrained by the three conditions

that they are orthogonal and have constant magnitudes. This parametrization reduces the derivation of any Fierz identity practically to inspection. Note, for example, that

$$\rho^2 = (\widetilde{\Psi}\Psi)^2 + (\widetilde{\Psi}\gamma_5\Psi)^2 = (\widetilde{\Psi}\gamma_\mu\Psi)(\widetilde{\Psi}\gamma^\mu\Psi) = -(\widetilde{\Psi}\gamma_\mu\gamma_5\Psi)(\widetilde{\Psi}\gamma^\mu\gamma_5\Psi)\,.$$

Evidently Table I tells us all we need to know about the bilinear covariants and makes further reference to Fierz identities superfluous.

Note that the factor $i'\hbar$ occurs explicitly in Table I only in those expressions involving electron spin. The conventional justification for including the i' is to make antihermitian operators hermitian so the bilinear covariants are real. We have seen however that this smuggles spin into the expressions. That can be made explicit by using (2.24) to derive the general identity

$$i'\hbar\widetilde{\Psi}\Gamma\Psi = \widetilde{\Psi}\Gamma\gamma_\alpha\gamma_\beta\Psi S^{\alpha\beta}\,, \tag{2.31}$$

where Γ is any matrix operator.

Perhaps the most significant thing to note about Table I is that only 7 of the 8 parameters in the wave function are involved. The missing parameter is the *phase* of the wave function. To understand the significance of this, note also that, in contrast to the vectors e_0 and e_3 representing velocity and spin directions, the vectors e_1 and e_2 do not appear in Table I except indirectly in the product e_2e_1. The missing parameter is one of the six parameters implicit in the rotor R determining the Lorentz rotation (2.18). We have already noted that 5 of these parameters are needed to determine the velocity and spin directions e_0 and e_3. By duality, these vectors also determine the direction $e_2e_1 = ie_3e_0$ of the "spin plane" containing e_1 and e_2. The remaining parameter therefore determines the directions of e_1 and e_2 in this plane. It is literally an angle of rotation in this plane and the spin bivector $\hat{S} = e_2e_1 = R\mathbf{i}\widetilde{R}$ is the generator of the rotation. Thus, we arrive at a *geometrical interpretation of the phase of the wave function* which is *inherent in the Dirac Theory.* But all of this is invisible in the conventional matrix formulation.

The purpose of Table I is to explicate the correspondence of the matrix formulation to the real (STA) formulation of the Dirac theory. Once it is understood that the two formulations are completely isomorphic, the matrix formulation can be dispensed with and Table I becomes superfluous. By revealing the geometrical meaning of the unit imaginary and the wave function phase along with this connection to spin, STA challenges us to ascertain the physical significance of these geometrical facts, a challenge that will be met in subsequent Sections.

3. OBSERVABLES AND CONSERVATION LAWS.

One of the miracles of the Dirac theory was the spontaneous emergence of spin in the theory when nothing about spin seemed to be included in the assumptions. This miracle has been attributed to Dirac's derivation of his linearized relativistic wave equation, so spin has been said to be "a relativistic phenomenon." However, we have seen that the Dirac operator (1.25) is equally suited to the formulation of Maxwell's equation (1.28), and we have concluded that the Dirac algebra arises from spacetime geometry rather than anything special about quantum theory. The origin of spin must be elsewhere.

Our objective here is to ascertain precisely what features of the Dirac theory are responsible for its extraordinary empirical success and to establish a coherent physical interpretation

which accounts for all its salient aspects. The geometric insights of STA provide us with a perspective from which to criticize some conventional beliefs about quantum mechanics and so leads us to some unconventional conclusions.

The first point to be understood is that there is more to the Dirac theory than the Dirac equation. Indeed, the Dirac wave function has no physical meaning at all apart from assumptions that relate it to *physical observables*. Now, there is a strong tradition in quantum mechanics to associate *H*ermitian *O*perators with *O*bservables and their eigenvalues with observed values. Let's call this the *HOO Principle*. There is no denying that impressive results have been achieved in quantum mechanics using the HOO Principle. However, we shall see that certain features of the Dirac theory conflict with the view that the HOO Principle is a universal principle of quantum mechanics. It is contended here that the successes of HOO Principle derive from one set of operators only, namely, the *kinetic energy-momentum operators* $\underline{p}_\mu$ defined in the convention matrix theory by

$$\underline{p}_\mu i = i'\hbar\partial_\mu - e\,A_\mu\,. \tag{3.1}$$

Moreover, it will be seen that STA leads us to a new view on why these operators are so significant in quantum mechanics.

In the approach taken here observables are defined quite literally as quantities which can be measured experimentally either directly or indirectly. Observables of the Dirac theory are associated directly with the Dirac wave function rather than with operators, though operators may be used to express the association. A set of observables is said to be complete if it supplies a *coherent* physical interpretation for all mathematical features of the wave function. *A complete set of observables is determined by the conservation laws for electron position, charge,* energy-*momentum and angular momentum.* The task now is to specify these observables and their conservation laws unambiguously.

We assume first of all that the Dirac theory describes the *electron as a point particle*, but the description is statistical and the *position probability current* is to be identified with the Dirac current (2.20). This interpretation can be upheld only if the Dirac current is rigorously conserved. To establish that, we follow Appendix B of Ref. 4, multiplying the Dirac equation (2.5) on the right by $i\gamma_0\gamma_3\gamma_\mu\widetilde{\psi}$ and using (2.18) to get

$$(\nabla\psi)\hbar\gamma_\mu\widetilde{\psi} = -im\rho e^{i\beta}e_3e_\mu + e\rho Ae_1e_2e_\mu\,.$$

The scalar part of this equation gives us

$$\nabla\cdot(\rho e_\mu) = \frac{2}{\hbar}\,\rho e_\mu\cdot(e_3m\sin\beta + (e_2e_1)\cdot A)\,. \tag{3.2}$$

Thus we have the four equations

$$\nabla\cdot(\rho v) = \partial_\mu(\rho v^\mu) = 0\,, \tag{3.3}$$

$$\nabla\cdot(\rho s) = -m\sin\beta\,, \tag{3.4}$$

$$\nabla\cdot(\rho e_1) = \frac{2}{\hbar}\,\rho A\cdot e_2\,, \tag{3.5}$$

$$\nabla\cdot(\rho e_2) = -\frac{2}{\hbar}\,\rho A\cdot e_1\,. \tag{3.6}$$

Equation (3.3) is the desired *position probability conservation law*. The meaning of the other equations remains to be determined.

There are other conserved currents besides the Dirac current, so further argument is needed to justify its interpretation as probability current. We must establish the internal and external validity of the interpretation, that is, we must how that internally it is logically coherent with the interpretation of other observables and externally it agrees with empirical observations.

The Dirac current ρv assigns a unit timelike vector $v(x)$ to each spacetime point x where $\rho \neq 0$. In keeping with the statistical interpretation of the Dirac current, we interpret $v(x)$ as the *expected* proper *velocity* of the electron at x, that is, the velocity predicted for the electron if it happens to be at x. In the γ_0-system, the probability that the electron actually is at x is given by

$$(\rho v) \cdot (\gamma_0 d^3x)\,. \tag{3.7}$$

It is normalized so that

$$\int d^3x\, \rho_0 = 1\,, \tag{3.8}$$

where $d^3x = |\, d^3x\,|$ and the integral is over the spacelike hyperplane defined by the equation $x \cdot \gamma_0 = t$, and

$$\rho_0 = \rho v_0 = (\rho v) \cdot \gamma_0 = (\psi\gamma_0\widetilde{\psi}\gamma_0)_{(0)} = (\psi\psi^\dagger)_{(0)} \tag{3.9}$$

is the *probability density* in the γ_0-system.

The velocity $v(x)$ defines a local reference frame at x called the *electron rest frame*. The *proper probability density* $\rho = (\rho v) \cdot v$ can be interpreted as the probability density in the rest frame. By a well known theorem, the probability conservation law (3.3) implies that through each spacetime point there passes a *unique integral curve* which is tangent to v at each of its points. Let us call these curves (*electron*) *streamlines*. In any spacetime region where $\rho \neq 0$, a solution of the Dirac equation determines a *family of streamlines* which fills the region with exactly one streamline through each point. The streamline through a specific point x_0 is the expected history of an electron at x_0, that is, it is the optimal prediction for the history of an electron which actually is at x_0 (with relative probability $\rho(x_0)$, of course!). Parametrized by proper time τ, the streamline $x = x(\tau)$ is determined by the equation

$$\frac{dx}{d\tau} = v(x(\tau))\,. \tag{3.10}$$

The physical significance of these predicted electron histories is discussed in the next Section.

Although our chief concern will be with observables describing the flow of conserved quantities along streamlines, we pause to consider the main theorem relating local flow to the time development of spatially averaged observables. The result is helpful for comparison with the standard operator approach to the Dirac theory. Let f be some observable in the Dirac theory represented by a multivector-valued function $f = f(x)$. The *average value* of f at time t in the γ_0-system is defined by

$$\langle\, f \,\rangle = \int d^3x\, \rho_0 f\,. \tag{3.11}$$

To determine how this quantity changes with time, we use

$$\partial_\mu(\rho v^\mu f) = \rho v \cdot \nabla f = \rho \frac{df}{d\tau} = \rho_0 \frac{df}{dt}\,, \tag{3.12}$$

with the derivative on the right taken along an electron streamline. Assuming that ρ_0 vanishes at spatial infinity, Gauss's theorem enables us to put (3.12) in the useful integral form

$$\frac{d}{dt}\langle\, f\,\rangle = \int d^3x\, \rho v \cdot \nabla f = \left\langle \frac{df}{dt} \right\rangle . \tag{3.13}$$

This result is known as "Reynold's Theorem" in hydrodynamics.

Taking the proper position vector x as observable, we have the *average position* of the electron given by

$$\langle\, x\,\rangle = \int d^3x\, \rho_0 x\,, \tag{3.14}$$

and application of (3.13) gives the *average velocity*

$$\frac{d}{dt}\langle\, x\,\rangle = \int d^3x\, \rho v = \left\langle \frac{dx}{dt} \right\rangle . \tag{3.15}$$

To see that this is a sensible result, use the space-time splits (1.19a) and (1.22a) to get

$$\langle\, x\,\rangle\gamma_0 = 1 + \langle\, \mathbf{x}\,\rangle \tag{3.16}$$

from (3.14), and

$$\frac{d}{dt}\langle\, x\,\rangle\gamma_0 = 1 + \langle\, \mathbf{v}\,\rangle \tag{3.17}$$

from, (3.15). Thus, we have

$$\frac{d}{dt}\langle\, \mathbf{x}\,\rangle = \langle\, \mathbf{v}\,\rangle = \left\langle \frac{d\mathbf{x}}{dt} \right\rangle . \tag{3.18}$$

These elementary results have been belabored here because there is considerable dispute in the literature on how to define position and velocity operators in the Dirac theory.[24] The present definitions of position and velocity (without operators!) are actually equivalent to the most straight-forward operator definitions in the standard formulation. To establish that we use Table I to relate the components of in (3.18) to the matrix formulation, with the result

$$\langle\, \mathbf{v}\,\rangle \cdot \boldsymbol{\sigma}_k = \langle\, \mathbf{v} \cdot \boldsymbol{\sigma}_k\,\rangle = \int d^3x\, \Psi^\dagger \alpha_k \Psi\,, \tag{3.19}$$

where, as noted before, $\alpha_k = \gamma_k\gamma_0 = \gamma_0\gamma^k$ is the matrix analog of $\boldsymbol{\sigma}_k = \gamma_k\gamma_0$ in STA.

The α_k are hermitian operators often interpreted as "velocity operators" in accordance with the HOO Principle. However, this leads to peculiar and ultimately unphysical conclusions.[25] STA resolves the difficulty by revealing that the commutation relations for the α_k have a geometrical meaning independent of any properties of the electron. It shows that the α_k are *"velocity operators"* in only a trivial sense. The role of the α_k in (3.19) is isomorphic to the role of basis vectors $\boldsymbol{\sigma}_k$ used to select components of the vector $\mathbf{v}$. The velocity

vector is inherent in the bilinear function $\Psi\Psi^\dagger$, not in the operators α_k. The α_k simply pick out its components in (3.19). Accordingly, the equivalence of STA representations to conventional operator representations exhibited in (3.19) and Table I leads to two important conclusions:[7] The hermiticity of the α_k is only incidental to their role in the Dirac theory, and their eigenvalues have no physical significance whatever! These concepts play no role in the STA formulation.

Having chosen a particle interpretation for the Dirac theory, the assumption that the particle is charged implies that the *charge current* (density) J must be proportional to the Dirac current; specifically,

$$J = e\psi\gamma_0\widetilde{\psi} = e\rho v\,. \tag{3.20}$$

Then charge conservation $\nabla \cdot J = 0$ is an immediate consequence of probability conservation. Later it will be seen that there is more to this story.

One more assumption is needed to complete the identification of observables in the Dirac theory. It comes from the interpretation of the $\underline{p}_\mu$ in (3.1) as kinetic energy-momentum operators. In the STA formulation they are defined by

$$\underline{p}_\mu = \underline{i}\hbar\partial_\mu - e\,A_\mu\,, \tag{3.21}$$

where the underbar signifies a "linear operator" and the operator $\underline{i}$ signifies right multiplication by the bivector $\mathbf{i} = \gamma_2\gamma_1$, as defined by

$$\underline{i}\,\psi = \psi\mathbf{i}\,. \tag{3.22}$$

The importance of (3.21) can hardly be overemphasized. Above all, it embodies the fruitful "minimal coupling" rule, a fundamental principle of gauge theory which fixes the form of electromagnetic interactions. In this capacity it plays a crucial heuristic role in the original formulation of the Dirac equation, as is clear when the equation is written in the form

$$\gamma^\mu\underline{p}_\mu\psi = \psi\gamma_0 m\,. \tag{3.23}$$

However, the STA formulation tells us even more. It reveals geometrical properties of the $\underline{p}_\mu$ which provide clues to a deeper physical meaning. We have already noted a connection of the factor $\mathbf{i}\hbar$ with spin in (2.22). We establish below that this connection is a consequence of the form and interpretation of the $\underline{p}_\mu$. Thus, *spin was inadvertently smuggled into the Dirac theory by the $\underline{p}_\mu$, hidden in the innocent looking factor $i'\hbar$.* Its sudden appearance was only incidentally related to relativity. History has shown that it is impossible to recognize this fact in the conventional formulation of the Dirac theory, with its emphasis on attributing physical meaning to operators and their commutation rules. The connection of $i'\hbar$ with spin is not inherent in the $\underline{p}_\mu$ alone. It appears only when the $\underline{p}_\mu$ operate on the wave function, as is evident in (2.24). This leads to the conclusion that the significance of the $\underline{p}_\mu$ lies in what they imply about the physical meaning of the wave function. Indeed, the STA formulation reveals the $\underline{p}_\mu$ have something important to tell us about the kinematics of electron motion.

The operators $\underline{p}_\mu$ or, equivalently, $\underline{p}^\mu = \gamma^\mu \cdot \gamma^\nu\underline{p}_\nu$ are given a physical meaning by using them to define the components $T^{\mu\nu}$ of the electron *energy-momentum tensor*:

$$T^{\mu\nu} = T^\mu \cdot \gamma^\nu = (\gamma_0\widetilde{\psi}\gamma^\mu\underline{p}^\nu\psi)_{(0)}\,. \tag{3.24}$$

TABLE II: Observables of the energy-momentum operator, relating real and matrix versions.

Energy-momentum tensor	$T^{\mu\nu} = T^\mu \cdot \gamma^\nu = (\gamma_0 \widetilde{\psi} \gamma^\mu \underline{p}^\nu \psi)_{(0)}$ $= \widetilde{\Psi} \gamma^\mu \underline{p}^\nu \Psi$
Kinetic energy density	$T^{00} = (\psi^\dagger \underline{p}^0 \psi)_{(0)} = \Psi^\dagger \underline{p}^0 \Psi$
Kinetic momentum density	$T^{0k} = (\psi^\dagger \underline{p}^k \psi)_{(0)} = \Psi^\dagger \underline{p}^k \Psi$
Angular Momentum tensor	$J^{\nu\alpha\beta} = \left[T^\nu \wedge x + i\rho(s \wedge \gamma^\nu) \right] \cdot (\gamma^\alpha \wedge \gamma^\beta)$ $= T^{\nu\alpha} x^\beta - T^{\nu\beta} x^\alpha - \frac{i'\hbar}{2} \widetilde{\Psi} \gamma_5 \gamma_\mu \Psi \epsilon^{\mu\nu\alpha\beta}$
Gordon current	$K_\mu = \frac{e}{m} (\widetilde{\psi} \underline{p}_\mu \psi)_{(0)} = \frac{e}{m} \widetilde{\Psi} \underline{p}_\mu \Psi$

Its matrix equivalent is given in Table II. As mentioned in the discussion of the electromagnetic energy-momentum tensor,

$$T^\mu = T(\gamma^\mu) = T^{\mu\nu} \gamma_\nu \tag{3.25}$$

is the energy-momentum flux through a hyperplane with normal γ^μ. The energy-momentum density in the electron rest system is

$$T(v) = v_\mu T^\mu = \rho p \,. \tag{3.26}$$

This defines the *"expected" proper momentum* $p = p(x)$. The observable $p = p(x)$ is the statistical prediction for the momentum of the electron at x. In general, the momentum p is not collinear with the velocity, because it includes a contribution from the spin. A measure of this noncollinearity is $p \wedge v$, which should be recognized as defining the *relative* momentum in the electron rest frame.

From the definition (3.24) of $T^{\mu\nu}$ in terms of the Dirac wave function, momentum and angular momentum conservation laws can be established by direct calculation from the Dirac equation. First, we find that[4] (See Appendix B for an alternative approach)

$$\partial_\mu T^\mu = J \cdot F \,, \tag{3.27}$$

where J is the Dirac charge current (3.20) and $F = \nabla \wedge A$ is the electromagnetic field. The right side of (3.27) is *exactly* the classical Lorentz force, so using (1.36) and denoting

the electromagnetic energy-momentum tensor (1.35) by T^μ_{EM}, we can rephrase (3.27) as *the total energy-momentum conservation law*

$$\partial_\mu(T^\mu + T^\mu_{EM}) = 0\,. \tag{3.28}$$

To derive the angular momentum conservation law, we identify $T^\mu \wedge x$ as the orbital angular momentum tensor (See Table II for comparison with more conventional expressions). Noting that $\partial_\mu x = \gamma_\mu$, we calculate

$$\partial_\mu(T^\mu \wedge x) = T^\mu \wedge \gamma_\mu - \partial_\mu T^\mu \wedge x\,. \tag{3.29}$$

To evaluate the first term on the right, we return to the definition (3.24) and find

$$\gamma_\mu T^{\mu\nu} = [\,(\underline{p}^\nu\psi)\gamma_0\widetilde{\psi}]_{(1)} = \tfrac{1}{2}\big[\,(\underline{p}^\nu\psi)\gamma_0\widetilde{\psi} + \psi\gamma_0(\underline{p}^\nu\psi)\tilde{}\,\big] = (\underline{p}^\nu\psi)\gamma_0\widetilde{\psi} - \partial^\nu(\tfrac{1}{2}\hbar\psi i\gamma_3\widetilde{\psi})\,.$$

Summing with γ_ν and using the Dirac equation (3.23) to evaluate the first term on the right while recognizing the spin vector (2.21) in the second term, we obtain

$$\gamma_\nu\gamma_\mu T^{\mu\nu} = m\psi\widetilde{\psi} + \nabla(\rho s i)\,. \tag{3.30}$$

By the way, the pseudoscalar part of this equation proves (3.4), and the scalar part gives the curious result

$$T^\mu{}_\mu = T^\mu \cdot \gamma_\mu = m\cos\beta\,. \tag{3.31}$$

However, the bivector part gives the relation we are looking for:

$$T^\mu \wedge \gamma_\mu = T^{\mu\nu}\gamma_\mu \wedge \gamma_\nu = \nabla \cdot (\rho s i) = -\partial_\mu(\rho S^\mu)\,, \tag{3.32}$$

where

$$S^\mu = (is)\cdot\gamma^\mu = i(s\wedge\gamma^\mu) \tag{3.33}$$

is the spin angular momentum tensor already identified in (2.26) and (2.28). Thus from (3.29) and (3.27) we obtain the angular momentum conservation law

$$\partial_\mu J^\mu = (F\cdot J)\wedge x\,, \tag{3.34}$$

where

$$J(\gamma^\mu) = J^\mu = T^\mu \wedge x + \rho S^\mu \tag{3.35}$$

is the angular momentum tensor, representing the total angular momentum flux in the γ^μ direction. In the electron rest system, therefore, the angular momentum density is

$$J(v) = \rho(p\wedge x + S)\,, \tag{3.36}$$

where recalling (2.12), $p\wedge x$ is recognized as the expected orbital angular momentum and as already advertised in (2.22), $S = isv$ can be indentified as an intrinsic angular momentum or spin. This completes the justification for interpreting S as spin. The task remaining is to dig deeper and understand its origin.

We now have a complete set of conservation laws for the *observables* $r,\ v,\ S$ and p, but we still need to ascertain precisely how p is related to the wave function. For that purpose we employ the invariant decomposition $\psi = (\rho e^{i\beta})^{\frac{1}{2}} R$. First we need some kinematics. By an argument used in Section 3, it is easy to prove that the derivatives of the rotor R must have the form

$$\partial_\mu R = \tfrac{1}{2}\,\Omega_\mu R\,, \tag{3.37}$$

where $\Omega_\mu = \Omega_\mu(x)$ is a bivector field. Consequently the derivatives of the e_ν defined by (2.18) have the form

$$\partial_\mu e_\nu = \Omega_\mu \cdot e_\nu\,. \tag{3.38}$$

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

Now, with the help of (2.23), the effect of $\underline{p}_\nu$ on ψ can be put in the form

$$\underline{p}_\nu\psi = [\,\partial_\nu(\ln\rho + i\beta) + \Omega_\nu\,]S\psi - eA_\nu\psi\,. \tag{3.39}$$

Whence

$$(\underline{p}_\nu\psi)\gamma_0\widetilde{\psi} = [\,\partial_\nu(\ln\rho + i\beta) + \Omega_\nu\,]i\rho s - eA_\nu v\,. \tag{3.40}$$

Inserting this in the definition (3.24) for the energy-momentum tensor, after some manipulations beginning with $is = Sv$, we get the explicit expression

$$T_{\mu\nu} = \rho\big[\,v_\mu(\Omega_\nu \cdot S - eA_\nu) + (\gamma_\mu \wedge v)\cdot(\partial_\nu S) - s_\mu\partial_\nu\beta\,\big]\,. \tag{3.41}$$

From this we find, by (3.26), the momentum components

$$p_\nu = \Omega_\nu \cdot S - eA_\nu\,. \tag{3.42}$$

This reveals that (apart from the A_ν contribution) the *momentum has a kinematical meaning related to the spin:* It is completely determined by the component of Ω_ν in the spin plane. In other words, it describes the *rotation rate* of the frame $\{e_\mu\}$ in the spin plane or, if you will "about the spin axis." But we have identified the angle of rotation in this plane with the phase of the wave function. Thus, the momentum describes the phase change in all directions of the wave function or, equivalently, of the frame $\{e_\mu\}$. A physical interpretation for this geometrical fact will be offered in Section 5.

The kinematical import of the operator $\underline{p}_\nu$ is derived from its action on the rotor R. To make that explicit, use (3.37) and (2.22) to get

$$(\partial_\nu R)i\hbar\widetilde{R} = \Omega_\nu S = \Omega_\nu \cdot S + \Omega_\nu \wedge S + \partial_\nu S\,, \tag{3.43}$$

where (2.22) was used to establish that

$$\partial_\nu S = \tfrac{1}{2}[\,\Omega_\nu, S\,] = \tfrac{1}{2}(\Omega_\nu S - S\Omega_\nu)\,. \tag{3.44}$$

Introducing the abbreviation

$$iq_\nu = \Omega_\nu \wedge S\,, \qquad \text{or} \qquad q_\nu = -(iS)\cdot\Omega_\nu\,, \tag{3.45}$$

we can put (3.43) in the form

$$(\underline{p}_\nu R)\widetilde{R} = p_\nu + iq_\nu + \partial_\nu S\,. \tag{3.46}$$

This shows explicitly how the operator $\underline{p}_\nu$ relates to kinematical observables, although the physical significance of q_ν is obscure. Note that both p_ν and $\partial_\nu S$ contribute to $T_{\mu\nu}$ in (3.41), but q_ν does not. By the way, it should be noted that the last two terms in (3.41) describe energy-momentum flux orthogonal to the v direction. It is altogether natural that this flux should depend on the component of $\partial_\nu S$ as shown. However, the significance of the parameter β in the last term remains obscure.

An auxiliary conservation law can be derived from the Dirac equation by decomposing the Dirac current as follows. Solving (3.23) for the Dirac charge current, we have

$$J = e\psi\gamma_0\widetilde{\psi} = \frac{e}{m}\,(\underline{p}_\mu\psi)\widetilde{\psi}\,. \tag{3.47}$$

The identity (3.46) is easily generalized to

$$(\underline{p}_\mu\psi)\widetilde{\psi} = (p_\mu + iq_\mu)\rho e^{i\beta} + \partial_\mu(\rho S e^{i\beta})\,. \tag{3.48}$$

The right side exhibits the scalar, pseudoscalar and bivector parts explicitly. From the scalar part we define the *Gordon current:*

$$K_\mu = \frac{e}{m}\,[\,(\underline{p}_\mu\psi)\widetilde{\psi}\,]_{(0)} = \frac{e}{m}\,(\widetilde{\psi}\underline{p}_\mu\psi)_{(0)} = \frac{e}{m}\,(p_\mu\rho\cos\beta - q_\mu\rho\sin\beta)\,. \tag{3.49}$$

Or in vector form,

$$K = \frac{e}{m}\,\rho(p\cos\beta - q\sin\beta)\,. \tag{3.50}$$

As anticipated in the last Section, from the last term in (3.48) we define the magnetization

$$M = \frac{e}{m}\,\rho S e^{i\beta}\,. \tag{3.51}$$

When (3.48) is inserted into (3.47), the pseudovector part must vanish, and vector part gives us the so-called *"Gordon decomposition"*

$$J = K + \nabla\cdot M\,. \tag{3.52}$$

This is ostensibly a decomposition into a *conduction current* K and a *magnetization current* $\nabla\cdot M$, both of which are separately conserved. But how does this square with the physical interpretation already ascribed to J? It suggests that there is a substructure to the charge flow described by J. Evidently, if we are to understand this substructure we must understand the role of the parameter β so prominently displayed in (3.50) and (3.51). A curious fact is that β does not contribute to the definition (2.20) for the Dirac current in terms of the wave function; β is related to J only indirectly through the Gordon Relation (3.52). This suggests that β characterizes some feature of the substructure.

So far we have supplied a physical interpretation for all parameters in the wave function (2.17) except *"duality parameter"* β. The physical interpretation of β is more problematic

than that of the other parameters. Let us refer to this as the *β-problem.* This problem has not been recognized in conventional formulations of the Dirac theory, because the structure of the theory was not analyzed in sufficient depth to identify it. The problem arose, however, in a different guise when it was noted that the Dirac equation admits negative energy solutions. The famous *Klein paradox* showed that negative energy states could not be avoided in matching boundary conditions at a potential barrier. This was interpreted as showing that electron-positron pairs are created at the barrier, and it was concluded that second quantization of the Dirac wave function is necessary to deal with the many particle aspects of such situations. However, recognition of the β-problem provides a new perspective which suggests that second quantization is unnecessary, though this is not to deny the reality of pair creation. An analysis of the Klein Paradox from this perspective has been given by Steve Gull.[26]

In the plane wave solutions of the Dirac equation (next Section), the parameter β unequivocally distinguishes electron and positron solutions. This suggests that β parametrizes an admixture of electron-positron states where $\cos\beta$ is the relative probability of observing an electron. Then, while $\rho = \rho(x)$ represents the relative probability of observing a particle at x, $\rho\cos\beta$ is the probability that the particle is an electron, while $\rho\sin\beta$ is the probability that it is an positron. On this interpretation, the Gordon current shows a redistribution of the current flow as a function of β. It leads also to a plausible interpretation for the β-dependence of the magnetization in (3.51). In accordance with (4.39), in the electron rest system determined by J, we can split M into

$$M = -\mathbf{P} + i\mathbf{M}\,, \tag{3.53}$$

where, since $v \cdot s = 0$,

$$i\mathbf{M} = \frac{e}{m}\, S\rho\cos\beta \tag{3.54}$$

is the *magnetic moment* density, while

$$\mathbf{P} = -\frac{e}{m}\, iS\rho\sin\beta \tag{3.55}$$

is the *electric dipole moment* density. The dependence of $\mathbf{P}$ on $\sin\beta$ makes sense, because pair creation produces electric dipoles. On the other hand, cancelation of magnetic moments by created pairs may account for the reduction of $\mathbf{M}$ by the $\cos\beta$ factor in (3.54). It is tempting, also, to interpret equation (3.4) as describing a creation of spin concomitant with pair creation.

Unfortunately, there are difficulties with this straight forward interpretation of β as an *antiparticle mixing parameter.* The standard Darwin solutions of the Dirac hydrogen atom exhibit a strange β dependence which cannot reasonably be attributed to pair creation. However, the solutions also attribute some apparently unphysical properties to the Dirac current; suggesting that they may be superpositions of more basic physical solutions. Indeed, Heinz Krüger has recently found hydrogen atom solutions with $\beta = 0$.[27]

It is easy to show that a superposition of solutions to the Dirac equation with $\beta = 0$ can produce a composite solution with $\beta \neq 0$. It may be, therefore, that β characterizes a more general class of statistical superpositions than particle-antiparticle mixtures. At any rate, since the basic observables v, S and p are completely characterized by the kinematical

factor R in the wave function, it appears that a statistical interpretation for β as well as ρ is appropriate.

4. ELECTRON TRAJECTORIES

In classical theory the concept of *particle* refers to an object of negligible size with a continuous trajectory. It is often asserted that it is meaningless or impossible in quantum mechanics to regard the electron as a particle in this sense. On the contrary, it asserted here that the particle concept is not only essential for a complete and coherent interpretation of the Dirac theory, it is also of practical value and opens up possibilitiesfor new physics at a deeper level. Indeed, in this Section it will be explained how particle trajectoriescan be computed in the Dirac theory and how this articulates perfectly with the classical theory formulated in Section 3.

David Bohm has long been the most articulate champion of the particle concept in quantum mechanics.[28] He argued that the difference between classical and quantum mechanics is not in the concept of particle itself but in the equation for particles trajectories. From Schroedinger's equation he derived an equation of motion for the electron which differs from the classical equation only in a stochastic term called the "Quantum Force." He was careful, however, not to commit himself to any special hypothesis about the origins of the Quantum Force. He accepted the form of the force dictated by Schroedinger's equation. However, he took pains to show that all implications of Schroedinger theory are compatible with a strict particle interpretation. The same general particle interpretation of the Dirac theory is adopted here, and the generalization of Bohm's equation derived below provides a new perspective on the Quantum Force.

We have already noted that each solution of the Dirac equation determines a family of nonintersecting streamlines which can be interpreted as "expected" electron histories. Here we derive equations of motion for observables of the electron along a single *history* $x = x(\tau)$. By a space-time split the history can always be projectedinto a particle *trajectory* $\mathbf{x}(\tau) = x(\tau) \wedge \gamma_0$ in a given inertial system. It will be convenient to use the terms 'history' and 'trajectory' almost interchangeably. The representation of motion by trajectories is most helpful in interpreting experiments, but histories are usually more convenient for theoretical purposes.

The main objection to a strict particle interpretation of the Dirac and Schroedinger theories is the claim that a wave interpretation is essential to explain diffraction. This claim is false, as should be obvious from the *fact* that, as we have noted, the wave function determines a unique family of electron trajectories. For double slit diffraction these trajectories have been calculated from Schroedinger's equation,[29] and, recently, from the Dirac equation.[44,45] Sure enough, after flowing uniformly through the slits, the trajectories bunch up at diffraction maxima and thin out at the minima. According to Bohm, the cause of this phenomenon is the Quantum Force rather than wave interference. This shows at least that the particle interpretation is not inconsistent with diffraction phenomena, though the origin of the Quantum Force remains to be explained. The obvious objections to this account of diffraction have been adequately refuted in the literature.[29,30] It is worth noting, though, that this account has the decided advantage of avoiding the paradoxical "collapse of the wave function" inherent in the conventional "dualist" explanation of diffraction. At

no time is it claimed that the electron spreads out like a wave to interfere with itself and then "collapse" when it is detected in a localized region. The claim is only that the electron is likely to travel on one of a family of possible trajectories consistent with experimental constraints; which trajectory is known initially only with a certain probability, though it can be inferred more precisely after detection in the final state. Indeed, it is possible then to infer which slit the electron passed through.[29] These remarks apply to the Dirac theory as well as to the Schroedinger theory, though there are some differences in the predicted trajectories,[45,46] because the Schroedinger current is the nonrelativistic limit of the Gordon current rather than the Dirac current.[9]

The probability density ρ_0 is *literally* an observable in a diffraction pattern, though not in intermediate states of a diffraction experiment. The same can be said for the velocities of detected electrons. This is justification for referring to ρ and v as "observables," though they are not associated with any operators save the Dirac wave function itself. But is it equally valid to regard them as "observables" in an atom? Though the Dirac theory predicts a family of orbits (or trajectories) in an atom, most physicists don't take this seriously, and it is often asserted that it is meaningless to say that the electron has a definite velocity in an atom. But here is some evidence to the contrary that should give the sceptics pause: The hydrogen s-state wave function is spherically symmetric and its Schroedinger (or Gordon) current vanishes, so no electron motion is indicated. However, the radial probability distribution has a maximum at Bohr radius. This would seem to be no more than a strange coincidence, except for the fact that the Dirac current does not vanish for an s-state, because the magnetization current is not zero. Moreover, the average angular momentum associated with this current is $\hbar$,[9] exactly as in the Bohr theory! Now comes the experimental evidence. When negative muons are captured in atomic s-states their lifetimes are increased by a time dilation factor corresponding to a velocity of — you guessed it! — the Bohr velocity. Besides the idea that an electron in an s-state has a definite velocity, this evidence supports the *major contention* that the electron velocity is more correctly described by the Dirac current than by the Gordon current.

Now let us investigate the equations for motion along a Dirac streamline $x = x(\tau)$. On this curve the kinematical factor in the Dirac wave function (2.17) can be expressed as a function of proper time

$$R = R(x(\tau))\,. \tag{4.1}$$

By (2.18), (2.20) and (3.10), this determines a *comoving frame*

$$e_\mu = R\,\gamma_\mu \widetilde{R} \tag{4.2}$$

on the streamline with velocity $v = e_0$, while the spin vector s and bivector S are given as before by (2.21) and (2.22). In accordance with (3.37), differentiation of (4.1) leads to

$$\dot{R} = v \cdot \nabla R = \tfrac{1}{2}\Omega R\,, \tag{4.3}$$

where the overdot indicates differentiation with respect to proper time, and

$$\Omega = v^\mu \Omega_\mu = \Omega(x(\tau)) \tag{4.4}$$

is the rotational velocity of the frame $\{e_\mu\}$. Accordingly,

$$\dot{e}_\mu = v \cdot \nabla\, e_\mu = \Omega \cdot e_\mu\,. \tag{4.5}$$

But these equations are identical in form to those for the classical theory of a relativistic rigid body with negligible size.[6] This is a consequence of the particle interpretation. In Bohmian terms, the only difference between classical and quantum theory is in the functional form of Ω. Our main task, therefore, is to investigate what the Dirac theory tells us about Ω. Among other things, that automatically gives us the classical limit formulated as in Ref. 6, a limit in which the electron still has a nonvanishing spin of magnitude $\hbar/2$.

From (3.42) we immediately obtain

$$\Omega \cdot S = (p + eA) \cdot v = \tfrac{1}{2}\hbar\omega \,. \tag{4.6}$$

This defines rate of rotation in the spin plane, $\omega = \omega(x(\tau))$, as a function of the electron momentum. For a free particle (considered below), we find that it "spins" with the ultrahigh frequency

$$\omega = \frac{2m}{\hbar} = 1.6 \times 10^{21}\,\mathrm{s}^{-1} \,. \tag{4.7}$$

According to (4.6), this frequency will be altered by external fields.

Equation (4.6) is part of a more general equation obtained from (3.43):

$$\Omega S = (p + eA) \cdot v + i(q \cdot v) + \dot{S} \,. \tag{4.8}$$

As an interesting aside, this can be solved for

$$\Omega = \dot{S}S^{-1} + (q \cdot v)iS^{-1} + (p + eA) \cdot vS^{-1} \,, \tag{4.9}$$

where $S^{-1} = is^{-1}v$. Whence,

$$\dot{v} = \Omega \cdot v = (\dot{S} \cdot v)S^{-1} - (q \cdot v)s^{-1} \,. \tag{4.10}$$

This shows something about the coupling of spin and velocity, but it is not useful for solving the equations of motion.

A general expression for Ω in terms of observables can be derived from the Dirac equation. This has been done in two steps in Ref. 4. The first step yields the interesting result

$$\Omega = -\nabla \wedge v + v \cdot (i\nabla\beta) + (m\cos\beta + eA \cdot v)S^{-1} \,. \tag{4.11}$$

But this tells us nothing about particle streamlines, since

$$\dot{v} = v \cdot (\nabla \wedge v) \tag{4.12}$$

is a mere identity, which can be derived from (1.12) and the fact that v^2 is constant. The second step yields

$$-\nabla \wedge v + v \cdot (i\nabla\beta) = m^{-1}(eFe^{i\beta} + Q) \,, \tag{4.13}$$

where Q has the complicated form

$$Q = -e^{i\beta}[\,\partial_\mu W^\mu + \gamma_\mu \wedge \gamma_\nu (W^\mu W^\nu) S^{-1}\,]_{(0)} \,, \tag{4.14}$$

with

$$W_\mu = (\rho e^{i\beta})^{-1}\partial_\mu(\rho e^{i\beta}S) = \partial_\mu S + S\partial_\mu(\ln\rho + i\beta) \,. \tag{4.15}$$

Inserting (4.13) in (4.11), we get from (4.5) and (3.44) the equations of motion for velocity and spin:

$$m\dot{v} = e(Fe^{i\beta})\cdot v + Q\cdot v\,, \tag{4.16}$$

$$\dot{S} = F\times\Big(\frac{e}{m}\,Se^{i\beta}\Big) + Q\times S\,, \tag{4.17}$$

where $A\times B = \frac{1}{2}(AB - BA)$ is the *commutator product.*

Except for the surprising factor $e^{i\beta}$, the first term on the right of (4.16) is the classical Lorentz force. The term $Q\cdot v$ is the generalization of Bohm's *Quantum Force.* A crucial fact to note from (4.15) is that the dependence of the Quantum Force on Plank's constant comes *entirely* from the spin S. This *spin dependence of the Quantum Force* is hidden in the Schroedinger approximation, but it can be shown to be implicit there nevertheless.[9] The *classical limit* can be characterized first by $\rho \to 0$ and $\partial_\mu \ln\rho \to 0$; second, by $\partial_\mu S = v_\mu \dot{S}$, which comes from assuming that only the variation of S along the history can affect the motion. Accordingly, (4.14) reduces to $Q = \ddot{S}$, and for the limiting classical equations of motion for a particle with intrinsic spin we obtain[13]

$$m\dot{v} = (eF - \ddot{S})\cdot v\,, \tag{4.18}$$

$$m\dot{S} = (eF - \ddot{S})\times S\,. \tag{4.19}$$

These coupled equations have not been seriously studied. Of course, they should be studied in conjunction with the spinor equation (4.3).

In the remainder of this Section we examine *classical solutions* of the Dirac equation, that is, solutions whose streamlines are classical trajectories. For a free particle ($A = 0$), the Dirac equation (2.5) admits *plane wave* solutions of the form

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R = \rho^{\frac{1}{2}} e^{i\beta/2} R_0 e^{-\mathbf{i}p\cdot x/\hbar}\,, \tag{4.20}$$

where the kinematical factor R has been decomposed to explicitly exhibit its spacetime dependence in a phase factor. Inserting this into (2.5) and using $\nabla p\cdot x = p$, we obtain

$$p\psi = \psi\gamma_0 m\,. \tag{4.21}$$

Solving for p we get

$$p = me^{i\beta} R\gamma_0 \widetilde{R} = mve^{-i\beta}\,. \tag{4.22}$$

This implies $e^{i\beta} = \pm 1$, so

$$e^{i\beta/2} = 1 \text{ or } i\,, \tag{4.23}$$

and $p = \pm mv$ corresponding to two distinct solutions. One solution appears to have negative energy $E = p\cdot\gamma_0$, but that can be rectified by changing the sign in the phase of the "trial solution" (4.20).

Thus we obtain two distinct kinds of plane wave solutions with positive energy $E = p\cdot\gamma_0$:

$$\psi_- = \rho^{\frac{1}{2}} R_0 e^{-\mathbf{i}p\cdot x/\hbar}\,, \tag{4.24}$$

$$\psi_+ = \rho^{\frac{1}{2}} iR_0 e^{+\mathbf{i}p\cdot x/\hbar}\,. \tag{4.25}$$

We can identify these as *electron* and *positron* wave functions. Indeed, the two solutions are related by charge conjugation. According to (2.15), the charge conjugate of (4.24) is

$$\psi_-^C = \psi_- \boldsymbol{\sigma}_2 = \rho^{\frac{1}{2}} i R_0' e^{-\mathbf{i}p\cdot x/\hbar}\,, \tag{4.26a}$$

where

$$R_0' = R_0(-i\boldsymbol{\sigma}_2)\,. \tag{4.26b}$$

As seen below, the factor $-i\boldsymbol{\sigma}_2$ represents a spatial rotation which just "flips" the direction of the spin vector. Evidently (4.25) and (4.26a) are both positron solutions, but with oppositely directed spins.

Determining the comoving frame (4.2) for the electron solution (4.24), we find that the velocity $v = R_0\gamma_0\widetilde{R}_0$ and the spin $s = \frac{1}{2}\hbar R_0\gamma_3\widetilde{R}_0$ are constant, but, for $k = 1, 2$,

$$e_k(\tau) = e_k(0)e^{-p\cdot x/S} = e_k(0)e^{e_2 e_1 \omega\tau}\,, \tag{4.27}$$

where $\tau = v \cdot x$ and ω is given by (4.7). Thus, the streamlines are straight lines along which the spin is constant and e_1 and e_2 rotate about the "spin axis" with the ultrahigh frequency (4.7) as the electron moves along the streamline. A similar result is found for the positron solution.

For applications, the constants in the solution must be specified in more detail. If the wave functions are normalized to one particle per unit volume V in the γ_0-system, then we have

$$\rho_0 = \gamma_0 \cdot (\rho v) = \frac{1}{V} \quad \text{or} \quad \rho = \frac{m}{EV} = \frac{1}{\gamma_0 \cdot vV}\,.$$

Following the procedure beginning with (2.13), we make the space-time split

$$R = LU \qquad \text{where} \qquad U = U_0 e^{-\mathbf{i}p\cdot x/\hbar}\,. \tag{4.28}$$

The result of calculating L from γ_0 and the momentum p has already been found in (2.24). As in (2.19) and (3.37), it is convenient to represent the spin direction by the relative vector

$$\boldsymbol{\sigma} = U\boldsymbol{\sigma}_3\widetilde{U}\,. \tag{4.29}$$

This is all we need to characterize spin. But to make contact with more conventional representations, we decompose it as follows: Choosing $\boldsymbol{\sigma}_3$ as *"quantization axis,"* we decompose U into spin up and spin down amplitudes denoted by U_+ and U_- respectively, and defined by

$$U_\pm\boldsymbol{\sigma}_3 = \pm\boldsymbol{\sigma}_3 U_\pm \tag{4.30}$$

or

$$U_\pm = \tfrac{1}{2}(U \pm \boldsymbol{\sigma}_3 U)\,. \tag{4.31}$$

Thus

$$U = U_+ + U_-\,. \tag{4.32}$$

It follows that

$$U\widetilde{U} = |\,U_+\,|^2 + |\,U_-\,|^2 = 1\,, \tag{4.33a}$$

$$U_+\widetilde{U}_- + U_-\widetilde{U}_+ = 0\,, \tag{4.33b}$$

$$\boldsymbol{\sigma} = U\boldsymbol{\sigma}_3\widetilde{U} = \big\{|\,U_+\,|^2 - |\,U_-\,|^2\big\}\boldsymbol{\sigma}_3 + 2U_-\widetilde{U}_+\boldsymbol{\sigma}_3\,. \tag{4.34}$$

Since $\boldsymbol{\sigma}\boldsymbol{\sigma}_3 = \boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_3 + i(\boldsymbol{\sigma}\times\boldsymbol{\sigma}_3)$,

$$\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_3 = |\,U_+\,|^2 - |\,U_-\,|^2\,, \tag{4.35a}$$

$$\boldsymbol{\sigma}_3\times\boldsymbol{\sigma} = 2iU_-\widetilde{U}_+\,. \tag{4.35b}$$

This decomposition into spin up and down amplitudes is usually given a statistical interpretation in quantum mechanics, but we see here its geometrical significance.

The classical limit is ordinarily obtained as an "eikonal approximation" to the Dirac equation. Accordingly, the wave function is set in the form

$$\psi = \psi_0 e^{-\mathbf{i}\varphi/\hbar}\,. \tag{4.36}$$

Then the "amplitude" ψ_0 is assumed to be slowly varying compared to the "phase" φ, so the derivatives of ψ_0 in the Dirac equation can be neglected to a good approximation. Thus, inserting (4.36) into the Dirac equation, say in the form (3.47), we obtain

$$(\nabla\varphi - eA)e^{i\beta} = mv\,. \tag{4.37}$$

As in the plane wave case (4.22) this implies $e^{i\beta} = \pm 1$, and the two values correspond to electron and positron solutions. For the electron case,

$$\nabla\varphi - eA = mv\,. \tag{4.38}$$

This defines a family of classical histories in spacetime. For a given external potential $A = A(x)$, the phase φ can be found by solving the "Hamilton-Jacobi equation"

$$(\nabla\varphi - eA)^2 = m^2\,, \tag{4.39}$$

obtained by squaring (4.38). On the other hand, the curl of (4.38) gives

$$m\nabla\wedge v = -e\nabla\wedge A = -eF\,. \tag{4.40}$$

Dotting this with v and using the identity (4.12), we obtain exactly the classical equation (3.6) for each streamline.

Inserting (4.40) into (4.11), we obtain

$$\Omega = \frac{e}{m}\,F + (m + eA\cdot v)S^{-1}\,. \tag{4.41}$$

Whence the rotor equation (4.3) assumes the explicit form

$$\dot{R} = \frac{e}{2m}\,FR - R\mathbf{i}(m + eA\cdot v)/\hbar\,. \tag{4.42}$$

This admits a solution by separation of variables:

$$R = R_0 e^{-\mathbf{i}\varphi/\hbar}\,, \tag{4.43}$$

where

$$\dot{R}_0 = \frac{e}{2m}\, F R_0 \tag{4.44}$$

and

$$\dot{\varphi} = v \cdot \nabla\varphi = m + eA \cdot v\,. \tag{4.45}$$

Equation (4.44) is identical with the classical equation in Ref. 6, while (4.45) can be obtained from (4.38).

Thus, in the eikonal approximation the quantum equation for a comoving frame differs from the classical equation only in having additional rotation in the spin plane. Quantum mechanics also assigns energy to this rotation, and an explicit expression for it is obtained by inserting (4.41) into (4.1), with the interesting result

$$p \cdot v = m + \frac{e}{m}\, F \cdot S\,. \tag{4.46}$$

This is what one would expect classically if there were some sort of localized motion in the spin plane. That possibility will be considered in the next Section.

The eikonal solutions characterized above are exact solutions of the Dirac equation when the ψ_0 in (4.38) satisfies

$$\nabla\psi_0 = 0\,. \tag{4.47}$$

This equation has a whole class of exact solutions where ψ_0 is not constant. This class is comparable in richness to the class of analytic functions in complex variable theory, for (4.47) can be regarded as a generalization of the Cauchy-Riemann equations.[15] Considering the exact correspondence of the eikonal equations with classical theory, we can regard wave functions of this class as *exact classical solutions* of the Dirac equation. An important member of this class is the so-called *Volkov solution* for an electron in the field of an electromagnetic plane wave.[32] Remarkably, this solution of the Dirac equation is *identical* to the solution of the classical spinor equation for a point charge in a plane wave field.[6]

As a final observation about the eikonal approximation, we note that it rules out the possibility of finding any spin dependence of the streamlines such as that exhibited in equation (4.18). Evidently the spin dependence appears when the $\nabla\varphi$ in (4.38) is generalized to a vector field with nonvanishing curl.

5. THE ZITTERBEWEGUNG INTERPRETATION.

Now that we have the geometrical and physical interpretation of the Dirac wave function well in hand, we are prepared to examine deeper possibilities of the Dirac theory. We have seen that the kinematics of electron motion is completely characterized by the "Dirac rotor" R in the invariant decomposition (2.17) of the wave function. The Dirac rotor determines a comoving frame $\{e_\mu = R\,\gamma_\mu \widetilde{R}\}$ which rotates at high frequency in the e_2e_1-plane, the "spin plane," as the electron moves along a streamline. Moreover, according to (4.6) and (4.46),

there is energy associated with this rotation, indeed, all the rest energy $p \cdot v$ of the electron. These facts suggest that the *electron mass, spin and magnetic moment are manifestation of a local circular motion of the electron.* Mindful that the velocity attributed to the electron is an independent assumption imposed on the Dirac theory from physical considerations, we recognize that this suggestion can be accommodated by giving the electron a component of velocity in the spin plane. Accordingly, we now *define the electron velocity* u *by*

$$u = v - e_2 = e_0 - e_2 \,. \tag{5.1}$$

The choice $u^2 = 0$ has the advantage that the electron mass can be attributed to kinetic energy of self interaction while the spin is the corresponding angular momentum.[11]

This new identification of electron velocity makes the plane wave solutions a lot more physically meaningful. For $p \cdot x = mv \cdot x = m\tau$, the kinematical factor for the solution (4.24) can be written in the form

$$R = e^{\frac{1}{2}\Omega\tau} R_0 \,, \tag{5.2}$$

where Ω is the constant bivector

$$\Omega = mc^2 S^{-1} = \frac{2mc^2}{\hbar}\, e_1 e_2 \,. \tag{5.3}$$

From (5.2) it follows that v is constant and

$$e_2(\tau) = e^{\Omega\tau} e_2(0) \,. \tag{5.4}$$

So $u = \dot{z}$ can be integrated immediately to get the electron history

$$z(\tau) = v\tau + (e^{\Omega\tau} - 1) r_0 + z_0 \,, \tag{5.5}$$

where $r_0 = \Omega^{-1} e_2(0)$. This is a lightlike helix centered on the Dirac streamline $x(\tau) = v\tau + z_0 - r_0$. In the electron "rest system" defined by v, it projects to a circular orbit of radius

$$|\, r_0 \,| = |\, \Omega^{-1} \,| = \frac{\hbar}{2m} = 1.9 \times 10^{-13}\mathrm{m} \,. \tag{5.6}$$

The diameter of the orbit is thus equal to an electron Compton wavelength. For $r(\tau) = e^{\Omega\tau} r_0$, the angular momentum of this circular motion is, as intended, the spin

$$(m\dot{r}) \wedge r = m\dot{r} r = m r^2 \Omega = m\Omega^{-1} = S \,. \tag{5.7}$$

Finally, if z_0 is varied parametrically over a hyperplane normal to v, equation (5.5) describes a 3-parameter family of spacetime filling lightlike helixes, each centered on a unique Dirac streamline. According to the Born statistical interpretation, the electron can be on any one of these helixes with uniform probability.

Let us refer to this localized helical motion of the electron by the name *zitterbewegung* (*zbw*) originally introduced by Schroedinger.[33] Accordingly, we call $\omega = \Omega \cdot S$ the *zbw frequency* and $\lambda = \omega^{-1}$ the *zbw radius*. The phase of the wave function can now be interpreted literally as the phase in the circular motion, so we can refer to that as the *zbw phase*.

Although the frequency and radius ascribed to the *zbw* are the same here as in Schroedinger's work, its role in the theory is quite different. Schroedinger attributed it to interference between positive and negative energy components of a wave packet,[33,34] whereas here it is associated directly with the complex phase factor of a plane wave. From the present point of view, wave packets and interference are not essential ingredients of the *zbw*, although the phenomenon noticed by Schroedinger certainly appears when wave packets are constructed. Of course, the present interpretation was not an option open to Schroedinger, because the association of the unit imaginary with spin was not established (or even dreamed of), and the vector e_2 needed to form the spacelike component of the *zbw velocity* u was buried out of sight in the matrix formalism. Now that it has been exhumed, we can see that the *zbw* must play a ubiquitous role in quantum mechanics. The present approach associates the *zbw* phase and frequency with the phase and frequency of the complex phase factor in the electron wave function. Henceforth, this will be referred to as *the zitterbewegung interpretation* of quantum mechanics.

The strength of the *zbw* interpretation lies first in its coherence and completeness in the Dirac theory and second in the intimations it gives of more fundamental physics. It will be noted that the *zbw* interpretation is completely general, because the definition (5.1) of the *zbw* velocity is well defined for any solution of the Dirac equation. It is also perfectly compatible with everything said about the interpretation of the Dirac theory in previous Sections. One need only recognize that the Dirac velocity can be interpreted as the average of the electron velocity over a *zbw* period, as expressed by writing

$$v = \bar{u} \, . \tag{5.8}$$

Since the period is on the order of 10^{-21}s, it is v rather than u that best describes electron motion in most experiments.

A possible difficulty with the interpretation of u as electron velocity is the fact that ρu is not necessarily a conserved current, for from (3.6) we have

$$\nabla \cdot (\rho u) = \frac{2}{\hbar} \, \rho A \cdot e_1 \, . \tag{5.9}$$

However, it is probably sufficient that ρv is conserved.

Perhaps the strongest theoretical support for the *zbw* interpretation is the fact that it is fundamentally geometrical; it completes the kinematical interpretation of R, so all components of R, even the complex phase factor, characterize features of the electron history. This kinematical interpretation is made most explicitly in Ref. 14, where the comoving frame $\{e_\mu\}$ is interpreted as a Frenet frame, with vectors e_1 and e_3 corresponding to first and third curvatures; the *zbw* radius is then seen as the radius of curvature for the particle history.

The *key ingredients* of the *zbw* interpretation are the complex phase factor and the energy-momentum operators $\underline{p}_\mu$ defined by (3.21). The unit imaginary $\mathbf{i}$ appearing in both of these has the dual properties of representing the plane in which *zbw* circulation takes place and generating rotations in that plane. The phase factor literally represents a rotation on the electron's circular orbit in the $\mathbf{i}$-plane. Operating on the phase factor, the $\underline{p}_\mu$ computes the phase rotation rates in all spacetime directions and associates them with the electron energy-momentum. Thus, the *zbw* interpretation explains the physical significance of the mysterious "quantum mechanical operators" $\underline{p}_\mu$.

The key ingredients of the *zbw* interpretation are preserved in the nonrelativistic limit and so provide a *zitterbewegung interpretation of Schroedinger theory.* The nonrelativistic approximation to the STA version of the Dirac theory, leading through the Pauli theory to the Schroedinger theory, has been treated in detail elsewhere.[15,13] But the essential point can be seen by a split of the Dirac wave function y into the factors

$$\psi = \rho^{\frac{1}{2}} e^{i\beta/2} L U e^{-\mathbf{i}(m/\hbar)t} \,. \tag{5.10}$$

In the nonrelativistic approximation three of these factors are neglected or eliminated and ψ is reduced to the Pauli wave function

$$\psi_P = \rho^{\frac{1}{2}} U_0 e^{-\mathbf{i}(\varphi/\hbar)} \,, \tag{5.11}$$

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 already been discussed in Ref. 9. In the Schroedinger approximation the factor U_0 is neglected so ψ_P reduces to the Schroedinger wave function

$$\psi_S = \rho^{\frac{1}{2}} e^{-\mathbf{i}(\varphi/\hbar)} \,. \tag{5.12}$$

It follows from this derivation of the Schroedinger wave function that just as in the Dirac theory, the phase $\varphi/\hbar$ describes the *zbw*, and $\partial_\mu \varphi$ describes the *zbw* energy and momentum. We see now the physical significance of the complex that phase factor $e^{-\mathbf{i}(\varphi/\hbar)}$ is kinematical rather than logical or statistical as so often claimed.

The *zbw* interpretation explains much more than the electron spin and magnetic moment. That is especially clear in the Schroedinger theory where spin is ignored but the complex phase factor is essential. Stationary state solutions of both the Schroedinger and Dirac equations reveal an important property of the *zbw*. The singlevaluedness of the wave function implies that the *orbital frequency is a harmonic of the zbw frequency shift in stationary states.* This opens the possibility of *zbw resonance* as a fundamental explanatory principle in quantum mechanics.[11,37,38] The Pauli principle may be a consequence of *zbw* resonance between electron pairs, since it is linked to stationary state conditions. Diffraction may be explained as *zbw resonant momentum exchange.* Thus we have the possibility, or better, the challenge of finding *zbw* explanations for all the familiar phenomena of quantum mechanics, including barrier penetration and the Aharonov-Bohm effect.

Further support for the *zbw* interpretation comes from recent successes of *semiclassical mechanics* in molecular dynamics and electronic structure,[36] often surpassing the results from standard quantum mechanical methods. Such success may be surprising from the conventional view of quantum mechanics, but from the *zbw* perspective, the semiclassical approach of imposing quantum conditions on classical dynamics is just of way of meeting the conditions for *zbw* resonances. This constitutes further evidence for the possibility that standard quantum mechanics is dealing with ensembles of particle orbits with *zbw* periodicity.

Quantum mechanics is characterized by phase coherence over distances very much larger than an electron Compton wavelength defining the dimensions of the *zbw*. By what causal mechanism might *zbw* coherence be established over such large distances? A tantalizing possibility arises by interpreting the circular *zbw* orbit literally as the orbit of a point charge.

For that implies that the electron must be the source of a (nonradiating) electromagnetic field which fluctuates with the *zbw* frequency. The observed Coulomb and magnetic dipole fields of the electron are averages of this field over times much longer than a *zbw* period. The *zbw* fluctuations are much too rapid to observe directly, though perhaps they have been observed indirectly all along in quantum coherence phenomena. This rapidly fluctuating field is a prime candidate for Bohm's Quantum Force. A speculative analysis of its quantum implications is given in Refs. 11, 37 and 38.

Considering how well the *zbw* interpretation fits the Dirac theory, we can regard the Dirac theory and all its successes as evidence that the *zbw* is a real physical phenomena. The Dirac theory, then, does not explain the *zbw*, it simply tells us that the zbw exists and describes some of its properties. To explain the *zbw* we must go beyond the Dirac theory to discover new physical mechanisms such as the fluctuating "Quantum Force" proposed in the preceding paragraph. However, the Dirac theory is not without clues as to what to look for.[39] One important clue concerns the origin of electron mass. The very form of the important equation (4.6) suggests that the electron's mass may be a consequence of *magnetic self-interaction*, as expressed by writing

$$m = S \cdot \Omega = \boldsymbol{\mu} \cdot \mathbf{B}_S \,, \tag{5.13}$$

where $\mathbf{B}_S$ is the self-magnetic field presumed to be the origin of the free particle Ω. This is a suggestive starting point for a *zbw* approach to quantum electrodynamics, but that must be deferred to another day.

6. ELECTROWEAK INTERACTIONS.

The STA formulation of the Dirac theory has indubitable implications for the Weinberg-Salam (W-S) theory of electroweak interactions. The W-S theory generalizes the electromagnetic (E-M) gauge group to the electroweak (E-W) gauge group SU(2)×U(1). However, this was done without realizing that *the imaginary unit* $\mathbf{i}$ *which generates E-M gauge transformations in the Dirac theory is a spacelike bivector* identified with the electron spin. This fact forces a strong geometrical constraint on the W-S theory: Since $\mathbf{i}$ has a spacetime interpretation, the generators of the larger E-W group which include it must have related spacetime interpretations. Remarkably, this constraint can be easily satisfied in the following way:

The Dirac current $\psi\gamma_0\widetilde{\psi}$ is a timelike vector field, so only 4 parameters are needed to specify it. However, 8 parameters are needed to specify the wave function ψ uniquely. Therefore, the Dirac current is invariant under a 4-parameter group of gauge transformations on the wave function:

$$\psi \rightarrow \psi G \,, \tag{6.1}$$

where $G = G(x)$ is an even multivector satisfying

$$G\gamma_0\widetilde{G} = \gamma_0 \,. \tag{6.2}$$

It follows that

$$G = Ue^{i\lambda} \,, \tag{6.3}$$

where $U\widetilde{U} = 1$. This exhibits explicitly the SU(2)×U(1) structure of the gauge group. Thus, *the invariance group of the Dirac current can be identified with the E-W gauge group.* The subgroup which leaves the spin density $\rho\hat{s} = \psi\gamma_3\widetilde{\psi}$ invariant is characterized by the additional condition

$$G\gamma_3\widetilde{G} = \gamma_3 \,. \tag{6.4}$$

The E-M gauge transformations belong to this subgroup. Note also that the U(1) factor in (6.3) is a duality factor exactly like the one parametrized by $\beta/2$ in the invariant decomposition of the Dirac wave function (2.17). Thus, it may be that the fundamental physical role of β is to serve as a gauge parameter in electroweak theory.

Of course, the Dirac equation is not invariant under the entire E-W gauge group $\{G\}$, but it is easily generalized to one that is by introducing a suitable "gauge invariant derivative" in the standard way. That has been done in Ref. 10, where the Weinberg-Salam model is completely reformulated in terms of STA with the E-W gauge group defined as above. This opens up possibilities for integrating the zitterbewegung idea with electroweak theory. Evidently that would obviate the need for including Higgs bosons in the theory, since the zitterbewegung provides an alternative mechanism to account for the electron mass.

7. CONCLUSIONS.

The objective of this work has been to understand what makes quantum mechanics so successful by analyzing the Dirac theory. The analysis has been developed progressively on three levels: reformulation, reinterpretation and modification. Let us take stock, now, to see how far we have progressed toward the objective.

A. REFORMULATION. We have seen that reformulation of the Dirac theory in terms of STA eliminated superfluous degrees of freedom in the Dirac algebra and reveals a hidden geometrical structure in the Dirac equation and its solutions. The main results are:

(1.) The Dirac wave function has the invariant decomposition

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} \, R \,. \tag{7.1}$$

(2.) The factor $\mathbf{i}\hbar$ in the Dirac equation is a spacelike bivector related to the spin by

$$S = \tfrac{1}{2} R (\mathbf{i}\hbar) \widetilde{R} \,. \tag{7.2}$$

(3.) The electron energy-momentum p_ν is related to the spin by

$$p_\nu = \Omega_\nu \cdot S - eA_\nu \,, \tag{7.3}$$

where $\partial_\nu R = \frac{1}{2}\Omega_\nu R$.

These results are mathematical facts inherent in the original Dirac theory. By making the geometric structure of the theory explicit, however, they suggest a new, more coherent and complete interpretation of the theory.

B. REINTERPRETATION. The new zitterbewegung interpretation is imposed on the Dirac theory simply by identifying the electron velocity with the lightlike vector $u = R(\gamma_0 - \gamma_2)\widetilde{R}$. It follows that the spin S in (7.2) is the angular momentum of the zitterbewegung, and (7.3) attributes energy-momentum to this motion. The general helical character of the zitterbewegung is completely determined by the Dirac equation without further assumption.

This approach has the great formal advantage of providing the entire rotor R with a kinematical interpretation. In particular, the complex phase factor is interpreted as a direct representation of the zitterbewegung itself. Thus, a physical explanation is given for the appearance of complex numbers in quantum mechanics. Moreover, the zitterbewegung interpretation of the phase factor carries over to Schroedinger theory and so suggests a reinterpretation of quantum mechanics generally. This has the great advantage over variants of the Copenhagen interpretation of being grounded in the Dirac theory.

Above all, the zitterbewegung interpretation presents us with an array of challenges. First, there is a theoretical challenge to see how far we can go in providing zitterbewegung interpretations for the standard results of quantum mechanics and even quantum electrodynamics. Second, there is a challenge to probe the zitterbewegung experimentally to see if it can be established as a "literally real" phenomenon. Finally, there is a challenge to see if the zitterbewegung can lead us beyond present quantum mechanics to deeper physical insights.

C. MODIFICATIONS. If indeed the zitterbewegung is physically real it is probably a consequence of electromagnetic or electroweak self-interaction, and it may be the source of an electromagnetic field which fluctuates with the zitterbewegung frequency. Thus it opens up the possibility of a new approach to the self-interaction problem and actually explaining the phenomenon of quantization rather than assuming it. Of course, such possibilities cannot be explored theoretically without going beyond the Dirac theory.

REFERENCES

1. D. Hestenes, **Space-Time Algebra** (Gordon & Breach, London, l966).
2. D. Hestenes, Real Spinor Fields, *J. Math. Phys.* **8**, 798–808 (1967).
4. D. Hestenes, Local Observables in the Dirac Theory, *J. Math. Phys.* **14**, 893–905 (1973).
5. D. Hestenes, Proper Particle Mechanics, *J. Math. Phys.* **15**, 1768–1777 (1974).
6. D. Hestenes, Proper Dynamics of a Rigid Point Particle, *J. Math. Phys.* **15**, 1778–1786 (1974).
7. D. Hestenes, Observables, Operators and Complex Numbers in the Dirac Theory, *J. Math. Phys.* **16**, 556-572 (1975).
8. R. Gurtler and D. Hestenes, Consistency in the formulation of the Dirac, Pauli and Schroedinger Theories, *J. Math. Phys.* **16**, 573–583 (1975).
9. D. Hestenes, Spin and Uncertainty in the Interpretation of Quantum Mechanics, *Am. J. Phys.*, **47**, 339–415 (1979).

10. D. Hestenes, Space-Time Structure of Weak and Electromagnetic Interactions, *Found. Phys.*, **12**, l53–l68 (1982).

11. D. Hestenes, Quantum Mechanics from Self-Interaction, *Found. Phys.* **15**, 63–87 (1985).

12. D. Hestenes, Clifford Algebra and the Interpretation of Quantum Mechanics. In **Clifford Algebras and their Applications in Mathematical Physics**, J.S.R. Chisholm & A. K. Common (eds.), (Reidel Publ. Co., Dordrecht/Boston, l986), p. 321–346.

13. D. Hestenes, On Decoupling Probability from Kinematics in Quantum Mechanics. In. **Maximum Entropy and Bayesian Methods**, Dartmouth College l989, P. Fougere (Ed.) Kluwer, Dordrecht/Boston (1990).

14. D. Hestenes, The Zitterbewegung Interpretation of Quantum Mechanics, *Found. Phys.* **20**, 1213–1232 (1990).

15. D. Hestenes and G. Sobczyk, **Clifford Algebra to Geometric Calculus, A Unified Language for Mathematics and Physics**, G Reidel Publ. Co., Dordrecht/Boston (1984).

16. D. Hestenes, **New Foundations for Classical Mechanics**, G. Reidel Publ. Co., Dordrecht/ Boston (1985).

17. P.A.M. Dirac, **Quantum Mechanics**, Oxford U. Press, London, 4th edition (1958).

18. D. Hestenes, Vectors, Spinors and Complex Numbers in Classical and Quantum Physics, *Am. J. Phys.* **39**, 1013–1028 (1971).

19. D. Hestenes, A Spinor Approach to Gravitational Motion and Precession, *Int. J. Theo. Phys.*, **25**, 589–598 (1986).

20. H.E. Moses, Eigenfunctions of the curl operator, rotationally invariant Helmholtz theorem, and applications to electromagnetic theory and fluid mechanics. *SIAM J. Appl. Math.* **21**, 114–144 (1971).

21. J. Bjorken and S. Drell, **Relativistic Quantum Mechanics**, McGraw-Hill, N.Y. (1964).

22. J. Jauch and F. Rohrlich, **The Theory of Photons and Electrons**, Addison-Wesley, Reading Mass. (1955).

23. J. P. Crawford, On the algebra of Dirac bispinor densities, *J. Math. Phys.* **26**, 1439–1441 (1985).

24. T. D. Newton and E. P. Wigner, Localized States for Elementary Systems, *Rev. Mod. Phys.* **21**, 400–406 (1949).

25. R. P. Feynman, **Quantum Electrodynamics**, Benjamin, N.Y. (1961), p. 47–49.

26. S. Gull, Charged particled in potential steps, In D. Hestenes and A. Weingarthofer (eds.), **The ELECTRON, new theory and experiment**, Kluwer-Dordrecht/Boston (1991), 37–48 (1991).

27. H. Krüger, private communication.

28. D. Bohm & B. Hiley, Unbroken Quantum Realism, from Microscopic to Macroscopic Levels, *Phys. Rev. Letters* **55**, 2511 (1985).

29. C. Philippidis, C. Dewdney and B. J. Hiley, Quantum Interference and the Quantum Potential, Nuovo Cimento **52B**, 15–28 (1979).

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

31. M. P. Silberman, Relativistic time dilatation of bound muons and the Lorentz invariance of charge, *Am. J. Phys.* **50**, 251–254 (1982).

32. L. S. Brown & T.W.G. Kibble, Interactions of Intense Laser beams with Electrons, *Phys. Rev. A* **133**, 705 (1964).

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

34. K. Huang, On the Zitterbewegung of the Electron, *Am. J. Phys.* **47**, 797 (1949).

35. A. O. Barut and A. J. Bracken, Zitterbewegung and the internal geometry of the electron, *Phys. Rev.* **D23**, 2454 (1981).

36. T. Uzer, D. Farrelly, J. Milligan, P. Raines & J. Skelton, Celestial Mechanics on a Microscopic Scale, *Science* **253**, 42–48 (1991).

37. D. Hestenes, Zitterbewegung in Radiative Processes. In D. Hestenes and A. Weingarthofer (eds.), **The ELECTRON, new theory and experiment**, Kluwer-Dordrecht/ Boston (1991) p. 21–36.

38. D. Hestenes, The Kinematic Origin of Complex Wave Functions. In W.T. Grandy & P.W. Miloni (eds.), **Physics and Probability: Essays in Honor of Edwin T. Jaynes**, Cambridge U. Press, Cambridge (1993), p. 153–160.

39. D. Hestenes, Zitterbewegung Modeling, *Found. Phys.* **28**, 365–387 (1993).

40. S. Gull, A. Lasenby and C. Doran, Imaginary Numbers Are Not Real—The Geometric Algebra of Spacetime, *Found. Phys.* **23**, 1175–1202 (1993).

41. P. Lounesto, Clifford Albegras and Hestenes Spinors, *Found. Phys.* **23**, 1203–1238 (1993).

42. S. Gull, A. Lasenby and C. Doran, States and Operators in the Spacetime Algebra, *Found. Phys.* **23**, 1239–1264 (1993).

43. H. Krüger, Classical Limit of Real Dirac Theory: Quantization of Relativistic Central Field Orbits, *Found. Phys.* **23**, 1265–1288 (1993).

44. A. Lasenby, C. Doran and S. Gull, A Multivector Derivative Approach to Lagrangian Field Theory, *Found. Phys.* **23**, 1295–1328 (1993).

45. S. Gull, A. Lasenby and C. Doran, Electron Paths, Tunneling, and Diffraction in the Spacetime Algebra, *Found. Phys.* **23**, 1329–1356 (1993).

46. C. Doran, A. Lasenby, S. Gull, S. Somaroo & A. Challinor, Spacetime Algebra and Electron Physics. In *Advances in Imaging and Electron Physics.* Academic Press. **95**, (1995).

APPENDIX A: TRANSFORMATIONS AND INVARIANTS.

This Appendix formulates general transformation laws for fields on spacetime and applies the results to establish Poincaré invariance of the field equations. The transformation law for spinor fields is shown to rest on a convention which can be chosen to make it identical with the transformation law for tensor fields.

Let f be a transformation of a 4-dimensional region (or manifold) $\mathcal{R} = \{x\}$ onto a region $\mathcal{R}' = \{x'\}$ in a spacetime; the pointwise transformation is thus

$$f : x \to x' = f(x)\,. \tag{A.1}$$

A transformation is understood to be a differentiable invertible mapping. It induces a transformation $\underline{f}$ of a vector field $a = a(x)$ on $\mathcal{R}$ into a vector field $a' = a'(x')$ on $\mathcal{R}'$ defined by

$$\underline{f} : a \to a' = \underline{f}a \equiv a \cdot \nabla f\,, \tag{A.2}$$

where the derivative is evaluated at $x = f^{-1}(x')$. A parenthesis has been dropped in writing $\underline{f}a$ in recognition that $\underline{f}$ is a linear operator on tangent vectors.

The transformation $\underline{f}$ is called the *differential* of f. It has a natural extension from vector fields to arbitrary multivector fields. Thus, for vector fields a_1, a_2, ... a_k, the differential of the k-vector field $a_1 \wedge a_2 \wedge \cdots \wedge a_k$ is defined by

$$\underline{f}(a_1 \wedge a_2 \wedge \cdots \wedge a_k) = (\underline{f}a_1) \wedge (\underline{f}a_2) \wedge \cdots \wedge (\underline{f}a_k)\,. \tag{A.3}$$

By linearity this determines the differential for any multivector field $M = M(x)$:

$$\underline{f}M = \sum_{k=0}^{4} \underline{f}(M)_{(k)}\,, \tag{A.4}$$

where it is understood that

$$(\underline{f}M)_{(0)} = (M)_{(0)}\,, \tag{A.5}$$

which is to say that every scalar field is an invariant of $\underline{f}$ and hence of f. An extensive treatment of the differential on differentiable manifolds is given in Ref. 15.

The differential of the unit pseudoscalar is given by

$$\underline{f}i = i \det \underline{f}\,, \tag{A.6}$$

where

$$\det \underline{f} = i^{-1}\,\underline{f}i = -i\,\underline{f}i \tag{A.7}$$

is the *Jacobian* of f.

As an application of general interest, let us calculate the form of the differential for an arbitrary *infinitesimal transformation*

$$f(x) = x + \epsilon(x)\,. \tag{A.8}$$

It is understood that, for any unit vector $\widehat{a}$, $\epsilon \cdot \widehat{a}$ is a small quantity. This is equivalent to the condition that ϵ^2 is small, except when ϵ is a null vector. For a vector field the corresponding induced transformation is

$$\underline{f}a = a \cdot \nabla(x + \epsilon) = a + a \cdot \nabla\epsilon \,. \tag{A.9}$$

Neglecting second order terms, therefore, for a bivector field $a \wedge b$, we have

$$\underline{f}(a \wedge b) = a \wedge b + (a \cdot \nabla\epsilon) \wedge b + a \wedge (b \cdot \nabla\epsilon) = a \wedge b + [\,(a \wedge b) \cdot \nabla\,] \wedge \epsilon \,. \tag{A.10}$$

This result generalizes easily to the differential for an arbitrary multivector field:

$$\underline{f}M = M + (M \cdot \nabla) \wedge \epsilon \,, \tag{A.11}$$

where

$$M \cdot \nabla = \sum_{k=1}^{4} (M)_{(k)} \cdot \nabla \,. \tag{A.12}$$

Note that $k = 0$ is not included in the sum since the scalar part of M is invariant.

As a significant example, we apply (A.11) to the pseudoscalar i and obtain

$$\underline{f}i = i + (i \cdot \nabla) \wedge \epsilon = i(1 + \nabla \cdot \epsilon) \,. \tag{A.13}$$

This gives us immediately the useful expression

$$\det \underline{f} = 1 + \nabla \cdot \epsilon \tag{A.14}$$

for the Jacobian of f.

Equation (A.11) determines a new function

$$M'(x') = M'(x + \epsilon) = M + [\,M(x) \cdot \nabla\,] \wedge \epsilon \,.$$

To exhibit the ϵ-dependence of the argument explicitly, we make a Taylor expansion of the argument and keep only first order terms to get

$$M'(x) = M(x) - \epsilon \cdot \nabla M(x) + [\,M(x) \cdot \nabla\,] \wedge \epsilon \,. \tag{A.15}$$

This tells us explicitly how the infinitesimal transformation changes the field M at a designated point x.

Now let us turn to the question of Poincaré invariance of the equations of physics. The *Poincaré group* is the group of transformations on spacetime which leave invariant the "interval" $(x_2 - x_1)^2$ between every pair of spacetime points x_1 and x_2. We are concerned here only with the *Restricted Poincaré Group* for physical reasons discussed below. This is the subgroup of Poincaré transformations continuously connected to the identity. Every such transformation is the composite of a Lorentz rotation and a translation, so, according to (1.24), it can be written in the canonical form

$$\underline{f}(x) = Rx\widetilde{R} + c \,, \tag{A.16}$$

where c is a constant vector and R is a constant rotor with $R\widetilde{R} = 1$. From (A.2) we find immediately the differential

$$a' = \underline{f}a = RaR\,. \tag{A.17}$$

For the product of two vector fields this gives the simple result

$$a'b' = (\underline{f}a)(\underline{f}b) = RabR\,. \tag{A.18}$$

By virtue of (1.3) this can be decomposed into a scalar part

$$a' \cdot b' = a \cdot b \tag{A.19}$$

and a bivector part

$$a' \wedge b' = \underline{f}(a \wedge b) = R(a \wedge b)\widetilde{R}\,. \tag{A.20}$$

According to (A.3), the outer product is an invariant of the differential for any transformation. The inner product and the geometric product are not generally invariant; however, (A.19) and (A.18) show that they are in the present case. It follows that for an arbitrary multivector field M the transformation law is simply

$$M' = \underline{f}M = RM\widetilde{R}\,. \tag{A.21}$$

From this the Poincaré invariance of the basic equations of physics is easily established. For example, for Maxwell's equation (1.28) we have

$$\nabla' F' = (\underline{f}\nabla)(\underline{f}F) = (R\nabla\widetilde{R})(RF\widetilde{R}) = R\nabla F\widetilde{R} = RJ\widetilde{R} = J'\,. \tag{A.22}$$

Thus, the relation of field F to current J is a Poincaré invariant.

The physical significance of Poincaré invariance deserves some comment, since the matter is frequently muddled in the literature. Poincaré transformations are commonly interpreted as relations among different inertial reference systems or observers. That is clearly not the correct interpretation here, for no reference system has even been mentioned either in the formulation of Maxwell's equation or of its induced transformation (A.22). Indeed, Maxwell's equation $\nabla F = J$ is manifestly independent of any coordinate system, so no argument at all is needed to establish its observer independence. The Poincaré invariance expressed by (A.22) should be interpreted as an equivalence of spacetime points rather than an equivalence of observers. It describes a physical property of the Minkowski model of spacetime. Translation invariance implies that *spacetime is homogeneous* in sense that the same laws of physics are the same at every spacetime point. Similarly, Lorentz rotation invariance implies that *spacetime is isotropic* in the sense that the laws of physics do not favor any particular timelike or spacelike directions. Poincaré invariance thus provides the theoretical basis for comparing the results of physical experiments and observations made at different times and places. It is the formal assertion that the laws of physics are the same everywherewhen.

The Poincaré invariance of the Dirac equation (2.5) can be established in the same way as that of Maxwell's equation. Thus,

$$\begin{aligned}\nabla'\psi'\mathbf{i}'\hbar - eA\psi' &= (R\nabla\widetilde{R})(R\psi\widetilde{R})(R\mathbf{i}'\widetilde{R})\hbar - e(RA\widetilde{R})(R\psi\widetilde{R}) \\ &= R(\nabla\psi\mathbf{i}\hbar - eA\psi)\widetilde{R} = R(m\psi\gamma_0)\widetilde{R} = m\psi'\gamma_0'\,.\end{aligned} \tag{A.23}$$

Note that the transformation law for the spinor wave function ψ has been taken to be

$$\psi' = R\psi\widetilde{R} \tag{A.24}$$

in accordance with (A.21). However, the usual transformation law for a Dirac spinor is

$$\psi' = R\psi \tag{A.25}$$

or, in the conventional matrix representation (2.2),

$$\Psi' = R\Psi\,. \tag{A.26}$$

Nevertheless, the transformation laws (A.24) and (A.25) are physically equivalent, because all observables are bilinear functions of the wave function. Indeed, as established by (2.7) through (2.11), the factor $\widetilde{R}$ on the right of (A.24) can be transformed away at will. The choice between the transformation laws (A.24) and (A.25) is therefore a matter of convention. Though (A.25) is simpler, (A.24) has the advantage of conformity with (A.21) and hence the more general transformation law (A.4).

In the conventional formulation,[21] spinors are *defined* by the transformation law (A.26), so the possibility of (A.24) does not arise. On the other hand, transformations are not employed to define spinors in the manifestly invariant STA formulation. Rather, the spinor ψ itself represents a transformation as defined by (2.19).

It is of interest to consider briefly the infinitesimal Poincaré transformations, since they play such a prominent role in the conventional approach to relativistic quantum theory. For an infinitesimal translation, we take $R = 1$ and $\epsilon = c$ in (A.16), so (A.15) reduces to

$$M'(x) = M(x - c) = (1 - c\cdot\nabla)M(x)\,. \tag{A.27}$$

This applies equally to the electromagnetic field and the Dirac wave function. Writing $c\cdot\nabla = c^\mu\partial_\mu$, we recognized the ∂_μ as generators of translations. It is noted that the ∂_μ can be made hermitian by the artifice of introducing a unit imaginary factor i', so the translation operators in quantum mechanics are usually defined to be $i'\partial_\mu$. These operators are then identified with momentum operators. However, our analysis of the Dirac theory in Section 3 reveals that the success of this formal procedure should be attributed to the physical interpretation of the Dirac wave function rather than to a general physical significance of hermitian operators.

For an infinitesimal Lorentz rotation, we take $c = 0$ in (A.16) and write

$$R = e^{\frac{1}{2}B} \approx 1 + \tfrac{1}{2}B\,, \tag{A.28}$$

where B is an infinitesimal bivector. Then (A.16) reduces to

$$x' = (1 + \tfrac{1}{2}B)x(1 - \tfrac{1}{2}B) \approx x + B\cdot x\,. \tag{A.29}$$

Hence, $\epsilon = B\cdot x$ in (A.15) and/or (A.30) in (A.21) gives

$$M'(x) = [\,1 - B\cdot(x\wedge\nabla) + B\times\,]M(x)\,, \tag{A.30}$$

where $B \times M$ is the commutator product. Alternatively, for a spinor ψ subject to the one-sided transformation law (A.27), the result is

$$\psi'(x) = [\,1 - B \cdot (x \wedge \nabla) + \tfrac{1}{2}B]\,\psi(x)\,. \tag{A.31}$$

This is expressed in a more conventional form by expanding B with respect to a basis to get

$$\psi'(x) = (1 + \tfrac{1}{2}B^{\mu\nu}J_{\mu\nu})\psi(x)\,, \tag{A.32}$$

where the

$$J_{\mu\nu} = (\gamma_\mu \wedge \gamma_\nu) \cdot (x \wedge \nabla) + \tfrac{1}{2}\gamma_\mu \wedge \gamma_\nu = x_\mu \partial_\nu - x_\nu \partial_\mu + \tfrac{1}{2}\gamma_\mu \wedge \gamma_\nu \tag{A.33}$$

are the usual "angular momentum operators" for a Dirac particle. In a similar way, angular momentum operators for the electromagnetic field can be read off (A.30).

APPENDIX B: LAGRANGIAN FORMULATION

This appendix is concerned with the Lagrangian formulation of the Dirac theory. The Lagrange approach has the advantages of ensuring consistency among the various field equations and directly relating them to conservation laws. Though this approach to the Dirac theory has been discussed many times in the literature, the STA formulation is sufficiently novel to merit one more version. A powerful generalization of the present method has since been presented in Ref. 44.

Let $\mathcal{L} = \mathcal{L}(x)$ be the Lagrangian for some field on spacetime. The associated action integral over any region $\mathcal{R}$ is

$$\mathcal{A} = \int_{\mathcal{R}} \mathcal{L}(x)\,|\,d^4x\,|\,, \tag{B.1}$$

where the oriented "volume element" for the region is the pseudoscalar

$$d^4x = d_1x \wedge d_2x \wedge d_3x \wedge d_4x = i\,|\,d^4x\,|\,. \tag{B.2}$$

A general variation of the action involves both a change in the functional form of $\mathcal{L}$ and an infinitesimal displacement of $\mathcal{R}$ producing a new action

$$\mathcal{A}' = \mathcal{A} + \delta\mathcal{A} = \int_{\mathcal{R}'} \mathcal{L}'(x')\,|\,d^4x'\,|\,. \tag{B.3}$$

For an infinitesimal displacement $x \to x' = x + \epsilon(x)$, (A.14) gives us

$$|\,d^4x'\,| = (1 + \nabla \cdot \epsilon)\,|\,d^4x\,|\,. \tag{B.4}$$

Hence, writing $\mathcal{L}' = \mathcal{L} + \delta\mathcal{L}$, to first order we have

$$\delta\mathcal{A} = \int_{\mathcal{R}'} (\delta\mathcal{L} + \mathcal{L}\nabla \cdot \epsilon)\,|\,d^4x\,|\,. \tag{B.5}$$

Now for a given $\mathcal{L}$, both field equations and conservation laws can be derived by requiring the invariance condition $\delta\mathcal{A} = 0$ subject to various constraints.

For the Dirac electron, we adopt the (nonunique) Lagrangian

$$\mathcal{L} = \langle\, \hbar(\nabla\psi)i\gamma_3\widetilde{\psi} - eA\psi\gamma_0\widetilde{\psi} - m\psi\widetilde{\psi}\,\rangle\,, \tag{B.6}$$

where $\langle\ldots\rangle = (\ldots)_{(0)}$ means "scalar part." We derive the Dirac equation by requiring $\delta\mathcal{A} = 0$ for an arbitrary variation $\delta\psi(x) = \psi'(x) - \psi(x)$ in the functional form of the wave function which vanishes on the boundary of $\mathcal{R}$. In this case the boundary is fixed and $\epsilon = 0$ in (B.5). The derivation employs the scalar-part properties $\langle\,\widetilde{M}\,\rangle = \langle\, M\,\rangle$ and $\langle\, MN\,\rangle = \langle\, NM\,\rangle$. Thus, using $(\delta\psi)\,\widetilde{\ } = \delta\widetilde{\psi}$ the variation of the last term in (A.6) can be put in the form

$$\delta\langle\,\psi\widetilde{\psi}\,\rangle = \langle\,(\delta\psi)\widetilde{\psi}\,\rangle + \langle\,\psi\delta\widetilde{\psi}\,\rangle = 2\langle\,\psi\delta\widetilde{\psi}\,\rangle\,.$$

Similarly, the variation of the second term in (B.6) involves

$$\langle\, A(\delta\psi)\gamma_0\widetilde{\psi}\,\rangle = \langle\,\psi\gamma_0(\delta\widetilde{\psi})A\,\rangle = \langle\, A\psi\gamma_0\delta\widetilde{\psi}\,\rangle\,.$$

To evaluate the variation of the first term in (B.6), we use $\delta(\nabla\psi) = \nabla(\delta\psi)$ and

$$\begin{aligned}\langle\,(\nabla\delta\psi)i\gamma_3\widetilde{\psi}\,\rangle &= \langle\,\nabla(\delta\psi i\gamma_3\widetilde{\psi})\,\rangle - \langle\,\delta\psi i\gamma_3(\nabla\psi)\,\widetilde{\ }\,\rangle \\ &= \langle\,(\nabla\psi)i\gamma_3\delta\widetilde{\psi}\,\rangle + \nabla\cdot(\delta\psi i\gamma_3\widetilde{\psi})_{(1)}\,.\end{aligned} \tag{B.7}$$

The last term here does not contribute to $\delta\mathcal{A}$ in (B.5), because $\delta\psi$ vanishes on the boundary. Thus, we arrive at

$$\delta\mathcal{L} = 2\langle\,(\hbar\nabla\psi i\gamma_3 - eA\psi\gamma_0 - m\psi)\delta\widetilde{\psi}\,\rangle\,. \tag{B.8}$$

This vanishes for all values of the arbitrary even multivector $\delta\widetilde{\psi}$ only if the Dirac equation (2.5) is satisfied.

CONSERVATION LAWS

Conservation Laws are derived by requiring invariance of the action under infinitesimal displacements preserving the field equations. For performing the calculation it is convenient to decompose $\delta\psi$ into a part

$$\delta^*\psi = \psi'(x) - \psi(x) \tag{B.9}$$

due to a change in the value of ψ and a part due to the shift $\epsilon = x' - x$ in the argument. This is easily done by writing

$$\delta\psi = \psi'(x') - \psi(x) = \psi'(x') - \psi(x') + \psi(x') - \psi(x)\,.$$

To first order in small quantities $\delta^*\psi(x') = \delta^*\psi(x)$ and we have

$$\delta\psi = \delta^*\psi(x) + \epsilon\cdot\nabla\psi(x)\,. \tag{B.10}$$

Applying the same argument to the integrand of (B.5), we have

$$\delta\mathcal{L} + \mathcal{L}\nabla\cdot\epsilon = \delta^*\mathcal{L} + \epsilon\cdot\nabla\mathcal{L} + \mathcal{L}\nabla\cdot\epsilon\,.$$

Thus, $\delta A = 0$ for any choice of the region $\mathcal{R}$ only if

$$\delta^*\mathcal{L} + \nabla\cdot(\epsilon\mathcal{L}) = 0\ . \tag{B.11}$$

This is a Conservation Law for specified ϵ.

To evaluate (B.11) for the electron Lagrangian (B.6), we note that $\delta^*\mathcal{L}$ will have the same form as (B.8) except that the perfect divergence term in (B.7) must be included and an additional term due to δ^*A must be added. However, since we require that the Dirac equation be satisfied, the result is simply

$$\delta^*\mathcal{L} = \nabla\cdot(\hbar\delta^*\psi i\gamma_3\widetilde{\psi})_{(1)} - e\langle\,\delta^*A\psi\gamma_0\widetilde{\psi}\,\rangle\,. \tag{B.12}$$

Inserting this into (B.11), we can express the general conservation law in the form

$$\nabla\cdot\big[\,\hbar(\delta\psi - \epsilon\cdot\nabla\psi)i\gamma_3\widetilde{\psi} + \epsilon\mathcal{L}\big]_{(1)} = e\langle\,(\delta A - \epsilon\cdot\nabla A)\psi\gamma_0\widetilde{\psi}\,\rangle\,. \tag{B.13}$$

It will be helpful to reformulate this in terms of the energy-momentum operators $\underset{\_}{p}_\mu$. From the definition (3.21) we have

$$\epsilon\cdot\underset{\_}{p}\,\psi = \epsilon\cdot\nabla\psi i\hbar\gamma_3\gamma_0 - e\,\epsilon\cdot A\,\psi\,. \tag{B.14}$$

So from the definition of the energy-momentum tensor T^μ in (3.24) or Table II, we obtain

$$\epsilon\cdot T^\mu = \langle\,\gamma^\mu(\epsilon\cdot\underset{\_}{p}\,\psi)\gamma_0\widetilde{\psi}\,\rangle\,. \tag{B.15}$$

Consequently,

$$\partial_\mu(\epsilon\cdot T^\mu) = \partial_\mu\langle\,\gamma^\mu(\epsilon\cdot\nabla\psi)i\hbar\gamma_3\widetilde{\psi}\,\rangle - \partial_\mu(e\epsilon\cdot A\langle\,\gamma^\mu\psi\gamma_0\widetilde{\psi}\,\rangle)\,, \tag{B.16}$$

which relates one term on the left of (B.13) to T^μ. The Lagrangian (B.6) can also be expressed in terms of T^μ, with the result

$$\mathcal{L} = T^\mu{}_\mu - \langle\,m\psi\widetilde{\psi}\,\rangle\,. \tag{B.17}$$

But we have already observed in (3.31) that this vanishes in consequence of the Dirac equation. Finally, we note that the last term in (B.16) can be written

$$\partial_\mu(\epsilon\cdot AJ^\mu) = J\cdot\nabla(\epsilon\cdot A)\,, \tag{B.18}$$

where $J = e\psi\gamma_0\widetilde{\psi}$ is the Dirac charge current. Hence, with the help of the identity

$$\epsilon\cdot F\cdot J = (\epsilon\cdot\nabla A)\cdot J - \epsilon\cdot(J\cdot\nabla A)\,, \tag{B.19}$$

where $F = \nabla \wedge A$, we can put (B.13) in the form

$$\partial_\mu(\epsilon \cdot T^\mu - \langle\, \gamma^\mu \delta\psi i\hbar\gamma_3 \widetilde{\psi}\,\rangle) = \epsilon \cdot F \cdot J - A \cdot (J \cdot \nabla\epsilon) - J \cdot (\delta A)\,. \tag{B.20}$$

This is the desired final form of the general conservation law. Now it is a simple matter to assess the implications of requiring Poincaré invariance.

A. Translation Invariance. For an infinitesimal translation ϵ is constant, $\delta\psi = 0$, and $\delta A = 0$. Hence, (B.20) reduces to

$$\epsilon \cdot (\partial_\mu T^\mu) = \epsilon \cdot (F \cdot J)\,. \tag{B.21}$$

Since ϵ is arbitrary, this implies the energy-momentum conservation law (3.27). Thus, energy-momentum conservation is a consequence of the homogeneity of spacetime.

B. Lorentz Invariance. For an infinitesimal Lorentz rotation, $\epsilon = B \cdot x$ by (A.29), $\delta A = B \times A = B \cdot A$ by (A.30), and $\delta\psi = (1/2)B\psi$ by (A.31). In consequence, note the following:

$$A \cdot [\, J \cdot \nabla(B \cdot x)\,] = (B \cdot J) \cdot A = -J \cdot (B \cdot A) = -J \cdot (\delta A)\,,$$

$$\epsilon \cdot T^\mu = (B \cdot x) \cdot T^\mu = B \cdot (x \wedge T^\mu)\,,$$

$$\langle\, \gamma^\mu \delta\psi i\gamma_3 \hbar\widetilde{\psi}\,\rangle = \langle\, Bi(\tfrac{1}{2}\hbar\psi\gamma_3\widetilde{\psi})\gamma^\mu\,\rangle = B \cdot (\rho S^\mu)\,,$$

where $S^\mu = i(s \wedge \gamma^\mu)$ is the spin angular momentum tensor of (3.33). Inserting these results into (B.20), we obtain

$$B \cdot [\, \partial_\mu(T^\mu \wedge x + \rho S^\mu)\,] = B \cdot [\, (F \cdot J) \wedge x\,]\,. \tag{B.22}$$

Since B is an arbitrary bivector, this implies the angular momentum conservation law (3.34). Thus, angular momentum conservation is a consequence of the isotropy of spacetime.

For the sake of completeness, we note that a complete Lagrangian for electron and E-M fields together is obtained by adding to the Dirac Lagrangian (B.6) the term $\frac{1}{2}\langle\, F^2\,\rangle = -\frac{1}{2}\langle\, F\widetilde{F}\,\rangle$, where $F = \nabla \wedge A$. The electromagnetic part of the Lagrangian is then

$$\mathcal{L}_{EM} = \tfrac{1}{2}\langle\, F^2\,\rangle - A \cdot J\,. \tag{B.23}$$

The signs have been chosen so that the interaction term agrees with (B.6) for $J = e\psi\gamma_0\widetilde{\psi}$. From this the E-M field equation can be derived by the general variational principle. Thus, we note that

$$\delta[\, \tfrac{1}{2}\langle\, F^2\,\rangle\,] = \langle\, F\nabla\delta A\,\rangle = \delta A \cdot (\nabla F) + \partial_\mu\langle\, F\gamma^\mu\delta A\,\rangle\,. \tag{B.24}$$

The last term vanishes for $\delta A = 0$ on the boundary, so we have

$$\delta\mathcal{L}_{EM} = \delta A \cdot (\nabla F - J) = 0\,.$$

Since δA is arbitrary, this implies Maxwell's equation $\nabla F = J$.

Conservation laws for the electromagnetic field can be obtained by inserting (B.23) into the general Conservation Law (B.11). Thus, using (B.24) with $\delta^* A = \delta A - \epsilon \cdot \nabla A$ and $\delta^* J = \delta J - \epsilon \cdot \nabla J$, we obtain

$$\begin{aligned} \partial_\mu \big[\langle\, F\gamma^\mu \epsilon \cdot \nabla A \,\rangle - \tfrac{1}{2}\gamma^\mu \cdot \epsilon \langle\, F^2 \,\rangle - \langle\, F\gamma^\mu \delta A \,\rangle \big] \\ = -J \cdot (\epsilon \cdot \nabla A) - A \cdot (\delta J) - (J \cdot A) \nabla \cdot \epsilon \,. \end{aligned} \tag{B.25}$$

Let us define the *canonical energy-momentum* tensor $T_c^\mu = T_c(\gamma^\mu)$ by

$$\begin{aligned} T_c(n) &\equiv -\tfrac{1}{2} FnF - n \cdot \grave{F} \cdot \grave{\nabla} \grave{A} \\ &= \grave{\nabla} \langle\, Fn\grave{A} \,\rangle - \tfrac{1}{2} n \langle\, F^2 \,\rangle + A \langle\, n \nabla F \,\rangle \,, \end{aligned} \tag{B.26}$$

where the reverse accents serve to indicate which functions are differentiated by ∇. Inserting this into (B.25), we get the Conservation Law in the form

$$\partial_\mu [\, \epsilon \cdot T_c^\mu - \langle\, F\gamma^\mu \delta A \,\rangle \,] = \epsilon \cdot (J \cdot F) + A \cdot (J \cdot \nabla \epsilon) - A \cdot (\delta J) - J \cdot A \, \nabla \cdot \epsilon \,. \tag{B.27}$$

As before, translation invariance yields the energy-momentum conservation law

$$\partial_\mu T_c^\mu = J \cdot F = -F \cdot J \,. \tag{B.28}$$

And Lorentz invariance yields the angular momentum conservation law

$$\partial_\mu [\, T_c^\mu \wedge x + S_c^\mu \,] = (J \cdot F) \wedge x \,, \tag{B.29}$$

where the E-M spin tensor $S_c^\mu = S_c(\gamma^\mu)$ is given by

$$S_c(n) = (F \cdot n) \wedge A \,. \tag{B.30}$$

Of course, both (B.28) and (B.29) can be obtained by direct differentiation of (B.26). Also note that when they are added to the corresponding equations (3.27) and (3.34) for the electron, the internal forces and torques cancel.