Why *i*?

W. E. Baylis, J. Huschilt, and Jiansu Wei
Department of Physics, University of Windsor, Windsor, Ontario N9B 3P4, Canada

(Received 7 October 1991; accepted 19 March 1992)

How is it that complex numbers, involving the imaginary $i=\sqrt{-1}$ play such an important role in physics, which always measures real quantities? An answer can be given in the framework of the vector algebra of three-dimensional space, in which an associative, invertible product of vectors is defined. In this mathematical structure, also known as the Pauli algebra, i arises naturally and carries geometrical significance. In particular, i enters as the unit volume element, and imaginary vectors are pseudovectors which represent planes, such as planes of rotation or reflection. The i from the vector algebra is related to common applications of imaginary numbers in physics, including rotations in a plane, electromagnetic waves, and phase factors and operator relations in quantum mechanics. Moreover, the same algebra of real three-dimensional vectors which yields complex numbers also forms the basis for a complex four-dimensional space with the Minkowski metric and provides a natural formalism for compact, covariant treatments of relativistic phenomena.

I. INTRODUCTION

Physics is a human discipline designed to describe the real physical world. It has been remarkably successful at modeling phenomena and their observation and measurement with mathematics. Through a single small set of universal laws, physics accurately relates events from the smallest scale of quark interactions to the largest scale of the structure of the universe. The beauty with which often highly abstract mathematics unifies the real world is a continuing source of joy and wonder for the practicing physicist.

In fact, it may work too well! One's wonder verges on puzzlement. A simple but important example of applicable abstraction is the use of imaginary numbers in physics. Although imaginary numbers were rejected by mathematicians as prominent as René Descartes (1596–1650),[1] Gottfried Leibniz (1646–1716) called them "a fine and wonderful refuge of the divine spirit, almost an amphibian between being and non-being."[2] The imaginary $i=\sqrt{-1}$ enters mathematics through solutions to quadratic equations such as $x^2+1=0$, but many of its uses in electronics, electromagnetic theory, and quantum mechanics transcend these modest beginnings to involve mathematical structures over the complex field $\mathbb{C}$. Yet the events physics seeks to describe occur, as far as we understand, in real three-dimensional space R^3, and the input and output of any physical theory should reflect measurements of only real quantities.

Often, it may be argued, complex numbers are introduced for mathematical convenience and require little explanation. Thus even though the physicist may ultimately be interested in a function of real values, she/he may be able to learn more about the asymptotic properties of the function and be able to integrate it more simply if it is "continued" into the complex field. Furthermore, a complex number can be represented as a point in two-dimensional space, and there is no problem reversing this procedure to represent the position in a plane by a complex number—no problem that is until we start using powers and other algebraic properties of the numbers and expect them still to have meaning in terms of the position in the plane. In many cases, imaginary numbers can be exorcised from the mathematics with conciseness as the only sacrifice, but sometimes, as in quantum theory, i plays such a central role that there is no way to make sense of equations separated into real and imaginary parts.

In this paper, we seek some of the origins of the ubiquitous i in the vector algebra of three-dimensional space. Imaginary numbers and the complex field $\mathbb{C}$ are seen to arise naturally when one attempts to develop more powerful methods of manipulating vectors. Most physicists have an acquaintance with the *geometric algebra* (also known as the "multivector" or "Clifford" algebra) of space-time through its representation in terms of the Dirac gamma matrices of relativistic quantum theory. Although the importance of geometric algebras has been recognized for some time in quantum field theory, there has recently also been a movement to exploit their power in classical physics.[3–13] We concentrate here on the relatively simple geometric algebra of three-dimensional space, also known as the Pauli algebra. The standard matrix representation of this algebra replaces basis vectors by 2×2 Pauli spin matrices, and many physicists will be familiar with some applications of the algebra in this form.[10] However, specific representations encumber the mathematics with unnecessary baggage, and we find it simpler to work directly in the algebra without reference to any matrices.

The present article is similar in many respects to an article two decades ago by Hestenes,[4] whose many contributions have pioneered the development and geometrical interpretation of the Dirac and Pauli algebras in physics. In his 1971 paper, Hestenes showed how complex numbers arise in the Pauli algebra and how the algebra unites them with vectors and spinors. The message bears repeating even today, but our emphasis is somewhat different. We emphasize here how common uses of i in traditional physics can be explained in terms of the geometric i of the Pauli algebra. However, a more important difference between the present contribution and Ref. 4 concerns the treatment and significance of inverse elements. We show the involution of "spatial reversal" to be proportional to the inverse for any invertible element of the algebra, and in the product of an element with its spatial reverse, we recognize the Minkowski-space norm of a four-vector. This leads to a natural formalism for relativistic phenomena: the Pauli algebra is able to provide a covariant description of physics while maintaining the qualitative difference between time and space variables in any given frame. Even in the quantum Dirac theory, the usual results can be efficiently obtained within the framework of the Pauli algebra. It is therefore possible to avoid the introduction of a more general geometric-algebra formulation, with its often unfamiliar products and its allowance for a noncommuting volume element (which in the Dirac algebra has more in common with a fifth dimension than with the imaginary i). Our treatment stays close to the traditional vector and scalar operations and, as a result, should be readily accessible to the average physics reader.

Section II introduces the Pauli algebra, and Sec. III relates the algebraic product of vectors to the usual dot and cross products, and shows how the trivector element of the algebra plays the role of i. Its geometrical interpretation is discussed in the following section, and its role in rotations and reflections is emphasized with examples in Sec. V. Some traditional uses of i are best explained by looking at the subalgebras of the Pauli algebra, which we do in Sec. VI, and in Secs. VII and VIII we then study applications in special relativity and electromagnetic theory. The paper finishes by showing how the i in quantum theory can be associated with spin.

II. VECTOR ALGEBRA

The description of motion in three-dimensional (Euclidean) space R^3 requires many vectors: relative positions, velocities, accelerations, forces, electric and magnetic fields, and others are part and parcel of the familiar merchandise of physics. Every vector of R^3 can be written as a linear combination

$$\mathbf{a}=a^j\mathbf{e}_j \tag{1}$$

(the convention of summing over repeated indices is adopted here) of the orthonormal basis vectors $\mathbf{e}_j$, $j=1,2,3$:

$$\mathbf{e}_j\cdot\mathbf{e}_k=\delta_{jk}. \tag{2}$$

Traditional vector notation in physics can be traced back to J. Willard Gibbs (1839–1903) and Oliver Heaviside (1850–1925);[14] it relies on the rather restricted manipulations of vectors required in a vector space with a Euclidean inner product: vector addition $\mathbf{a}+\mathbf{b}$, the dot or inner product $\mathbf{a}\cdot\mathbf{b}=a^jb^k\mathbf{e}_j\cdot\mathbf{e}_k=|\mathbf{a}|\,|\mathbf{b}|\cos\theta$, and multiplication $\lambda\mathbf{a}$ of vectors $\mathbf{a}$ by scalars λ. One additional product has been added to the standard repertoire of vector calculations, namely the cross product

$$\mathbf{a}\times\mathbf{b}=a^jb^k\epsilon_{jkl}\mathbf{e}_l, \tag{3}$$

where ϵ_{jkl} is the usual Levi-Civita symbol which is $+1(-1)$ if j, k, l is a cyclic (anticyclic) permutation of 1,2,3 and vanishes if any two of the indices are equal. Other products of vectors are traditionally represented by tensors of Cartesian components, in which, for example, components of the cross product are elements of an antisymmetric second-rank tensor $a^jb^k-b^ja^k$. Of course the use of components requires the specification, at least implicitly, of an axis system and defeats some of the beauty of the component-free Gibbsian vectors.

The manipulations defined in a vector space are quite limited compared to the those available for scalars. William Rowan Hamilton (1805–1865), in his development of quaternions,[15] sought an algebra which would give to vector computations much of the manipulative power one has with real numbers. We follow here the somewhat more general approach of William Kingdon Clifford (1845–1879), who was able to unite the efforts of Hamilton and the German mathematician Herman Grassmann (1809–1877).

The essential ingredient missing from the vector calculations of Gibbs and Heaviside is an associative, invertible product of vectors which is distributive over vector addition. Neither the dot nor the cross product by itself is invertible, since knowing both $\mathbf{a}$ and $\mathbf{a}\cdot\mathbf{b}$ (or both $\mathbf{a}$ and $\mathbf{a}\times\mathbf{b}$) does not allow one to determine $\mathbf{b}$. Further, neither product is associative. Let us postulate an associative, invertible product, which for simplicity we indicate by $\mathbf{ab}$, and try to work out its properties. We will *not* assume the product to be commutative: $\mathbf{ab}$ may not equal $\mathbf{ba}$. It is sufficient to find the relations for products of the three unit basis vectors $\mathbf{e}_j$ since those for linear combinations thereof will then follow easily. Later, we can make the notation component free.

It's useful to anticipate some results. In three-dimensional space, we can divide elements into four geometric classes: scalars (elements of zero dimensions), lines (one-dimensional elements), surfaces (two-dimensional elements), and volumes (three-dimensional elements). Directed lines are, of course, just vectors. Algebraic products of two vectors could contain information about the plane in which the vectors lie and an area determined by them. Products of three vectors could contain information about the volume, say, of the parallelepiped whose edges are given by the vectors.

III. ALGEBRAIC PRODUCTS OF VECTORS

To determine the basic rule for vector products, it suffices to consider the product of a vector with itself. The square of a real vector $\mathbf{a}$ cannot contain a vector part, because the only direction singled out is that of $\mathbf{a}$ itself, but $\mathbf{a}^2=(-\mathbf{a})^2$ so that there is no reason to choose $\mathbf{a}$ over $-\mathbf{a}$. Furthermore, no plane or volume containing $\mathbf{a}$ is uniquely specified by $\mathbf{a}^2$. We must evidently choose $\mathbf{a}^2$ to be a scalar, and the obvious choice is

$$\mathbf{a}^2\equiv a^ja^k\mathbf{e}_j\mathbf{e}_k=\mathbf{a}\cdot\mathbf{a}=a^ja^j. \tag{4}$$

This constraint gives the basic multiplication rule for the vector algebra. Since it holds for any vector components a^ja^k, we can equate the factors multiplying $a^ja^k(=a^ka^j)$:

$$\mathbf{e}_j\mathbf{e}_k+\mathbf{e}_k\mathbf{e}_j=2\delta_{jk}. \tag{5}$$

The product of any number of vectors is completely determined by the rule (5). The product of two orthogonal vectors is called a *bivector* and, as discussed below, gives information about the plane in which the vectors lie. There are just three linearly independent bivectors which can be formed from products of the $\mathbf{e}_j$: $\mathbf{e}_1\mathbf{e}_2$, $\mathbf{e}_2\mathbf{e}_3$, and $\mathbf{e}_3\mathbf{e}_1$. In the product of three orthonormal basis vectors, if any two of the vectors are the same, the result is $\pm$ the remaining one:

$$\mathbf{e}_1\mathbf{e}_1\mathbf{e}_2=\mathbf{e}_2=-\mathbf{e}_1\mathbf{e}_2\mathbf{e}_k \tag{6}$$

and so on. In three dimensions, the only triple products of basis vectors which are not linear combinations of the basis vectors themselves are permutations of the trivector $\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3\equiv\eta$, which may be interpreted as the signed volume of the unit cube of sides $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$. That's it! That exhausts the possible products of basis vectors. Any product of four basis vectors of three-dimensional space must contain at least two identical factors and can, with the help of (5), be re-expressed as a product of two basis vectors. Similarly, all higher-order products of the $\mathbf{e}_j$ can be reduced to linear combinations of the 8 *basis forms* $\{1;\ \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3;\ \mathbf{e}_1\mathbf{e}_2, \mathbf{e}_2\mathbf{e}_3, \mathbf{e}_3\mathbf{e}_1;\ \eta\}$.

Note that the three distinct bivectors can be written as products of the trivector η with a vector, for example

$$\mathbf{e}_1\mathbf{e}_2=\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3\mathbf{e}_3=\eta\mathbf{e}_3=\eta\epsilon_{12k}\mathbf{e}_k.$$

More generally,

$$\mathbf{e}_j\mathbf{e}_k-\mathbf{e}_k\mathbf{e}_j=2\eta\epsilon_{jkl}\mathbf{e}_l. \tag{7}$$

The vector $\epsilon_{jkl}\mathbf{e}_l$ is called the (*Hodge*) *dual* of the bivector $\mathbf{e}_j\mathbf{e}_k$, $j\neq k$ and is normal to the surface represented by the bivector.[16] Although the vectors and bivectors do not generally commute with each other, one can verify that the trivector η does commute with all the basis forms in $\mathscr{P}$, for example by (5)

$$\mathbf{e}_1\eta=\mathbf{e}_1\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3=-\mathbf{e}_1\mathbf{e}_2\mathbf{e}_1\mathbf{e}_3=\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3\mathbf{e}_1=\eta\mathbf{e}_1. \tag{8}$$

Similarly, η commutes with any vector and hence with any product of vectors. Furthermore,

$$\eta^2=\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3=\mathbf{e}_1\mathbf{e}_2\mathbf{e}_1\mathbf{e}_2=-1. \tag{9}$$

In the vector algebra of three-dimensional space, the trivector η has the same properties as the imaginary i, and we can therefore use the symbols interchangeably:[17] $\eta=i$. This is how the imaginary i arises in the Pauli algebra.[18]

From products of three-dimensional vectors, we have thus generated an eight-dimensional real vector space $\mathscr{V}$ whose elements are real linear combinations of eight basis forms in four subspaces:

one scalar $(1\epsilon\mathscr{V}_0)$,

three basis vectors $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\epsilon\mathscr{V}_1)$,

three bivectors $(i\mathbf{e}_1,i\mathbf{e}_2,i\mathbf{e}_3\epsilon\mathscr{V}_2)$,

and

the trivector $(\eta=i\epsilon\mathscr{V}_3)$.

Since we can identify $\mathscr{V}_3=i\mathscr{V}_0$ and $\mathscr{V}_2=i\mathscr{V}_1$, the eight-dimensional real space $\mathscr{V}=\mathscr{V}_0\oplus\mathscr{V}_1\oplus\mathscr{V}_2\oplus\mathscr{V}_3$ can also be considered a four-dimensional complex vector space. Every element is then a complex linear combination of the four basis elements $\{1,\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ of $\mathscr{V}_0\oplus\mathscr{V}_1$. The algebra of real three-dimensional vectors thus generates a complex four-dimensional space and provides a natural introduction of complex numbers into the real world of physics.

By adding (5) and (7) we can also express the product rule by

$$\mathbf{e}_j\mathbf{e}_k=\delta_{jk}+i\epsilon_{jkl}\mathbf{e}_l. \tag{10}$$

Readers may recognize this rule as a relation obeyed by the 2×2 Pauli spin matrices

$$\underline{\sigma}_1=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad \underline{\sigma}_2=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\quad \underline{\sigma}_3=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \tag{11}$$

Indeed, the Pauli matrices provide a lowest-dimensional faithful matrix representation of the algebra and give the algebra its name: the *Pauli algebra*. The representation is realized simply by replacing every unit vector $\mathbf{e}_k$ by the corresponding matrix $\underline{\sigma}_k$. Of course, many other representations are possible. The algebra is more general than any one of them, and by avoiding any specific matrix representation, we can express the elements of the algebra in a coordinate-free form, independent of any coordinate axes. Thus our use of the term "Pauli algebra" refers to the algebraic structure, the geometric algebra of three-dimensional Euclidean space, and not to any particular matrix representation.

IV. GEOMETRICAL SIGNIFICANCE

Thus any element p of the Pauli algebra $\mathscr{P}$ can be written as the linear combinations

$$p=p^\mu e_\mu+q^\nu ie_\nu=(p^\mu+iq^\mu)e_\mu, \tag{12}$$

where for simplicity we have put $e_k=\mathbf{e}_k$, $k=1,2,3$, and $e_0=1$. The association of the imaginary i with the volume element of the vector algebra of three-dimensional space lends geometrical meaning to imaginary numbers and vectors, and the physical significance of (12) is clearer if we consider p to be the sum $p=p_0+\mathbf{p}+i\mathbf{q}+iq_0$ of a scalar $p_0\equiv p^0$, a vector $\mathbf{p}=p^k\mathbf{e}_k$, a bivector or *pseudovector* $i\mathbf{q}=iq^j\mathbf{e}_j$, and a trivector or *pseudoscalar* $iq_0\equiv iq^0$. These four parts are distinguished by their behavior under rotation and spatial inversion. Rotation affects both vectors and pseudovectors in the same way and leaves scalars and pseudoscalars invariant. Spatial inversion, on the other hand, replaces the unit vectors $\mathbf{e}_j$ by $-\mathbf{e}_j$ and thus changes $i=\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ to $-i$; it consequently changes right-handed coordinate systems into left-handed ones and reverses the sign of vectors and pseudoscalars but not of scalars and pseudovectors.

Scalars include common physical quantities like charge, mass, length, density, time, and energy which are unchanged by either rotations or inversions.[19] Relative positions, momenta, velocities, forces, acceleration, and electric fields are examples of vectors. Pseudovectors in $\mathscr{P}$ are bivectors and represent planes, such as planes of rotation and planes of reflection. Cross products of vectors, such as angular momenta and magnetic fields, are the duals of pseudovectors: they point in a direction normal to the surface represented by the pseudovector. The dual vector to a plane of rotation lies along the axis of rotation. We call the dual to a bivector (in three-dimensional space) an *axial* vector.[20] Pseudoscalars represent signed volumes in three-dimensional space. For example, the dot product of the average velocity (a vector) in m/s at some position down a pipe of a confined fluid with the cross-sectional area (a pseudovector) in m^2 of the pipe at that position gives a pseudoscalar representing the volume flow in m^3/s.

V. EXAMPLES

A. Dot and cross products

Consider some simple examples. (These and many others have been presented in detail elsewhere; see especially Ref. 5.) Let $\mathbf{a}$, $\mathbf{b}$ be any two vectors. From the general multiplication rule (10) of the algebra, it follows that

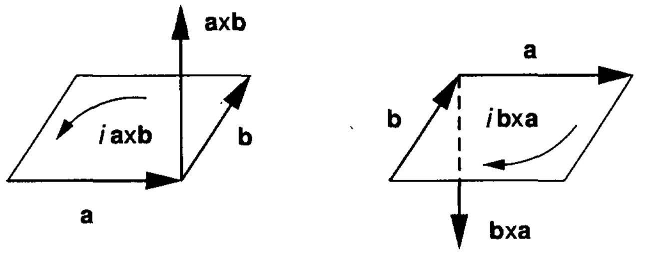

Fig. 1. The bivectors $i\mathbf{a}\times\mathbf{b}$ and $i\mathbf{b}\times\mathbf{a}=-i\mathbf{a}\times\mathbf{b}$ represent plane areas of opposite orientation whose edges are parallel to the vectors **a** and **b**. In three dimensions, they are also pseudovectors and their duals are axial vectors $\mathbf{a}\times\mathbf{b}$ and $\mathbf{b}\times\mathbf{a}$ normal to the plane. The orientation of the plane is related to the sense of circulation of **a** and **b** around its periphery.

$$\begin{aligned}\mathbf{ab}&=\mathbf{a}\cdot\mathbf{b}+i\mathbf{a}\times\mathbf{b}\\&=|\mathbf{a}|\,|\mathbf{b}|(\cos\theta+i\hat{\theta}\sin\theta)\\&=|\mathbf{a}|\,|\mathbf{b}|\exp(i\boldsymbol{\theta}),\end{aligned}\tag{13}$$

where θ is the angle from **a** to **b** and the direction $\hat{\theta}$ (the "hat" indicates a unit vector) is parallel to $\mathbf{a}\times\mathbf{b}$. The scalar part of the product is the dot product, which is symmetric in the two factors **a** and **b**:

$$\mathbf{a}\cdot\mathbf{b}=(\mathbf{ab}+\mathbf{ba})/2.\tag{14}$$

The bivector or pseudovector part gives the orientation of the plane formed by **a** and **b** and its magnitude is equal to the area of the parallelogram bordered by **a** and **b** (see Fig. 1). It is the antisymmetric part of the product:

$$i\mathbf{a}\times\mathbf{b}=(\mathbf{ab}-\mathbf{ba})/2.\tag{15}$$

The vector dual to the plane is the cross product $\mathbf{a}\times\mathbf{b}$. From (13), parallel vectors commute whereas perpendicular ones anticommute:

$$\begin{aligned}&\mathbf{a}\parallel\mathbf{b}\Leftrightarrow\mathbf{ab}=\mathbf{ba},\\&\mathbf{a}\perp\mathbf{b}\Leftrightarrow\mathbf{ab}+\mathbf{ba}=0.\end{aligned}\tag{16}$$

B. Algebraic functions of vectors

The exponential expression in (13) is meaningful because in the vector algebra we can easily calculate any power of a vector, and thus any analytic function f of it:

$$\mathbf{a}^n=\begin{cases}|\mathbf{a}|^n, & n \text{ even},\\ |\mathbf{a}|^n\hat{\mathbf{a}}, & n\text{ odd},\end{cases}$$

$$f(\mathbf{a})=f_+(|\mathbf{a}|)+\hat{\mathbf{a}}f_-(|\mathbf{a}|),\tag{17}$$

where $f_\pm(x)=\frac{1}{2}[f(x)\pm f(-x)]$ is the even or odd part of $f(x)$. Such a function is an element of the algebra with, in general, both scalar and vector parts, both of which may be complex. The scalar part is the even part of the function, and the vector part, which is parallel to the vector argument, is the odd part.[21]

Functions of vectors are often useful even when the function cannot be expanded in a series of positive powers. An important example is the *inverse* $\mathbf{a}^{-1}$ of a vector. Since $\mathbf{a}^2$ is a scalar, the vector

$$\mathbf{a}^{-1}=\mathbf{a}/\mathbf{a}^2\tag{18}$$

is evidently the inverse of **a**. As long as $\mathbf{a}^2\neq 0$, relation (17) can be extended to negative integer power.

C. Rotations in a plane

The physical significance of $\widehat{\mathbf{a}\mathbf{b}}=\exp(i\boldsymbol{\theta})$ becomes clear if we multiply it from the left by any vector **r** lying in the $i\boldsymbol{\theta}$ plane:

$$\begin{aligned}\mathbf{r}\exp(i\boldsymbol{\theta})&=\exp(-i\boldsymbol{\theta})\mathbf{r}\\&=(\cos\theta-i\hat{\theta}\sin\theta)\mathbf{r}\\&=\mathbf{r}\cos\theta+\hat{\theta}\times\mathbf{r}\sin\theta.\end{aligned}\tag{19}$$

The result describes the vector **r** after a rotation: one can regard $\widehat{\mathbf{a}\mathbf{b}}=\exp(i\boldsymbol{\theta})$ as the operator which rotates any vector in the plane $i\boldsymbol{\theta}$ by the angle θ which separates $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ in the plane. As a special case, it rotates $\hat{\mathbf{a}}$ into $\hat{\mathbf{b}}$:

$$\hat{\mathbf{a}}\exp(i\boldsymbol{\theta})=\hat{\mathbf{a}}\widehat{\mathbf{a}\mathbf{b}}=\hat{\mathbf{b}},\tag{20}$$

but more generally, it rotates any vector in the plane of $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ by θ.

The axial vector dual to the plane of rotation $i\hat{\theta}$ is $\hat{\theta}$, the *axis* of rotation. Successive rotations about the same axis $\hat{\theta}$ commute and combine by adding the angles: $\exp(i\theta_1\hat{\theta})\exp(i\theta_2\hat{\theta})=\exp[i(\theta_1+\theta_2)\hat{\theta}]$. The usual trigonometric identities for the sine and cosine of a sum of angles is obtained by expanding both sides of this result as in (19). This natural algebraic formalism for rotating vectors in a plane should look familiar, since it is closely related to the Wessel–Argand diagrams of complex scalars, which is frequently used to describe such rotations. The precise relationship between the two formalisms will be established in the next section, but first we want to show how the algebraic formalism, because it contains information about the rotation axis and the orientation of the rotation plane, can be extended to three-dimensional rotations.

D. Rotations in three dimensions

If the vector **r** does not lie in the rotation plane $i\hat{\theta}$, then we rotate only that part $\mathbf{r}_\perp\equiv\mathbf{r}-\mathbf{r}_\parallel$ perpendicular to $\hat{\theta}$:

$$\mathbf{r}\to\mathbf{r}_\parallel+\exp(-i\boldsymbol{\theta})\mathbf{r}_\perp=\exp(-i\boldsymbol{\theta}/2)\mathbf{r}\exp(i\boldsymbol{\theta}/2).\tag{21}$$

This is the general algebraic expression for rotations in three-dimensional space. Any three-dimensional rotation can be represented by elements $\pm\exp(-i\boldsymbol{\theta}/2)$ of the group SU(2). The product of any two rotations is another rotation, whose angle and rotation axis can be found directly by expanding the exponentials as in (19). The noncommutivity of rotations about different axes is seen to be a result of the noncommutivity of vectors in different directions.

E. Rotating frames

The algebraic formalism for rotations is well suited to handling transformations to and from rotating systems, where $\boldsymbol{\theta}=\boldsymbol{\omega}t$. Thus if **r** is any vector in a frame rotating at constant angular velocity $\boldsymbol{\omega}$, the corresponding vector in the laboratory frame is

$$\mathbf{r}'=\exp(-i\boldsymbol{\omega}t/2)\mathbf{r}\exp(i\boldsymbol{\omega}t/2)\tag{22}$$

and its time derivative is

$$\dot{\mathbf{r}}' = \exp(-i\omega t/2)[-(i/2)[\omega,\mathbf{r}]+\dot{\mathbf{r}}]\exp(i\omega t/2)$$

$$=\exp(-i\omega t/2)(\omega\times\mathbf{r}+\dot{\mathbf{r}})\exp(i\omega t/2). \quad (23)$$

Although this relation is traditionally written without the explicit rotation operators, their presence accurately describes the relation between the frames and is essential for unambiguous treatments. Higher-order derivatives are similarly calculated.

The algebra of rotations can also give concrete descriptions of, say, cycloid motion:

$$\mathbf{r}(t)=\mathbf{V}t+\exp(-i\omega t)\mathbf{r}_0, \quad (24)$$

where the constant vectors $\mathbf{V}$ and $\mathbf{r}_0$ lie in the plane $i\omega$. The term $\exp(-i\omega t)\mathbf{r}_0$ describes uniform circular motion about the axis ω, and $\mathbf{V}t$ gives the position of the axis. Such motion, for example, solves the Lorentz-force equation for a charge in crossed electric and magnetic fields; $\mathbf{V}$ is then the drift velocity and ω the angular cyclotron frequency.

F. Reflection in planes

A vector $\mathbf{r}$ is reflected in a plane $i\hat{\mathbf{a}}$ by the simple transformation

$$\mathbf{r}\to(i\hat{\mathbf{a}})\mathbf{r}(i\hat{\mathbf{a}})=-\hat{\mathbf{a}}\mathbf{r}\hat{\mathbf{a}}=\mathbf{r}-2\mathbf{r}\cdot\hat{\mathbf{a}}\hat{\mathbf{a}}. \quad (25)$$

The last equality follows directly from (16) when $\mathbf{r}$ is split into parts parallel and perpendicular to $\hat{\mathbf{a}}$. Because $i\hat{\mathbf{a}}=\exp(i\pi\hat{\mathbf{a}}/2)$, the reflection (25) is seen to be equivalent to a 180° rotation in the plane $i\hat{\mathbf{a}}$ [see (21)] together with an inversion $\mathbf{r}\to-\mathbf{r}$.

Two successive reflections in planes $i\hat{\mathbf{a}}$ and $i\hat{\mathbf{b}}$

$$\mathbf{r}\to i\hat{\mathbf{b}}(i\hat{\mathbf{a}})\mathbf{r}(i\hat{\mathbf{a}})i\hat{\mathbf{b}}=(\hat{\mathbf{b}}\hat{\mathbf{a}})\mathbf{r}(\hat{\mathbf{a}}\hat{\mathbf{b}})=\exp(-i\boldsymbol{\theta})\mathbf{r}\exp(i\boldsymbol{\theta}) \quad (26)$$

is equivalent to a rotation by $2\boldsymbol{\theta}$, twice the angle between the two planes (see Secs. V C and D) about an axis along the intersection of the planes. The combination of three reflections in orthogonal planes is easily seen to be the same as spatial inversion. For example, successive reflections of $\mathbf{r}$ in the $i\mathbf{e}_1$, $i\mathbf{e}_2$, and $i\mathbf{e}_3$ planes transforms $\mathbf{r}$ into

$$(i\mathbf{e}_3)(i\mathbf{e}_2)(i\mathbf{e}_1)\mathbf{r}(i\mathbf{e}_1)(i\mathbf{e}_2)(i\mathbf{e}_3)=-\mathbf{e}_3\mathbf{e}_2\mathbf{e}_1\mathbf{r}\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3=-\mathbf{r}, \quad (27)$$

since $\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3=i=-\mathbf{e}_3\mathbf{e}_2\mathbf{e}_1$. This, of course, is the principle of the corner-cube reflector.

G. Angular momentum

A further example illustrating the use of planes in algebraic products is the orbital angular momentum $\mathbf{L}$, where $i\mathbf{L}$ is just the pseudovector part of the product $\mathbf{rp}$:

$$i\mathbf{L}=(\mathbf{rp}-\mathbf{pr})/2=i\mathbf{r}\times\mathbf{p}. \quad (28)$$

Now, $\mathbf{L}$ is the axial vector normal to every vector in the orbital plane, and because of the geometrical significance of i, $i\mathbf{L}$ is the bivector representing the orbital plane itself. Its magnitude is the area of the parallelogram formed by $\mathbf{r}$ and $\mathbf{p}$ and hence $2m$ times the rate at which the orbital area is swept out. Of course it is just this relationship that connects the conservation of angular momentum to Kepler's second law.

The product $\mathbf{abc}=(\mathbf{ab})\mathbf{c}=\mathbf{a}(\mathbf{bc})$ of three vectors has both a vector and a trivector part, and in three dimensions the trivector is also a pseudoscalar $i\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}$ which gives the volume of the parallelepiped with edges parallel to $\mathbf{a},\mathbf{b},\mathbf{c}$ (see Fig. 2).[22] Since the dot product of vectors or pseudovectors is symmetric with respect to interchange of the two factors [see (14)], $(\mathbf{ab})\mathbf{c},\mathbf{c}(\mathbf{ab})=(\mathbf{ca})\mathbf{b}$, and $\mathbf{b}(\mathbf{ca})=(\mathbf{bc})\mathbf{a}$ all have the same pseudoscalar parts. The relations

$$\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}=\mathbf{c}\times\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\times\mathbf{c}\cdot\mathbf{a}=\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$$

follow immediately.

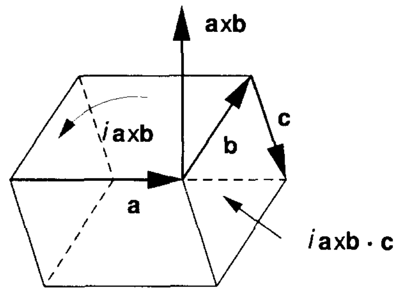

Fig. 2. The oriented volume of the parallelepiped whose edges are parallel to vectors **a**, **b**, **c** is given by the trivector part of **abc**, and in three dimensions this trivector is the pseudoscalar $i\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}$, whose magnitude gives the numerical size of the volume. In the figure, the $\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}$ is negative because $\mathbf{a}\times\mathbf{b}$ is roughly opposite to **c**.

Note that in contrast to the field of real numbers, the algebra of vectors contains a noncommutative product (13) as well as some *zero divisors* (divisors or factors of zero): nonzero elements whose product vanishes, for example $(1+\mathbf{e}_3)(1-\mathbf{e}_3)=0$. Therefore, the vectors do not form a field, and one must pay attention to the order of vector factors and be careful not to divide by a zero divisor. It is precisely these added complications which endow the Pauli algebra $\mathscr{P}$ with its rich mathematical structure capable of describing a wealth of physical phenomena.

VI. STRUCTURE OF $\mathscr{P}$

So far, we have seen how complex scalars and vectors arise naturally in the vector algebra $\mathscr{P}$ of three-dimensional space. The general element (12) of $\mathscr{P}$ contains both. As Hestenes[3,5] has emphasized, the imaginary i carries important geometrical information in the algebra. We still need to show how this geometrical i is related to the many occurrences of i in traditional formulations of physics. Some of the relations are most clearly seen if the structure of $\mathscr{P}$ is understood. Thus before we investigate further examples of i in relativity, electromagnetism, and quantum theory, we should pause to look at the shape, especially the subalgebras and involutory transformations of $\mathscr{P}$.

There are three basic involutory transformations $p\to p'$ in $\mathscr{P}$ which take elements into other elements. They are said to be *involutory* because when applied a second time, they return the elements to their original identities $p'\to p$. *Spatial inversion*, mentioned in Sec. IV, is such a transformation. *Reversal*, effected by reversing the order of multiplication of all vectors, is another. Because it simply changes the sign of the pseudoscalar $i=\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$, it is also referred to as *Hermitean conjugation* and is denoted by a dagger: $p\to p^{+}$. If both operations of spatial inversion and reversal are performed, the result is *spatial reversal*; it amounts to changing the signs of vector and pseudovector

Table I. The basic involutory transformations of $\mathscr{P}$.

Operation	Action	Summary
Identity	$p\to p=p_0+\mathbf{p}+i\mathbf{q}+iq_0$	$++++$
Spatial inversion	$p\to\bar{p}^{+}=p_0-\mathbf{p}+i\mathbf{q}-iq_0$	$+-+-$
Reversal = Hermitean conj	$p\to p^{+}=p_0+\mathbf{p}-i\mathbf{q}-iq_0$	$++--$
Spatial reversal	$p\to\bar{p}=p_0-\mathbf{p}-i\mathbf{q}+iq_0$	$+--+$

parts and is denoted by a bar: $p\to\bar{p}$. Table I defines the action of these transformations on an arbitrary element p and summarizes the results by specifying the relative signs of the scalar, vector, bivector, and trivector parts of the transformed elements.[23]

An element p is said to be *even* iff (if and only if) $p=\bar{p}^{+}$; it is a *(complex) scalar* (scalar and/or pseudoscalar) iff $p=\bar{p}$; and it is *real* iff $p=p^{+}$. The matrix representation of any real element is *Hermitean.* From the effect of these transformations on elements, it is seen that when they are combined, the order of their application is immaterial.

There are several important subalgebras of the Pauli algebra $\mathscr{P}$. One is the *center* $\mathscr{P}^{c}=\mathscr{V}_0\oplus\mathscr{V}_3$ of $\mathscr{P}$, that is, the part of $\mathscr{P}$ which commutes with all elements. The inverse c^{-1} of any nonzero element $c\in\mathscr{P}^{c}$ is identified by noting that the product cc^{+} is a (real) scalar. Thus $c^{-1}=c^{+}/(cc^{+})$. Because the elements of $\mathscr{P}^{c}$ all commute and every nonzero element has an inverse, the subalgebra $\mathscr{P}^{c}$ is said to be a *field;* indeed it is the field $\mathbb{C}$ of complex numbers. A closely related commutative subalgebra can be formed from the vectors in a plane by multiplying them by one of the unit vectors in that plane. Consider for example the real vector $\mathbf{r}=x\mathbf{e}_1+y\mathbf{e}_2$ in the plane $i\mathbf{e}_3$. Multiplying each vector $\mathbf{r}$ by $\mathbf{e}_1$ we obtain $\mathbf{e}_1\mathbf{r}=x+y(i\mathbf{e}_3)$, a real linear combination of the basis elements $\{1,i\mathbf{e}_3\}$. The set of such elements is closed under addition and multiplication and forms a subalgebra with a one-to-one linear correspondence to the complex field $\mathbb{C}$.[24] This correspondence justifies the common use of $\mathbb{C}$ not only to represent positions in a plane with respect to a given coordinate system, but also to describe rotations in the plane. Since $\mathbf{r}^2=(\mathbf{e}_1\mathbf{r})^{+}(\mathbf{e}_1\mathbf{r})$ and $2\mathbf{r}\cdot\mathbf{r}'=(\mathbf{e}_1\mathbf{r})^{+}(\mathbf{e}_1\mathbf{r}')+(\mathbf{e}_1\mathbf{r}')^{+}(\mathbf{e}_1\mathbf{r})$ it also explains why the square of a vector is represented by $z^{+}z$ and why the dot product of two vectors takes the form $\Re\{z^{+}z'\}$. Of course $\mathscr{P}$ has the advantage of giving a coordinate-free representation of positions and rotations in three dimensions.

Because of the correspondence between positions in a plane and the field of complex numbers, uniform circular motion (see Sec. V D) is described in $\mathbb{C}$ by the phase factor $\exp(-i\omega t)$. Since the simple harmonic motion of oscillators is the projection of uniform circular motion, it is only a small step to the use of such factors in oscillating circuits. Including an exponential decay $\exp(-\gamma t)$ of the amplitude leads one to speak of "complex (angular) frequencies" $\omega-i\gamma$.

Another subalgebra is the even Pauli algebra $\mathscr{P}^{+}=\mathscr{V}_0\oplus\mathscr{V}_2$ which consists of all scalars and bivectors of $\mathscr{P}$. Since the product of two bivectors is a linear combination of a scalar and a bivector, $\mathscr{P}^{+}$ is closed under multiplication as well as under addition and thus does indeed form a subalgebra of $\mathscr{P}$. It is isomorphic to Hamilton's quaternion algebra $\mathscr{H}$ which is a *division algebra* because it contains no zero divisors.

VII. SPECIAL RELATIVITY

In special relativity, the four-vector is a basic covariant entity, and products of four-vectors give other covariant quantities. Many three-dimensional vectors, such as position and momentum, arise as the spatial parts of four-vectors in Minkowski space-time. Others, like angular momentum and electric and magnetic fields, result from products of four-vectors. We seek a covariant formulation of relativity which treats four-vectors and their products as single entities and which can avoid, as the Pauli algebra does for vectors in R^3, cumbersome component notation. While the notation must be covariant, it should also maintain close ties to physical measurements. Although space and time components are mixed by Lorentz transformations, the time dimension is qualitatively distinct from spatial dimensions in any given frame: no observer would confuse the two in measurements performed in his laboratory. Ideally our notation, while fully covariant, should allow a qualitative distinction between time and space in any given frame. We shall see below how these apparently contradictory goals of covariance, on the one hand, and of the distinction of space and time in any given frame, on the other, can be reconciled in the Pauli algebra.

A natural way to extend the advantages of a multivector algebra to special relativity is to use the geometric algebra based on Minkowski space-time instead of on R^3. Considerable work along these lines has been published by Hestenes, Salingaros, and others, and the geometric algebra of space-time has a familiar representation based on the Dirac gamma matrices.[3,25] This multivector algebra, known also as the Dirac algebra, has the somewhat awkward feature that the volume element η of the algebra anticommutes with vectors and pseudovectors and so is obviously distinct from the usual i.

However, as surprising as it may seem, there is no need to leave the relatively simple Pauli algebra in order to handle problems in relativity. We saw in Sec. III that the algebra of real three-dimensional vectors generates a complex four-dimensional space. Now we shall show that the Minkowski metric arises naturally in this space when one finds the multiplicative inverse of a general Pauli element.

An element p has an inverse if there exists another element, say p' whose product with p is a nonvanishing complex scalar: $pp'=\overline{pp'}=\bar{p}'\bar{p}\neq 0$. Indeed, the inverse is then simply $p'^{-1}=p/(pp')$. The square of any real vector $\mathbf{p}$ is a scalar, so that, as seen in Sec. V B, the inverse of $\mathbf{p}$ is $\mathbf{p}^{-1}=\mathbf{p}/(\mathbf{p}^2)$. However, the square of a general element p does not necessarily belong to $\mathscr{P}^{c}=\mathbb{C}$. Only the product $p\bar{p}$ is always its own spatial reverse and is therefore a complex scalar. Therefore, as long as $p\bar{p}$ does not vanish, p has the inverse

$$p^{-1}=\bar{p}/(p\bar{p}). \tag{29}$$

Evidently the (complex) scalar $p\bar{p}$ plays the same role for an element p of the complex four-dimensional vector space of $\mathscr{P}$ that the scalar $\mathbf{r}^2$ plays for a vector $\mathbf{r}$ in real three-dimensional space: we will refer to $p\bar{p}$ as the *norm* of p. If the element $p=p_0+\mathbf{p}$ is identified with the real four-vector $(p_0,\mathbf{p})$, its norm

$$p\bar{p}=(p_0+\mathbf{p})(p_0-\mathbf{p})=p_0^2-\mathbf{p}^2 \tag{30}$$

is also real, and its form gives the Minkowski metric of the four-dimensional space. The scalar part of p is the zero or "time" component of the four-vector (here we use units with the speed of light $c=1$). Four-vectors with a vanishing norm are said to be *light-like:* $p_0^2=\mathbf{p}^2$. They are seen to be zero divisors and hence noninvertible elements of the algebra. Transformations which leave $p\bar{p}$ invariant for any four-vector p are called *Lorentz transformations.* Spatial inversion is an example of an *improper* Lorentz transformation. Any scalar invariant under Lorentz transformations is said to be a *Lorentz scalar.*

Lorentz transformations that can be realized physically by a sequence of infinitesimal transformations are called *restricted Lorentz transformations* (proper and orthochronous); they are represented by a simple generalization of rotations (21):

$$p \to LpL^+, \tag{31}$$

where L is any *unimodular* element: $L\bar{L}=1$, which can be written as the product $L=\exp(\mathbf{w}/2)\exp(-i\boldsymbol{\theta}/2)$.[26] When $\boldsymbol{\theta}=0$, $L=\exp(\mathbf{w}/2)$ and is real; its application (31) gives a boost with the boost parameter equal to the vector $\mathbf{w}$. When $\mathbf{w}=0$, then L rotates the four-vector by the angle θ about the axis $\hat{\theta}$. As before (see 21), the rotation plane is $i\theta$. We thus see the familiar relation that a boost takes the form of an imaginary rotation, but now there is added physical meaning to the fact that the rotation parameter $i\boldsymbol{\theta}$ is a bivector whereas the boost parameter $\mathbf{w}$ is a vector.

As a simple example, consider the transformation of the four-velocity u from its rest-frame value $u_{\text{rest}}=1$:

$$u=Lu_{\text{rest}}L^+=e^{\mathbf{w}}. \tag{32}$$

The scalar and vector parts of this relation provide the usual relations between the boost parameter $\mathbf{w}$ and the four-velocity u achieved by the boost: $u_0\equiv\gamma=\cosh|\mathbf{w}|$ and $\mathbf{u}\equiv\gamma\mathbf{v}=\hat{\mathbf{w}}\sinh|\mathbf{w}|$ The resultant four-velocity u is independent of the initial rotation angle $\boldsymbol{\theta}$ and can be identified with the square of the boost element $\exp(\mathbf{w}/2)$. The application of two successive boosts gives the rule for velocity composition:

$$U\equiv\exp(\mathbf{w}'/2)u\exp(\mathbf{w}'/2)=u'(u_0+\mathbf{u}_\parallel)+\mathbf{u}_\perp, \tag{33}$$

where $\mathbf{u}_\parallel$ and $\mathbf{u}_\perp$ are the parts of $\mathbf{u}$ parallel and perpendicular to $\mathbf{u}'$, respectively.[11] The result is particularly simple when the boost parameters $\mathbf{w}$ and $\mathbf{w}'$ are parallel: $U=u'u$. The composition of parallel four-velocities is thus given not by their sum but by their product. We can expand the relation into scalar and vector parts $U_0\equiv\Gamma=\gamma'\gamma(1+\mathbf{v}'\cdot\mathbf{v})$ and $\mathbf{U}\equiv\Gamma\mathbf{V}=\gamma'\gamma(\mathbf{v}+\mathbf{v}')$ in order to obtain the traditional form of the relationship. It is not much more difficult to expand the noncollinear case (33). Since the product of real elements is real only if the elements commute, the composition of boosts is a pure boost only if the boost directions are collinear. Otherwise, the product is the combination of a boost and a rotation.[11]

The four-vectors, like p or u, transform as covariant elements as in Eq. (31) but may, in any given frame, be expanded into scalar (time) and vector (spatial) parts. The distinction between scalars and vectors mirrors the qualitative difference between time and space components. In this way, the Pauli algebra is able to accommodate both covariance and the distinctiveness of space and time components in any given frame. Further examples can be found in Refs. 3, 12, and 27.

VIII. ELECTROMAGNETIC THEORY

The transformation properties of products of four-vectors are easily found from the four-vector transformation (31) and the unimodularity of L. An important example is $\partial\bar{A}$, where $\partial=\partial/\partial t-\boldsymbol{\nabla}$ is the four-vector differential operator and $A=\phi+\mathbf{A}$ is the four-potential. Let us first define a generalization of the vector dot product. If $a=a^\mu e_\mu=a_0+\mathbf{a}$ and $b=b^\nu e_\nu=b_0+\mathbf{b}$ are any two elements of $\mathscr{P}$, where the components a^μ, b^ν may be complex, then the dot product is simply the complex scalar part of the product $\mathbf{ab}$:

$$a\cdot b=(a\bar{b}+\overline{a\bar{b}})/2=a_0b_0+\mathbf{a}\cdot\mathbf{b}. \tag{34}$$

Products of four-vectors transform simply if vectors are alternated with spatially reversed vectors. The product $\partial\bar{A}$ transforms under restricted Lorentz transformations as

$$\partial\bar{A}\to L\partial L^+\overline{(LAL^+)}=L\partial L^+\bar{L}^+\bar{A}\,\bar{L}=L\partial\bar{A}\,\bar{L}. \tag{35}$$

Its scalar part is a Lorentz scalar:

$$\partial\cdot\bar{A}\to L\partial\cdot\bar{A}\,\bar{L}=\partial\cdot\bar{A}L\bar{L}=\partial\cdot\bar{A}. \tag{36}$$

The remaining vector plus bivector together is called a six-vector and although it transforms like any vector under rotations, its characteristic boost transformation is seen to be distinct.

The scalar, vector, pseudovector (bivector), and pseudoscalar (trivector) forms of $\mathscr{P}$, which belong to the four subspaces $\mathscr{V}_0$, $\mathscr{V}_1$, $\mathscr{V}_2$, $\mathscr{V}_3$, respectively, can thus be combined to make various covariant quantities:

Lorentz scalar	$\epsilon\mathscr{V}_0$,
four-vector	$\epsilon\mathscr{V}_0\oplus\mathscr{V}_1$,
six-vector	$\epsilon\mathscr{V}_1\oplus\mathscr{V}_2$,
pseudo-four-vector	$\epsilon\mathscr{V}_2\oplus\mathscr{V}_3$,
Lorentz pseudoscalar	$\epsilon\mathscr{V}_3$.

The product of 2 four-vectors is generally a Lorentz scalar plus a six-vector. Similarly, the product of 3 four-vectors or of a four-vector and a six-vector can be shown to be a four-vector plus a pseudo four-vector, whereas the product of 2 six-vectors is a Lorentz scalar plus a Lorentz pseudoscalar plus another six-vector.

By writing out the individual terms, one sees that the six-vector part of $\partial\bar{A}$ gives the electromagnetic field:

$$\partial\bar{A}-\partial\cdot\bar{A}=-\left(\frac{\partial\mathbf{A}}{\partial t}+\boldsymbol{\nabla}\phi\right)+i\boldsymbol{\nabla}\times\mathbf{A}=\mathbf{E}+i\mathbf{B}\equiv\mathbf{F}. \tag{37}$$

The electric field $\mathbf{E}$ is a vector but the magnetic field $\mathbf{B}$ enters as a pseudovector $i\mathbf{B}$, emphasizing the role of the plane $i\mathbf{B}$ perpendicular to $\mathbf{B}$. The scalar, vector, pseudovector, and pseudoscalar parts of the single field equation

$$\bar{\partial}\mathbf{F}=4\pi K\bar{j} \tag{38}$$

are exactly Maxwell's usual four microscopic equations:

$$\left(\frac{\partial}{\partial t}+\boldsymbol{\nabla}\right)(\mathbf{E}+i\mathbf{B})=\boldsymbol{\nabla}\cdot\mathbf{E}+\left(\frac{\partial\mathbf{E}}{\partial t}-\boldsymbol{\nabla}\times\mathbf{B}\right)+i\left(\frac{\partial\mathbf{B}}{\partial t}+\boldsymbol{\nabla}\times\mathbf{E}\right)+i\boldsymbol{\nabla}\cdot\mathbf{B}=4\pi K(\rho-\mathbf{j}). \tag{39}$$

Here, $j = \rho + \mathbf{j}$ is the current density and K is a constant depending on units ($=1$ in Gaussian units, $=1/4\pi\epsilon_0$ in SI units, and $=1/4\pi$ in Heaviside–Lorentz units). Magnetic monopoles can be added by making j and A complex.[28]

The general plane-wave solution to the wave equation (38) in source-free space ($j=0$) is[13]

$$\mathbf{F}(x) = \frac{k}{2\omega}\hat{\xi} f(\bar{k}\cdot x) = \frac{1}{2}(\hat{\xi} + i\hat{\eta}) f(\bar{k}\cdot x), \tag{40}$$

where $k = \omega + \mathbf{k}$ is a constant propagation four-vector which satisfies $k\bar{k} = \omega^2 - \mathbf{k}^2 = 0$, $\hat{\xi}$ is any unit vector in the plane normal to $\hat{\mathbf{k}}$, $\hat{\eta} = \hat{\mathbf{k}} \times \hat{\xi}$, and f is any scalar function. Rotations of F about $\hat{\mathbf{k}}$ are equivalent to multiplications by a phase factor:

$$\begin{aligned}\mathbf{F} &\to \exp(-i\phi\hat{\mathbf{k}}/2)\mathbf{F}\exp(i\phi\hat{\mathbf{k}}/2)\\ &= \exp(-i\phi\hat{\mathbf{k}})\mathbf{F} = \exp(-i\phi)\mathbf{F},\end{aligned} \tag{41}$$

where we noted that $\bar{k}\mathbf{F} = \omega(1-\hat{\mathbf{k}})\mathbf{F} = 0$. The phase factor belongs to the group U(1) of "duality rotations," which mix the vector and pseudovector parts of a six-vector, in this case the electric and magnetic fields. Thus if the function $f(\bar{k}\cdot x)$ varies as $\exp(\mp i\bar{k}\cdot x)$ where $\bar{k}\cdot x = \omega t - \mathbf{k}\cdot\mathbf{x}$, then at fixed $\mathbf{x}$, the phase of $\mathbf{F}$ advances at the rate $\pm\omega$, which is equivalent to a rotation at the angular rate ω about $\pm\hat{\mathbf{k}}$. The plane-wave solution then represents a circularly polarized wave. On the other hand, if the phase of $f(\bar{k}\cdot x)$ is constant, the solution is plane polarized.

A similar complex form is familiar to most physicists; one usually takes the real part to represent the electric field of a circularly polarized plane wave. The phase of $f(\bar{k}\cdot x)$ then gives the direction of $\mathbf{E}$ in the plane $i\mathbf{k}$. In $\mathscr{P}$, the polarization is not necessarily circular, and the real and imaginary parts of $\mathbf{F}$ have definite meaning: the real part is a vector and hence the electric field, and the imaginary part is a pseudovector identified with i times the magnetic field. Further examples may be found in Refs. 12 and 29.

IX. QUANTUM THEORY AND SPINORS

The imaginary i seems to be used in quantum theory in a fundamentally new way. Before, when terms with i appeared, we could separate the equations into real and imaginary parts, or into scalar, vector, pseudovector, and pseudoscalar parts. However, the basic equations

$$\begin{aligned}&[p_x, x] = -i\hbar,\\ &i\hbar\partial_t\psi = H\psi,\end{aligned} \tag{42}$$

seem to make no sense when separated in the same way.[30] One may question the very structure of these equations. Thus p_x and x are both spatial components of four-vectors; how can their commutator be a pseudoscalar? One would expect it to contain only scalar and six-vector parts. Furthermore, both ∂_t and H are time components of four-vectors; their ratio should be a scalar, not a pseudoscalar or volume element. The i from quantum operators and commutation relations seems to be distinct from the i for the volume element of geometric algebras or, equivalently, from that of the duality transformations of electromagnetism.

The resolution of this puzzle lies in the properties of the wave function ψ. The operator equations (42) only have physical meaning when applied to such a wave function. In relativistic quantum theory, fermion wave functions ψ are constructed from 2 two-component *spinor* functions $\Lambda_\pm = \Lambda(1\pm\mathbf{e}_3)/2$ which together form a spinor $\Lambda = \Lambda_+ + \Lambda_-$ in $\mathscr{P}$ that represents the Lorentz transformation of the particle from its rest frame to the observer's frame.[31–35] Spinors are more fundamental objects than four- or six-vectors in that four- and six-vectors can be constructed from spinors but not *vice versa.* If Λ_1 and Λ_2 are two Lorentz spinors in $\mathscr{P}$, they are defined to transform under Lorentz transformations according to the simple rule

$$\Lambda_j \to L\Lambda_j, \quad j=1,2. \tag{43}$$

If e is any invariant quantity, the spinors may be combined into the *bilinear covariants* $\Lambda_1 e\Lambda_2^+ \pm \Lambda_1 e^+\Lambda_2^+$ to form four-vectors (+ sign) and pseudo-four-vectors (−sign), and into $\Lambda_1 e\Lambda_2^{-1} \pm \Lambda_2\bar{e}\Lambda_1^{-1}$ to form complex scalars (+ sign) and six-vectors (− sign), since these covariant classes are defined by their behavior under Lorentz transformations, spatial reversal, and Hermitean conjugation. The bilinear covariant $\Lambda\Lambda^+$ is the probability current density, whereas $-i\Lambda\mathbf{e}_3\Lambda^{-1}$, where $\mathbf{e}_3$ is a fixed unit vector in the fermion rest frame, is the spin six-vector $\mathbf{S}$. The two-component spinors $\Lambda_\pm$ may be seen to be eigenelements of $\mathbf{S}$:

$$\mathbf{S}\Lambda_\pm = \mp i\Lambda_\pm. \tag{44}$$

In the traditional formulation of the Dirac theory, the Dirac wave function ψ combines components of Λ_- with those of $(\bar{\Lambda}_+)^+$. As a consequence, multiplying ψ by a phase factor $\exp(i\theta)$ is equivalent to multiplying $\Lambda_\pm$ by $\exp(\mp i\theta)$ and hence to rotating the system by an angle 2θ in the plane of the six-vector $\mathbf{S}$ about the spin direction. Such a "rotation" generally includes a boost component for a moving particle; however, it is equivalent to a pure rotation of its rest frame by $2\theta\mathbf{e}_3$. Thus the time dependence $\exp(-iEt/\hbar)$ of a stationary state of eigenenergy E corresponds to the rotation of the system about its spin axis at the angular frequency $2E/\hbar$ equal, for a free particle, to the *Zitterbewegung* frequency.[36]

Expressing the basic quantum relations (42) in terms of the spinor Λ, we find

$$\begin{aligned}&[p_x, x]\Lambda = -\hbar\mathbf{S}\Lambda\\ &\hbar\partial_t(\mathbf{S}\Lambda) = H\Lambda.\end{aligned} \tag{45}$$

The quantum use of i in relations (42) has been replaced by the spin $\mathbf{S}$ in (45), where $\mathbf{S}$ is a six-vector which reduces to a pseudovector (representing a spatial plane) in the rest frame. In these forms, since $\mathbf{S}\Lambda$ is another spinor, the equations make structural sense. It thus appears that the i in the quantum theory of fermions is associated with the spin of the particle.[32,37]

X. SUMMARY

In seeking a better understanding of the success of mathematical abstraction in physics and in particular of the wide applicability of imaginary numbers in theories of physical phenomena, we found that the algebra of real three-dimensional vectors generates a complex four-dimensional vector space with a Minkowski metric. The algebra, known as the Pauli algebra or the geometric algebra of three-dimensional space, thereby provides both a source of complex numbers in physics and also a covariant formulation of relativistic phenomena, one, however, which is able to maintain the qualitative distinction between time and space dimensions in any one frame. The imaginary i in the algebra carries geometric meaning: when multiplied by a scalar, it is a trivector or a pseudoscalar which represents a volume element, whereas when multiplied by a vector it is a bivector or pseudovector which represents a plane. We have been able to unify many of the varied uses of i in physics and to give them physical interpretations by relating them to the geometrical i of the Pauli algebra. The i in quantum expressions for fermions can be understood within the framework of the Pauli algebra as representing the spin six-vector.

ACKNOWLEDGMENTS

Thoughtful and stimulating comments by Dr. D. Hestenes are gratefully acknowledged. The authors also thank Dr. F. W. Lemire and Dr. N. E. Hedgecock for their careful reading of the manuscript and fruitful discussions are acknowledged with G. Jones and D. Kosokowsky. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

[1]P. E. B. Jourdain, in *The World of Mathematics*, edited by James R. Newman (Simon and Schuster, New York, 1956), Vol. I.

[2]F. Klein, *Elementary Mathematics from an Advanced Standpoint*, translated by E. R. Hedrick and C. A. Noble (Dover, New York, 1924), part I of the third German ed., p. 56.

[3]D. Hestenes, *Spacetime Algebra* (Gordon and Breach, New York, 1966).

[4]D. Hestenes, "Vectors, spinors, and complex numbers in classical and quantum physics," Am. J. Phys. **39**, 1013–1027 (1971).

[5]D. Hestenes, *New Foundations for Classical Mechanics* (Reidel, Dordrecht, 1987).

[6]R. S. Ingarden and A. Jamiolkowski, *Classical Electrodynamics* (Elsevier, Amsterdam, 1985).

[7]J. S. R. Chisholm and A. K. Common (Ed.), *Clifford Algebras and their Applications in Mathematical Physics* (Reidel, Dordrecht, 1986).

[8]B. Jancewicz, *Multivectors and Clifford Algebras in Electrodynamics* (World Scientific, Singapore, 1989).

[9]P. Bamberg and S. Sternberg, *A Course in Mathematics for Students of Physics* (Cambridge U. P., Cambridge, 1989), Vols. 1 and 2.

[10]W. E. Baylis, "Special relativity with 2×2 matrices," Am. J. Phys. **48**, 918–925 (1980).

[11]W. E. Baylis and G. Jones, "Special relativity with Clifford algebras and 2×2 matrices, and the exact product of two boosts," J. Math. Phys. **29**, 57–62 (1988).

[12]W. E. Baylis and G. Jones, "The Pauli algebra approach to special relativity," J. Phys. A: Math. Gen. **22**, 1–16 (1989).

[13]W. E. Baylis and G. Jones, "Relativistic dynamics of charges in external fields: the Pauli algebra approach," J. Phys. A: Math. Gen. **22**, 17–29 (1989).

[14]P. J. Nahin, "Oliver Heaviside," Sci. Am. **262** (6), 122–129 (1990).

[15]W. R. Hamilton, *Elements of Quaternions*, edited by C. J. Jolly (Chelsea, New York, 1969), Vols. I and II; 3rd ed.

[16]Only in three dimensions can the orientation of a plane be specified by a vector, since only there is the dual to a two-surface element a vector. In four dimensions, for example, the directions normal to a plane define another, orthogonal, plane, and the dual of a bivector is another bivector.

[17]Clearly another choice is $\eta=-i$. The change from $\eta=i$ is equivalent to the transformation from a right-handed coordinate system to a left-handed one.

[18]More generally, such vector algebras, called *Clifford algebras*, can be found in spaces of n dimensions and various signatures. The "volume" basis form η is then the product of all n basis vectors: $\eta=\mathbf{e}_1\mathbf{e}_2\cdots\mathbf{e}_n$. Whether or not it commutes with the other elements of the algebra depends on the dimensionality of the original vector space. Thus, in two dimensions for example, η commutes with scalars but anticommutes with vectors, and in four dimensions, η commutes with scalars and bivectors but anticommutes with vectors and pseudovectors.

[19]However, note that length, density, time, and energy are *not* invariant under boosts and are therefore not *Lorentz* scalars (see Sec. VII).

[20]The terms *axial vector* and *pseudovector* are often treated as synonymous, but it is quite useful to be able to distinguish a bivector (the pseudovector) from its dual (the axial vector). The distinction reveals a subtlety about their behavior under spatial inversion. If $\mathbf{a}$ and $\mathbf{b}$ are regular polar vectors, the bivector representing the plane containing them is the pseudovector $(\mathbf{ab}-\mathbf{ba})/2=i(\mathbf{a}\times\mathbf{b})$, which is also known as the *outer* or *Grassman* product $\mathbf{a}\wedge\mathbf{b}$. The axial vector is the dual vector normal to the plane, namely $\mathbf{a}\times\mathbf{b}=a^jb^k\epsilon_{jkl}\mathbf{e}_l$. If the vectors are inverted $\mathbf{a}\to-\mathbf{a},\mathbf{b}\to-\mathbf{b}$, by changing the signs of their components while leaving the basis vectors fixed, both the pseudovector and the axial vector are invariant. On the other hand, if the components are held fixed and the basis vectors $\mathbf{e}_k$ inverted, then the pseudovector is invariant but the axial vector changes sign. This behavior of axial vectors is consistent with their definition as vector-like elements which, in contrast to polar vectors, change sign under a transformation from a right-handed coordinate system to a left-handed one [see, for example, R. Baierlein, *Newtonian Dynamics* (McGraw-Hill, New York, 1983), p. 316]. Such a transformation can be realized by the simultaneous change in the signs of components (of polar vectors) and the directions of basis vectors.

[21]Many scalar functions of vectors are not the scalar part of any algebraic function. An important example of such a scalar function is the linear functional $\mathbf{r}\cdot\mathbf{a}=r^ja^j$ which defines an inner product of the argument $\mathbf{r}$ with a fixed vector $\mathbf{a}$.

[22]To keep the proliferation of parentheses in check, we impose the following priority rule: when dot, cross, and algebraic products appear within the same level of parenthesis, perform the cross product first, then the dot product, and finally the algebraic product.

[23]As long as $\mathscr{P}$ is thought of as a real algebra with eight basis forms, spatial inversion is an automorphism; it preserves the order of the products: $\overline{(ab)}^+=\bar{a}^+\bar{b}^+$. Hermitean conjugation and spatial reversal, on the other hand are involutory antiautomorphisms, or more simply, *involutions*, because they reverse the order of multiplication: $\overline{(ab)}=\bar{b}\bar{a}$ and $(ab)^+=b^+a^+$.

[24]The two-dimensional vector $\mathbf{r}=x\mathbf{e}_1+y\mathbf{e}_2$ is mapped onto the complex number $\zeta=x+iy$ according to

$$\mathbf{r}\to\zeta=(\mathbf{e}_1+i\mathbf{e}_2)\cdot\mathbf{r}=\left(\frac{1+\mathbf{e}_3}{2}\right)\mathbf{e}_1\mathbf{r}+\overline{\left(\frac{1+\mathbf{e}_3}{2}\right)}\mathbf{e}_1\mathbf{r}.$$

The inverse map is

$$\zeta\to\mathbf{r}=\mathscr{R}[(\mathbf{e}_1-i\mathbf{e}_2)\zeta]=\mathbf{e}_1\left(\frac{1+\mathbf{e}_3}{2}\right)\zeta+\left[\mathbf{e}_1\left(\frac{1+\mathbf{e}_3}{2}\right)\zeta\right]^+.$$

These mappings give the formal relation between complex numbers and their Wessel (Argand) diagrams. In particular, a rotation $\exp(-i\phi\mathbf{e}_3)$ in the $i\mathbf{e}_3$ plane maps onto the phase factor $\exp(i\phi)$:

$$\exp(-i\phi\mathbf{e}_3)\mathbf{r}\to\left(\frac{1+\mathbf{e}_3}{2}\right)\mathbf{e}_1\exp(-i\phi\mathbf{e}_3)\mathbf{r}+\text{s.r.}$$

$$=\exp(i\phi\mathbf{e}_3)\left(\frac{1+\mathbf{e}_3}{2}\right)\mathbf{e}_1\mathbf{r}+\text{s.r.}=\exp(i\phi)\zeta,$$

where s.r. stands for the spatial reverse of the previous term. The mapping describes a vector-space isomorphism between vectors in a plane and the complex field $\mathbb{C}$. However, it does *not* establish an *algebraic* isomorphism since a product of two vectors never corresponds to the product of the complex numbers to which those vectors map.

[25]N. Salingaros and M. Dresden, "Physical algebras in four dimensions. I. The Clifford algebra in Minkowski spacetime," Adv. Appl. Math. **4**, 1–30 (1983).

[26]In fact, any element can be expressed as the product of a real element and a unitary element. The group of unimodular elements L is SL(2,C), the double covering group for restricted Lorentz transformations. Its unitary subgroup SU(2) covers rotations. The group SL(2,C) is said to be a six-parameter group, since every Lorentz transformation is uniquely determined by the choice of six parameters: the three real components of the vector $\mathbf{w}$ and the three imaginary components of the bivector $i\boldsymbol{\theta}$.

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

[28]J. Wei and W. E. Baylis, "Monopoles without strings: a conflict be-

tween the one-photon condition and duality invariance," Found. Phys. Lett. 4, 537–556 (1991).
[29]D. Hestenes, "Proper dynamics of a rigid point particle," J. Math. Phys. 15, 1778–1786 (1974).
[30]C. N. Yang, in *Centenary Celebration of a Polymath*, edited by C. W. Kilmister (Cambridge U. P., Cambridge, 1987), pp. 53–64.
[31]F. Gürsey, "Relativistic kinematics of a classical point particle in spinor form," Nuovo Cimento 5, 784–809 (1957).
[32]D. Hestenes, "Observables, operators, and complex numbers in the Dirac theory," J. Math. Phys. 16, 556–572 (1975).
[33]D. Hestenes, "The Zitterbewegung interpretation of quantum mechanics," Found. Phys. 20, 1213–1232 (1990).
[34]W. E. Baylis, "Classical eigenspinors and the Dirac equation," Phys. Rev. A 45, 293—302 (1992).
[35]Lorentz spinors may have components in all four subspaces $\mathscr{V}_j$, $j=0$, 1, 2, 3 of $\mathscr{P}$. They form a group under addition, the even subgroup of which is the Clifford spin group of *rotational spinors* in R^3. Two-component spinors are formed by projecting Lorentz spinors onto minimal left ideals of $\mathscr{P}$, that is by multiplying them from the right by primitive idempots of the form $\frac{1}{2}(1\pm \mathbf{e}_3)$, which project out one column of the 2×2-matrix representations of the Lorentz spinors.
[36]K. Huang, "On the *Zitterbewegung* of the Dirac electron," Am. J. Phys. 20, 479–484 (1952).
[37]What about spin-0 bosons? It is the Dirac theory of spin$-1/2$ fermions that falls so simply out of the covariant Pauli-algebra (see Ref. 34) and for which i seems inextricably tied to the spin: all interference phenomena derive from the spin. This may imply a similar relation for bosons which comprise a pair of elementary fermions, but there would seem to be a problem with elementary bosons. Perhaps there is no elementary boson with spin 0 (the Higgs has not been found yet!) or perhaps the formalism is simply not up to the task of describing it.

A solid-state low-voltage Tesla coil demonstrator

Donald G. Bruns
5195 Hearthstone Lane, Colorado Springs, Colorado 80919

(Received 18 October 1991; accepted 19 March 1992)

A low-voltage demonstration Tesla coil using a solid-state photovoltaic relay to replace the conventional spark gap has been analyzed and then built. This relay incorporates an isolated LED to illuminate a silicon photovoltaic stack which drives a bidirectional FET. Component values for the inductances and capacitances have been determined theoretically from measured parameters. Computer simulation by integrating the coupled circuit equations shows excellent agreement with oscilloscope traces. Energy transfer between the primary and secondary circuits is demonstrated, along with continuous secondary oscillations after the primary circuit is interrupted. This low-voltage design is easier to build and diagnose than high-voltage Tesla coils.

I. INTRODUCTION

Nikola Tesla invented the Tesla coil late in the nineteenth century, exploring many high-power variations in his Colorado Springs laboratory.[1] They were all basically air-cored high-frequency transformers, generating very high voltages. Many of his experiments were complicated, using large coils made with heavy copper wires to conduct very high currents. His high-voltage capacitors used hundreds of salt water-filled Leyden jars made from the local Manitou Springs mineral water bottling plant. Tesla documented his achievements with multiple-exposure photographs which show his small wooden building filled with curved sparks up to 40 m in length. Using spark length is how Tesla often diagnosed his experiments.

Tesla's ultimate goal was to generate high enough voltages that he could transmit useful electrical power freely through the atmosphere. One contemporary account claimed he succeeded in sending enough power to energize a bank of light bulbs 40 km away.[2] However, he never completed his final and largest experiment on Long Island, New York, which he designed inside a 60-m-high wooden tower. Although lack of funding was the primary reason the tower was torn down, in the light of today's knowledge, it never would have succeeded in the manner he envisioned. While Tesla was advanced for his time, he didn't have electronic diagnostic tools to study his experiments.

Even though Tesla's grandiose plans would not have worked, we remain fascinated with high-voltage Tesla coils. Generating fiery arcs and lighting fluorescent tubes at a distance are always exciting demonstrations. In the last 60 years, instructions for building high-voltage Tesla coils appeared occasionally in popular magazines, journals, newsletters, and books.[3–13] A useful instrument in many physics laboratories is the hand-held Tesla coil used to excite gas discharges and find leaks in vacuum systems. There has also been some interest in using very large Tesla coils to test military aircraft with simulated lightning,[14] and using smaller coils to generate electron beams.[15] While any of these Tesla coils can be experimental subjects, detailed measurements to compare with theoretical predictions requires sophisticated equipment to deal with high voltages.

A conventional Tesla coil consists of tuned primary and secondary circuits. An interrupter in the primary circuit stimulates oscillations from the charge stored in a large capacitor. The primary circuit interrupter design is critical to maximize power transfer to the secondary circuit. High-power Tesla coils use variations on rotating spark gaps to extinguish the high-voltage spark,[16] a technique that hasn't

