SPACETIME CALCULUS

David Hestenes
Arizona State University

Abstract: This book provides a synopsis of spacetime calculus with applications to classical electrodynamics, quantum theory and gravitation. The calculus is a coordinate-free mathematical language enabling a unified treatment of all these topics and bringing new insights and methods to each of them.

CONTENTS

PART I: Mathematical Fundamentals

1. SPACETIME ALGEBRA.

We represent Minkowski *spacetime* as a real 4-dimensional vector space $\mathcal{M}^4$. The two properties of scalar multiplication and vector addition in $\mathcal{M}^4$ provide only a partial specification of spacetime geometry. To complete the specification we introduce an associative *geometric product* among vectors $a, b, c, \ldots$ with the property that the square of any vector is a (real) scalar. Thus for any vector a we can write

$$a^2 = aa = \epsilon |\, a \,|^2 \,, \tag{1.1}$$

where ϵ is the *signature* of a and $|\, a \,|$ is a (real) positive scalar. As usual, we say that a is *timelike*, *lightlike* or *spacelike* if its signature is positive ($\epsilon = 1$), null ($\epsilon = 0$), or negative ($\epsilon = -1$). To specify the signature of $\mathcal{M}^4$ as a whole, we adopt the axioms: (a) $\mathcal{M}^4$ contains at least one timelike vector; and (b) Every 2-plane in $\mathcal{M}^4$ contains at least one spacelike vector.

The vector space $\mathcal{M}^4$ is not closed under the geometric product just defined. Rather, by multiplication and (distributive) addition it generates a real, associative (but noncommutative), geometric algebra of dimension $2^4 = 16$, called the *spacetime algebra* (STA). The name is appropriate because all the elements and operations of the algebra have a geometric interpretation, and it suffices for the representation of any geometric structure on spacetime.

From the *geometric product* ab of two vectors it is convenient to define two other products. The *inner product* $a \cdot b$ is defined by

$$a \cdot b = \tfrac{1}{2}(ab + ba) = b \cdot a \,, \tag{1.2}$$

and the *outer product* $a \wedge b$ is defined by

$$a \wedge b = \tfrac{1}{2}(ab - ba) = -b \wedge a \,. \tag{1.3}$$

The three products are therefore related by the important identity

$$ab = a \cdot b + a \wedge b \,, \tag{1.4}$$

which can be seen as a decomposition of the product ab into symmetric and antisymmetric parts.

From (1.1) it follows that the inner product $a \cdot b$ is scalar-valued. The outer product $a \wedge b$ is neither scalar nor vector but a new entity called a *bivector*, which can be interpreted geometrically as an oriented plane segment, just as a vector can be interpreted as an oriented line segment.

Inner and outer products can be generalized. The outer product and the notion of *k-vector* can be defined iteratively as follows: Scalars are regarded as 0-vectors, vectors as 1-vectors and bivectors as 2-vectors. For a given k-vector K the positive integer k is called the *grade* (or *step*) of K. The outer product of a vector a with K is a $(k+1)$-vector defined in terms of the geometric product by

$$a \wedge K = \tfrac{1}{2}(aK + (-1)^k Ka) = (-1)^k K \wedge a \,, \tag{1.5}$$

The corresponding inner product is defined by

$$a \cdot K = \tfrac{1}{2}(aK + (-1)^{k+1} Ka) = (-1)^{k+1} K \cdot a, \tag{1.6}$$

and it can be proved that the result is a $(k-1)$-vector. Adding (1.5) and (1.6) we obtain

$$aK = a \cdot K + a \wedge K \,, \tag{1.7}$$

which obviously generalizes (1.4). The important thing about (1.7) is that it decomposes aK into $(k-1)$-vector and $(k+1)$-vector parts.

Manipulations and inferences involving inner and outer products are facilitated by a host of theorems and identities given in [1], of which the most important are recorded here. The outer product is associative, and

$$a_1 \wedge a_2 \wedge \ldots \wedge a_k = 0 \tag{1.8}$$

if and only if the vectors $a_1, a_2, \ldots, a_k$ are linearly dependent. Since $\mathcal{M}^4$ has dimension 4, (1.8) is an identity in STA for $k > 4$, so the generation of new elements by multiplication with vectors terminates at $k = 4$. A nonvanishing k-vector can be interpreted as a directed volume element for $\mathcal{M}^4$ spanned by the vectors $a_1, a_2, \ldots, a_k$. In STA 4-vectors are called *pseudoscalars*, and for any four such vectors we can write

$$a_1 \wedge a_2 \wedge ... \wedge a_k = \lambda i\,, \tag{1.9}$$

where i is the unit pseudoscalar and λ is a scalar which vanishes if the vectors are linearly dependent.

The *unit pseudoscalar* is so important that the special symbol i is reserved for it. It has the properties

$$i^2 = -1\,, \tag{1.10}$$

and for every vector a in $\mathcal{M}^4$

$$ia = -ai\,. \tag{1.11}$$

Of course, i can be interpreted geometrically as *the* (unique) unit oriented volume element for spacetime. A convention for specifying its orientation is given below. Multiplicative properties of the unit pseudoscalar characterize the geometric concept of *duality*. The *dual* of a k-vector K in STA is the $(4-k)$-vector defined (up to a sign) by iK or Ki. Trivially, every pseudoscalar is the dual of a scalar. Every 3-vector is the dual of a vector; for this reason 3-vectors are often called *pseudovectors*. The inner and outer products are dual to one another. This is easily proved by using (1.7) to expand the associative identity $(aK)i = a(Ki)$ in two ways:

$$(a \cdot K + a \wedge K)i = a \cdot (Ki) + a \wedge (Ki)\,.$$

Each side of this identity has parts of grade $(4-k\pm1)$ and which can be separately equated, because they are linearly independent. Thus, one obtains the *duality identities*

$$a \cdot (K)i = a \wedge (Ki)\,, \tag{1.12a}$$

$$a \wedge (K)i = a \cdot (Ki)\,, \tag{1.12b}$$

Note that (1.12b) can be solved for

$$a \cdot K = [a \wedge (Ki)]i^{-1}\,, \tag{1.13}$$

which could be used to define the inner product from the outer product and duality.

Unlike the outer product, the inner product is not associative. Instead, it obeys various identities, of which the following involving vectors, k-vector K and s-vector B are most important:

$$(b \wedge a) \cdot K = b \cdot (a \cdot K) = (K \cdot b) \cdot a = K \cdot (b \wedge a) \qquad \text{for grade } k \geq 2\,, \tag{1.14}$$

$$a \cdot (K \wedge B) = (a \cdot K) \wedge B + (-1)^k K \wedge (a \cdot B)\,. \tag{1.15}$$

The latter implies the following identity involving vectors alone:

$$\begin{aligned} a \cdot (a_1 \wedge a_2 \wedge \ldots \wedge a_k) = (a \cdot a_1)a_2 \wedge \ldots \wedge a_k - (a \cdot a_2)a_1 \wedge a_3 \ldots \wedge a_k + \\ \cdots + (-1)^{k-1}(a \cdot a_k)a_1 \wedge a_2 \ldots \wedge a_{k-1}\,. \end{aligned} \tag{1.16}$$

This is called the *Laplace expansion*, because it generalizes and implies the familiar expansion for determinants. The simplest case is used so often that it is worth writing down:

$$a \cdot (b \wedge c) = (a \cdot b)c - (a \cdot c)b = a \cdot b\, c - a \cdot c\, b\,. \tag{1.17}$$

Parentheses have been eliminated here by invoking a *precedence convention*, that in ambiguous algebraic expressions inner products are to be formed before outer products, and both of those before geometric products. This convention allows us to drop parentheses on the right side of (1.16) but not on the left.

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

$$M = \alpha + a + F + bi + \beta i = \sum_{k=0}^{4} \langle M \rangle_k\,, \tag{1.18}$$

where a and b are scalars, a and b are vectors and F is a bivector. This is a decomposition of M into its "k-vector parts" $\langle M \rangle_k$, where

$$\langle M \rangle_0 = \langle M \rangle = \alpha \tag{1.19}$$

is the scalar part, $\langle M \rangle_1 = a$ is the vector part, $\langle M \rangle_2 = F$ is the bivector part, $\langle M \rangle_3 = bi$ is the pseudovector part and $\langle M \rangle_4 = \beta i$ is the pseudoscalar part. Duality has been used to simplify the form of the trivector part in (1.18) by expressing it as the dual of a vector. Like the decomposition of a complex number into real and imaginary parts, the decomposition (1.18) is significant because the parts with different grade are linearly independent of one another and have distinct geometric interpretations. On the other hand, multivectors of mixed grade often have geometric significance that transcends that of their graded parts.

Any multivector M can be uniquely decomposed into parts of even and odd grade as

$$M_\pm = \tfrac{1}{2}(M \mp iMi)\,. \tag{1.20}$$

In terms of the expanded form (1.18), the *even part* can be written

$$M_+ = \alpha + F + \beta i\,, \tag{1.21}$$

while the *odd part* becomes

$$M_- = a + bi\,. \tag{1.22}$$

The set $\{M_+\}$ of all even multivectors forms an important subalgebra of STA called the *even subalgebra*. Spinors can be represented as elements of the even subalgebra.

Computations are facilitated by the operation of reversion, defined for arbitrary multivectors M and N by

$$(MN)^{\sim} = \widetilde{N}\widetilde{M}\,, \tag{1.23a}$$

with

$$\widetilde{a} = a \tag{1.23b}$$

for any vector a. For M in the expanded form (1.18), the reverse $\widetilde{M}$ is given by

$$\widetilde{M} = \alpha + a - F - bi + \beta i\,. \tag{1.24}$$

Note that bivectors and trivectors change sign under reversion while scalars and pseudoscalars do not.

STA makes it possible to formulate and solve fundamental equations *without using coordinates*. However, it also facilitates the manipulation of coordinates, as shown below and in later sections. Let $\{\gamma_\mu; \mu = 0, 1, 2, 3\}$ be a *right-handed orthonormal frame* of vectors in $\mathcal{M}^4$ with γ_0 in the *forward light cone*. In accordance with (1.2), we can write

$$\eta_{\mu\nu} \equiv \gamma_\mu \cdot \gamma_\nu = \tfrac{1}{2}(\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu)\,. \tag{1.25}$$

This may be recognized as the defining condition for the "Dirac algebra," which is a matrix representation of STA over a complex number field instead of the reals. Although the present interpretation of the $\{\gamma_\mu\}$ as an orthonormal frame of vectors is quite different from their usual interpretation as matrix components of a single vector, it can be shown that these alternatives are in fact compatible.

The *orientation* of the *unit pseudoscalar* i relative to the frame $\{\gamma_\mu\}$ is set by the equation

$$i = \gamma_0\gamma_1\gamma_2\gamma_3 = \gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3\,. \tag{1.26}$$

We shall refer to $\{\gamma_\mu\}$ as a *standard frame*, without necessarily associating it with the reference frame of a physical observer. For manipulating coordinates it is convenient to introduce the *reciprocal frame* $\{\gamma^\mu\}$ defined by the equations

$$\gamma_\mu = \eta_{\mu\nu}\gamma^\nu \qquad \text{or} \qquad \gamma_\mu \cdot \gamma^\nu = \delta_\mu^\nu\,. \tag{1.27}$$

(summation convention assumed). "Rectangular coordinates" $\{x^\nu\}$ of a spacetime point x are then given by

$$x^\nu = \gamma^\nu \cdot x \qquad \text{and} \qquad x = x^\nu\gamma_\nu\,. \tag{1.28}$$

The γ_μ generate by multiplication a complete basis for STA, consisting of the $2^4 = 16$ linearly independent elements

$$1, \quad \gamma_\mu, \quad \gamma_\mu \wedge \gamma_\nu, \quad \gamma_\mu i, \quad i\,. \tag{1.29}$$

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

$$F = \tfrac{1}{2}F^{\mu\nu}\gamma_\mu \wedge \gamma_\nu\,, \tag{1.30}$$

with its "scalar components" given by

$$F^{\mu\nu} = \gamma^\mu \cdot F \cdot \gamma^\nu = \gamma^\nu \cdot (\gamma^\mu \cdot F) = (\gamma^\nu \wedge \gamma^\mu) \cdot F\,. \tag{1.31}$$

However, such an expansion is seldom needed.

Besides the inner and outer products defined above, many other products can be defined in terms of the geometric product. The *commutator product* $A \times B$ is defined for any A and B by

$$A \times B \equiv \tfrac{1}{2}(AB - BA) = -B \times A\,. \tag{1.32}$$

Mathematicians classify this product as a "derivation" with respect to the geometric product, because it has the "distributive property"

$$A \times (BC) = (A \times B)C + B(A \times C)\,. \tag{1.33}$$

This implies the *Jacobi Identity*

$$A \times (B \times C) = (A \times B) \times C + B \times (A \times C)\,, \tag{1.34}$$

which is a derivation on the commutator product. The relation of the commutator product to the inner and outer products is grade dependent; thus, for a vector a,

$$a \times \langle M\rangle_k = a \wedge \langle M\rangle_k \qquad \text{if } k \text{ is odd}\,, \tag{1.35a}$$

$$a \times \langle M \rangle_k = a \cdot \langle M \rangle_k \qquad \text{if } k \text{ is even}\,. \tag{1.35b}$$

The commutator product is especially useful in computations with bivectors. With any bivector A this product is grade preserving:

$$A \times \langle M \rangle_k = \langle A \times M \rangle_k\,. \tag{1.36}$$

In particular, this implies that the space of bivectors is closed under the commutator product. It therefore forms a Lie algebra — which is, in fact, the Lie algebra of the Lorentz group. The geometric product of bivector A with M has the expanded form

$$AM = A \cdot M + A \times M + A \wedge M \qquad \text{for grade } M \geq 2\,. \tag{1.37}$$

This should be compared with the corresponding expansion (1.4) for the product of vectors.

If ψ is an even multivector, then $\psi\widetilde{\psi}$ is also even but its bivector part must vanish according to (1.24), since $(\psi\widetilde{\psi})\,\widetilde{\ } = \psi\widetilde{\psi}$. Therefore, we can write

$$\psi\widetilde{\psi} = \rho e^{i\beta}\,, \tag{1.38a}$$

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

$$R\widetilde{R} = \widetilde{R}R = 1\,. \tag{1.38b}$$

Then ψ can be put in the *canonical form*

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R \tag{1.38c}$$

We shall see that this *invariant decomposition* has a fundamental physical significance in the Dirac Theory.

An important special case of the decomposition (1.38c) is its application to a bivector F, for which it is convenient to write $f = \rho^{\frac{1}{2}} R$. Thus, for any bivector F which is not null ($F^2 \neq 0$) we have the invariant *canonical form*

$$F = fe^{i\varphi} = f(\cos\varphi + i\sin\varphi)\,, \tag{1.39}$$

where $f^2 = |\,f\,|^2$, so f is a "timelike bivector." The dual if is a "spacelike bivector," since $(if)^2 = |\,f\,|^2$. Thus the right side of (1.39) is the unique decomposition of F into a sum of mutually commuting timelike and spacelike parts. Equation (1.39) can be solved directly for φ and f in terms of F, with the results

$$e^{i\varphi} = \frac{\left(F^2\right)^{\frac{1}{2}}}{\left[F^2\left(F^2\right)^{\dagger}\right]^{\frac{1}{4}}} = \frac{(F\cdot F + F\wedge F)^{\frac{1}{2}}}{\left[(F\cdot F)^2 + |\,F\wedge F\,|^2\right]^{\frac{1}{4}}}\,, \tag{1.40a}$$

and

$$f = Fe^{-i\varphi} = \frac{F(F\cdot F - F\wedge F)^{\frac{1}{2}}}{\left[(F\cdot F)^2 + |\,F\wedge F\,|^2\right]^{\frac{1}{4}}}\,, \tag{1.40b}$$

In later sections it will be seen that the decomposition (1.39) has important physical and practical implications, especially when applied to an electromagnetic field.

When $F^2 = 0$, F can still be written in the form (1.39) with

$$f = k \wedge e = ke\,, \tag{1.40}$$

where k is a null vector and e is spacelike. In this case, the decomposition is not unique, and the exponential factor can always be absorbed in the definition of f.

2. Vector Derivatives and Differentials

To extend spacetime algebra into a complete *spacetime calculus*, suitable definitions for derivatives and integrals are required. In terms of the coordinates (1.28), an operator $\nabla \equiv \partial_x$ interpreted as the *derivative* with respect to a spacetime point x can be defined by

$$\nabla = \gamma^\mu \partial_\mu \tag{2.1}$$

where

$$\partial_\mu = \frac{\partial}{\partial x^\mu} = \gamma_\mu \cdot \nabla \,. \tag{2.2}$$

The square of ∇ is the d'Alembertian

$$\nabla^2 = \eta^{\mu\nu} \partial_\mu \partial_\nu \,. \tag{2.3}$$

The matrix representation of (2.1) can be recognized as the "Dirac operator," originally discovered by Dirac by seeking a "square root" of the d'Alembertian (2.3) in order to find a first order "relativistically invariant" wave equation for the electron. In STA however, where the γ^μ are vectors rather than matrices, it is clear that ∇ is a vector operator; indeed, it provides an appropriate definition for the derivative with respect to any spacetime vector variable.

Contrary to the impression given by conventional accounts of relativistic quantum theory, the operator ∇ is not specially adapted to spin-$\frac{1}{2}$ wave equations. 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*

$$\nabla F = J \,. \tag{2.4}$$

Since ∇ is a vector operator the identity (1.7) applies, so that we can write

$$\nabla F = \nabla \cdot F + \nabla \wedge F \,, \tag{2.5}$$

where $\nabla \cdot F$ is the *divergence* of F and $\nabla \wedge F$ is the *curl*. We can accordingly separate (2.4) into vector and trivector parts:

$$\nabla \cdot F = J \,, \tag{2.6}$$

$$\nabla \wedge F = 0 \,. \tag{2.7}$$

As in conventional scalar calculus, the derivatives of elementary functions are often needed for computations. The most useful derivatives are listed here:

Table of vector derivatives: (2.8)

$$\nabla(x \cdot a) = a \cdot \nabla x = a \,, \qquad \nabla(x \cdot A) = A \cdot \nabla x = rA$$

$$\nabla |\, x \,|^2 = \nabla x^2 = 2x \,, \qquad \nabla(x \wedge A) = A \wedge \nabla x = (4 - r)A$$

$$\nabla \wedge x = 0 \,, \qquad \nabla(Ax) = \gamma^\mu A \gamma_\mu = (-1)^r (4 - 2r) A \,,$$

$$\nabla x = \nabla \cdot x = 4$$

$$\nabla|\,x\,|^k = k|\,x\,|^{k-2}x\,, \qquad\qquad \nabla\left(\frac{x}{|\,x\,|^k}\right) = \frac{4-k}{|\,x\,|^k}\,.$$

In the table, $\nabla = \partial_x$, and obvious singularities at $x = 0$ are excluded; a is a "*free vector*" variable (i.e. independent of x), while A is a free r-vector.

The *directional derivative* $a\cdot\nabla$, that is, the "derivative in the direction of vector a" can be obtained from ∇ by applying the inner product. Alternatively, the directional derivative can be defined directly by

$$a\cdot\nabla F = a\cdot\partial_x F(x) = \frac{d}{d\tau}\,F(x+a\tau)\Big|_{\tau=0} = \lim_{\tau\to 0}\frac{F(x+a\tau)-F(x)}{\tau}\,, \tag{2.9}$$

where $F = F(x)$ is any multivector valued function. Then the general vector derivative can be obtained from the directional derivative by using (2.8): thus,

$$\nabla F = \partial_x F(x) = \partial_a a\cdot\partial_x F(x)\,. \tag{2.10}$$

This relation can serve as an alternative to the partial derivative in (2.1) for defining the vector derivative. Of course, the directional derivative has the same limit properties as the partial derivative, including the rules for differentiating sums and products, but the explicit display of the vector variable is advantageous in concept and calculation.

Equation (2.10) and the preceding equations above define the derivative ∂_a with respect to any spacetime vector a. As already indicated, we reserve the symbol "x" for a vector representing a position in spacetime, and we use the special symbol $\nabla \equiv \partial_x$ for the derivative with respect to this "spacetime point." When differentiating with respect to any other vector variable a, we indicate the variable explicitly by writing ∂_a. Mathematically, there is no difference between ∇ and ∂_a. However, there is often an important difference in physical or geometrical interpretation associated with these operators.

The directional derivative (2.9) produces from F a tensor field called the *differential* of F, denoted variously by

$$\underline{F}(a) = F_a \equiv a\cdot\nabla F\,. \tag{2.11}$$

As explained in the next section, the underbar notation serves to emphasize that $\underline{F}(a)$ is a linear function of a, though it is not a linear transformation unless it is vector valued. The argument a may be a free variable or a vector field $a = a(x)$.

The *second differential* $\underline{F}_b(a) = F_{ab}$ is defined by

$$\underline{F}_b(a) \equiv b\cdot\nabla\underline{F}(a) - \underline{F}(b\cdot\nabla a) = b\cdot\grave{\nabla}\grave{\underline{F}}(a)\,, \tag{2.12}$$

where the accents over $\grave{\nabla}$ and $\grave{F}$ serve to indicate that only F is differentiated and a is not. Of course, the accents can be dropped if a is a free variable. The second differential has the symmetry property

$$\underline{F}_b(a) = \underline{F}_a(b)\,. \tag{2.13}$$

Using (1.14) and (1.17), this can be expressed as an operator identity:

$$(a\wedge b)\cdot(\nabla\wedge\nabla) = [\,a\cdot\nabla, b\cdot\nabla\,] = 0\,, \tag{2.14}$$

where the bracket denotes the commutator. Differentation by ∂_a and ∂_b puts this identity in the form

$$\nabla\wedge\nabla = 0\,. \tag{2.15}$$

These last three equations are equivalent formulations of the *integrability condition* for vector derivatives.

Since the derivative ∇ has the algebraic properties of a vector, a large assortment of "differential identities" can be derived by replacing some vector by ∇ in any algebraic identity. The only caveat is to take proper account of the product rule for differentiation. For example, the product rule gives

$$\nabla \cdot (a \wedge b) = \grave{\nabla} \cdot (\grave{a} \wedge b) + \grave{\nabla} \cdot (a \wedge \grave{b}) \,,$$

whence the algebraic identity (1.17) yields

$$a \cdot \nabla b - b \cdot \nabla a = \nabla \cdot (a \wedge b) + a \nabla \cdot b - b \nabla \cdot a \,, \tag{2.16}$$

The left side of this identity may be identified as a Lie bracket; a more general concept of the Lie bracket is introduced later on. Other identities will be derived as needed.

3. Linear Algebra

The computational and representational power of linear algebra is greatly enhanced by integrating it with geometric algebra. In fact, geometric calculus makes it possible to perform coordinate-free computations in linear algebra without resorting to matrices. Integration of the two algebras can be achieved with the few basic concepts, notations and theorems reviewed below. However, linear algebra is a large subject, so we restrict our attention to the essentials needed for gravitation theory.

Though the approach works for vector spaces of any dimension, we will be concerned only with linear transformations of Minkowski space, which map spacetime vectors into vectors. We need a notation which clearly distinguishes linear operators and their products from vectors and their products. To this end, we distinguish symbols representing a linear transformation (or operator) by affixing them with an underbar (or overbar). Thus, for a linear operator $\underline{f}$ acting on a vector a, we write

$$\underline{f}a = \underline{f}(a) \,. \tag{3.1}$$

As usual in linear algebra, the parenthesis around the argument of $\underline{f}$ can be included or omitted, either for emphasis or to remove ambiguity.

Every linear transformation $\underline{f}$ on Minkowski space has a unique extension to a linear function on the whole STA, called the *outermorphism* of f because it preserves outer products. It is convenient to use the same notation $\underline{f}$ for the outermorphism and the operator that "induces" it, distinguishing them when necessary by their arguments. The outermorphism is defined by the property

$$\underline{f}(A \wedge B) = (\underline{f}A) \wedge (\underline{f}B) \tag{3.2}$$

for arbitrary multivectors A, B, and

$$\underline{f}\alpha = \alpha \tag{3.3}$$

for any scalar α. It follows that, for any factoring $A = a_1 \wedge a_2 \wedge \ldots \wedge a_r$ of an r-vector A into vectors,

$$\underline{f}(A) = (\underline{f}a_1) \wedge (\underline{f}a_2) \wedge \ldots \wedge (\underline{f}a_r) \,. \tag{3.4}$$

This relation can be used to compute the outermorphism directly from the inducing linear operator.

Since the outermorphism preserves the outer product, it is grade preserving, that is

$$\underline{f}\langle M \rangle_k = \langle \underline{f}M \rangle_k \tag{3.5}$$

for any multivector M. This implies that $\underline{f}$ alters the pseudoscalar i only by a scalar multiple. Indeed

$$\underline{f}(i) = (\det \underline{f})i \qquad \text{or} \qquad \det \underline{f} = -i\underline{f}(i) \,, \tag{3.6}$$

which defines the *determinant* of $\underline{f}$. Note that the outermorphism makes it possible to define (and evaluate) the determinant without introducing a basis or matrices.

The "product" of two linear transformations, expressed by

$$\underline{h} = \underline{g}\,\underline{f}\,, \tag{3.7}$$

applies also to their outermorphisms. In other words, the outermorphism of a product equals the product of outermorphisms. It follows immediately from (3.6) that

$$\det(\underline{g}\,\underline{f}) = (\det \underline{g})(\det \underline{f})\,, \tag{3.8}$$

from which many other properties of the determinant follow, such as

$$\det(\underline{f}^{-1}) = (\det \underline{f})^{-1} \tag{3.9}$$

whenever $\underline{f}^{-1}$ exists.

Every linear transformation $\underline{f}$ has an *adjoint* transformation $\bar{f}$ which can be extended to an outermorphism denoted by the same symbol. The adjoint outermorphism can be defined in terms of $\underline{f}$ by

$$\langle M \bar{f} N\rangle = \langle N \underline{f} M\rangle\,, \tag{3.10}$$

where M and N are arbitrary multivectors and the bracket as usual indicates "scalar part." For vectors a, b this can be written

$$b \cdot \bar{f}(a) = a \cdot \underline{f}(b)\,. \tag{3.11}$$

Differentiating with respect to b we obtain, with the help of (2.8),

$$\bar{f}(a) = \partial_b\, a \cdot \underline{f}(b)\,. \tag{3.12}$$

This is the most useful formula for obtaining $\bar{f}$ from $\underline{f}$. Indeed, it might well be taken as the preferred definition of $\bar{f}$.

Unlike the outer product, the inner product is not generally preserved by outermorphisms. However, it obeys the fundamental transformation law

$$\bar{f}[\,\underline{f}(A) \cdot B\,] = A \cdot \bar{f}(B) \tag{3.13}$$

for (grade A) $\leq$ (grade B). Of course, this law also holds with an interchange of overbar and underbar. If $\underline{f}$ is invertible, it can be written in the form

$$\bar{f}[\,A \cdot B\,] = \underline{f}^{-1}(A) \cdot \bar{f}(B)\,. \tag{3.14}$$

For $B = i$, since $A \cdot i = Ai$, this immediately gives the general formula for the inverse outermorphism:

$$\underline{f}^{-1}A = [\,\bar{f}(Ai)\,](\bar{f}i)^{-1} = (\det \underline{f})^{-1}\bar{f}(Ai)i^{-1}\,. \tag{3.15}$$

This relation shows explicitly the double duality involved in computing the inverse.

Although all linear transformations preserve the outer product (by definition of class preserves the inner product. This is called the *Lorentz group*, and its members are called *Lorentz transformations*. The defining property for a Lorentz transformation $\underline{L}$ is

$$(\underline{L}a) \cdot (\underline{L}b) = a \cdot (\bar{\underline{L}}\,\underline{L}b) = a \cdot b\,. \tag{3.16}$$

This is equivalent to the operator condition

$$\overline{\underline{L}} = \underline{L}^{-1} \tag{3.17}$$

STA makes it possible to express any $\underline{L}$ in the simple *canonical form*

$$\underline{L}(a) = \epsilon L a L^{-1}\,, \tag{3.18}$$

where the multivector L is either *even* with $\epsilon = 1$ or *odd* with $\epsilon = -1$. This defines a double-valued homomorphism between Lorentz transformations $\{\underline{L}\}$ and multivectors $\{\pm L\}$, where the composition of two Lorentz transformations $\underline{L}_1\underline{L}_2$ corresponds to the geometric product $\pm L_1 L_2$. Thus, the Lorentz group has a double-valued representation as a multiplicative group of multivectors. This *multivector representation* of the Lorentz group greatly facilitates the analysis and application of Lorentz transformations in STA.

From (3.18) it follows immediately that, for arbitrary multivectors A and B,

$$\underline{L}(AB) = (\underline{L}A)(\underline{L}B)\,. \tag{3.19}$$

Lorentz transformations therefore preserve the geometric product. This implies that (3.16) generalizes to

$$\underline{L}(A \cdot B) = (\underline{L}A) \cdot (\underline{L}B)\,. \tag{3.20}$$

in agreement with (3.14) when (3.17) is satisfied.

From (3.18) it follows easily that

$$\underline{L}(i) = \epsilon i\,, \qquad \text{where} \qquad \epsilon = \det \underline{L} = \pm 1\,. \tag{3.21}$$

A Lorentz transformation $\underline{L}$ is said to be *proper* if $\epsilon = 1$, and *improper* if $\epsilon = -1$. It is said to be *orthochronous* if, for any timelike vector v,

$$v \cdot \underline{L}(v) > 0\,. \tag{3.22}$$

A *proper, orthochronous* Lorentz transformation is called a *Lorentz rotation* (or a *restricted* Lorentz transformation). For a Lorentz rotation $\underline{R}$ the canonical form can be written

$$\underline{R}(a) = R a \widetilde{R}\,, \tag{3.23}$$

where the even multivector R is called a *rotor* and is normalized by the condition

$$R\widetilde{R} = 1\,. \tag{3.24}$$

The rotors form a multiplicative group called the *Rotor group*, which is a double-valued representation of the Lorentz rotation group (also called the restricted Lorentz group).

The most elementary kind of Lorentz transformation is a *reflection* $\underline{n}$ by a (non-null) vector n, according to

$$\underline{n}(a) = -nan^{-1}\,. \tag{3.25}$$

This is a reflection with respect to a hyperplane with normal n. A reflection

$$\underline{v}(a) = -vav \tag{3.26}$$

with respect to a timelike vector $v = v^{-1}$ is called a *time reflection*. Let n_1, n_2, n_3 be spacelike vectors which compose the trivector

$$n_3 n_2 n_1 = iv\,. \tag{3.27}$$

A *space inversion* $\underline{v}_s$ can then be defined as the composite of reflections with respect to these three vectors, so it can be written

$$\underline{v}_s(a) = n_3 n_2 n_1 a n_1 n_2 n_3 = ivavi = vav\,. \tag{3.28}$$

Note the difference in sign between the right sides of (3.26) and (3.28). Although $\underline{v}_s$ is determined by v alone on the right side of (3.28), the multivector representation of $\underline{v}_s$ must be the trivector iv in order to agree with (3.18). The composite of the time reflection (3.26) with the space inversion (3.28) is the *spacetime inversion*

$$\underline{v}_{st}(a) = \underline{v}_s v(a) = -iai^{-1} = -a\,, \tag{3.29}$$

which is represented by the pseudoscalar i. Note that spacetime inversion is proper but not orthochronous.

Two basic types of Lorentz rotation can be obtained from the product of two reflections, namely *timelike rotations* (or *boosts*) and *spacelike rotations*. For a *boost*

$$\underline{V}(a) = Va\widetilde{V}\,, \tag{3.30}$$

the rotor V can be factored into a product

$$V = v_2 v_1 \tag{3.31}$$

of two unit timelike vectors v_1 and v_2. The boost is a rotation in the timelike plane containing v_1 and v_2. The factorization (3.31) is not unique. Indeed, for a given V any timelike vector in the plane can be chosen as v_1, and v_2 then computed from (3.31). Similarly, for a *spacelike rotation*

$$\underline{Q}(a) = Qa\widetilde{Q}\,, \tag{3.32}$$

the rotor Q can be factored into a product

$$Q = n_2 n_1 \tag{3.33}$$

of two unit spacelike vectors in the *spacelike plane* of the rotation. Note that the product, say $n_2 v_1$, of a spacelike vector with a timelike vector is not a rotor, because the corresponding Lorentz transformation is not orthochronous. Likewise, the pseudoscalar i is not a rotor, even though it can be expressed as the product of two pairs of vectors, for it does not satisfy the rotor condition (3.24).

Any Lorentz rotation $\underline{R}$ can be decomposed into the product

$$\underline{R} = \underline{V}\,\underline{Q} \tag{3.34}$$

of a boost $\underline{V}$ and spacelike rotation $\underline{Q}$ with respect to a given timelike vector $v_0 = v_0^{-1}$. To compute $\underline{V}$ and $\underline{Q}$ from $\underline{R}$, first compute the vector

$$v = \underline{R}v_0 = Rv_0\widetilde{R}\,. \tag{3.35}$$

the timelike vectors v and v_0 determine the timelike plane for the boost, which can therefore be defined by

$$v = \underline{V}v_0 = Vv_0\widetilde{V} = V^2 v_0\,. \tag{3.36}$$

This can be solved for

$$V = \big(vv_0\big)^{\frac{1}{2}} = vw = wv_0\,, \tag{3.37a}$$

where the unit vector

$$w = \frac{v + v_0}{|\, v + v_0 \,|} = \frac{v + v_0}{\Big[\, 2(1 + v \cdot v_0) \,\Big]^{\frac{1}{2}}} \tag{3.37b}$$

"bisects the angle" between v and v_0. The rotor Q can then be computed from

$$Q = \widetilde{V} R \,, \tag{3.38}$$

so that the spacelike rotation satisfies

$$\underline{Q} v_0 = Q v_0 \widetilde{Q} = v_0 \,. \tag{3.39}$$

This makes (3.36) consistent with (3.35) by virtue of (3.34).

Equations (3.31) and (3.32) show how to parametrize boosts and spacelike rotations by vectors in the plane of rotation. More specifically, (3.37a,b) parametrizes a boost in terms of initial and final velocity vectors. This is especially useful, because the velocity vectors are often given, or are of direct relevance, in a physical problem. Another useful parametrization is in terms of angle (Appendix B of [2]). Any rotor R can be expressed in the exponential form

$$\pm R = e^{\frac{1}{2}F} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\tfrac{1}{2} F\right)^k \,, \tag{3.40}$$

where F is a bivector parametrizing the rotation. The positive sign can always be selected when $F^2 \neq 0$, and, according to (1.39), F can be written in the canonical form

$$F = (\alpha + i\beta) f \qquad \text{where} \qquad f^2 = 1, \tag{3.41}$$

α and β being scalar parameters. Since the timelike unit bivector f commutes with its dual if, which is a spacelike bivector, the rotor R can be decomposed into a product of commuting timelike and spacelike rotors. Thus

$$R = VQ = QV \,, \tag{3.42}$$

where

$$V = e^{\frac{1}{2}\alpha f} = \cosh \tfrac{1}{2}\alpha + f \sinh \tfrac{1}{2}\alpha \,, \tag{3.43}$$

and

$$Q = e^{\frac{1}{2} i\beta f} = \cos \tfrac{1}{2}\beta + if \sin \tfrac{1}{2}\beta \,. \tag{3.44}$$

The parameter α is commonly called the *rapidity* of the boost. The parameter β is the usual angular measure of a spatial rotation.

When $F^2 = 0$, equation (3.40) can be reduced to the form

$$R = e^{\frac{1}{2}\alpha f} = 1 + \tfrac{1}{2}\alpha f \,, \tag{3.45}$$

where f is a null bivector, and it can be expressed in the factored form (1.40). The two signs are inequivalent cases. There is no choice of null F which can eliminate the minus sign. The *lightlike rotor* in (3.45) represents a *lightlike Lorentz rotation*.

The spacelike rotations that preserve a timelike vector v_0 are commonly called *spatial rotations* without mentioning the proviso (3.38). The set of such rotations is the 3-parameter *spatial rotation group* (of v_0). More generally, for any given vector n, the subgroup of Lorentz transformations $\underline{N}$ satisfying

$$\underline{N}(n) = N n \widetilde{N} = n \tag{3.46}$$

is called the *little group* of n. The little group of a lightlike vector can be parametrized as a lightlike rotor (3.40) composed with timelike and spacelike rotors.

The above classification of Lorentz transformations can be extended to more general linear transformations. For any linear transformation $\underline{f}$ the composite $\bar{\underline{f}}\underline{f}$ is a symmetric transformation. If the latter has a well defined square root $\underline{S} = (\bar{\underline{f}}\underline{f})^{\frac{1}{2}} = \overline{S}$, then $\underline{f}$ admits to the *polar decomposition*

$$\underline{f} = \underline{R}\underline{S} = \underline{S}'\underline{R}\,, \tag{3.47}$$

where $\underline{R}$ is a Lorentz rotation and $\underline{S}' = \underline{R}\underline{S}\underline{R}^{-1}$. Symmetric transformations are, of course, defined by the condition

$$\overline{S} = \underline{S}\,. \tag{3.48}$$

On Euclidean spaces every linear transformation has a polar decomposition, but on Minkowski space there are symmetric transformations (with null eigenvectors) which do not possess square roots and so do not admit a polar decomposition. A complete classification of symmetric transformations is given in the next section.

4. Spacetime Splits.

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

$$x\gamma_0 = t + \mathbf{x}\,, \tag{4.1a}$$

where

$$t = x \cdot \gamma_0 \tag{4.1b}$$

and

$$\mathbf{x} = x \wedge \gamma_0\,. \tag{4.1c}$$

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 (4.1b) assigns a unique time t to every event x; indeed, (4.1b) is the equation for a one parameter family of spacelike hyperplanes with normal γ_0.

Equation (4.1c) assigns to each event x a unique *position vector* $\mathbf{x}$ in the γ_0 system. Thus, to each event x the single equation (4.1a) assigns a unique time and position in the γ_0-system. Note that the reverse of (4.1a) is

$$\gamma_0 x = \gamma_0 \cdot x + \gamma_0 \wedge x = t - \mathbf{x}\,,$$

so, since $\gamma_0^2 = 1$,

$$x^2 = (x\gamma_0)(\gamma_0 x) = (t - \mathbf{x})(t + \mathbf{x}) = t^2 - \mathbf{x}^2\,. \tag{4.2}$$

The form and value of this equation is independent of the chosen observer; thus we have proved that the expression $t^2 - \mathbf{x}^2$ is *Lorentz invariant* without even mentioning a Lorentz transformation. Henceforth, we use the term Lorentz invariant to mean independent of a chosen spacetime split. In contrast to (4.2), equation (4.1a) is not Lorentz invariant; indeed, for a different observer γ_0' we get the split

$$x\gamma_0' = t' + \mathbf{x}' \,. \tag{4.3}$$

Mostly we shall work with *manifestly Lorentz invariant* equations, which are independent of even an indirect reference to an inertial system.

The set of all position vectors (4.1c) 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

$$\boldsymbol{\sigma}_k = \gamma_k \wedge \gamma_0 = \gamma_k\gamma_0 \,. \tag{4.4a}$$

These generate a basis for the relative bivectors:

$$\boldsymbol{\sigma}_i \wedge \boldsymbol{\sigma}_j = \boldsymbol{\sigma}_i\boldsymbol{\sigma}_j = i\boldsymbol{\sigma}_k = \gamma_j\gamma_i \,, \tag{4.4b}$$

where the allowed values of the indices $\{i, j, k\}$ are cyclic permutations of 1,2,3. The right sides of (4.4a) and (4.4b) 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,

$$\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2\boldsymbol{\sigma}_3 = i = \gamma_0\gamma_1\gamma_2\gamma_3 \,. \tag{4.4c}$$

The geometrical operation of *reversion* in the algebra of $\mathcal{P}^3$ can be defined by

$$\boldsymbol{\sigma}_k^\dagger = \boldsymbol{\sigma}_k \qquad \text{and} \qquad (\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j)^\dagger = \boldsymbol{\sigma}_j\boldsymbol{\sigma}_i \,. \tag{4.4d}$$

For an arbitrary multivector M, this is related to reversion (1.24) in the entire STA by

$$M^\dagger \equiv \gamma_0\widetilde{M}\gamma_0 \,. \tag{4.5}$$

The explicit appearance of the timelike vector γ_0 here shows the dependence of $M^\dagger$ on a particular spacetime split.

Now let us rapidly survey the space-time splits of some important physical quantities. 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 (4.1a); whence

$$v\gamma_0 = v_0(1 + \mathbf{v}) \,, \tag{4.6a}$$

where

$$v_0 = v \cdot \gamma_0 = \frac{dt}{d\tau} = \left(1 - \mathbf{v}^2\right)^{-\frac{1}{2}} \tag{4.6b}$$

is the "time dilation" factor, and

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} = \frac{d\tau}{dt}\frac{d\mathbf{x}}{d\tau} = \frac{v \wedge \gamma_0}{v \cdot \gamma_0} \tag{4.6c}$$

is the *relative velocity* in the γ_0-system. The last equality in (4.6b) was obtained from

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

Let p be the *proper momentum* (i.e. energy-momentum vector) of a particle. The space-time split of p into *energy* (or relative mass) E and *relative momentum* $\mathbf{p}$ is given by

$$p\gamma_0 = E + \mathbf{p}\,, \tag{4.7a}$$

where

$$E = p\cdot\gamma_0\,, \tag{4.7b}$$

$$\mathbf{p} = p\wedge\gamma_0\,. \tag{4.7c}$$

Of course

$$p^2 = (E+\mathbf{p})(E-\mathbf{p}) = E^2 - p^2 = m^2\,, \tag{4.8}$$

where m is the *proper mass* of the particle.

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

$$F = \mathbf{E} + i\mathbf{B}\,, \tag{4.9a}$$

where

$$\mathbf{E} = (F\cdot\gamma_0)\gamma_0 = \tfrac{1}{2}(F + F^\dagger)\,, \tag{4.9b}$$

$$i\mathbf{B} = (F\wedge\gamma_0)\gamma_0 = \tfrac{1}{2}(F - F^\dagger)\,, \tag{4.9c}$$

and, in accordance with (4.5), $F^\dagger = \mathbf{E} - i\mathbf{B}$. Equation (4.9a) 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, (4.9a) 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" [3,4].

Now consider the relation of the split (4.9a) to the invariant decomposition $F = fe^{i\beta}$ given by (1.39). If γ_0 lies in the plane of the timelike bivector f, as expressed by $f\wedge\gamma_0 = 0$ so $f = (f\cdot\gamma_0)\gamma_0$, then (4.9a, b) gives us $\mathbf{E} = f\cos\varphi$ and $\mathbf{B} = f\sin\varphi$. Thus, to any such observer F consists of parallel electric and magnetic fields, and these fields are the same for all such observers. Consequently, we can interpret $F = fe^{i\varphi}$ physically as a Lorentz invariant decomposition of the field F into parallel (commuting) electric and magnetic parts. The decomposition is invariant, because it is independent of any observer; it characterizes an intrinsic structural property of the field.

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

$$\mathbf{EB} = \mathbf{E}\cdot\mathbf{B} = i(\mathbf{E}\times\mathbf{B})\,, \tag{4.10a}$$

where

$$\mathbf{E}\cdot\mathbf{B} = \tfrac{1}{2}(\mathbf{EB} + \mathbf{BE}) \tag{4.10b}$$

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

$$\mathbf{E}\times\mathbf{B} = \frac{1}{2i}(\mathbf{EB} - \mathbf{BE}) = i^{-1}(\mathbf{E}\wedge\mathbf{B}) \tag{4.10c}$$

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 (4.10) of the dot and cross products into the single geometric product simplifies many aspects of classical nonrelativistic physics, as demonstrated at length in Ref. l6.

The cross product (4.10c) is commonly used to represent relative angular momentum. To relate that to proper angular momentum, consider a particle with proper momentum p at a spacetime point x. Employing the splits (4.1a) and (4.7a) we find

$$px = (E + \mathbf{p})(t - \mathbf{x}) = Et + \mathbf{p}t - E\mathbf{x} - \mathbf{p}\mathbf{x}\,.$$

The scalar part of this gives the familiar split

$$p \cdot x = Et - \mathbf{p} \cdot \mathbf{x}\,, \tag{4.11}$$

so often employed in the phase of a wave function. The bivector part gives us the proper angular momentum

$$p \wedge x = \mathbf{p}t - E\mathbf{x} - i(\mathbf{x}\times\mathbf{p})\,, \tag{4.12}$$

where (4.10c) has been used. Note that the split of (4.12) into relative vector and bivector parts corresponds exactly to the split (4.9a) of the electromagnetic field into vector and bivector parts.

Our final application of the space-time split is to a Lorentz rotation. In this case, the split is done differently than in the previous examples. The Lorentz rotation (3.23) transforms a standard frame $\{\gamma_\mu\}$ into a new frame of vectors $\{e_\mu\}$ given by

$$e_\mu = R\gamma_\mu\widetilde{R}\,. \tag{4.13}$$

A *space-time split* of the Lorentz rotation (4.13) by γ_0 is accomplished by a split of the rotor R into the product

$$R = LU\,, \tag{4.14}$$

where $U^\dagger = \gamma_0\widetilde{U}\gamma_0 = \widetilde{U}$ or

$$U\gamma_0\widetilde{U} = \gamma_0\,, \tag{4.15}$$

and $L^\dagger = \gamma_0\widetilde{L}\gamma_0 = L$ or

$$\gamma_0\widetilde{L} = L\gamma_0\,. \tag{4.16}$$

This determines a split of (4.13) into a sequence of two Lorentz rotations determined by U and L respectively; thus,

$$e_\mu = R\gamma_\mu\widetilde{R} = L(U\gamma_\mu\widetilde{U})\widetilde{L}\,. \tag{4.17}$$

In particular, by (4.15) and (4.16),

$$e_0 = R\gamma_0\widetilde{R} = L\gamma_0\widetilde{L} = L^2\gamma_0\,. \tag{4.18}$$

Hence,

$$L^2 = e_0\gamma_0\,. \tag{4.19}$$

This determines L uniquely in terms of the timelike vectors e_0 and γ_0, which, in turn, uniquely determines the split (4.14) of R, since U can be computed from $U = \widetilde{L}R$. Note that the split (4.14) is a special case of the decomposition (3.34).

Equation (4.15) for variable U defines the *"little group"* of Lorentz rotations which leave γ_0 invariant; This is the group of *"spatial rotations"* in the γ_0-system. Each such rotation takes a frame of proper vectors γ_k (for $k = 1, 2, 3$) into a new frame of vectors $U\gamma_k\widetilde{U}$ in the γ_0-system.

Multiplication by γ_0 expresses this as a rotation of relative vectors $\boldsymbol{\sigma}_k = \gamma_k\gamma_0$ into relative vectors $\mathbf{e}_k$; thus,

$$\mathbf{e}_k = U\boldsymbol{\sigma}_k U^\dagger = U\boldsymbol{\sigma}_k\widetilde{U}\,. \tag{4.20}$$

From (4.12) it follows that U can be parametrized in the exponential form

$$U = e^{-\frac{1}{2}i\mathbf{a}}\,, \tag{4.21}$$

where $\mathbf{a}$ is a relative vector specifying the axis and angle of rotation. This approach to spatial rotations is treated exhaustively with many applications to mechanics in Ref. 5.

Since (4.18) has the same form as (3.36), it can be solved for L in the form of (3.37a, b). If $e_0 = v$ is the proper velocity of a particle of mass m, then (4.6a) and (4.7a) enable us to write (4.19) in the alternative forms

$$L^2 = v\gamma_0 = \frac{p\gamma_0}{m} = \frac{E+\mathbf{p}}{m}\,, \tag{4.22}$$

so (3.37a, b) gives

$$L = (v\gamma_0)^2 = \frac{1+v\gamma_0}{\big[2(1+v\cdot\gamma_0)\big]^{\frac{1}{2}}} = \frac{m+p\gamma_0}{\big[2m(m+p\cdot\gamma_0)\big]^{\frac{1}{2}}} = \frac{m+E+\mathbf{p}}{\big[2m(m+E)\big]^{\frac{1}{2}}}\,. \tag{4.23}$$

Then L represents a boost of a particle from rest in the γ_0-system to a relative momentum $\mathbf{p}$.

5. Rigid Bodies and Charged Particles.

The equation

$$e_\mu = R\gamma_\mu\widetilde{R} \tag{5.1}$$

can be used to describe the relativistic kinematics of a rigid body (with negligible dimensions) traversing a world line $x = x(\tau)$ with proper time τ, provided we identify e_0 with the proper velocity v of the body, so that

$$\frac{dx}{d\tau} = \dot{x} = v = e_0 = R\gamma_0\widetilde{R}\,. \tag{5.2}$$

Then $\{e_\mu = e_\mu(\tau); \mu = 0,1,2,3\}$ is a *comoving frame* traversing the world line along with the particle, and the spinor R must also be a function of proper time, so that, at each time τ, equation (5.1) describes a Lorentz rotation of some arbitrarily chosen fixed frame $\{\gamma_\mu\}$ into the comoving frame $\{e_\mu = e_\mu(\tau)\}$. Thus, we have a rotor-valued function of proper time $R = R(\tau)$ determining a 1-parameter family of Lorentz rotations $e_\mu(\tau) = R(\tau)\gamma_\mu(\tau)$.

The spacelike vectors $e_k = R\gamma_k\widetilde{R}$ (for $k = 1,2,3$) can be identified with the principle axes of the body. But the same equations can be used for modeling a particle with an intrinsic angular momentum or *spin*, where e_3 is identified with the spin direction $\hat{s}$; so we write

$$\hat{s} = e_3 = R\gamma_3\widetilde{R}\,. \tag{5.3}$$

Later it will be seen that this corresponds exactly to the spin vector in the Dirac theory where the magnitude of the spin has the constant value $|\,s\,| = \hbar/2$.

The *rotor equation of motion for* $R = R(\tau)$ has the form

$$\dot{R} = \tfrac{1}{2}\Omega R\,, \tag{5.4}$$

where $\Omega = \Omega(\tau)$ is a bivector-valued function. The fact that $\Omega = 2\dot{R}\widetilde{R} = -\widetilde{\Omega}$ is necessarily a bivector is easily proved by differentiating $R\widetilde{R} = 1$. Differentiating (5.1) and using (5.4), we see that the equations of motion for the comoving frame have the form

$$\dot{e}_\mu = \Omega \cdot e_\mu \,. \tag{5.5}$$

Clearly Ω can be interpreted as a generalized *rotational velocity* of the comoving frame.

The dynamics of the rigid body, that is, the action of external forces and torques on the body is completely characterized by specifying Ω as a definite function of proper time. The single rotor equation (5.4) is equivalent to the set of four frame equations (5.5). Besides the theoretical advantage of being closely related to the Dirac equation, it has the practical advantage of being simpler and easier to solve. The corresponding nonrelativistic rotor equation for a spinning top has been analyzed at length in Ref. 4. It should be noted that nonrelativistic rotor equation describes only rotational motion, while its relativistic generalization (5.4) describes rotational and translational motion together.

For a classical particle with mass m and charge e in an electromagnetic field F, the dynamics is specified by

$$\Omega = \frac{e}{m} F \,.$$

So (5.5) gives the particle equation of motion

$$m\dot{v} = eF \cdot v \,. \tag{5.6}$$

This will be recognized as the classical Lorentz force with tensor components $m\dot{v}^\mu = eF^{\mu\nu}v_\nu$. If self-interaction is neglected, it describes the motion of a "test charge" in an external field F.

Though (5.6) can be solved directly, it is usually much simpler to solve the rotor equation (5.4). For example, if F is a uniform field on spacetime, then $\dot{\Omega} = 0$ and (5.4) has the solution

$$R = e^{\frac{1}{2}\Omega\tau} R_0 \,, \tag{5.7}$$

where $R_0 = R(0)$ specifies the initial conditions. When this is substituted into (5.7) we get the explicit τ dependence of the proper velocity v. The integration of (5.2) for the history $x(t)$ is most simply accomplished in the general case of arbitrary non-null F by exploiting the invariant decomposition $F = fe^{i\varphi}$ given by (1.22). This separates Ω into commuting parts $\Omega_1 = f(e/m)\cos\varphi$ and $\Omega_2 = f(e/m)\sin\varphi$, so

$$e^{\frac{1}{2}\Omega\tau} = e^{\frac{1}{2}(\Omega_1+\Omega_2)\tau} = e^{\frac{1}{2}\Omega_1\tau} e^{\frac{1}{2}\Omega_2\tau} \,. \tag{5.8}$$

It also determines an invariant decomposition of the initial velocity $v(0)$ into a component v_1 in the f-plane and a component v_2 orthogonal to the f-plane; thus,

$$v(0) = f^{-1}(f \cdot v(0)) + f^{-1}(f \wedge v(0)) = v_1 + v_2 \,. \tag{5.9}$$

When this is substituted in (5.2) and (5.8) is used, we get

$$\frac{dx}{d\tau} = v = e^{\frac{1}{2}\Omega_1\tau} v_1 + e^{\frac{1}{2}\Omega_2\tau} v_2 \,. \tag{5.10}$$

Note that this is an invariant decomposition of the motion into "electriclike" and "magneticlike" components. It integrates easily to give the history

$$x(\tau) - x(0) = (e^{\frac{1}{2}\Omega_1\tau} - 1)\,\Omega_1^{-1} v_1 + e^{\frac{1}{2}\Omega_2\tau}\Omega_2^{-1} v_2 \,. \tag{5.11}$$

This general result, which applies for arbitrary initial conditions and arbitrary uniform electric and magnetic fields, has such a simple form because it is expressed in terms of invariants. It looks far more complicated when subjected to a space-time split and expressed directly as a function of "laboratory fields" in an inertial system. Details are given in Ref. 6.

As a second example with important applications, we integrate the rotor equation for a "classical test charge" in an electromagnetic plane wave. Any plane wave field $F = F(x)$ with proper propagation vector k can be written in the canonical form

$$F = fz\,, \tag{5.12a}$$

where f is a constant null bivector ($f^2 = 0$), and the x-dependence of F is exhibited explicitly by

$$z(k\cdot x) = \alpha_+ e^{i(k\cdot x)} + \alpha_- e^{-i(k\cdot x)}\,, \tag{5.12b}$$

with

$$\alpha_\pm = \rho_\pm e^{\pm i\delta_\pm}\,, \tag{5.12c}$$

where $\alpha_\pm$ and $\rho_\pm \geq 0$ are scalars. It is crucial to note that the "imaginary" i here is the unit pseudoscalar, because it endows these solutions with geometrical properties not possessed by conventional "complex solutions." Indeed, as explained in Ref. 3, the pseudoscalar property of i implies that the two terms on the right side of (5.12b) describe right and left circular polarizations. Thus, the orientation of i determines handedness of the solutions.

For the plane wave (5.12a,b,c), Maxwell's equation (2.4) reduces to the algebraic condition,

$$kf = 0\,. \tag{5.13}$$

This implies $k^2 = 0$ as well as $f^2 = 0$. To integrate the rotor equation of motion

$$\dot{R} = \frac{e}{2m}\,FR\,, \tag{5.14}$$

it is necessary to express F as a function of τ. This can be done by using special properties of F to find constants of motion. Multiplying (5.14) by k and using (5.13) we find immediately that kR is a constant of the motion. So, with the initial condition $R(0) = 1$, we obtain $k = kR = Rk = k\widetilde{R}$; whence

$$Rk\widetilde{R} = k\,. \tag{5.15}$$

Thus, the one parameter family of Lorentz rotations represented by $R = R(\tau)$ lies in the little group of the lightlike vector k. Multiplying (5.15) by (5.1), we find the constants of motion $k\cdot e_\mu = k\cdot\gamma_\mu$. This includes the constant

$$\omega = k\cdot v\,, \tag{5.16}$$

which can be interpreted as the frequency of the plane wave "seen by the particle." Since $v = dx/d\tau$, we can integrate (5.16) immediately to get

$$k\cdot(x(\tau) - x(0)) = \omega\tau\,. \tag{5.17}$$

Inserting this into (5.12b) and absorbing $k\cdot x(0)$ in the phase factor, we get $z(k\cdot x) = z(\omega\tau)$, expressing the desired τ dependence of F. Equation (5.14) can now be integrated directly, with the result

$$R = \exp\left(efz_1/2m\right) = 1 + \frac{e}{2m}\,fz_1\,, \tag{5.18a}$$

where

$$z_1 = \frac{2}{\omega}\,\sinh\left(\omega\tau/2\right)\left[\alpha_+ e^{i\omega\tau/2} + \alpha_- e^{-i\omega\tau/2}\right]. \tag{5.18b}$$

This gives the velocity v and, by integrating (5.2), the complete particle history. Again, details are given in Ref. 6, where the problem of motion in a Coulomb field is also solved by the same method.

We have established that specification of kinematics by the rotor equation (5.4) and dynamics by $\Omega = (e/m)F$ is a geometrically perspicuous and analytically efficient means of characterizing the motion of a classical charged particle. But it is more than that! It also provides us automatically with a classical model of spin precession simply by assuming that the particle has an intrinsic spin characterized by (5.3). Moreover, any dynamics of spin precession can be characterized by specifying a functional form for Ω. This includes gravitational precession [7] and electron spin precession in the Dirac theory. To facilitate the analysis for any given dynamical model, we shall first carry the analysis as far as possible for arbitrary Ω. Then we shall give a specific application to measurement of the g-factor for a Dirac particle.

The rotor equation of motion (5.4) determines both translational and rotational motions of the comoving frame (5.1), whatever the frame models physically. It is of interest to separate translational and rotational modes, though they are generally coupled. This can be done by a space-time split by the particle velocity v or by the reference vector γ_0. We shall consider both ways and how they are related.

To split the rotational velocity Ω by the velocity v, we write

$$\Omega = \Omega v^2 = (\Omega \cdot v)v + (\Omega \wedge v)v \,.$$

This produces the split

$$\Omega = \Omega_+ + \Omega_- \,, \tag{5.19a}$$

where

$$\Omega_+ = \tfrac{1}{2}(\Omega + v\widetilde{\Omega}v) = (\Omega \cdot v)v = \dot{v}v \,, \tag{5.19b}$$

and

$$\Omega_- = \tfrac{1}{2}(\Omega - v\widetilde{\Omega}v) = (\Omega \wedge v)v \,. \tag{5.19c}$$

Note that $\Omega \cdot v = \dot{v}$ was used in (5.19b) to express Ω_+ entirely in terms of the proper acceleration $\dot{v}$ and velocity v. This split has exactly the same form as the split (2.9 a,b,c) of the electromagnetic bivector into electric and magnetic parts corresponding here to Ω_+ and Ω_- respectively. However, it is a split with respect the instantaneous "rest frame" of the particle rather than a fixed inertial frame. In the rest frame the relative velocity of the particle itself vanishes, of course, so the particle's acceleration is entirely determined by the "electriclike force" Ω_+, as (5.19b) shows explicitly. The "magneticlike" part Ω_- is completely independent of the particle motion; it is the *Larmor Precession* (frequency) of the spin for a particle with a magnetic moment, so let us refer to it as the Larmor precession in the general case.

Unfortunately, (5.19a) does not completely decouple precession form translation because Ω_+ contributes to both. Also, we need a way to compare precessions at different points on the particle history. These difficulties can be resolved by adopting the γ_0-split

$$R = LU \,, \tag{5.20}$$

exactly as defined by (2.15) and subsequent equations. At every time τ, this split determines a "deboost" of relative vectors $e_k e_0 = R\gamma_k\gamma_0\widetilde{R} = R\boldsymbol{\sigma}_k\widetilde{R}$ $(k = 1, 2, 3)$ into relative vectors

$$\mathbf{e}_k = \widetilde{L}(e_k e_0)L = U\boldsymbol{\sigma}_k\widetilde{U} \tag{5.21}$$

in the fixed reference system of γ_0. The particle is brought to rest, so to speak, so we can watch it precess (or spin) in one place. The precession is described by an equation of the form

$$\frac{dU}{dt} = -\tfrac{1}{2}i\boldsymbol{\omega}U \,, \tag{5.22}$$

so diffentiation of (5.21) and use of (2.10c) to define the cross product yields the familiar equations for a rotating frame

$$\frac{d\mathbf{e}_k}{dt} = \boldsymbol{\omega}\times\mathbf{e}_k\,. \tag{5.23}$$

The problem now is to express $\boldsymbol{\omega}$ in terms of the given Ω and determine the relative contributions of the parts Ω_+ and Ω_-. To do that, we use the time dilation factor $v_0 = v\cdot\gamma_0 = dt/d\tau$ to change the time variable in (5.22) and write

$$\omega = -i\boldsymbol{\omega}v_0 \tag{5.24}$$

so (5.22) becomes $\dot{U} = \frac{1}{2}\omega U$. Then differentiation of (5.20) and use of (5.4) gives

$$\Omega = 2\dot{R}R = 2\dot{L}\widetilde{L} + L\omega\widetilde{L}\,.$$

Solving for ω and using the split (5.19a), we get

$$\omega = \widetilde{L}\Omega_- L + \widetilde{L}\dot{v}vL - 2\dot{L}\widetilde{L}\,.$$

Differentiation of (2.18) leads to

$$\widetilde{L}(\dot{v}v)L = \widetilde{L}\dot{L} + \dot{L}\widetilde{L}\,, \tag{5.25}$$

while differentiation of (2.22) gives

$$2\dot{L}\widetilde{L} = \frac{\dot{v}\wedge(v+\gamma_0)}{1+v\cdot\gamma_0}\,. \tag{5.26}$$

These terms combine to give the well-known *Thomas precession* frequency

$$\omega_T = \big((2\dot{L}\widetilde{L})\wedge\gamma_0\big)\gamma_0 = \dot{L}\widetilde{L} - \widetilde{L}\dot{L} = \frac{(\dot{v}\wedge v\wedge\gamma_0)\gamma_0}{1+v\cdot\gamma_0} = i\left(\frac{v_0^2}{1+v_0}\right)\mathbf{v}\times\dot{\mathbf{v}}\,. \tag{5.27}$$

The last step here, expressing the proper vectors in terms of relative vectors, was carried out by differentiating (2.6) to get

$$\dot{v}v = \dot{v}\wedge v = v_0^2(\dot{\mathbf{v}} + i(\mathbf{v}\times\dot{\mathbf{v}}))\,.$$

Finally, writing

$$\omega_L = \widetilde{L}\,\Omega_- L \tag{5.28}$$

for the transformed Larmor precession, we have the desired result

$$\omega = \omega_T + \omega_L\,. \tag{5.29}$$

The Thomas term describes the effect of motion on the precession explicitly and completely. More details are given in Ref. 8, but some improvements have been introduced in the present account.

Now let us apply the rotor approach to a practica problem of spin precession. In general, for a charged particle with an intrinsic magnetic moment in a *uniform* electromagnetic field $F = F_+ + F_-$,

$$\Omega = \frac{e}{mc}\big(F_+ + \frac{g}{2}F_-\big) = \frac{e}{mc}\left[\,F + \tfrac{1}{2}(g-2)F_-\,\right]. \tag{5.30}$$

where as defined by (5.19c) F_- is the magnetic field in the instantaneous rest frame of the particle, and g is the gyromagnetic ratio. This yields the classical equation of motion (5.6) for the velocity, but by (5.3) and (5.5) the equation of motion for the spin is

$$\dot{s} = \frac{e}{m}\left[\,F + \tfrac{1}{2}(g-2)F_-\,\right]\cdot s\,. \tag{5.31}$$

This is the well-known Bargmann-Michel-Telegdi (BMT) equation, which has been applied to high precision measurements of the g-factor for the electron and muon.

To apply the BMT equation, it must be solved for the rate of spin precession. The general solution for an arbitrary combination $F = \mathbf{E} + i\mathbf{B}$ of uniform electric and magnetic fields is easily found by replacing the BMT equation by the rotor equation

$$\dot{R} = \frac{e}{2m} FR + R\tfrac{1}{2}(g-2)\left(\frac{e}{2m}\right) i\mathbf{B}_0 \,, \tag{5.32}$$

where

$$i\mathbf{B}_0 = \widetilde{R}F_-R = \tfrac{1}{2}\big[\widetilde{R}F_-R - (\widetilde{R}F_-R)^\dagger\big] \,. \tag{5.33}$$

is the "effective magnetic field" in the γ_0-system. With initial conditions $R(0) = L_0$, $U(0) = 1$, for a boost without spatial rotation, the solution of (5.32) is

$$R = \exp\left[\frac{e}{2m} F\tau\right] \, L_0 \, \exp\left[\tfrac{1}{2}(g-2)\left(\frac{e}{2m}\right) i\mathbf{B}_0\tau\right] \,, \tag{5.34}$$

where $\mathbf{B}_0$ has the constant value

$$\mathbf{B}_0 = \frac{1}{2i}\big[\,\widetilde{L}_0 F L_0 - (\widetilde{L}_0 F L_0)^\dagger\,\big] = \mathbf{B} + \frac{v_{00}^2}{1+v_{00}} \mathbf{v}_0 \times (\mathbf{B}\times\mathbf{v}_0) + v_{00}\mathbf{E}\times\mathbf{v}_0 \,, \tag{5.35}$$

where $v_{00} = v(0)\cdot\gamma_0 = (1-\mathbf{v}^2)^{-\frac{1}{2}}$. The first factor in (5.34) has the same effect on both the velocity v and the spin s, so the last factor gives directly the change in the relative directions of the relative velocity $\mathbf{v}$ and the spin $\mathbf{s}$. This can be measured experimentally.

To conclude this Section, some general remarks about the description of spin will be helpful in applications and in comparisons with more conventional approaches. We have represented the spin by the proper vector $s = |\,s\,|e_3$ defined by (5.3) and alternatively by the relative vector $\mathbf{s} = \mathbf{e}_3$ where $\boldsymbol{\sigma} = |\,s\,|\mathbf{e}_3$ is defined by (5.21). For a particle with proper velocity $v = L^2\gamma_0$, these two representations are related by

$$sv = L\boldsymbol{\sigma}\widetilde{L} \tag{5.36}$$

or, equivalently, by

$$\boldsymbol{\sigma} = \widetilde{L}(sv)L = \widetilde{L}sL\gamma_0 \,. \tag{5.37}$$

A straightforward spacetime split of the proper spin vector s, like (4.6a) for the velocity vector, gives

$$s\gamma_0 = s_0 + \mathbf{s} \,, \tag{5.38a}$$

where

$$\mathbf{s} = s \wedge \gamma_0 \tag{5.38b}$$

is the relative spin vector, and $s\cdot v = 0$ implies that

$$s_0 = \mathbf{v}\cdot\mathbf{s} \,. \tag{5.38c}$$

From, (5.36) and (5.38a), the relation of $\mathbf{s}$ to $\boldsymbol{\sigma}$ is found to be

$$\mathbf{s} = \boldsymbol{\sigma} + (v_0 - 1)(\boldsymbol{\sigma}\cdot\hat{\mathbf{v}})\hat{\mathbf{v}} \,, \tag{5.39}$$

where v_0 is given by (2.7) and $\hat{\mathbf{v}} = \mathbf{v}/|\,\mathbf{v}\,|$. Both vectors $\mathbf{s}$ and $\boldsymbol{\sigma}$ are sometimes used in the literature, and some confusion results from a failure to recognize that they come from two different kinds of spacetime split. Of course either one can be used, since one determines the other, but $\boldsymbol{\sigma}$ is usually

simpler because its magnitude is constant. Note from (5.39) that they are indistinguishable in the non-relativistic approximation.

6. ELECTROMAGNETIC FIELDS.

As observed in Section 2, STA makes it possible to reduce pairs of equations for the electromagnetic field $F = F(x)$ to a single "Maxwell's Equation"

$$\nabla F = J\,. \tag{6.1}$$

This reduction brings many simplifications to electromagnetic theory. For example, as shown in Section 9, the operator ∇ has an inverse so (6.1) can be solved for

$$F = \nabla^{-1} J\,, \tag{6.2}$$

Actually, ∇^{-1} is an integral operator which depends on boundary conditions on F for the region on which it is defined, so (6.2) is an integral form of Maxwell's equation. However, if the "current" $J = J(x)$ is the sole source of F, then (6.2) provides the unique solution to (6.1).

This section surveys other simplifications to the formulation and analysis of electromagnetic equations. Differentiating (6.1) we obtain

$$\nabla^2 F = \nabla J = \nabla \cdot J + \nabla \wedge J\,, \tag{6.3}$$

where ∇^2 is the d'Alembertian (2.3). Separately equating scalar and bivector parts of (6.3), we obtain the *charge conservation* law

$$\nabla \cdot J = 0 \tag{6.4}$$

and an alternative equation for the E-M field

$$\nabla^2 F = \nabla \wedge J\,. \tag{6.5}$$

A different field equation is obtained by using the fact that, under general conditions, a bivector field $F = F(x)$ can be expressed as a derivative with the specific form

$$F = \nabla(A + Bi)\,, \tag{6.6}$$

where $A = A(x)$ and $B = B(x)$ are vector fields, so F *has* a "vector potential" A and a "trivector potential" Bi. This is a generalization of the well-known "Helmholz theorem" in vector analysis [9]. Since $\nabla A = \nabla \cdot A + \nabla \wedge A$ with a similar equation for ∇B, the bivector part of (6.6) can be written

$$F = \nabla \wedge A + (\nabla \wedge B)i\,, \tag{6.7}$$

while the scalar and pseudoscalar parts yield the so-called "Lorentz condition"

$$\nabla \cdot A = 0\,, \qquad \nabla \cdot B = 0\,. \tag{6.8}$$

Inserting (6.6) into Maxwell's equation (6.1) and separating vector and trivector parts, we obtain the usual wave equation for the vector potential

$$\nabla^2 A = J\,, \tag{6.9}$$

as well as

$$\nabla^2 Bi = 0\,. \tag{6.10}$$

The last equation shows that B is independent of the source J, so it can be set to zero in (6.6). However, in a theory with magnetic charges, Maxwell's equation takes the form

$$\nabla F = J + iK, \tag{6.11}$$

where $K = K(x)$ is a vector field, the "magnetic current density." On substituting (6.6) into (6.11) we obtain in place of (6.10),

$$\nabla^2 Bi = iK\,. \tag{6.12}$$

The pseudoscalar i can be factored out to make (6.12) appear symmetrical with (6.9), but this symmetry between the roles of electric and magnetic currents is deceptive, because one is vectorial while the other is actually trivectorial.

The separation of Maxwell's equation (6.11) into electric and magnetic parts can be achieved directly by using (2.5) again to write

$$\nabla F = \nabla \cdot F + \nabla \wedge F\,. \tag{6.13}$$

Then (6.11) can be separated into a vector part

$$\nabla \cdot F = J \tag{6.14}$$

and a trivector part

$$\nabla \wedge F = iK\,. \tag{6.15}$$

This last equation can be made to look similar to (6.14) by using the duality relation (1.12a); thus, (6.15) becomes

$$\nabla \cdot (Fi) = K\,. \tag{6.16}$$

Note that the dual Fi of the bivector F is also a bivector. Hereafter we shall restrict our attention to the "physical case" $K = 0$.

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

$$J = J^C + \nabla \cdot M\,. \tag{6.17}$$

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 (6.17), equation (6.14) can be put in the form

$$\nabla \cdot G = J^C \tag{6.18}$$

where we have defined a new field

$$G = F - M\,. \tag{6.19}$$

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 (6.18) 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

$$T(n) = \tfrac{1}{2} Fn\widetilde{F} = -\tfrac{1}{2} FnF\,. \tag{6.20}$$

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

$$T^\mu \equiv T(\gamma^\mu) = \tfrac{1}{2} F \gamma^\mu \widetilde{F} \,. \tag{6.21}$$

The divergence of $T(n)$ can be evaluated by using Maxwell's equation (6.1), with the result

$$\partial_\mu T^\mu = T(\nabla) = J \cdot F \,. \tag{6.22}$$

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.

The compact, invariant form (6.20) enables us to solve easily the eigenvector problem for the E-M energy-momentum tensor. If F is not a null field, it has the invariant decomposition $F = fe^{i\varphi}$ given by (1.39), which, when inserted in (6.20), gives

$$T(n) = -\tfrac{1}{2} fnf \tag{6.23}$$

This is simpler than (6.20) because f is simpler than F. Note also that it implies that all fields differing only by an arbitrary "duality factor" $e^{i\varphi}$ have the same energy-momentum tensor. The eigenvalues can be found from (6.26) by inspection. The bivector f determines a timelike plane. Any vector n in that plane satisfies $n \wedge f = 0$, or equivalently, $nf = -fn$. On the other hand, if n is orthogonal to the plane, then $n \cdot f = 0$ and $nf = fn$ For these two cases, (6.23) gives us

$$T(n) = \pm \tfrac{1}{2} f^2 n \,. \tag{6.24}$$

Thus $T(n)$ has a pair of doubly degenerate eigenvalues $\pm\frac{1}{2}f^2$ corresponding to "eigenbivectors" f and if, all expressible in terms of F by (1.40b). This approach should be compared with conventional matrix methods to appreciate the simplifications achieved by STA.

The versatility of STA is also illustrated by the ease with which the above invariant formulation of "Maxwell Theory" can be related to more conventional formulations. The tensor components $F^{\mu\nu}$ of the E-M field F are given by (1.31), whence, using (2.2), we find

$$\partial_\mu F^{\mu\nu} = J \cdot \gamma^\nu = J^\nu \tag{6.25}$$

for the tensor components of Maxwell's equation (6.14). Similarly, the tensor components of (6.15) are

$$\partial_{[\nu} F_{\alpha\beta]} = K^\mu \epsilon_{\mu\nu\alpha\beta} \,, \tag{6.26}$$

where the brackets indicate antisymmetrization and $\epsilon_{\mu\nu\alpha\beta} = i \cdot (\gamma_\mu \gamma_\nu \gamma_\alpha \gamma_\beta)$. The tensor components of the energy-momentum tensor (6.21) are

$$\begin{aligned} T^{\mu\nu} = \gamma^\mu \cdot T^\nu = -\tfrac{1}{2}(\gamma^\mu F \gamma^\nu F)_{(0)} &= (\gamma^\mu \cdot F) \cdot (F \cdot \gamma^\nu) - \tfrac{1}{2}\gamma^\mu \cdot \gamma^\nu (F^2)_{(0)} \\ &= F^{\mu\alpha} F^\nu_\alpha - \tfrac{1}{2} \mathrm{g}^{\mu\nu} F_{\alpha\beta} F^{\alpha\beta} \end{aligned} \tag{6.27}$$

A space-time split of Maxwell's equation (6.1) puts it in the standard relative vector form for an inertial system. Thus, following the procedure in Section 4,

$$J\gamma_0 = J_0 + \mathbf{J} \tag{6.28}$$

splits the current J into a charge density $j_0 = J \cdot \gamma_0$ and a relative current $\mathbf{J} = J \wedge \gamma_0$ in the γ_0-system. Similarly,

$$\gamma_0 \nabla = \partial_t + \nabla \tag{6.29}$$

splits $\bigtriangledown = \partial_x$ into a time derivative $\partial_t = \gamma_0 \cdot \bigtriangledown$ and spatial derivative $\nabla = \gamma_0 \wedge \bigtriangledown = \boldsymbol{\partial}_{\mathbf{x}}$ with respect to the relative position vector $\mathbf{x} = x \wedge \gamma_0$. Combining this with the split (4.9a) of F into electric and magnetic parts, we get Maxwell's equation (6.1) in the split form

$$(\partial_t + \nabla)(\mathbf{E} + i\mathbf{B}) = J_0 - \mathbf{J}\,. \tag{6.30}$$

This can be separated into relative even and odd parts:

$$i\partial_t \mathbf{B} + \nabla \mathbf{E} = J_0\,, \tag{6.31a}$$

$$\partial_t \mathbf{E} + i\nabla \mathbf{B} = -\mathbf{J}\,. \tag{6.31b}$$

Equation (4.10a) gives us the decomposition

$$\nabla \mathbf{E} = \nabla \cdot \mathbf{E} + i(\nabla \times \mathbf{E})\,. \tag{6.32}$$

This enables us to split (6.31a) into two familiar Maxwell equations. In a similar way the other two equations are obtained from (6.31b).

The vector field $T^0 = T(\gamma^0) = T(\gamma_0)$ is the *energy-momentum density* in the γ_0-system. The split

$$T^0\gamma^0 = T^0\gamma_0 = T^{00} + \mathbf{T}^0 \tag{6.33}$$

separates it into an energy density $T^{00} = T^0 \cdot \gamma^0$ and a momentum density $\mathbf{T}^0 = T^0 \wedge \gamma^0$. Using the fact that γ_0 anticommutes with relative vectors, from (6.26) we obtain

$$T^0\gamma^0 = \tfrac{1}{2} F F^\dagger = \tfrac{1}{2}(\mathbf{E} + i\mathbf{B})(\mathbf{E} - i\mathbf{B})\,. \tag{6.34}$$

Whence, the familiar results

$$T^{00} = \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2)\,, \tag{6.35a}$$

$$\mathbf{T}^0 = \mathbf{E} \times \mathbf{B}\,. \tag{6.35b}$$

Mathematical advantages of writing the E-M field in the complex form $F = \mathbf{E} + i\mathbf{B}$ have been noted and exploited by many investigators (e.g. [10]), but without recognizing its geometrical basis where the imaginary is the unit pseudoscalar.

The spacetime split helps us with physical interpretation. Corresponding to the split $F = \mathbf{E} + i\mathbf{B}$, the magnetization field M splits into

$$M = -\mathbf{P} + i\mathbf{M}\,, \tag{6.36}$$

where $\mathbf{P}$ is the electric *polarization density* and $\mathbf{M}$ is the *magnetic moment density*. Writing

$$G = \mathbf{D} + i\mathbf{H}\,, \tag{6.37}$$

we see that (6.19) gives us the familiar relations

$$\mathbf{D} = \mathbf{E} + \mathbf{P}\,, \tag{6.38a}$$

$$\mathbf{H} = \mathbf{B} - \mathbf{M}\,. \tag{6.38b}$$

7. Transformations on spacetime

This section describes the apparatus of geometric calculus for handling transformations of spacetime and the induced transformations of multivector fields on spacetime. We concentrate on the mappings of 4-dimensional regions, including the whole of spacetime, but the apparatus applies with only minor adjustments to the mapping of any submanifold in spacetime. Throughout, we assume whatever differentiability is required to perform indicated operations, so that we might as well assume that the transformations are diffeomorphisms and defer the analysis of discontinuities in derivatives. We therefore assume that all transformations are invertible unless otherwise indicated.

Let f be a diffeomorphism which transforms each point x in some region of spacetime into another point x', as expressed by

$$f:\ x \quad \to \quad x' = f(x)\,. \tag{7.1}$$

This induces a linear transformation of tangent vectors at x to tangent vectors at x', given by the differential

$$\underline{f}:\quad a \quad \to \quad a' = \underline{f}(a) = a\cdot\nabla f\,. \tag{7.2}$$

More explicitly, this determines the transformation of a vector field $a = a(x)$ into a vector field

$$a' = a'(x') \equiv \underline{f}[\,a(x);\,x\,] = \underline{f}[a(f^{-1}(x'));\,f^{-1}(x')\,]\,. \tag{7.3}$$

The outermorphism of $\underline{f}$ determines an induced transformation of specified multivector fields. In particular,

$$\underline{f}(i) = J_f i, \qquad \text{where} \qquad J_f = \det\underline{f} = -i\,\underline{f}i \tag{7.4}$$

is the *Jacobian* of f.

The transformation f also induces an *adjoint* transformation $\bar{f}$ which takes tangent vectors at x' back to tangent vectors at x, as defined by

$$\bar{f}:\quad b' \quad \to \quad b = \bar{f}(b') \equiv \dot{\nabla}\dot{f}\cdot b' = \partial_x f(x)\cdot b'\,. \tag{7.5}$$

More explicitly, for vector fields

$$\bar{f}:\quad b'(x') \quad \to \quad b(x) = \bar{f}\,[\,b'(x');\,x\,] = \bar{f}\,[\,b'(f(x));\,x\,]\,. \tag{7.6}$$

The differential is related to the adjoint by

$$b'\cdot\underline{f}(a) = a\cdot\bar{f}(b')\,. \tag{7.7}$$

According to (7.15), $\bar{f}$ determines the inverse transformation

$$\underline{f}^{-1}(a') = \bar{f}(a'i)(J_f i)^{-1} = a\,. \tag{7.8}$$

Also, however,

$$\underline{f}^{-1}(a') = a'\cdot\partial_{x'}f^{-1}(x')\,. \tag{7.9}$$

Thus, the inverse of the differential equals the differential of the inverse.

Since the adjoint maps "backward" instead of "forward," it is often convenient to deal with its inverse

$$\overline{f^{-1}}:\quad a(x) \quad \to \quad a'(x') = \overline{f^{-1}}\,[\,a(f^{-1}(x'))\,]\,. \tag{7.10}$$

This has the advantage of being directly comparable to $\underline{f}$. Note that it is not necessary to distinguish between $\overline{f^{-1}}$ and $\bar{f}^{-1}$.

Thus, we have two kinds of induced transformations for multivector fields: The first, by $\underline{f}$, is commonly said to be *contravariant*, while the second, by $\bar{f}$ or $\overline{f^{-1}}$, is said to be *covariant*. The first is said to "transform like a vector," while the second is said to "transform like a *covector*." The term "vector" is thus associated with the differential while "covector" is associated with the adjoint. This linking of the vector concept to a transformation law is axiomatic in ordinary tensor calculus. In geometric calculus, however, the two concepts are kept separate. The algebraic concept of vector is determined by the axioms of geometric algebra without reference to any coordinates or transformations. Association of a vector or a vector field with a particular transformation law is a separate issue.

The transformation of a multivector field can also be defined by the rule of *direct substitution:* A field $F = F(x)$ is transformed to

$$F'(x') \equiv F'(f(x)) = F(x)\,. \tag{7.11}$$

Thus, the values of the field are unchanged — although they are associated with different points by changing the functional form of the field. It is very important to note here that the alternative definition $F'(x) \equiv F(x')$ is adopted in [11]. Each of these two alternatives has something to recommend it.

Directional derivatives of the two different functions in (7.11) are related by the *chain rule:*

$$a \cdot \nabla F = a \cdot \partial_x F'(f(x)) = (a \cdot \nabla_x f(x)) \cdot \nabla_{x'} F'(x') = \underline{f}(a) \cdot \nabla' F' = a' \cdot \nabla' F'\,. \tag{7.12}$$

The chain rule is more simply expressed as an operator identity

$$a \cdot \nabla = a \cdot \bar{f}(\nabla') = \underline{f}(a) \cdot \nabla' = a' \cdot \nabla'\,. \tag{7.13}$$

Differentiation with respect to the vector a yields the general transformation law for the vector derivative:

$$\nabla = \bar{f}(\nabla') \qquad \text{or} \qquad \nabla' = \overline{f^{-1}}(\nabla)\,. \tag{7.14}$$

This is the most basic formulation of the chain rule, from which its many implications are most easily derived. All properties of induced transformations are essentially implications of this rule, including the transformation law for the differential, as (7.13) shows.

The rule for the induced transformation of the curl is derived by using the integrability condition (2.15) to prove that the adjoint function has vanishing curl; thus, for the adjoint of a vector field,

$$\dot{\nabla} \wedge \dot{\bar{f}}(a') = \nabla_b \wedge \bar{f}_b(a') = \nabla_b \wedge \nabla_c f_{cb} \cdot a' = \nabla \wedge \nabla f \cdot a' = 0\,. \tag{7.15}$$

The transformation rule for the curl of a vector field $a = \bar{f}(a')$ is therefore

$$\nabla \wedge a = \nabla \wedge \bar{f}(a') = \bar{f}(\nabla' \wedge a')\,. \tag{7.16}$$

To extend this to multivector fields, note that the differential of an outermorphism is not itself an outermorphism; rather it satisfies the "product rule"

$$\bar{f}_b(A' \wedge B') = \bar{f}_b(A') \wedge \bar{f}(B') + \bar{f}(A') \wedge \bar{f}_b(B')\,. \tag{7.17}$$

Therefore, it follows from (7.15) that the curl of the adjoint outermorphism vanishes, and (7.16) generalizes to

$$\nabla \wedge A = \bar{f}(\nabla' \wedge A') \qquad \text{or} \qquad \nabla' \wedge A' = \overline{f^{-1}}(\nabla \wedge A)\,, \tag{7.18}$$

where $A = \bar{f}(A')$. Thus, the outermorphism of the curl is the curl of an outermorphism.

The transformation rule for the divergence is more complex, but it can be derived from that of the curl by exploiting the duality of inner and outer products (1.12) and the transformation law (7.14) relating them. Thus,

$$\bar{f}\,(\nabla' \wedge (A'i)) = \bar{f}\,[(\nabla' \cdot A')i] = \overline{f^{-1}}(\nabla' \cdot A')\bar{f}\,(i)\,.$$

Then, using (7.18) and (7.4) we obtain

$$\nabla \wedge \bar{f}\,(A'i) = \nabla \wedge [\,\underline{f}^{-1}(A')\bar{f}\,(i)] = \nabla \cdot (J_f A)i\,.$$

For the divergence, therefore, we have the transformation rule

$$\nabla' \cdot A' = \nabla' \cdot \underline{f}(A) = {J_f}^{-1}\underline{f}[\,\nabla \cdot (J_f A)\,] = \underline{f}[\nabla \cdot A + (\nabla \ln J_f) \cdot A]\,, \tag{7.19}$$

where $A' = \underline{f}(A)$. This formula can be separated into two parts:

$$\grave{\nabla}'\,\grave{\underline{f}}(A) = \underline{f}[(\nabla' \ln\, J_f) \cdot A] = (\nabla' \ln\, J_f) \cdot \underline{f}(A)\,, \tag{7.20}$$

$$\grave{\nabla}' \cdot \underline{f}(\grave{A}) = \underline{f}(\nabla \cdot A)\,. \tag{7.21}$$

The whole may be recovered from the parts by using the following generalization of (7.13) (which can also be derived from (7.14)):

$$\underline{f}(A) \cdot \nabla' = \underline{f}(A \cdot \nabla)\,. \tag{7.22}$$

8. Directed Integrals and the Fundamental Theorem

In the theory of integration, geometric calculus absorbs, clarifies and generalizes the calculus of differential forms. Only the essentials are sketched here; details are given in [1], and [12] discusses the basic concepts at greater length with applications to physics.

The integrand of any integral over a k-dimensional manifold is a *differential k-form*

$$L = L\,(d^k x) = L[d^k x;\, x\,], \tag{8.1}$$

where $d^k x$ is a k-vector-valued measure on the manifold. If the surface is not null at x, we can write

$$d^k x = I_k\,|\,d^k x\,|\,, \tag{8.2}$$

where $I_k = I_k(x)$ is a unit k-vector field tangent to the manifold at x, and $|\,d^k x\,|$ is an ordinary scalar-valued measure. Thus, $d^k x$ describes the direction of the tangent space to the manifold at each point. For this reason it is called a *directed measure*. Since the integrals are defined from weighted sums, the integrand $L(d^k x)$ must be a linear function of its k-vector argument; accordingly it is a *k-form* as defined in Section 4. Of course, the values of L may vary with x, as indicated by the explicit x-dependence shown on the right side of (2.17).

The *exterior differential* of a k-form L is a $(k+1)$-form dL defined by

$$dL = \grave{L}[\,(d^{k+1}x) \cdot \grave{\nabla}\,] = L[\,(d^{k+1}x) \cdot \grave{\nabla};\, \grave{x}\,]\,, \tag{8.3}$$

where the accent indicates that only the implicit dependence of L on x is differentiated. The exterior derivative of any "k-form" which is already the exterior derivative of another form necessarily vanishes, as is expressed by

$$d^2 L = 0. \tag{8.4}$$

This is an easy consequence of the integrability condition (2.15); thus,

$$d^2L = d\grave{L}[(d^{k+1}x)\cdot\grave{\nabla}] = \grave{L}[(d^{k+1}x)\cdot(\grave{\nabla}\wedge\grave{\nabla})] = 0\,.$$

The *Fundamental Theorem of Integral Calculus* (also known as the "Boundary Theorem" or the "Generalized Stokes' Theorem") can now be written in the compact form

$$\int dL = \int dL\,(d^{k+1}x) = \oint L(d^kx) = \oint L\,. \tag{8.5}$$

This says that the integral of any k-form L over a *closed* k-dimensional manifold is equal to the integral of its exterior derivative over the enclosed $(k+1)$-dimensional manifold. It follows from (8.4) that this integral vanishes if $L = dN$ where N is a $(k-1)$-form.

To emphasize their dependence on a *directed measure*, the integrals in (8.5) may be called *directed integrals*. In conventional approaches to differential forms this dependence is disguised and all forms are scalar-valued. For that special case we can write

$$L = \langle A\,d^kx\rangle = (d^kx)\cdot A(x)\,, \tag{8.6a}$$

where $A = A(x)$ is a k-vector field. Then

$$dL = [(d^{k+1}x)\cdot\nabla]\cdot A = (d^{k+1}x)\cdot(\nabla\wedge A)\,. \tag{8.6b}$$

In this case, therefore, the exterior derivative is equivalent to the curl.

An alternative form of the Fundamental Theorem called "Gauss's Theorem" is commonly used in physics. If L is a 3-form, its 3-vector argument can be written as the dual of a vector, and a *tensor field* $T(n) = T[\,n(x);\,x\,]$ can be defined by

$$T(n) = L(in)\,. \tag{8.7}$$

According to (8.2) we can write

$$d^4x = i\,|\,d^4x\,| \qquad \text{and} \qquad d^3x = in^{-1}\,|\,d^3x\,|\,, \tag{8.8}$$

where n is the *outward unit normal* defined by the relation $I_3n = I_4 = i$. Substitution into (8.5) then gives *Gauss's Theorem:*

$$\int \grave{T}(\grave{\nabla})\,|\,d^4x\,| = \oint T(n^{-1})\,|\,d^3x\,|\,. \tag{8.9}$$

where $n^{-1} = \epsilon n$ with signature ϵ. Though $\grave{T}(\grave{\nabla})$ may be called the "divergence of the tensor T," it is not generally equivalent to the divergence as defined earlier for multivector fields. However, if L is scalar-valued as in (8.6a), then (8.7) implies that

$$T(n) = n\cdot a\,, \tag{8.10a}$$

where $a = a(x) = A(x)i$ is a vector field. In this case, we do have the divergence

$$\grave{T}(\grave{\nabla}) = \nabla\cdot a\,. \tag{8.10b}$$

Note that duality has changed the curl in (8.6b) into the divergence in (8.10b).

A *change of integration variables* in a directed integral is a transformation on a differential form by direct substitution. Thus, for the k-form defined in (8.1) we have

$$L'(d^k x') = L(d^k x)\,, \tag{8.11}$$

where

$$d^k x' = \underline{f}(d^k x) \qquad \text{or} \qquad d^k x = \underline{f}^{-1}(d^k x') \tag{8.12}$$

In other words, $L' = L\underline{f}^{-1}$ or, more explicitly,

$$L'(d^k x';\, x') = L[\,\underline{f}^{-1}(d^k x);\, f^{-1}(x)\,] = L(d^k x; x)\,. \tag{8.13}$$

The value of the integral of (8.11) is thus unaffected by the change of variables,

$$\int L'(d^k x') = \int L[\,\underline{f}^{-1}(d^k x')\,] = L(d^k x)\,. \tag{8.14}$$

The exterior derivative and hence the fundamental theorem are likewise unaffected. In other words,

$$dL' = dL\,. \tag{8.15}$$

This follows from

$$(d^k x')\cdot\nabla' = \underline{f}(d^k x)\cdot\bar{f}\,(\nabla) = \underline{f}[\,(d^k x)\cdot\nabla\,] \tag{8.16}$$

and

$$d\,\underline{f}(d^k x) = \dot{\underline{f}}\,[\,(d^k x)\cdot\dot{\nabla}\,] = 0\,. \tag{8.17}$$

Like (8.4), the last equation is a consequence of the integrability condition.

It has recently become popular to formulate Maxwell's equations in terms of differential forms, so it is important to understand how that articulates with STA. Let $dx = \gamma_\mu dx^\mu$ be an arbitrary "line element," that is, a tangent vector at x. From the current J we can construct a 1-form

$$\alpha = J\cdot dx = J_\mu dx^\mu\,. \tag{8.18}$$

Line elements d_1x and d_2x determine a directed area element $d^2x = d_1x\wedge d_2x$. Projection of the bivector field F onto an arbitrary directed area defines the *electromagnetic 2-form*

$$\boldsymbol{\omega} = F\cdot d^2x = F_{\mu\nu}d_1x^\mu d_2x^\nu\,. \tag{8.19}$$

The *exterior derivative* $d\omega$ of the form ω can be defined in terms of the curl as the 3-form

$$d\omega = d^3x\cdot(\nabla\wedge F)\,, \tag{8.20}$$

Defining a *dual form* ${}^*\omega$ for ω by

$${}^*\omega = d^2x\cdot(Fi)\,, \tag{8.21}$$

with the help of the duality relation (1.12a), we obtain the exterior derivative

$$d{}^*\omega = d^3x\cdot(\nabla\wedge(Fi)) = d^3x\cdot((\nabla\cdot F)i)\,. \tag{8.22}$$

From the trivector Ji we get a current 3-form ${}^*\alpha = d^3x\cdot(Ji)$. Now it should be evident that the two Maxwell equations (6.17) and (6.18) with $K=0$ map into the differential form equations

$$d{}^*\omega = {}^*\alpha\,, \tag{8.23a}$$

$$d\omega = 0\,. \tag{8.23b}$$

The special notations from standard differential form theory (such as are used in (8.23a,b)) add nothing of value to the STA formalism. (Indeed, they detract from it, so we will not use them.) As (8.20) indicates, the exterior differential is completely equivalent to the curl; moreover, the d^3x just gets in the way unless one is performing an integration. Most important: standard form theory does not allow one to combine the two Maxwell forms (8.23a,b) into the single equation (6.4).

As an application of the Fundamental Theorem (8.5), we use it to derive scalar-valued integral forms for Maxwell's equation. Inserting (8.19) and (8.20) into (8.5) and using $\nabla \wedge F = 0$, we deduce that

$$\oint d^2x \cdot F = 0 \tag{8.24}$$

for any closed 2-dimensional submanifold $\mathcal{B}$ in spacetime. A spacetime split shows that this integral formula is equivalent to Faraday's Law or to "the absence of magnetic poles," or a mixture of the two, depending on the choice of $\mathcal{B}$.

To derived a similar integral formula for the vector part $\nabla \cdot F = J$ of Maxwell's equation, define a unit normal n by writing

$$d^3x = in\,|\,d^3x\,|\,, \tag{8.25}$$

where i is the unit dextral pseudoscalar for spacetime, and write (8.22) in the form

$$d^3x \cdot (\nabla \wedge (Fi)) = d^3x \cdot (Ji) = J \cdot n\,|\,d^3x\,|\,. \tag{8.26}$$

The different definitions of normal n in (8.8) and (8.25) are deliberate and noteworthy. Insertion of (8.26) into (8.5) yields the integral equation

$$\oint d^2x \cdot (Fi) = \int J \cdot n\,|\,d^3x\,|, \tag{8.27}$$

which, like (8.24), holds for *any* closed 2-manifold $\mathcal{B}$, where the integral on the right is over any 3-manifold with boundary $\mathcal{B}$. Again a spacetime split reveals that (8.27) is equivalent to Ampere's Law, Gauss' Law, or a combination of the two, depending on the choice of $\mathcal{B}$.

The two integral equations (8.24) and (8.27) are fully equivalent to the vector and trivector parts of Maxwell's equation $\nabla F = J$. They can be combined into a single equation. First multiply (8.27) by i and use (8.26) to put it in the less familiar form

$$\oint (d^2x) \wedge F = \int (d^3x) \wedge J. \tag{8.28}$$

Adding (8.24) to (8.28), we can write the integral version of the whole Maxwell's equation in the form

$$\oint \langle d^2x F\rangle_I = \int \langle d^3x J\rangle_I, \tag{8.29}$$

where $\langle \ldots \rangle_I$ selects only the "invariant (= scalar+pseudoscalar) parts." It remains to be seen if this form has some slick physical applications.

9. Integral Equations and Conservation Laws

The main use of Gauss's Theorem (8.9) are (1) to convert (local) field equations into integral equations and (2) to convert local or (differential) conservation laws into global (or integral) conservation laws. This section shows how.

Let $\mathcal{M}$ be a 4-dimensional region in spacetime. For any two spacetime points x and y, a Green's function $G = G(y, x)$ for Maxwell's equation is defined on $\mathcal{M}$ as a solution to the differential equation.

$$\partial_y G(y, x) = -G(y, \dot{x})\dot{\partial}_x = \delta^4(y - x)\,, \tag{9.1}$$

where the right side is the 4-dimensional delta function. Let $T(n)$ in (8.9) be given by

$$T(n) = GnF\,, \tag{9.2}$$

where $F = F(x)$ is any differentiable function. If y is an interior point of $\mathcal{M}$, substitution of (9.2) into (8.9) yields

$$F(y) = \int_{\mathcal{M}} G(y, x)\partial_x F(x)|\, d^4x\,| - \oint_{\partial\mathcal{M}} G(y, x)n^{-1}F(x)\,|\, d^3x\,|\,. \tag{9.3}$$

This great formula allows us to calculate F at any point y inside $\mathcal{M}$ from its derivative $\nabla F = \partial_x F(x)$ and its values on the boundary if G is known.

Now let u be a constant, unit, timelike vector (field) directed in the forward light cone. The vector u determines a 1-parameter family of spacetime hyperplanes $\mathcal{S}(t)$ satisfying the equation

$$u \cdot x = t\,. \tag{9.4}$$

The vector u thus determines an inertial frame with time variable t, so $\mathcal{S}(t)$ is a surface of simultaneous t.

Let $\mathcal{V}(t)$ be a convex 3-dimensional region (submanifold) in $\mathcal{S}(t)$ which sweeps out a 4-dimensional region $\mathcal{M}$ in the time interval $t_1 \le t \le t_2$. In this interval the 2-dimensional boundary $\partial\mathcal{V}(t)$ sweeps out a 3-dimensional wall $\mathcal{W}$, so $\mathcal{M}$ is bounded by $\partial\mathcal{M} = \mathcal{V}(t_1) + \mathcal{V}(t_2) + \mathcal{W}$. We can use the integral formula (9.3) to solve *Maxwell's equation*

$$\nabla F = J \tag{9.5}$$

in the region $\mathcal{M}$ for the *electromagnetic field* $F = F(x)$ "produced by" the *charge current* (density) $J = J(x)$. The field F is bivector-valued while the current J is vector-valued. For simplicity, let us enlarge $\mathcal{V}(t)$ to coincide with $\mathcal{S}(t)$ and assume that the integral of F over $\partial\mathcal{V}$ is vanishingly small at spatial infinity. Then $\mathcal{M}$ is the entire region between the hyperplanes $\mathcal{S}_1 = \mathcal{S}(t_1)$ and $\mathcal{S}_2 = \mathcal{S}(t_2)$, and (9.3) gives us

$$F(y) = \int_{\mathcal{M}} G(y, x)J(x)\,|\, d^4x\,| + F_1 - F_2\,, \tag{9.6}$$

where

$$F_k(y) = \int_{\mathcal{S}_k} G(y, x)\, u\, F(x)\,|\, d^3x\,|\,. \tag{9.7}$$

Because of the condition (9.1) on the Green's function, the F_k satisfy the homogeneous equation

$$\nabla F_k = 0\,. \tag{9.8}$$

A *retarded Green's function* G_k can be found which vanishes on $\mathcal{S}_2$, in which case F_1 solves the *Cauchy problem* for the homogeneous Maxwell equation (9.8).

Note that the right side of (9.6) can be regarded as defining an inverse ∇^{-1} to the vector derivative ∇, as asserted previously in writing (6.5). Indeed, by operating on (9.6) with $\nabla = \partial_y$ and using (9.1), it is easily shown that $F(y)$ satisfies Maxwell's equation (9.5).

Green's functions for spacetime have been extensively studied by physicists and the results, contained in many books, are easily adapted to the present formulation. Thus, from [13] we find the following Green's function for (9.6) and (9.7):

$$G(r) = \frac{1}{4\pi}\,\partial_r \delta(r^2) = \frac{1}{2\pi}\, r\,\delta'(r^2)\,, \tag{9.10}$$

where $r = x - y$ and δ denotes a 1-dimensional delta function with derivative δ'. The analysis of retarded and advanced parts of G and their implications is standard, so it need not be discussed here.

Taking $\mathcal{M}$ to be all of spacetime so F_1 and F_2 can be set to zero, equation (9.6) with (9.10) can be integrated to get the field produced by point charge. For a particle with charge q and world line $z = z(\tau)$ with proper time τ, the charge current can be expressed by

$$J(x) = q\int_{-\infty}^{\infty} d\tau\, v\,\delta^4(x - z(\tau))\,, \tag{9.11}$$

where $v = v(\tau) = dz/d\tau$. Inserting this into (9.6) and integrating, we find that the retarded field can be expressed in the following explicit form

$$F(x) = \frac{q}{4\pi}\frac{r\wedge[\,v + r\cdot(v\wedge\dot v)\,]}{(r\cdot v)^3} = \frac{q}{4\pi(r\cdot v)^2}\left[\,\frac{r\wedge v}{|r\wedge v|} + \frac{1}{2}\frac{r\dot v v r}{r\cdot v}\,\right], \tag{9.12}$$

where $r = x - z$ satisfies $r^2 = 0$ and z, v, $\dot v = dv/d\tau$ are all evaluated at the intersection of the backward light cone with vertex at x. This is an invariant form for the classical *Lienard-Wiechart field*.

As the other major application of Gauss's Theorem (8.9), we show that it gives us an immediate integral formulation of any physics conservation law with a suitable choice of $T(n)$. Introducing the notations

$$\grave{T}(\grave{\nabla}) = f \tag{9.13}$$

and

$$\mathcal{I} = \int_{\mathcal{M}} f\,|\,d^4x\,| = \int_{t_1}^{t_2} dt \int_{\mathcal{V}(t)} f|\,d^3x\,|\,, \tag{9.14}$$

for the region $\mathcal{M}$ defined above, we can write (8.9) in the form

$$\mathcal{I} = P(t_2) - P(t_1) - \int_{t_1}^{t_2} dt \oint_{\partial\mathcal{V}(t)} T(n)\,|\,d^2x\,|\,, \tag{9.15}$$

where

$$P(t) = \int_{\mathcal{V}(t)} T(u)\,|\,d^3x\,|\,. \tag{9.16}$$

Now for some applications.

Energy-Momentum Conservation:

We first suppose that $T(n)$ is the energy-momentum tensor for some *physical system*, which could be a material medium, an electromagnetic field, or some combination of the two, and it could be either classical or quantum mechanical. For example, it could be the energy-momentum tensor (6.23) for the electromagnetic field. In general, the tensor $T(n)$ represents the flux of energy-momentum through a hypersurface with normal n.

For the vector field $f = f(x)$ specified independently of the tensor field $T(n) = T(x, n(x))$, equation (9.13) is the *local energy-momentum conservation law*, where the *work-force density* f characterizes the effect of external influences on the system in question. Equation (9.15) is then the *integral energy-momentum conservation law* for the system. The vector $P(t)$ given by (9.16) is the *total energy-momentum* of the system contained in $\mathcal{V}(t)$ at time t. The quantity $\mathcal{I}$ is the total *Impulse* delivered to the system in the region $\mathcal{M}$.

In the limit $t_2 \to t_1 = t$, the conservation law (9.15) can be written

$$\frac{dP}{dt} = \mathcal{F} + \oint_{\partial\mathcal{V}} T(n) \,|\, d^2x \,|\,, \tag{9.17}$$

where

$$\mathcal{F}(t) = \int_{\mathcal{V}(t)} f \,|\, d^3x \,|\, \tag{9.18}$$

is the total work-force on the system. We can decompose (9.17) into separate energy and momentum conservation laws by using a spacetime split: we write

$$Pu = E + \mathbf{p}\,, \tag{9.19}$$

where $E = P \cdot u$ is the energy and $\mathbf{p} = P \wedge u$ is the momentum of the system. Similarly we write

$$\mathcal{F}u = W + \mathbf{F}\,, \tag{9.20}$$

where $W = \mathcal{F} \cdot u$ is the work done on the system and $\mathbf{F} = \mathcal{F} \wedge u$ is the force exerted on it. We write

$$T(n)u = \mathbf{n} \cdot \mathbf{s} + \mathbf{T}(\mathbf{n})\,, \tag{9.21}$$

where $\mathbf{n} \cdot \mathbf{s} = u \cdot T(n)$ is the *energy flux*, $\mathbf{T}(\mathbf{n}) = T(n) \wedge u$ is the *stress tensor*, and $\mathbf{n} = n \wedge u = nu$ represents the normal as a "relative vector." We also note that

$$xu = t + \mathbf{x} \tag{9.22}$$

splits x into a time $t = x \cdot u$ and a position vector $\mathbf{x} = x \wedge u$. Finally, we multiply (9.17) by u and separate scalar and relative vector parts to get the *energy conservation law*

$$\frac{dE}{dt} = W + \oint \mathbf{s} \cdot \mathbf{n} \,|\, d^2\mathbf{x} \,|\, \tag{9.23}$$

and the *momentum conservation* law

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} + \oint \mathbf{T}(\mathbf{n}) \,|\, d^2\mathbf{x} \,|\,. \tag{9.24}$$

These are universal laws applying to all physical systems.

Angular Momentum Conservation:

The "generalized orbital angular momentum tensor" for the system just considered is defined by

$$L(n) = T(n) \wedge x\,. \tag{9.25}$$

With (4.9), its divergence is

$$\grave{L}(\grave{\nabla}) = f \wedge x + T(\grave{\nabla}) \wedge \grave{x}\,. \tag{9.26}$$

For a symmetric tensor such as (5.13) the last term vanishes. But, in general, there exits a bivector-valued tensor $S(n)$, the *spin tensor* for the system, which satisfies

$$\dot{S}(\dot{\nabla}) = \dot{x} \wedge T(\dot{\nabla})\,. \tag{9.27}$$

Now define the *total angular momentum tensor*

$$M(n) = T(n) \wedge x + S(n)\,. \tag{9.28}$$

Then the *local angular momentum conservation* law for the system is

$$\dot{M}(\dot{\nabla}) = f \wedge x\,. \tag{9.29}$$

Replacing (9.13) by (9.29), we can reinterpret (9.15) as an integral law for angular momentum and analyze it the way we did energy-momentum.

Charge Conservation:

From Maxwell's equation we derive the local charge conservation law

$$\nabla \cdot J = \nabla \cdot (\nabla \cdot F) = (\nabla \wedge \nabla) \cdot F = 0\,. \tag{9.30}$$

Now write $T(n) = n \cdot J$ and change the notion of (5.12) to

$$Q(t) = \int_{\mathcal{V}(t)} u \cdot J \,|\, d^3x \,|\,, \tag{9.31}$$

an expression for the total charge contained in $\mathcal{V}(t)$. Then (9.15) becomes

$$Q(t_2) - Q(t_1) = \int_{t_1}^{t_2} dt \oint_{\partial\mathcal{V}(t)} n \cdot J \,|\, d^2x \,|\,. \tag{9.23}$$

This is the charge conservation equation, telling us that the total charge in $\mathcal{V}(t)$ changes only by flowing through the boundary $\partial\mathcal{V}(t)$.

References

[1] D. Hestenes and G. Sobczyk. *CLIFFORD ALGEBRA to GEOMETRIC CALCULUS, A Unified Language for Mathematics and Physics*, G. Reidel Publ. Co., Dordrecht/Boston (1984), paperback (1985). Third printing 1992.

[2] D. Hestenes. *Space-Time Algebra*. Gordon & Breach, New York, (1966).

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

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

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

[6.] D. Hestenes, Proper Dynamics of a Rigid Point Particle, *J. Math. Phys.* **15**, 1778–1786 (1974).

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

[8] D. Hestenes, Proper Particle Mechanics, *J. Math. Phys.* **15**, 1768–1777 (1974).

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

[10] H.E. Moses, SIAM *J. Appl. Math.* **21**, 14–144 (1971).

[11] A. Lasenby, C. Doran & S. Gull. Gravity, Gauge Theories and Geometric Algebra. *Phys. Rev.* (1995) submitted.

[12] D. Hestenes. Differential Forms in Geometric Calculus. In F. Brackx *et al.* (eds), *Clifford Algebras and their Applications in Mathematical Physics*. Kluwer: Dordrecht/Boston (1993). p. 269–285.

[13] Barut, A.: 1980, *Electrodynamics and the classical theory of fields and particles*, Dover, New York.

[5] D. Hestenes. Curvature Calculations with Spacetime Algebra. *Int. J. Theo. Phys.*, **25**, 581–88 (1986).

[6] D. Hestenes. The Design of Linear Algebra and Geometry. *Acta Applicandae Mathematicae* **23**, 65–93 (1991).

[7] D. Hestenes. Differential Forms in Geometric Calculus. In F. Brackx *et al.* (eds), *Clifford Algebras and their Applications in Mathematical Physics*. Kluwer: Dordrecht/Boston (1993). p. 269–285.

[8] D. Hestenes. *New Foundations for Classical Mechanics*. G. Reidel Publ. Co., Dordrecht/Boston (1985), paperback (1987). Third printing with corrections (1993).

[9] Y. Choquet-Bruhat, C. De Witte-Morette & M. Dillard-Bleick. *Analysis, Manifolds and Physics*. North-Holland, New York (1977).

[10] H. Poincaré. *Science and Hypothesis*. (English trans. 1905). Dover, New York (1951).

[11] D. Kramer, H. Stephani, E. Herlt & M. MacCallum. E. Schmutzer (Ed.). *Exact Solutions of Einstein's Field Equations*. Cambridge U. Press, Cambridge (1980).

[12] W. R. Davis. The Role of Tetrad and Flat-Metric Fields in the Framework of the General Theory of Relativity. *Nuovo Cimento* **43B**, 2816–20 (1966).

[13] D. Hestenes. Spinor Approach to Gravitational Motion and Precession. *Int. J. Theo. Phys.*, **25**, 589–98 (1986).

[14] G. Sobczyk. Killing Vectors and Embedding of Exact Solutions in General Relativity. In J. Chisholm & A. Common (eds.), *Clifford Algebras and their Applications in Mathematical Physics*. D. Reidel, Dordrecht/Boston, p. 227–244 (1986).

[15] L. Eisenhart, *Continuous Groups of Transformations*. Dover, New York (1961).

[16] S. Hawking and G. Ellis, *The Large Scale Structure of Space-time*. Cambridge U. Press, Cambridge (1973).

PART II: QUANTUM THEORY

10. 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

$$u^\dagger u = 1\,, \tag{7.1a}$$

$$\gamma_0 u = u\,, \tag{7.1b}$$

$$\gamma_2\gamma_1 u = i'u\,. \tag{7.1c}$$

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

$$\Psi = \psi u\,, \tag{7.2}$$

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 (7.1c) enables us to make it real without altering (7.2) by replacing i' in the term by $\gamma_2\gamma_1$ on the right. Furthermore, the polynomial can be taken to be an even multivector, for if any term is odd, then (7.1b) allows us to make it even by multiplying on the right by γ_0. Thus, in (7.2) 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

$$\gamma^\mu(\partial_\mu\psi\gamma_2\gamma_1\hbar - eA_\mu\psi) = m\psi\gamma_0\,, \tag{7.3}$$

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 (7.1a, b, c) and (5.2) to get the standard form

$$\gamma^\mu(i'\hbar\partial_\mu - eA_\mu)\Psi = m\Psi\,. \tag{7.4}$$

This completes the proof. Alternative proofs are given elsewhere.[4,7] The original converse derivation of (7.3) from (7.4) was much more difficult.[2]

Henceforth, we can work with the *real Dirac equation* (7.3) without reference to its matrix representation (7.4). We know that computations in STA can be carried out without introducing a basis, so let us use (4.1) to write the real Dirac equation in the coordinate-free form

$$\nabla\psi\mathbf{i}\hbar - eA\psi = m\psi\gamma_0\,, \tag{7.5}$$

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

$$\mathbf{i} \equiv \gamma_2\gamma_1 = i\gamma_3\gamma_0 = i\boldsymbol{\sigma}_3 \tag{7.6}$$

emphasizes that this bivector plays the role of the imaginary i' that appears explicitly in the matrix form (7.4) 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 (7.5) 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

$$\gamma_\mu \to \gamma'_\mu = U\gamma_\mu\widetilde{U}\,, \tag{7.7}$$

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

$$\gamma'_0 = U\gamma_0\widetilde{U} \qquad \text{and} \qquad \mathbf{i}' = U\mathbf{i}\widetilde{U}\,. \tag{7.8}$$

A corresponding change in the wave function,

$$\psi \to \psi' = \psi\widetilde{U}\,, \tag{7.9}$$

induces a mapping of the Dirac equation (7.5) into an equation of the same form:

$$\nabla\psi\mathbf{i}'\hbar - eA\psi' = m\psi'\gamma'_0\,. \tag{7.10}$$

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 (7.2) gives

$$\Psi = \psi u = \psi' u'\,, \tag{7.11}$$

where $u' = Uu$.

For the special case

$$U = e^{i\varphi_0}\,, \tag{7.12}$$

where φ_0 is a scalar constant, (7.8) gives $\gamma'_0 = \gamma_0$ and $\mathbf{i}' = \mathbf{i}$, so ψ and

$$\psi' = \psi e^{i\varphi_0} \tag{7.13}$$

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 (7.5) on the right by $\boldsymbol{\sigma}_2$ yields

$$\nabla\psi^C\mathbf{i}\hbar - eA\psi^C = m\psi^C\gamma_0\,, \tag{7.14}$$

where

$$\psi^C = \psi\boldsymbol{\sigma}_2\,. \tag{7.15}$$

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 (7.13). The main thing to notice here is that in (7.15) 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.

Since the real Dirac wave function $\psi = \psi(x)$ is an even multivector, we know from (1.21c) that it has the Lorentz *invariant decomposition*

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R\,, \tag{7.16}$$

where

$$R\widetilde{R} = \widetilde{R}R = 1\,. \tag{7.17}$$

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

$$e_\mu = R\gamma_\mu\widetilde{R} \tag{7.18}$$

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 (7.18), 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 (7.16) 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 (7.16) and (7.18) we find that

$$\psi\gamma_\mu\widetilde{\psi} = \psi'\gamma_\mu\widetilde{\psi}' = \rho\, e_\mu\,. \tag{7.19}$$

Note that that we have here a set of four linearly independent vector fields which are invariant under the transformation specified by (7.7) and (7.8). Thus these fields do not depend on any coordinate system, despite the appearance of γ_μ on the left side of (7.19). Note also that the factor $e^{i\beta/2}$ in (7.16) does not contribute to (7.19), because the pseudoscalar i anticommutes with the γ_μ.

Two of the vector fields in (7.19) are given physical interpretations in the standard Dirac theory. First, the vector field

$$\psi\gamma_0\widetilde{\psi} = \rho\, e_0 = \rho v \tag{7.20}$$

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 (7.20) to the conventional definition of the Dirac current is displayed in Table I.

The second vector field

$$\tfrac{1}{2}\hbar\psi\gamma_3\widetilde{\psi} = \rho\tfrac{1}{2}\hbar e_3 = \rho s \tag{7.21}$$

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

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 $$\frac{e}{m}\,\frac{i'\hbar}{2}\,\widetilde{\Psi}\frac{1}{2}\big(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu\big)\Psi = \frac{e\hbar}{2m}\left(\gamma_\mu\gamma_\nu\psi\gamma_2\gamma_1\widetilde{\psi}\right)_{(0)}$$
$$= (\gamma_\mu\wedge\gamma_\nu)\cdot(M) = M_{\mu\nu} = \frac{e}{m}\,\rho(ie^{i\beta}sv)\cdot(\gamma_\mu\wedge\gamma_\nu)$$

Pseudovector* $$\tfrac{1}{2}i'\hbar\widetilde{\Psi}\gamma_\mu\gamma_5\Psi = \tfrac{1}{2}\hbar(\gamma_\mu\psi\gamma_3\widetilde{\psi})_{(0)} = \gamma_\mu\cdot(\rho s) = \rho s_\mu$$

Pseudoscalar* $$\widetilde{\Psi}\gamma_5\Psi = (i\psi\widetilde{\psi})_{(0)} = -\rho\sin\beta$$

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

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

$$S = isv = \tfrac{1}{2}\hbar ie_3e_0 = \tfrac{1}{2}\hbar R\gamma_2\gamma_1\widetilde{R} = \tfrac{1}{2}R(\mathbf{i}\hbar)\widetilde{R}. \tag{7.22}$$

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 (7.5). 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.

The hidden relation of spin to the imaginary i' in the Dirac theory can be made manifest in another way. Multiplying (7.21) on the right by ψ and using (7.16) we obtain

$$S\psi = \tfrac{1}{2}\psi\mathbf{i}\hbar\,. \tag{7.23}$$

Then using (7.1c) and (7.2) to translate this into the matrix formalism, we obtain

$$S\Psi = \tfrac{1}{2}\mathbf{i}\hbar\Psi\,. \tag{7.24}$$

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

$$S = \tfrac{1}{2}S^{\alpha\beta}\gamma_\alpha\gamma_\beta\,. \tag{7.25}$$

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 (7.24) is completely general, applying to any Dirac wave function whatsoever.

The identification of $S^{\alpha\beta}$ in (7.25) 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

$$\rho S^{\nu\alpha\beta} = \frac{i'\hbar}{2}\,\widetilde{\Psi}\gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta \Psi = \frac{i'\hbar}{2}\,\widetilde{\Psi}\gamma_5\gamma_\mu\Psi\epsilon^{\mu\nu\alpha\beta} = \rho s_\mu \epsilon^{\mu\nu\alpha\beta}\,, \tag{7.26}$$

where use has been made of the identity

$$\gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta = \gamma_5\gamma_\mu\epsilon^{\mu\nu\alpha\beta} \tag{7.27a}$$

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

$$\begin{aligned}\epsilon^{\mu\nu\alpha\beta} &= i(\gamma^\mu \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta) = (i\gamma^\mu\gamma^\nu\gamma^\alpha\gamma^\beta)_{(0)} \\ &= (\gamma_3 \wedge \gamma_2 \wedge \gamma_1 \wedge \gamma_0)\cdot(\gamma^\mu \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta)\,.\end{aligned} \tag{7.27b}$$

Alternatively,

$$\gamma^\mu \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta = -i\epsilon^{\mu\nu\alpha\beta}\,. \tag{7.27c}$$

From (5.26) and (5.27b) we find

$$S^{\nu\alpha\beta} = s_\mu\epsilon^{\mu\nu\alpha\beta} = i(s \wedge \gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta) = (is)\cdot(\gamma^\nu \wedge \gamma^\alpha \wedge \gamma^\beta)\,. \tag{5.28}$$

The last expression shows that the $S^{\nu\alpha\beta}$ are simply tensor components of the pseudovector is. Contraction of (5.28) with $v_\nu = v\cdot\gamma_\nu$ and use of duality (1.16b) gives the desired relation between $S^{\nu\alpha\beta}$ and $S^{\alpha\beta}$:

$$v_\nu S^{\nu\alpha\beta} = i(s \wedge v \wedge \gamma^\alpha \wedge \gamma^\beta) = [\,i(s\wedge v)\,]\cdot(\gamma^\alpha \wedge \gamma^\beta) = S^{\alpha\beta}\,. \tag{5.29}$$

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

$$M = \frac{e\hbar}{2m}\,\psi\gamma_2\gamma_1\widetilde{\psi} = \frac{e\hbar}{2m}\,\rho e^{i\beta}e_2e_1 = \frac{e}{2m}\,\rho S e^{i\beta}\,. \tag{5.30}$$

The interpretation of M as magnetization comes from the Gordon decomposition considered in the next Section. Equation (5.30) 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 (5.30) 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 (5.16) 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 (5.24) to derive the general identity

$$i'\hbar\widetilde{\Psi}\Gamma\Psi = \widetilde{\Psi}\Gamma\gamma_\alpha\gamma_\beta\Psi S^{\alpha\beta}\,, \tag{5.31}$$

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 (5.18). 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.

6. 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 (4.1) is equally suited to the formulation of Maxwell's equation (4.4), 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 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

$$\underline{p}_\mu i = i'\hbar\partial_\mu - e\,A_\mu\,. \tag{6.1}$$

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 (5.20). 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 (5.5) on the right by $i\gamma_0\gamma_3\gamma_\mu\widetilde{\psi}$ and using (5.18) 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

$$\nabla\cdot(\rho e_\mu) = \frac{2}{\hbar}\,\rho e_\mu\cdot(e_3m\sin\beta + (e_2e_1)\cdot A)\,. \tag{6.2}$$

Thus we have the four equations

$$\nabla\cdot(\rho v) = \partial_\mu(\rho v^\mu) = 0\,, \tag{6.3}$$

$$\nabla\cdot(\rho s) = -m\sin\beta\,, \tag{6.4}$$

$$\nabla\cdot(\rho e_1) = \frac{2}{\hbar}\,\rho A\cdot e_2\,, \tag{6.5}$$

$$\nabla\cdot(\rho e_2) = -\frac{2}{\hbar}\,\rho A\cdot e_1\,. \tag{6.6}$$

Equation (6.3) 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 show 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

$$(\rho v) \cdot (\gamma_0 d^3 x) \,. \tag{6.7}$$

It is normalized so that

$$\int d^3x \, \rho_0 = 1 \,, \tag{6.8}$$

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

$$\rho_0 = \rho v_0 = (\rho v) \cdot \gamma_0 = (\psi \gamma_0 \widetilde{\psi} \gamma_0)_{(0)} = (\psi \psi^\dagger)_{(0)} \tag{6.9}$$

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 (6.3) 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

$$\frac{dx}{d\tau} = v(x(\tau)) \,. \tag{6.10}$$

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

$$\langle f \rangle = \int d^3x \, \rho_0 f \,. \tag{6.11}$$

To determine how this quantity changes with time, we use

$$\partial_\mu (\rho v^\mu f) = \rho v \cdot \nabla f = \rho \frac{df}{d\tau} = \rho_0 \frac{df}{dt} \,, \tag{6.12}$$

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 (6.12) in the useful integral form

$$\frac{d}{dt} \langle f \rangle = \int d^3x \, \rho v \cdot \nabla f = \left\langle \frac{df}{dt} \right\rangle \,. \tag{6.13}$$

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

$$\langle x \rangle = \int d^3x \, \rho_0 x \,, \tag{6.14}$$

and application of (6.13) gives the *average velocity*

$$\frac{d}{dt}\langle x\rangle = \int d^3x\,\rho v = \left\langle \frac{dx}{dt} \right\rangle . \tag{6.15}$$

To see that this is a sensible result, use the space-time splits (2.1a) and (2.6) to get

$$\langle x\rangle\gamma_0 = 1 + \langle \mathbf{x}\rangle \tag{6.16}$$

from (6.14), and

$$\frac{d}{dt}\langle x\rangle\gamma_0 = 1 + \langle \mathbf{v}\rangle \tag{6.17}$$

from, (6.15). Thus, we have

$$\frac{d}{dt}\langle \mathbf{x}\rangle = \langle \mathbf{v}\rangle = \left\langle \frac{d\mathbf{x}}{dt} \right\rangle . \tag{6.18}$$

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 (6.18) to the matrix formulation; with the result

$$\langle \mathbf{v}\rangle \cdot \boldsymbol{\sigma}_k = \langle \mathbf{v}\cdot\boldsymbol{\sigma}_k\rangle = \int d^3x\,\Psi^\dagger\alpha_k\Psi\,, \tag{6.19}$$

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 (6.19) 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 (6.19). Accordingly, the equivalence of STA representations to conventional operator representations exhibited in (6.19) 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,

$$J = e\psi\gamma_0\widetilde{\psi} = e\rho v\,. \tag{6.20}$$

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 (6.1) as kinetic energy-momentum operators. In the STA formulation they are defined by

$$\underline{p}_\mu = \underline{i}\hbar\partial_\mu - e\,A_\mu\,, \tag{6.21}$$

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

$$\underline{i}\,\psi = \psi\mathbf{i}\,. \tag{6.22}$$

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	$K_\mu = \frac{e}{m} (\widetilde{\psi} \underline{p}_\mu \psi)_{(0)} = \frac{e}{m} \widetilde{\Psi} \underline{p}_\mu \Psi$

The importance of (6.21) 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

$$\gamma^\mu \underline{p}_\mu \psi = \psi \gamma_0 m \,. \tag{6.23}$$

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 (5.22). 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 (5.24). 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:*

$$T^{\mu\nu} = T^\mu \cdot \gamma^\nu = (\gamma_0 \widetilde{\psi} \gamma^\mu \underline{p}^\nu \psi)_{(0)} \,. \tag{6.24}$$

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

$$T^\mu = T(\gamma^\mu) = T^{\mu\nu} \gamma_\nu \tag{6.25}$$

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

$$T(v) = v_\mu T^\mu = \rho p \,. \tag{6.26}$$

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, by comparison with (2.7c), will be recognized as defining the *relative* momentum in the electron rest frame.

From the definition (6.24) 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)

$$\partial_\mu T^\mu = J \cdot F \,, \tag{6.27}$$

where J is the Dirac charge current (6.20) and $F = \nabla \wedge A$ is the electromagnetic field. The right side of (6.27) is *exactly* the classical Lorentz force, so using (4.25) and denoting the electromagnetic energy-momentum tensor (4.24) by T^μ_{EM}, we can rephrase (6.27) as *the total energy-momentum conservation law*

$$\partial_\mu (T^\mu + T^\mu_{EM}) = 0 \,. \tag{6.28}$$

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

$$\partial_\mu (T^\mu \wedge x) = T^\mu \wedge \gamma_\mu - \partial_\mu T^\mu \wedge x \,. \tag{6.29}$$

To evaluate the first term on the right, we return to the definition (6.24) and find

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

Summing with γ_ν and using the Dirac equation (6.23) to evaluate the first term on the right while recognizing the spin vector (5.21) in the second term, we obtain

$$\gamma_\nu \gamma_\mu T^{\mu\nu} = m\psi\widetilde{\psi} + \nabla(\rho s i) \,. \tag{6.30}$$

By the way, the pseudoscalar part of this equation proves (6.4), and the scalar part gives the curious result

$$T^\mu{}_\mu = T^\mu \cdot \gamma_\mu = m \cos\beta \,. \tag{6.31}$$

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

$$T^\mu \wedge \gamma_\mu = T^{\mu\nu}\gamma_\mu \wedge \gamma_\nu = \nabla \cdot (\rho s i) = -\partial_\mu (\rho S^\mu) \,, \tag{6.32}$$

where

$$S^\mu = (is) \cdot \gamma^\mu = i(s \wedge \gamma^\mu) \tag{6.33}$$

is the spin angular momentum tensor already identified in (5.26) and (5.28). Thus from (6.29) and (6.27) we obtain the angular momentum conservation law

$$\partial_\mu J^\mu = (F \cdot J) \wedge x \,, \tag{6.34}$$

where

$$J(\gamma^\mu) = J^\mu = T^\mu \wedge x + \rho S^\mu \tag{6.35}$$

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

$$J(v) = \rho(p \wedge x + S)\,, \tag{6.36}$$

where recalling (2.12), $p \wedge x$ is recognized as the expected orbital angular momentum and as already advertised in (5.22), $S = isv$ 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

$$\partial_\mu R = \tfrac{1}{2}\,\Omega_\mu R\,, \tag{6.37}$$

where $\Omega_\mu = \Omega_\mu(x)$ is a bivector field. Consequently the derivatives of the e_ν defined by (5.18) have the form

$$\partial_\mu e_\nu = \Omega_\mu \cdot e_\nu\,. \tag{6.38}$$

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

$$\underline{p}_\nu \psi = [\,\partial_\nu(\ln\rho + i\beta) + \Omega_\nu\,]S\psi - eA_\nu\psi\,. \tag{6.39}$$

Whence

$$(\underline{p}_\nu\psi)\gamma_0\widetilde{\psi} = [\,\partial_\nu(\ln\rho + i\beta) + \Omega_\nu\,]i\rho s - eA_\nu v\,. \tag{6.40}$$

Inserting this in the definition (6.24) for the energy-momentum tensor, after some manipulations beginning with $is = Sv$, we get the explicit expression

$$T_{\mu\nu} = \rho\big[\,v_\mu(\Omega_\nu \cdot S - eA_\nu) + (\gamma_\mu \wedge v)\cdot(\partial_\nu S) - s_\mu\partial_\nu\beta\,\big]\,. \tag{6.41}$$

From this we find, by (6.26), the momentum components

$$p_\nu = \Omega_\nu \cdot S - eA_\nu\,. \tag{6.42}$$

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 8.

The kinematical import of the operator $\underline{p}_\nu$ is derived from its action on the rotor R. To make that explicit, use (6.37) and (5.22) to get

$$(\partial_\nu R)i\hbar\widetilde{R} = \Omega_\nu S = \Omega_\nu \cdot S + \Omega_\nu \wedge S + \partial_\nu S\,, \tag{6.43}$$

where (5.22) was used to establish that

$$\partial_\nu S = \tfrac{1}{2}[\,\Omega_\nu, S\,] = \tfrac{1}{2}(\Omega_\nu S - S\Omega_\nu)\,. \tag{6.44}$$

Introducing the abbreviation

$$iq_\nu = \Omega_\nu \wedge S\,, \qquad \text{or} \qquad q_\nu = -(iS)\cdot\Omega_\nu\,, \tag{6.45}$$

we can put (6.43) in the form

$$(\underline{p}_\nu R)\widetilde{R} = p_\nu + iq_\nu + \partial_\nu S\,. \tag{6.46}$$

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 (6.41), but q_ν does not. By the way, it should be noted that the last two terms in (6.41) 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 (6.23) for the Dirac charge current, we have

$$J = e\psi\gamma_0\widetilde{\psi} = \frac{e}{m}\,(\underline{p}_\mu\psi)\widetilde{\psi}\,. \tag{6.47}$$

The identity (6.46) is easily generalized to

$$(\underline{p}_\mu\psi)\widetilde{\psi} = (p_\mu + iq_\mu)\rho e^{i\beta} + \partial_\mu(\rho S e^{i\beta})\,. \tag{6.48}$$

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

$$K_\mu = \frac{e}{m}\,[\,(\underline{p}_\mu\psi)\widetilde{\psi}\,]_{(0)} = \frac{e}{m}\,(\widetilde{\psi}\underline{p}_\mu\psi)_{(0)} = \frac{e}{m}\,(p_\mu\rho\cos\beta - q_\mu\rho\sin\beta)\,. \tag{6.49}$$

Or in vector form,

$$K = \frac{e}{m}\,\rho(p\cos\beta - q\sin\beta)\,. \tag{6.50}$$

As anticipated in the last Section, from the last term in (6.48) we define the magnetization

$$M = \frac{e}{m}\,\rho S e^{i\beta}\,. \tag{6.51}$$

When (6.48) is inserted into (6.47), the pseudovector part must vanish, and vector part gives us the so-called *"Gordon decomposition"*

$$J = K + \nabla\cdot M\,. \tag{6.52}$$

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 (6.50) and (6.51). A curious fact is that β does not contribute to the definition (5.20) for the Dirac current in terms of the wave function; β is related to J only indirectly through the Gordon Relation (6.52). 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 (5.16) 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. A resolution of the Klein Paradox from this perspective has been given by Steven 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 (6.51). In accordance with (4.39), in the electron rest system determined by J, we can split M into

$$M = -\mathbf{P} + i\mathbf{M}\,, \tag{6.53}$$

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

$$i\mathbf{M} = \frac{e}{m}\, S\rho\cos\beta \tag{6.54}$$

is the *magnetic moment* density, while

$$\mathbf{P} = -\frac{e}{m}\, iS\rho\sin\beta \tag{6.55}$$

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 (6.54). It is tempting, also, to interpret equation (6.4) 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.

7. 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 possibilities for new physics at a deeper level. Indeed, in this Section it will be explained how particle trajectories can 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 argues 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 derives an equation of motion for the electron which differs from the classical equation only in a stochastic term called the "Quantum Force." He is careful, however, not to commit himself to any special hypothesis about the origins of the Quantum Force. He accepts the form of the force dictated by Schroedinger's equation. However, he takes 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 projected into 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] 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 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, 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 (5.16) can be expressed as a function of proper time

$$R = R(x(\tau))\,. \tag{7.1}$$

By (5.18), (5.20) and (6.10), this determines a *comoving frame*

$$e_\mu = R\,\gamma_\mu \widetilde{R} \tag{7.2}$$

on the streamline with velocity $v = e_0$, while the spin vector s and bivector S are given as before by (5.21) and (5.22). In accordance with (6.37), differentiation of (7.1) leads to

$$\dot{R} = v \cdot \nabla R = \tfrac{1}{2}\Omega R\,, \tag{7.3}$$

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

$$\Omega = v^\mu \Omega_\mu = \Omega(x(\tau)) \tag{7.4}$$

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

$$\dot{e}_\mu = v \cdot \nabla\, e_\mu = \Omega \cdot e_\mu\,. \tag{7.5}$$

But these equations are identical in form to those for the classical theory in Section 3. 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 Section 3, a limit in which the electron still has a nonvanishing spin of magnitude $\hbar/2$.

From (6.42) we immediately obtain

$$\Omega \cdot S = (p + eA) \cdot v = \tfrac{1}{2}\hbar\omega\,. \tag{7.6}$$

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

$$\omega = \frac{2m}{\hbar} = 1.6 \times 10^{21}\,\mathrm{s}^{-1}\,. \tag{7.7}$$

According to (7.6), this frequency will be altered by external fields.

Equation (7.6) is part of a more general equation obtained from (6.43):

$$\Omega S = (p + eA) \cdot v + i(q \cdot v) + \dot{S}\,. \tag{7.8}$$

As an interesting aside, this can be solved for

$$\Omega = \dot{S}S^{-1} + (q \cdot v)iS^{-1} + (p + eA) \cdot vS^{-1}\,, \tag{7.9}$$

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

$$\dot{v} = \Omega \cdot v = (\dot{S} \cdot v)S^{-1} - (q \cdot v)s^{-1}\,. \tag{7.10}$$

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

$$\Omega = -\nabla \wedge v + v \cdot (i\nabla\beta) + (m\cos\beta + eA \cdot v)S^{-1} \,. \tag{7.11}$$

But this tells us nothing about particle streamlines, since

$$\dot{v} = v \cdot (\nabla \wedge v) \tag{7.12}$$

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

$$-\nabla \wedge v + v \cdot (i\nabla\beta) = m^{-1}(eFe^{i\beta} + Q) \,, \tag{7.13}$$

where Q has the complicated form

$$Q = -e^{i\beta}[\partial_\mu W^\mu + \gamma_\mu \wedge \gamma_\nu (W^\mu W^\nu) S^{-1}]_{(0)} \,, \tag{7.14}$$

with

$$W_\mu = (\rho e^{i\beta})^{-1}\partial_\mu(\rho e^{i\beta}S) = \partial_\mu S + S\partial_\mu(\ln\rho + i\beta) \,. \tag{7.15}$$

Inserting (7.13) in (7.11), we get from (7.5) and (6.44) the equations of motion for velocity and spin:

$$m\dot{v} = e(Fe^{i\beta}) \cdot v + Q \cdot v \,, \tag{7.16}$$

$$\dot{S} = F \times \Big(\frac{e}{m}\, S e^{i\beta}\Big) + Q \times S \,, \tag{7.17}$$

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 (7.16) is the classical Lorentz force. The term $Q \cdot v$ is the generalization of Bohm's *Quantum Force.* A crucial fact to note from (7.15) 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, (7.14) reduces to $Q = \ddot{S}$, and for the limiting classical equations of motion for a particle with intrinsic spin we obtain[13]

$$m\dot{v} = (eF - \ddot{S}) \cdot v \,, \tag{7.18}$$

$$m\dot{S} = (eF - \ddot{S}) \times S \,. \tag{7.19}$$

These coupled equations have not been seriously studied. Of course, they should be studied in conjunction with the spinor equation (7.3).

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 (5.5) admits *plane wave* solutions of the form

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R = \rho^{\frac{1}{2}} e^{i\beta/2} R_0 e^{-\mathbf{i}p\cdot x/\hbar} \,, \tag{7.20}$$

where the kinematical factor R has been decomposed to explicitly exhibit its spacetime dependence in a phase factor. Inserting this into (5.5) and using $\nabla p \cdot x = p$, we obtain

$$p\psi = \psi\gamma_0 m \,. \tag{7.21}$$

Solving for p we get

$$p = me^{i\beta} R\gamma_0 \widetilde{R} = mve^{-i\beta}\,. \tag{7.22}$$

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

$$e^{i\beta/2} = 1 \text{ or } i\,, \tag{7.23}$$

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" (7.20).

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

$$\psi_- = \rho^{\frac{1}{2}} R_0 e^{-\mathbf{i}p\cdot x/\hbar}\,, \tag{7.24}$$

$$\psi_+ = \rho^{\frac{1}{2}} i R_0 e^{+\mathbf{i}p\cdot x/\hbar}\,. \tag{7.25}$$

We can identify these as *electron* and *positron* wave functions. Indeed, the two solutions are related by charge conjugation. According to (5.15), the charge conjugate of (7.24) is

$$\psi_-^C = \psi_- \boldsymbol{\sigma}_2 = \rho^{\frac{1}{2}} i R_0' e^{-\mathbf{i}p\cdot x/\hbar}\,, \tag{7.26a}$$

where

$$R_0' = R_0(-i\boldsymbol{\sigma}_2)\,. \tag{7.26b}$$

As seen below, the factor $-i\boldsymbol{\sigma}_2$ represents a spatial rotation which just "flips" the direction of the spin vector. Evidently (7.25) and (7.26a) are both positron solutions, but with oppositely directed spins.

Determining the comoving frame (7.2) for the electron solution (7.24), 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 $k = 1, 2$,

$$e_k(\tau) = e_k(0)e^{-p\cdot x/S} = e_k(0)e^{e_2 e_1 \omega\tau}\,, \tag{7.27}$$

where $\tau = v \cdot x$ and ω is given by (7.7). 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 (7.7) 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

$$R = LU \qquad \text{where} \qquad U = U_0 e^{-\mathbf{i}p\cdot x/\hbar}\,. \tag{7.28}$$

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

$$\boldsymbol{\sigma} = U\boldsymbol{\sigma}_3\widetilde{U}\,. \tag{7.29}$$

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

$$U_\pm \boldsymbol{\sigma}_3 = \pm\boldsymbol{\sigma}_3 U_\pm \tag{7.30}$$

or

$$U_\pm = \tfrac{1}{2}(U \pm \boldsymbol{\sigma}_3 U)\,. \tag{7.31}$$

Thus

$$U = U_+ + U_- \tag{7.32}$$

It follows that

$$U\widetilde{U} = |\,U_+\,|^2 + |\,U_-\,|^2 = 1\,, \tag{7.33}$$

$$U_+\widetilde{U}_- + U_-\widetilde{U}_+ = 0\,, \tag{7.34}$$

$$\boldsymbol{\sigma} = U\boldsymbol{\sigma}_3\widetilde{U} = \big\{|\,U_+\,|^2 - |\,U_-\,|^2\big\}\boldsymbol{\sigma}_3 + 2U_-\widetilde{U}_+\boldsymbol{\sigma}_3\,. \tag{7.35}$$

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

$$\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_3 = |\,U_+\,|^2 - |\,U_-\,|^2\,, \tag{7.36}$$

$$\boldsymbol{\sigma}_3\times\boldsymbol{\sigma} = 2iU_-\widetilde{U}_+\,. \tag{7.37}$$

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

$$\psi = \psi_0 e^{-\mathbf{i}\varphi/\hbar}\,. \tag{7.38}$$

Then the "amplitude" ψ_0 is assumed to be slowly varying compared to "phase" φ, so the derivatives of ψ_0 in the Dirac equation can be neglected to a good approximation. Thus, inserting (7.38) into the Dirac equation, say in the form (6.47), we obtain

$$(\nabla\varphi - eA)e^{i\beta} = mv\,. \tag{7.39}$$

As in the plane wave case (7.22) this implies $e^{i\beta} = \pm 1$, and the two values correspond to electron and positron solutions. For the electron case,

$$\nabla\varphi - eA = mv\,. \tag{7.38}$$

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"

$$(\nabla\varphi - eA)^2 = m^2\,, \tag{7.39}$$

obtained by squaring (7.38). On the other hand, the curl of (7.38) gives

$$m\nabla\wedge v = -e\nabla\wedge A = -eF \tag{7.40}$$

Dotting this with v and using the identity (7.12), we obtain exactly the classical equation (3.6) for each streamline.

Inserting (7.40) into (7.11), we obtain

$$\Omega = \frac{e}{m}\,F + (m + eA\cdot v)S^{-1}\,. \tag{7.41}$$

Whence the rotor equation (7.3) assumes the explicit form

$$\dot{R} = \frac{e}{2m}\,FR - R\mathbf{i}(m + eA\cdot v)/\hbar\,. \tag{7.42}$$

This admits a solution by separation of variables:

$$R = R_0 e^{-\mathbf{i}\varphi/\hbar}\,, \tag{7.43}$$

where

$$\dot{R} = \frac{e}{2m}\,FR_0 \tag{7.44}$$

and

$$\dot{\varphi} = v \cdot \nabla\varphi = m + eA \cdot v\,. \tag{7.45}$$

Equation (7.44) is identical with the classical equation in Section 3, while (7.45) can be obtained from (7.38).

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 (7.41) into (7.1), with the interesting result

$$p \cdot v = m + \frac{e}{m}\,F \cdot S\,. \tag{7.46}$$

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

The eikonal solutions characterized above are exact solutions of the Dirac equation when the ψ_0 in (7.38) satisfies

$$\nabla\psi_0 = 0\,. \tag{7.47}$$

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 (7.47) 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] We have already found the classical solution for this case, namely, the rotor (3.18a, b). Identifying this solution with R_0 in (7.43), with the help of (3.17) it is readily verified that $\nabla R_0 = 0$. All that remains, then, is to determine the phase factor $\varphi(x)$. This is easily done by integrating (7.45) or solving the Hamilton-Jacobi equation (7.39).

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 (7.18). Evidently the spin dependence appears when the $\nabla\varphi$ in (7.38) is generalized to a vector field with nonvanishing curl.

8. 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 (5.16) 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 (7.6) and (7.46), 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*

$$u = v - e_2 = e_0 - e_2\,. \tag{8.1}$$

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 (7.24) can be written in the form

$$R = e^{\frac{1}{2}\Omega\tau}R_0\,, \tag{8.2}$$

where Ω is the constant bivector

$$\Omega = mc^2S^{-1} = \frac{2mc^2}{\hbar}\,e_1e_2\,. \tag{8.3}$$

From (8.2) it follows that v is constant and

$$e_2(\tau) = e^{\Omega\tau}e_2(0)\,. \tag{8.4}$$

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

$$z(\tau) = v\tau + (e^{\Omega\tau} - 1)r_0 + z_0\,, \tag{8.5}$$

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

$$|\,r_0\,| = |\,\Omega^{-1}\,| = \frac{\hbar}{2m} = 1.9\times10^{-13}\mathrm{m}\,. \tag{8.6}$$

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

$$(m\dot{r})\wedge r = m\dot{r}r = mr^2\Omega = m\Omega^{-1} = S\,. \tag{8.7}$$

Finally, if z_0 is varied parametrically over a hyperplane normal to v, equation (8.5) 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 (8.1) 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

$$v = \bar{u} \,. \tag{8.8}$$

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 (6.6) we have

$$\nabla \cdot (\rho u) = \frac{2}{\hbar}\, \rho A \cdot e_1 \,. \tag{8.9}$$

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 $\underset{\_}{p}_\mu$ defined by (6.21). 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 $\underset{\_}{p}_\mu$ 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" $\underset{\_}{p}_\mu$.

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

$$\psi = \rho^{\frac{1}{2}} e^{i\beta/2} L U e^{-\mathbf{i}(m/\hbar)t} \,. \tag{8.10}$$

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

$$\psi_P = \rho^{\frac{1}{2}} U_0 e^{-\mathbf{i}(\varphi/\hbar)} , \tag{8.11}$$

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

$$\psi_S = \rho^{\frac{1}{2}} e^{-\mathbf{i}(\varphi/\hbar)} . \tag{8.12}$$

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 in stationary states.* This opens the possibility of *zbw resonance* as a fundamental explanatory principle in quantum mechanics. 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 (Ref. Uzer etc.), often surpasing 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.

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

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 Ref. 11.

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. One important clue concerns the origin of electron

mass. The very form of the important equation (7.6) suggests that the electron's mass may be a consequence of *magnetic self-interaction*, as expressed by writing

$$m = S \cdot \Omega = \boldsymbol{\mu} \cdot \mathbf{B}_S \,, \tag{8.13}$$

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.

9. 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* **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 **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:

$$\psi \to \psi G \,, \tag{9.1}$$

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

$$G\gamma_0\widetilde{G} = \gamma_0 \,. \tag{9.2}$$

It follows that

$$G = Ue^{i\lambda} \,, \tag{9.3}$$

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

$$G\gamma_3\widetilde{G} = \gamma_3 \,. \tag{9.4}$$

The E-M gauge transformations belong to this subgroup. Note also that the U(1) factor in (9.3) is a duality factor exactly like the one parametrized by $\beta/2$ in the invariant decomposition of the Dirac wave function (5.16). 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.

10. 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

$$\psi = (\rho e^{i\beta})^{\frac{1}{2}} R\,. \tag{10.1}$$

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

$$S = \tfrac{1}{2} R(\mathbf{i}\hbar)\widetilde{R}\,. \tag{10.2}$$

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

$$p_\nu = \Omega_\nu \cdot S - eA_\nu\,, \tag{10.3}$$

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 (10.2) is the angular momentum of the zitterbewegung, and (10.3) 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 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.* (October, 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, SIAM *J. Appl. Math.* **21**, 14–144 (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, The Klein Paradox, (to be published).

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, *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).

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 f of a vector field $a = a(x)$ on $\mathcal{R}$ into a vector field $a' = a'(x')$ on $\mathcal{R}'$ defined by

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

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

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

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

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

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

$$M\cdot\nabla = \sum_{k=0}^{4}\underline{f}(M)_{(k)}\nabla\,. \tag{A.12}$$

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, using the duality relations (1.16 a,b), we 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

$$a' = \underline{f}a = Ra\widetilde{R}\,. \tag{A.17}$$

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

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

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 (4.4) 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 (5.5) 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 (5.2),

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

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 (5.7) through (5.11), 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,

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 Part II reveals that the success of this formal procedure should be attributed to the physical interpretation of the Dirac wave function rather than a general physical significance of hermitian operators.

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

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

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 advantage of directly relating equations of motion 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.

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 $x \to x' + \epsilon(x)$, (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

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

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

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

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

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

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

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

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 (5.5) 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 (A.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

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

It will be helpful to reformulate this in terms of the energy-momentum operators $\underline{p}_\mu$. From the definition (6.21) we have

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

So from the definition of the energy-momentum tensor T^μ in (6.24) or Table II, we obtain

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

Consequently,

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

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 (6.31) that this vanishes in consequence of the Dirac equation. Finally, we note that the last term in (B.16) can be written

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

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

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

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

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

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 (6.27). 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 (B.31), $\delta A = B \times A = B \cdot A$ by (A.32), and $\delta \psi = (1/2) B \psi$ by (A.33). In consequence, note the following:

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

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

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

where $S^\mu = i(s \wedge \gamma^\mu)$ is the spin angular momentum tensor of (6.33). 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 (6.34). 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 $\frac{1}{2} \langle F^2 \rangle$, where $F = \nabla \wedge A$. The electromagnetic part of the Lagrangian is then

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

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

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

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

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

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 conseration 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 (6.27) and (6.34) for the electron, the internal forces and torques cancel.