Spacetime algebra as a powerful tool for electromagnetism

Justin Dressel [a,b,*], Konstantin Y. Bliokh [b,c], Franco Nori [b,d]

[a] *Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA*
[b] *Center for Emergent Matter Science (CEMS), RIKEN, Wako-shi, Saitama, 351-0198, Japan*
[c] *Interdisciplinary Theoretical Science Research Group (iTHES), RIKEN, Wako-shi, Saitama, 351-0198, Japan*
[d] *Physics Department, University of Michigan, Ann Arbor, MI 48109-1040, USA*

ARTICLE INFO

ABSTRACT

We present a comprehensive introduction to *spacetime algebra* that emphasizes its practicality and power as a tool for the study of electromagnetism. We carefully develop this natural (Clifford) algebra of the Minkowski spacetime geometry, with a particular focus on its intrinsic (and often overlooked) *complex structure*. Notably, the scalar imaginary that appears throughout the electromagnetic theory properly corresponds to the unit 4-volume of spacetime itself, and thus has *physical* meaning. The electric and magnetic fields are combined into a single complex and frame-independent bivector field, which generalizes the *Riemann–Silberstein complex vector* that has recently resurfaced in studies of the single photon wavefunction. The complex structure of spacetime also underpins the emergence of electromagnetic waves, circular polarizations, the normal variables for canonical quantization, the distinction between electric and magnetic charge, complex spinor representations of Lorentz transformations, and the *dual* (electric–magnetic field exchange) *symmetry* that produces helicity conservation in vacuum fields. This latter symmetry manifests as an arbitrary global phase of the complex field, motivating the use of a complex vector potential, along with an associated transverse and gauge-invariant bivector potential, as well as complex (bivector and scalar) Hertz potentials. Our detailed treatment aims to encourage the use of spacetime algebra as a readily available and mature extension to existing vector calculus and tensor methods that can *greatly simplify* the analysis of fundamentally relativistic objects like the electromagnetic field.

© 2015 Elsevier B.V. All rights reserved.

Contents

* Corresponding author at: Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA.
E-mail address: prof.justin.dressel@gmail.com (J. Dressel).

If we include the biological sciences, as well as the physical sciences, Maxwell's paper was second only to Darwin's "Origin of Species". But the importance of Maxwell's equations was not obvious to his contemporaries. Physicists found it hard to understand because the equations were complicated. Mathematicians found it hard to understand because Maxwell used physical language to explain it.

[Freeman J. Dyson [1]]

This shows that the mathematical language has more to commend it than being the only language which we can speak; it shows that it is, in a very real sense, the correct language. … The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve.

[Eugene P. Wigner [2]]

1. Introduction

For the rest of my life I will reflect on what light is.

[Albert Einstein [3]]

1.1. Motivation

At least two distinct threads of recent research have converged upon an interesting conclusion: the electromagnetic field has a more natural representation as a *complex* vector field, which then behaves in precisely the same way as a relativistic quantum field for a single particle (i.e., a *wavefunction*). It is our aim in this report to clarify why this conclusion is warranted from careful considerations of the (often overlooked) intrinsic complex structure implied by the *geometry of spacetime*. We also wish to draw specific parallels between these threads of research, and unify them into a more comprehensive picture. We find that the coordinate-free, geometrically motivated, and manifestly reference-frame-independent *spacetime algebra* of Hestenes [4–9] is particularly well suited for this task, so we provide a thorough introduction to this formalism in order to encourage its use and continued development.

The first thread of preceding research concerns the development of a *first-quantized approach to the single photon*. Modern quantum optical experiments routinely consider situations involving only a single photon or pair of photons, so there has been substantial motivation to reduce the full quantized field formalism to a simpler '*photon wave function*' approach that more closely parallels the Schrödinger or Dirac equations for a single electron [10–31]. Such a first-quantization approach is equivalent to classical electromagnetism: it represents Maxwell's equations in the form of a relativistic wave equation for massless spin-1 particles. The difficulty with this pursuit has stemmed from the fact that the photon, being a manifestly relativistic and massless spin-1 boson, is *nonlocalizable*. More precisely, the position of a single photon is ill-defined [32–37], and photons have no localizable number density [19,20,26,36], which makes it impossible to find a position-resolved probability amplitude in the same way as for electrons.[1] Nevertheless, the *energy* density of a single photon in a particular inertial frame *is* localizable, and closely corresponds to what is actually measured by single photon counters.

Following this line of reasoning, a suitable 'photon wave function' can be derived from general considerations as an energy-density amplitude in a particular inertial reference frame [18,19,24,25], which directly produces a complex form of the usual electric and magnetic fields $\vec{E}$ and $\vec{B}$ in that frame

$$\vec{F} = [\vec{E}/c + \vec{B}i]/\sqrt{\mu_0}. \tag{1.1}$$

Furthermore, the quantum mechanical derivation that produces (1.1) also produces Maxwell's vacuum equations in an intriguing form that resembles the Dirac equation for massless spin-1 particles [10–31] (shown here in both the momentum-operator notation and the equivalent coordinate representation)

$$i\hbar\,\partial_t \vec{F} = c|p|\,\hat{\chi}\,\vec{F}, \qquad i\hbar\,\partial_t \vec{F} = c\hbar\,\vec{\nabla}\times\vec{F}, \tag{1.2a}$$

$$\hat{\vec{p}}\cdot\vec{F} = 0, \qquad -i\hbar\,\vec{\nabla}\cdot\vec{F} = 0. \tag{1.2b}$$

Here $c|p|$ is the energy of a single photon, and $\hat{\chi} = \hat{\vec{s}}\cdot\hat{\vec{p}}/|p|$ is the *helicity* operator that projects the spin-1 matrix operator $\hat{\vec{s}}$ onto the direction of the momentum operator $\hat{\vec{p}} = -i\hbar\vec{\nabla}$, which produces the curl $(\hat{\vec{s}}\cdot\vec{\nabla})\vec{F} = i\vec{\nabla}\times\vec{F}$. The second equation (1.2b) appears as an auxiliary transversality condition satisfied by radiation fields far from matter [36,39].

Treating the complex vector (1.1) as a single-particle wave function in the first-quantized quantum mechanical sense produces consistent results, and suggests that this vector should be a more natural representation for classical electromagnetic fields as well.[2] Indeed, the practical classical applications of (1.1) have been emphasized and reviewed

---

[1] The absence of a probability density for photons follows from the Weinberg–Witten theorem [38] that forbids conserved four-vector currents for relativistic massless fields with spins higher than 1/2 [22,26].

[2] The factor of $\hbar$ cancels in Maxwell's equations (1.2) because the photon is massless. This explains the absence of this obviously quantum factor in classical electromagnetism.

by Bialynicki-Birula [19,40], who noted that this complex vector was also historically considered by Riemann [41] and Silberstein [42,43] in some of the earliest investigations of electromagnetism. Remarkably, performing the standard second-quantization procedure using the complex Riemann–Silberstein wave function (1.1) identically reproduces the correctly quantized electromagnetic field [26,28]. Moreover, if the number of photons in (1.2) is increased to two using the standard bosonic symmetrization procedure, the resulting equations of motion precisely reproduce the Wolf equations from classical coherence theory [26,44] in a consistent way.

The explicit appearance of the momentum, spin, and helicity operators in (1.2) connects the complex field vector (1.1) to the second thread of research into the *local momentum, angular momentum, and helicity properties of structured optical fields* (such as optical vortices, complex interference fields, and near fields) [45–85]. The development of nano-optics employing structured light and local light–matter interactions has prompted the careful consideration of nontrivial local dynamical properties of light, even though such properties are usually considered to be unobservable in orthodox quantum-mechanical and field-theory approaches.

Indeed, it was traditionally believed that *spin and orbital angular momenta* of light are not meaningful separately, but only make sense when combined into the total angular momentum [39,86,87]. However, when small probe particles or atoms locally interact with optical fields that carry spin and orbital angular momenta, they clearly acquire separately observable intrinsic and extrinsic angular momenta that are proportional to the local expectation values of the spin $\hat{\vec{s}}$ and orbital $\hat{\vec{r}} \times \hat{\vec{p}}$ angular momentum operators for photons, respectively [47,52–58,74]. (Akin to this, the spin and orbital angular-momentum contributions of gluon fields in QCD are currently considered as measurable in spite of the field-theory restrictions [88].) Similarly, the local *momentum density* of the electromagnetic field (as well as any local current satisfying the continuity equation) is not uniquely defined in field theory, so only the integral momentum value has been considered to be measurable. Nevertheless, when a small probe particle is placed in an optical field it acquires momentum proportional to the canonical momentum density of the field, i.e., the local expectation value of $\hat{\vec{p}}$ [64,69,85]. Remarkably, this same canonical momentum density of the optical field (and *not* the Poynting vector, see [64,68,74,85]) is also recovered using other methods to measure the local field momentum [59–62,70–72], including the so-called *quantum weak measurements* [66,69,89–91]. Such local weak measurements of the momentum densities in light fields (i.e., analogues of the probability current for photons in the above 'photon wave function' approach) simultaneously corroborate predictions of the Madelung hydrodynamic approach to the quantum theory [92,93], the Bohmian causal model [94,95], and the *relativistic energy–momentum tensor in field theory* [68,69,74,96]. In the quantum case, these local densities correspond to the associated classical mean (background) field, so can be observed only by averaging many measurements of the individual quantum particles [91].

Most importantly for our study, the reconsideration of the momentum and angular momentum properties of light has prompted discussion of the *helicity* density of electromagnetic fields. Recently it was shown that this helicity density naturally appears in local light–matter interaction with chiral particles [76–81,85,97]. However, while in the quantum operator formalism the helicity $\hat{\chi}$ is an intuitive combination of the momentum and spin operators, the *field-theory* picture of the electromagnetic helicity is not so straightforward.

Namely, electromagnetic helicity [68,98–108] is a curious physical electromagnetic-field property that is only conserved during vacuum propagation. Unlike the momentum and angular momentum, which are related to the Poincaré spacetime symmetries according to Noether's theorem [109], the helicity conservation corresponds to the more abstract *dual symmetry* that involves the exchange of electric and magnetic fields [68,99,101,102,105–108] (which was first discussed by Heaviside and Larmor [110,111], and has been generalized to field and string theories beyond standard electromagnetism [112–125]). Due to this intimate relation with the field-exchange symmetry, the definition of the helicity must involve the electric and magnetic fields and their two associated vector-potentials on equal footing. Furthermore, the consideration of the dual symmetry, which is inherent to the vacuum Maxwell's equations (1.2), recently reignited discussion regarding the proper dual-symmetric description for all vacuum-field properties, including the canonical momentum and angular-momentum densities [64,68,106,126]. These quantities are dual *asymmetric* in the traditional electromagnetic field theory [68,127]. Notably, the dual electric–magnetic symmetry become particularly simple and natural when the electric and magnetic fields and potentials are combined into the complex Riemann–Silberstein form (1.1) [68,115]. Indeed, the continuous dual symmetry of the vacuum electromagnetic field takes the form of a simple $U(1)$ global phase-rotation

$$\vec{F} \mapsto \vec{F} \exp(i\theta). \tag{1.3}$$

It is this continuous 'gauge symmetry of the photon wave function' that produces the conservation of optical helicity as a consequence of Noether's theorem. Moreover, the whole Lagrangian electromagnetic field theory in vacuum, including its fundamental local Noether currents (energy–momentum, and angular momentum), becomes more natural and self-consistent with a properly complex and dual-symmetric Lagrangian [68]. This provides further evidence that the complex form of (1.1) is a more fundamental representation for the electromagnetic field.

In this report we clarify why the complex Riemann–Silberstein representation (1.1) has fundamental significance for the electromagnetic field. To accomplish this goal, we carefully construct the natural (Clifford) algebra of spacetime [4] and detail its intrinsic complex structure. We then derive the entirety of the traditional electromagnetic theory as an inevitable consequence of this *spacetime algebra*, paying close attention to the role of the complex structure at each stage. The complex vector (1.1) will appear naturally as a relative 3-space expansion of a *bivector field*, which is a proper geometric (and thus frame-independent) object in spacetime. Moreover, this bivector field will have a crucial difference from (1.1): the scalar

imaginary *i* will be replaced with an algebraic pseudoscalar *I* (also satisfying $I^2 = -1$) that is intrinsically meaningful to and required by the geometric structure of spacetime. This replacement makes the complex form of (1.1) *reference-frame-independent*, which is *impossible* when the mathematical representation is restricted to the usual scalar imaginary *i*. We find this replacement of the scalar *i* with *I* to be systematic throughout the electromagnetic theory, where it appears in electromagnetic waves, elliptical polarization, the normal variables for canonical field mode quantization, and the dual-symmetric internal rotations of the vacuum field. Evidently, the intrinsic complex structure inherent to the geometry of spacetime has deep and perhaps under-appreciated consequences for even our classical field theories.

This report is organized as follows. In Section 1.2 we briefly summarize some of the most interesting insights that we will uncover from spacetime algebra in the remainder of the report. In Section 2 we present a brief history of the formalism for the electromagnetic theory to give context for how the spacetime algebra approach relates. In Section 3 we give a comprehensive introduction to spacetime algebra, with a special focus on the emergent complex structure. In Section 4 we briefly extend this algebra to a spacetime manifold on which fields and calculus can be defined. In Section 5 we derive Maxwell's equations in vacuum from a *single*, simple, and inevitable equation. In Section 6 we consider three different potential representations, which are all naturally complex due to the dual symmetry of the vacuum field. In Section 7 we add a source to Maxwell's equation and recover the theory of electric and magnetic charges, along with the appropriately symmetrized Lorentz force and its associated energy–momentum and angular momentum tensors. In Section 8 we revisit the electromagnetic field theory. We start with the field Lagrangian and modify it to preserve dual symmetry in two related ways, derive the associated Noether currents, including helicity, and remark upon the breaking of dual symmetry that is inherent to the gauge mechanism. We conclude in Section 9.

1.2. Insights from the spacetime algebra approach

To orient the reader, we summarize a few of the interesting insights here that will arise from the spacetime algebra approach. These lists indicate what the reader can expect to understand from a more careful reading of the longer text. Note that more precise definitions of the quantities mentioned here are contained in the main text.

The Clifford product for spacetime algebra is initially motivated by the following basic benefits of its resulting *algebraic structure*:

- The primary reference-frame-independent objects of special relativity (i.e., scalars, 4-vectors, bivectors/anti-symmetric rank-2 tensors, pseudo-4-vectors, and pseudoscalars) are unified as distinct grades of a *single object known as a multivector*. An element of grade *k geometrically* corresponds to an oriented surface of dimension *k* (e.g., 4-vectors are line segments, while bivectors are oriented plane segments).
- The associative Clifford product combines the dot and wedge vector-products into a *single and often invertible product* (e.g., for a 4-vector $a^{-1} = a/a^2$).
- The symmetric part of the Clifford product is the *dot product* (Minkowski metric): $a \cdot b = (ab + ba)/2$. Thus, the square of a vector is its scalar pseudonorm $a^2 = \epsilon_a |a|^2$ with signature $\epsilon_a = \pm 1$.
- The antisymmetric part of the Clifford product is the *wedge product* $a \wedge b = (ab - ba)/2$ (familiar from differential forms), which generalizes the 3-vector cross product ($\times$).
- The sign ambiguity of the *spacetime metric signature* (dot product) is fixed to $(+, -, -, -)$ from general considerations that embed spacetime into a larger sequence of (physically meaningful) nested Clifford subalgebras.

From this structure we obtain a variety of useful and enlightening *mathematical unifications*:

- The *Dirac matrices* γ_μ (usually associated with quantum mechanics) appear as a matrix representation of an orthonormal basis for the 4-vectors in spacetime algebra, emphasizing that they are not intrinsically quantum mechanical in nature. The matrix product in the matrix representation simulates the Clifford product. One does *not* need this matrix representation, however, and can work directly with γ_μ as purely algebraic 4-vector elements, which dramatically simplifies calculations in practice.
- The *Dirac differential operator* $\nabla = \sum_\mu \gamma^\mu \partial_\mu$ appears as the proper vector derivative on spacetime, with *no* matrix representation required.
- The usual 3-*vectors* from nonrelativistic electromagnetism are *spacetime bivectors* that depend on a particular choice of inertial frame, specified by a timelike unit vector γ_0. A unit 3-vector $\vec{\sigma}_i = \gamma_i \gamma_0 = \gamma_i \wedge \gamma_0$ that is experienced as a spatial axis by an inertial observer thus has the geometric meaning of a *plane-segment* that is obtained by dragging the spatial unit 4-vector γ_i along the chosen proper-time axis γ_0.
- The *Pauli matrices* are a matrix representation of an orthonormal basis of relative 3-vectors $\vec{\sigma}_i$, emphasizing that they are also not intrinsically quantum mechanical in nature. The matrix product in the matrix representation again simulates the Clifford product.
- Factoring out a particular timelike unit vector γ_0 from a 4-vector $v = \sum_\mu v^\mu \gamma_\mu = (v_0 + \vec{v})\gamma_0$ produces a *paravector* $(v_0 + \vec{v})$ as a sum of a relative scalar v_0 and relative vector $\vec{v} = \sum_i v_i \vec{\sigma}_i$. Thus relative paravectors and proper 4-vectors are *dual* representations under right multiplication by γ_0.
- The usual 3-*gradient vector derivative* $\vec{\nabla} = \gamma_0 \wedge \nabla = \sum_i \vec{\sigma}_i \partial_i$ is the spatial part of the Dirac operator in a particular inertial frame specified by γ_0.

- The *d'Alembertian* is the square of the Dirac operator $\nabla^2 = \partial_0^2 - \vec{\nabla}^2$.
- *Directed integration* can be defined using a Riemann summation on a spacetime manifold, where the measure $d^k x$ becomes an oriented k-dimensional geometric surface that is represented within the algebra. This integration reproduces and generalizes the standard results from vector analysis, complex analysis (including Cauchy's integral theorems), and differential geometry.

All scalar components of the objects in spacetime algebra are *purely real numbers*. Nevertheless, an intrinsic *complex structure* emerges within the algebra due to the geometry of spacetime itself:

- The *"imaginary unit"* naturally appears as the pseudoscalar (unit 4-volume) $I = \gamma_0 \gamma_1 \gamma_2 \gamma_3 = \vec{\sigma}_1 \vec{\sigma}_2 \vec{\sigma}_3$, such that $I^2 = -1$. This pseudoscalar plays the role of the scalar imaginary i throughout the electromagnetic theory, without the need for any *ad hoc* introduction of a complex scalar field.
- The *Hodge-star duality operation* from differential forms is simply the right multiplication of any element by $I^{-1} = -I$. This operation transforms a geometric surface of dimension k into its orthogonal complement of dimension $4-k$ (e.g., the Hodge dual of a 4-vector v is its orthogonal 3-volume vI^{-1}, or pseudo-4-vector).
- The *quaternion algebra of Hamilton* $(1, \mathbf{i}, \mathbf{j}, \mathbf{k})$ that satisfies the defining relations, $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$ also appears as the (left-handed) set of spacelike planes (bivectors) in a relative inertial frame: $\mathbf{i} = \vec{\sigma}_1 I^{-1}, \mathbf{j} = -\vec{\sigma}_2 I^{-1}, \mathbf{k} = \vec{\sigma}_3 I^{-1}$. The bivectors $\vec{\sigma}_i I^{-1}$ directly describe the planes in which spatial rotations can occur, which explains why quaternions are particularly useful for describing spatial rotations.

In addition to the complex structure, *Lie group* and *spinor* structures also emerge:

- The bivector basis forms the *Lie algebra of the Lorentz group* under the Lie (commutator) bracket relation $[\mathbf{F}, \mathbf{G}] = (\mathbf{FG} - \mathbf{GF})/2$. Thus, bivectors directly generate *Lorentz transformations* when exponentiated.
- The *Lorentz group generators* are the 6 bivectors $\mathbf{S}_i = \vec{\sigma}_i I^{-1}$ and $\mathbf{K}_i = \vec{\sigma}_i$. Thus, writing the 15 brackets of the Lorentz group $[\mathbf{S}_i, \mathbf{S}_j] = \epsilon_{ijk} \mathbf{S}_k, [\mathbf{S}_i, \mathbf{K}_j] = \epsilon_{ijk} \mathbf{K}_k$ and $[\mathbf{K}_i, \mathbf{K}_j] = -\epsilon_{ijk} \mathbf{S}_k$ in terms of a particular reference frame produces simple variations of the three fundamental commutation relations $[\vec{\sigma}_i, \vec{\sigma}_j] = \epsilon_{ijk} \vec{\sigma}_k I$ that are usually associated with quantum mechanical spin (using a Pauli matrix representation). Indeed, the bivectors $\vec{\sigma}_i I^{-1}$ are spacelike planes that generate spatial rotations when exponentiated (just as in quantum mechanics), while $\vec{\sigma}_i$ are the timelike planes, and thus generate Lorentz boosts when exponentiated.
- The usual 3-*vector cross product* is the Hodge-dual of the bivector Lie bracket: $\mathbf{F} \times \mathbf{G} = [\mathbf{F}, \mathbf{G}] I^{-1}$. Thus, the set of commutation relations $[\vec{\sigma}_i, \vec{\sigma}_j] = \epsilon_{ijk} \vec{\sigma}_k I$ between the 3-vectors $\vec{\sigma}_i$ are simply another way of expressing the usual cross-product relations $\vec{\sigma}_i \times \vec{\sigma}_j = \epsilon_{ijk} \vec{\sigma}_k$.
- General *spinors* ψ appear as the (closed) even-graded subalgebra of spacetime algebra (i.e., scalars α, bivectors $\mathbf{F}$, and pseudoscalars αI). Lorentz and other group transformations acquire a simplified form as double-sided products with these spinors (e.g., exponentiated generators like $\psi = \exp(-\alpha \vec{\sigma}_3 I - \phi I) = [\cos \alpha - \sin \alpha \, \vec{\sigma}_3 I][\cos \phi - \sin \phi I]$), just as is familiar from unitary transformations in quantum mechanics and spatial rotations expressed using quaternions.

All these mathematical benefits produce a considerable amount of *physical* insight about the *electromagnetic theory* in a straightforward way:

- The *electromagnetic field* is a bivector field $\mathbf{F}$ with the same components $F^{\mu\nu}$ as the usual antisymmetric tensor. This tensor is the corresponding multilinear function $\underline{F}(v, w) = v \cdot \mathbf{F} \cdot w = \sum_{\mu\nu} v_\mu F^{\mu\nu} w_\nu$ that contracts its two 4-vector arguments (v and w) with the bivector $\mathbf{F}$.
- The electromagnetic field is an *irreducibly complex* object with an intrinsic phase $\mathbf{F} = \mathbf{f} \exp(\varphi I)$. This phase necessarily involves the intrinsic pseudoscalar (unit 4-volume) I of spacetime, and is intimately related to the appearance of *electromagnetic waves* and *circular polarizations*, with no need for any *ad hoc* addition of a complex scalar field.
- The continuous *dual (electric–magnetic exchange) symmetry* of the vacuum field is a U(1) internal gauge-symmetry (phase rotation) of the field: $\mathbf{F} \mapsto \mathbf{F} \exp(\theta I)$. This symmetry produces the conservation of helicity via Noether's theorem.
- Bivectors decompose in a relative inertial frame into a complex pair of 3-vectors $\mathbf{F} = \vec{E} + \vec{B} I$, which clarifies the origin of the *Riemann–Silberstein vector*. Both real and imaginary parts in a particular frame are thus physically meaningful. Moreover, the geometric properties of I keep $\mathbf{F}$ reference-frame-independent, even though the decomposition into relative fields $\vec{E} = (\mathbf{F} \cdot \gamma_0)\gamma_0$ and $\vec{B} I = (\mathbf{F} \wedge \gamma_0)\gamma_0$ still implicitly depends upon the chosen proper-time axis γ_0 in the relative 3-vectors $\vec{\sigma}_i = \gamma_i \gamma_0$ that form the basis for $\vec{E}$ and $\vec{B}$.
- All of *Maxwell's equations in vacuum* reduce to a single equation: $\nabla \mathbf{F} = 0$.
- All of *Maxwell's equations with sources* (both electric and magnetic) also reduce to a single equation: $\nabla \mathbf{F} = j$, where $j = j_e + j_m I$ is a complex representation of both types of source, $j_e = (c\rho_e + \vec{J}_e)\gamma_0$ and $j_m = (c\rho_m + \vec{J}_m)\gamma_0$.
- The scalar *Lorentz invariants* of the electromagnetic field are the invariant parts of its square: $\mathbf{F}^2 = (|\vec{E}|^2 - |\vec{B}|^2) + 2(\vec{E} \cdot \vec{B})I$.
- A *circularly polarized plane wave* is intrinsically complex with the simple exponential form $\mathbf{F}(x) = (sk) \exp[\pm(k \cdot x)I]$, with spacelike unit vector $s = \vec{E}_0 \gamma_0$, coordinates $x = (ct + \vec{x})\gamma_0$, and null wavevector $k = (\omega/c + \vec{k})\gamma_0$ such that $k^2 = |\omega/c|^2 - |\vec{k}|^2 = 0$ is the usual dispersion relation. The sign of I corresponds to the invariant handedness (i.e., *helicity*) of the wave. Expanding the exponential in the relative frame γ_0 yields $\mathbf{F} = \vec{E} + \vec{B} I$ with the relative fields $\vec{E} = \vec{E}_0 \cos(\vec{k} \cdot \vec{x} - \omega t/c) \pm \vec{\kappa} \times \vec{E}_0 \sin(\vec{k} \cdot \vec{x} - \omega t/c)$ and $\vec{B} = \vec{E}_0 \sin(\vec{k} \cdot \vec{x} - \omega t/c) \mp \vec{\kappa} \times \vec{E}_0 \cos(\vec{k} \cdot \vec{x} - \omega t/c)$, with unit vector $\vec{\kappa} = \vec{k}/|\vec{k}|$ indicating the propagation direction.

- The *vector-potential* representation is $\mathbf{F} = \nabla z = (\nabla \wedge a_e) + (\nabla \wedge a_m)I$, where $z = a_e + a_m I$ is a complex representation of both electric $a_e = (\phi_e + \vec{A})\gamma_0$ and magnetic $a_m = (\phi_m + \vec{C})\gamma_0$ 4-vector potentials, each satisfying the Lorenz–FitzGerald gauge conditions $\nabla \cdot a_e = \nabla \cdot a_m = 0$. Maxwell's equation then has the simple wave equation form $\nabla^2 z = j$ with the correspondingly complex source current $j = j_e + j_m I$.
- A *transverse and gauge-invariant bivector potential* representation can be defined in a particular inertial frame $\mathbf{Z} = z_\perp \gamma_0 = \vec{A}_\perp - \vec{C}_\perp I$, where $\mathbf{F} = (\nabla \mathbf{Z})\gamma_0$. This complex potential appears in the definition of the conserved optical helicity.
- The complex scalar *Hertz potential* $\Phi = \Phi_e + \Phi_m I$ for vacuum fields appears as $\mathbf{Z} = (\vec{\nabla} \times \vec{\Pi})I\Phi$ with a chosen unit direction vector $\vec{\Pi}$, and satisfies $\nabla^2 \Phi = 0$.
- The proper *Lorentz force* is $d(mw)/d\tau = \mathbf{F} \cdot (qw) = \langle \mathbf{F}(qw) \rangle_1$, where w is a proper 4-velocity of a particle with charge q and mass m.
- Making the charge q complex with a dual-symmetry phase rotation $q \mapsto q e^{\theta I} = q_e + q_m I$ produces an equivalent *magnetic monopole* description using the corresponding phase-rotated field $\mathbf{F} \mapsto \mathbf{F} e^{\theta I}$. The proper Lorentz force that includes these magnetic monopoles is still $d(mw)/d\tau = \langle \mathbf{F}(qw) \rangle_1$. Choosing electric sources is an arbitrary convention that essentially fixes this duality gauge freedom and *breaks the dual symmetry* of the fields.
- The *symmetric (Belinfante) energy–momentum tensor* is a bilinear function of the electromagnetic field and a chosen proper-time direction γ_0 that has a simple quadratic form: $\overleftrightarrow{T}_{\text{sym}}(\gamma_0) = \mathbf{F}\gamma_0\tilde{\mathbf{F}}/2 = (\vec{E} + \vec{B}I)(\vec{E} - \vec{B}I)\gamma_0/2 = (\varepsilon + \vec{P})\gamma_0$. This tensor recovers the usual energy density $\varepsilon = (|\vec{E}|^2 + |\vec{B}|^2)/2$ and Poynting vector $\vec{P} = \vec{E} \times \vec{B}$ in the relative frame γ_0, and has the same invariant mathematical form as the symmetric energy–momentum current for the Dirac theory of electrons.
- The corresponding (Belinfante) *angular momentum tensor* is a wedge product $\overleftrightarrow{M}_{\text{sym}}(\gamma_0) = x \wedge \overleftrightarrow{T}_{\text{sym}}(\gamma_0) = [\varepsilon \vec{x} - (ct)\vec{P}] + \vec{x} \times \vec{P}I^{-1}$ of a radial coordinate x with the energy–momentum tensor. The resulting angular momentum is a *bivector* that includes both boost and spatial-rotation angular momentum, which are distinguished by the extra factor of I in the spatial-rotational part (exactly as $\vec{\sigma}_i I^{-1}$ indicates a plane of rotation, while $\vec{\sigma}_i$ indicates a boost plane).
- The *dual-symmetric Lagrangian density* for the vacuum electromagnetic field is the scalar $\mathscr{L}_{\text{dual}}(x) = \langle (\nabla z)(\nabla z^*) \rangle_0/2$, which is a simple kinetic energy term for the complex vector potential field $z = a_e + a_m I$. This expands to the sum of independent terms for the electric and magnetic vector potentials: $\mathscr{L}_{\text{dual}} = \langle (\nabla a_e)^2 \rangle_0/2 + \langle (\nabla a_m)^2 \rangle_0/2$. Each term is identical in form to the traditional electromagnetic Lagrangian that only includes the electric part a_e.
- The *conserved Noether currents* of the dual-symmetric Lagrangian produce the proper conserved dual-symmetric *canonical energy–momentum tensor*, *orbital and spin angular momentum tensors*, and the *helicity pseudovector*.
- In the presence of only *electric* sources j_e, the *dual symmetry of the Lagrangian is broken*. The symmetry-breaking of this neutral vector boson doublet is entirely analogous to the boson doublet that is broken in the electroweak theory by the Brout–Englert–Higgs–Guralnik–Hagen–Kibble (BEHGHK, or Higgs) mechanism. According to this analogy, the second independent magnetic vector potential a_m that appears in the simple dual-symmetric electromagnetic Lagrangian may become related to the neutral Z_0 boson when additional interaction terms of the Standard Model are included.

2. A brief history of electromagnetic formalisms

> *In Science, it is when we take some interest in the great discoverers and their lives that it becomes endurable, and only when we begin to trace the development of ideas that it becomes fascinating.*
>
> [James Clerk Maxwell [128]]

The electromagnetic theory has a meandering mathematical history, largely due to the fact that the appropriate mathematics was being developed in parallel with the physical principles. As a result of this confusing evolution, the earlier and more general algebraic methods have been rediscovered and applied to electromagnetism only more recently. We summarize the highlights of this history in Table 1 for reference.

The original nonrelativistic unification of the physical concepts by Maxwell [129–133] was described as a set of 20 coupled differential equations of independent *scalar* variables. To conceptually simplify these equations down to 2 primary coupled equations that were easier to understand and solve, Maxwell later adopted and heavily advocated [134] the algebra of *quaternions*, which was developed by Hamilton [135,136] as a generalization of the algebra of complex numbers to include three independent "imaginary" axes that can describe rotations in three-dimensional space. The quaternionic formulation of electromagnetism was heavily attacked by Heaviside [110,137], who reformulated Maxwell's treatment into 4 coupled equations using an algebra of *3-vectors* (with the familiar dot and cross products), which was developed by Gibbs [138] as a simple subset of the algebraic work of Grassmann [139] and its extensions by Clifford [140]. This 3-vector reformulation of Heaviside has since remained the most widely known and used formulation of electromagnetism in practice, with the electric field $\vec{E}$ as a polar 3-vector and the magnetic field $\vec{B}$ as an axial 3-vector.

The relativistic treatment of electromagnetism has subsequently followed a rather different mathematical path, however, and now deviates substantially from these earlier foundations. Lorentz [141–143] first applied his eponymous transformations to moving electromagnetic bodies using the Heaviside formulation. His initial results were corrected and symmetrized by Poincaré [144,145], who suggested that the Lorentz transformation could be considered as a geometric rotation of a *4-vector* that had an *imaginary* time coordinate $(ct)i$ scaled by a constant c (corresponding to the speed of

Table 1
A brief history of the development of mathematical formalisms for representing the electromagnetic theory, showing the purely mathematical developments in italic font and their use in electromagnetism unitalicized. The bolded formalisms are the two most commonly used today: the nonrelativistic 3-Vectors of Gibbs preferred by Heaviside, and the relativistic tensor scalar components preferred by Einstein. Spacetime algebra was introduced by Hestenes in 1966 following Clifford's 1878 generalization of Grassmann's original 1844 exterior algebra that directly describes oriented geometric surfaces. Importantly, this spacetime algebra contains and generalizes *all* the other formalisms in a simple and powerful way.

History of electromagnetic formalisms		
Year	Nonrelativistic	Relativistic
1844	*Exterior algebra (Grassmann)*	
1853	*Quaternions (Hamilton)*	
1861	Scalar Components (Maxwell)	
1878	*Geometric algebra (Clifford)*	
1881	Quaternions (Maxwell)	
1892	**3-Vectors** (Gibbs & Heaviside)	
1899	*Differential forms (Cartan)*	
1901	Complex 3-Vectors (Riemann)	
1905		*4-Vectors with imaginary time (Poincaré)*
1907	Complex 3-Vectors (Silberstein)	
1908		*4-Vectors (Minkowski)*
1910		Complex 6-Vectors (Sommerfeld)
1911		Exterior Algebra (Wilson & Lewis)
1916		**Tensor scalar components** (Einstein)
1918		Differential forms (Weyl)
1966	**Spacetime algebra** (Hestenes)	

light in vacuum). This was the first unification of space and time into a single 4-vector entity of *spacetime*. The need for this invariant constant c was independently noticed by Einstein [146], and elevated to a postulate for his celebrated special theory of relativity that removed the need for a background aether. The 4-vector formalism suggested by Poincaré was fully developed by Minkowski [147], but Einstein himself argued against this construction [148] until his later work on gravitation made it necessary [149].

The 4-vector formalism of Poincaré and Minkowski was meanwhile developed by Sommerfeld [150,151], who emphasized that the electromagnetic field was not a 4-vector or combination of 4-vectors, but instead was a different type of object entirely that he called a "*6-vector*". This 6-vector became intrinsically *complex* due to the imaginary time (ct)i; specifically, it had both electric and magnetic components that differed by a factor of i and transformed into one another upon Lorentz boosts. Riemann [41] and Silberstein [42,43] independently noted this intrinsic complexity while working in the Heaviside formalism, which prompted them to write the total nonrelativistic field as the single complex vector (1.1) in agreement with the relativistic 6-vector construction of Sommerfeld.

Historically, however, the further development of the spacetime formalism of Poincaré led in a different direction. Minkowski [152] dropped the explicit scalar imaginary i attached to the time coordinate in favor of a different definition of the 4-vector dot product that produced the needed factor of -1 directly. This change in 4-vector notation removed the *ad hoc* scalar imaginary, but also effectively discouraged the continued development of the 6-vector of Sommerfeld (and the Riemann–Silberstein vector) by making the complex structure of spacetime implicit. In the absence of an explicit complex structure to distinguish the electric and magnetic field components, these components were reassembled into a *rank-2 antisymmetric tensor* (i.e., a multilinear antisymmetric function taking two vector arguments), whose characteristic components could be arranged into a 4×4 antisymmetric matrix

$$[F^{\mu\nu}] = \begin{bmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{bmatrix}. \tag{2.1}$$

Although the same components are preserved, this tensor formulation is more difficult to directly relate to the 3-vector formalism.

Of particular concern for practical applications, the familiar 3-vector cross product of Gibbs could no longer be used with the updated Minkowski 4-vector formalism, and its component representation with tensors like (2.1) was less conceptually clear. Without this cross product, laboratory practitioners had reduced physical intuition about the formalism, which hampered derivations and slowed the adoption of the relativistic formulation. In fact, an algebraic solution to this problem in the form of the *wedge product* had already been derived by Grassmann [139] and Clifford [140] (as we shall see shortly), and had even been adopted by Cartan's theory of differential forms [153]. However, this solution remained obscure to the physics community at the time (though it was partly rediscovered by Wilson and Lewis [154]). As such, the readily available methods for practical calculations with 4-vectors and tensors were either

(a) to convert the equations back into the nonrelativistic 3-vector notation and abandon manifest Lorentz covariance, or
(b) to work in *component-notation* like $F^{\mu\nu}$ above (and akin to the original papers by Maxwell), effectively obscuring the implicit algebraic structure of the relativistic theory.

The choice between these two practical alternatives has essentially created a cultural divide in the development and understanding of electromagnetism: the approach using nonrelativistic 3-vectors is still dominant in application-oriented fields like optics where physical intuition is required (e.g., [155]), while the component notation is dominant in theoretical high energy physics and gravitational communities that cannot afford to hide the structural consequences of relativity (though Cartan's differential forms do occasionally make an appearance, e.g., [156]). To make the component notation less cumbersome, Einstein developed an index summation convention for general relativity [157] (where repeated component indices are summed[3]: $v \cdot w = v^\mu w_\mu$), which has now been widely adopted by most practitioners who use component-based manipulations of relativistic quantities. Indeed, the formal manipulations of components using index-notation have essentially become synonymous with tensor analysis.

An elegant solution to this growing divide in formalisms was proposed a half-century later by Hestenes [4], who carefully revisited and further developed the geometric and algebraic work of Grassmann [139] and Clifford [140] to construct a full *spacetime algebra*. In modern mathematical terms, this spacetime algebra is the *orthogonal Clifford algebra* [158] that is uniquely constructed from the spacetime metric (dot product) proposed by Minkowski. This modern formalism has the great benefit of preserving and embracing all the existing algebraic approaches in use (including the complex numbers, the quaternions, and the 3-vectors of Gibbs), but it also extends them in a natural way to allow simple manipulations of relativistically invariant objects. For example, using spacetime algebra the electromagnetic field can be written in a relativistically invariant way (as we shall soon see) as

$$\mathbf{F} = [\vec{E}/c + \vec{B}I]/\sqrt{\mu_0}, \tag{2.2}$$

which has the same complex form as the Riemann–Silberstein vector (1.1). Importantly, however, the algebraic element I satisfying $I^2 = -1$ is *not* the scalar imaginary i added in an *ad hoc* way as in (1.1), but is instead an intrinsic *geometric* object of spacetime itself (i.e., the unit 4-volume) that emerges automatically alongside the 3-vector formalism of Gibbs. The geometric significance of the factor I makes the expression (2.2) a proper geometric object that is invariant under reference-frame changes, unlike (1.1). Moreover, the scalar components of (2.2) are precisely equivalent to the tensor components (2.1) even though $\mathbf{F}$ is not understood as an antisymmetric tensor. It is a geometric object in spacetime called a *bivector*, which is a modern refinement of the 6-vector concept introduced by Sommerfeld. The manifestly geometric significance of the electromagnetic field made manifest in this approach reaffirms and augments the topological fiber bundle foundations of modern gauge-field theories that was observed by Yang and Mills [159–162], Chern and Simons [163], and many others in theoretical high energy physics [164–170].

The spacetime algebra of Hestenes has been heavily developed by a relatively small community [6–8,171–178], but has recently been growing in popularity as a useful tool, and has been cross-pollinating with other modern mathematical investigations of Clifford algebras [158,179,180]. Indeed, spacetime algebra has been mentioned in a growing number of recent textbooks about algebraic approaches to physics and mathematics based on the work of Grassmann, Clifford, and Hestenes more generally [5,9,181–188]. Nevertheless, many of these treatments have under-emphasized certain features of spacetime that will be important for our discussion (such as its intrinsic complex structure) so we include our own introduction to this formalism that is tuned for applications in electromagnetism in what follows, assuming no prior background.

3. Spacetime algebra

Mathematics is taken for granted in the physics curriculum—a body of immutable truths to be assimilated and applied. The profound influence of mathematics on our conceptions of the physical world is never analyzed. The possibility that mathematical tools used today were invented to solve problems in the past and might not be well suited for current problems is never considered.

[David Hestenes [6]]

Physically speaking, *spacetime algebra* [4–7] is a complete and natural algebraic language for compactly describing physical quantities that satisfy the postulates of special relativity. Mathematically speaking, it is the largest associative algebra that can be constructed with the vector space of spacetime equipped with the Minkowski metric. It is an orthogonal Clifford algebra [158], which is a powerful tool that enables manifestly frame-independent and coordinate-free manipulations of geometrically significant objects. Unlike the dot and cross products used in standard vector analysis, the Clifford product between vectors is often *invertible* and not constrained to three dimensions. Unlike the component manipulations used in tensor analysis, spacetime algebra permits compact and component-free derivations that make the intrinsic geometric significance of physical quantities transparent.

When spacetime algebra is augmented with calculus then it subsumes many disparate mathematical techniques into a single comprehensive formalism, including (multi)linear algebra, vector analysis, complex analysis, quaternion analysis,

---

[3] Note that we will not use this convention in this report for clarity.

tensor analysis, spinor analysis, group theory, and differential forms [5,171–173]. Moreover, for those who are unfamiliar with any of these mathematical techniques, spacetime algebra provides an encompassing framework that encourages seamless transitions from familiar techniques to unfamiliar ones as the need arises. As such, spacetime algebra is also a useful tool for pedagogy [6,7].

During our overview we make an effort to illustrate how spacetime algebra contains and generalizes all the standard techniques for working with electromagnetism. Hence, one can appreciate spacetime algebra not as an obscure mathematical curiosity, but rather as a principled, practical, and powerful *extension* to the traditional methods of analysis. As such, all prior experience with electromagnetism is applicable to the spacetime algebra approach, making the extension readily accessible and primed for immediate use.

3.1. Spacetime

Recall that in special relativity [146] one postulates that the scalar time t and vector spatial coordinates $\vec{x}$ in a particular inertial reference frame always construct an invariant interval $(ct)^2 - |\vec{x}|^2$ that does not depend on the reference frame, where c is the speed of light in vacuum. An elegant way of encoding this physical postulate is to combine the scaled time and spatial components into a proper 4-vector $x = (ct, x_1, x_2, x_3)$ with a squared length equal to the invariant interval [144,145,147]. The proper notion of length is defined using the relativistic scalar (dot) product between two 4-vectors, which can be understood as a symmetric bilinear function $\eta(a, b)$ that takes two vector arguments a and b and returns their scalar shared length. This dot product is known as the *Minkowski metric* [152]. In terms of components $a = (a_0, a_1, a_2, a_3)$ and $b = (b_0, b_1, b_2, b_3)$ in a particular inertial frame, this metric has the form $a \cdot b \equiv b \cdot a \equiv \eta(a, b) = a_0 b_0 - a_1 b_1 - a_2 b_2 - a_3 b_3$. The proper squared length of a 4-vector x is therefore $x \cdot x$. Since there are one positive sign and three negative signs in this dot product, we say that it has a mixed *signature* $(+, -, -, -)$ and denote the vector space of 4-vectors as $\mathcal{M}_{1,3}$ to make this signature explicit.[4]

All physical vector quantities in special relativity are postulated to be 4-vectors in $\mathcal{M}_{1,3}$ that satisfy the Minkowski metric. These vectors are geometric quantities that do not depend on the choice of reference frame, so they shall be called *proper* relativistic objects in what follows. Importantly, the Minkowski metric is qualitatively different from the standard Euclidean dot product since it does not produce a positive length. Indeed, the mixture of positive and negative signs can make the length of a vector positive, negative, or zero, which produces three qualitatively different classes of vectors. We call these classes of vectors *timelike*, *spacelike*, and *lightlike*, respectively. As such, we can write the length, or *pseudonorm*, of a 4-vector a as $a \cdot a = \epsilon_a |a|^2$ in terms of a positive magnitude $|a|^2$ and a *signature* $\epsilon_a = \pm 1$ that is $+1$ for timelike vectors and -1 for spacelike vectors. For lightlike vectors the magnitude $|a|^2$ vanishes; thus, unlike for Euclidean vector spaces, a zero magnitude vector need not be the zero vector.

The Minkowski metric η and the vector space $\mathcal{M}_{1,3}$ are sufficient for describing proper vector quantities in special relativity, such as time–space $(ct, \vec{x})$ and energy–momentum $(\mathcal{E}/c, \vec{p})$. Since the laws of physics do not depend on the choice of reference frame, we expect that all physical quantities should be similarly represented by proper geometric objects like vectors. However, the electromagnetic field presents us with a conundrum: the polar 3-vector $\vec{E}$ of the electric field and axial 3-vector $\vec{B}$ of the magnetic field in a particular reference frame do not combine into proper 4-vectors. Relativistic angular momentum suffers a similar dilemma: the polar 3-vector $\vec{N} = (ct)\vec{p} - (\mathcal{E}/c)\vec{x}$ of the boost angular momentum (also known as the dynamic mass moment) and the axial 3-vector $\vec{L} = \vec{x} \times \vec{p}$ of the orbital angular momentum in a particular reference frame do not combine into proper 4-vectors. To resolve these dilemmas, the components of these vectors are typically assembled into the components $F^{\mu\nu}$ and $M^{\mu\nu}$ of rank-2 antisymmetric *tensors*, as we noted in (2.1) for the electromagnetic field [127,189]. This solution, while formally correct at the component level, is conceptually opaque. Why does a single rank-2 tensor $F^{\mu\nu}$ decompose into two 3-vector quantities $\vec{E}$ and $\vec{B}$ in a relative frame? How does a rank-2 tensor, which is mathematically defined as a multilinear *function* with two vector arguments, conceptually correspond to a physical quantity like the electromagnetic field or angular momentum? Is there some deeper significance to the mathematical space in which the tensors $F^{\mu\nu}$ and $M^{\mu\nu}$ reside? Do the proper tensor descriptions have any geometric significance in Minkowski space? Evidently, the vector space $\mathcal{M}_{1,3}$ does not contain the complete physical picture implied by special relativity, since it must be augmented by quantities like $F^{\mu\nu}$ and $M^{\mu\nu}$.

3.2. Spacetime product

To obtain the complete picture of special relativity in a systematic and principled way, we make a critical observation: any physical manipulation of vector quantities uses not only addition, but also *vector multiplication*. Indeed, standard treatments of electromagnetism involving relative 3-vectors use both the symmetric dot product and the antisymmetric vector cross

---

[4] Note that there is a sign-ambiguity in the spacetime interval, so one can seemingly choose either $(+, -, -, -)$ or $(-, +, +, +)$ for the metric signature, the latter being the initial choice of Poincaré. However, this choice is not completely arbitrary from an algebraic standpoint: it produces geometrically distinct spacetime algebras [158]. We choose the signature here that will produce the spacetime algebra that correctly contains the relative Euclidean 3-space as a proper subalgebra, which we detail in Section 3.8.
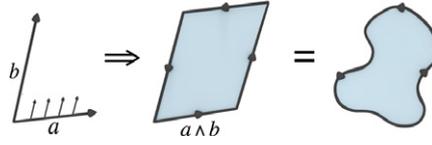
Fig. 1. The wedge product $a \wedge b$ between spacelike vectors a and b produces an oriented plane segment known as a bivector. Conceptually, the vector a slides along the vector b, sweeping out a parallelogram with area $|a \wedge b|$. The orientation of this area follows the right-hand rule, and may be visualized as a circulation around the boundary of the plane segment. Importantly, a bivector is characterized entirely by its magnitude and orientation, so the plane segment resulting from a wedge product may be deformed into any shape that preserves these two properties. Conversely, a bivector (of definite signature) may be factored into a wedge product between any two vectors that produce the same area and orientation $a \wedge b = c \wedge d$; if the two chosen factors are also orthogonal (i.e., $c \cdot d = 0$), then the bivector is a simple product of the orthogonal factors $c \wedge d = cd$.

product to properly discuss the physical implications of the theory. The vector space $\mathcal{M}_{1,3}$ only specifies the relativistic version of the dot product in the form of the Minkowski metric. Without introducing the proper relativistic notion of the cross product the physical picture of spacetime is incomplete.

Mathematically, the introduction of a product on a vector space creates an *algebra*. Hence, we seek to construct the appropriate algebra for spacetime from the vector space $\mathcal{M}_{1,3}$ by introducing a suitable vector product. We expect this vector product to be generally *noncommutative*, since the familiar cross product is also noncommutative.[5] We also expect the vector product to enlarge the mathematical space in order to properly accommodate quantities like the electromagnetic field tensor $F^{\mu\nu}$.

To accomplish these goals, we define the appropriate spacetime product to satisfy the following four properties for any vectors $a, b, c \in \mathcal{M}_{1,3}$:

$$a(bc) = (ab)c \quad \text{(Associativity)} \tag{3.1a}$$

$$a(b + c) = ab + ac \quad \text{(Left Distributivity)} \tag{3.1b}$$

$$(b + c)a = ba + ca \quad \text{(Right Distributivity)} \tag{3.1c}$$

$$a^2 = \eta(a, a) = \epsilon_a |a|^2 \quad \text{(Contraction)} \tag{3.1d}$$

Note that we omit any special product symbol for brevity. The contraction property (3.1d) distinguishes the resulting spacetime algebra as an *orthogonal Clifford algebra* [158] that is generated by the metric η and the vector space $\mathcal{M}_{1,3}$. This Clifford algebra is the largest associative algebra that can be constructed solely from spacetime, so it will contain all other potentially relevant algebras as subalgebras. Indeed, this nesting of algebras will be quite useful for practical calculations, as we shall see.

Decomposing the resulting associative vector product into symmetric and antisymmetric parts produces the proper spacetime generalizations to the 3-vector dot and cross products that we were seeking [140]:

$$ab = a \cdot b + a \wedge b. \tag{3.2}$$

The symmetric part of the product,

$$a \cdot b \equiv \frac{1}{2}(ab + ba) = b \cdot a = \eta(a, b), \tag{3.3}$$

is precisely the scalar (dot) product inherited from the spacetime structure of $\mathcal{M}_{1,3}$. The last equivalence follows from the contraction relation $(a + b)^2 = \eta(a + b, a + b)$ demanded by property (3.1d).

The antisymmetric part of the product,

$$a \wedge b \equiv \frac{1}{2}(ab - ba) = -b \wedge a, \tag{3.4}$$

is called the *wedge product* and is the proper generalization of the vector cross product to relativistic 4-vectors.[6] It produces a qualitatively new type of object called a *bivector* that does not exist *a priori* in $\mathcal{M}_{1,3}$, as anticipated. A bivector $(a \wedge b)$ produced from spacelike vectors a and b has the geometric meaning of a *plane segment* with magnitude equal to the area of the parallelogram bounded by a and b, and a surface orientation (handedness) determined by the right-hand rule; this construction is illustrated in Fig. 1. Hence, the wedge product generalizes the cross product by directly producing an oriented plane segment, rather than a vector normal to that surface. This generalization is important in spacetime since there is no unique normal vector to a plane in four dimensions. We will see in Sections 3.4, 3.6 and 5 that the electromagnetic field is properly expressed as precisely such a bivector.

The vector product (3.2) combines the nonassociative dot and wedge products into a single *associative* product. The result of the product thus decomposes into the sum of distinct scalar and bivector parts, which should be understood as

---

[5] Note that algebraic noncommutativity has nothing *a priori* to do with the noncommutativity in quantum mechanics.

[6] This is precisely Grassmann's exterior wedge product [139], adopted by Cartan when defining differential forms [153].

Grade	Orthonormal Basis						Blade Type	Geometry
4			γ_{0123}				Pseudoscalars	4-volumes
3		γ_{123}	γ_{230}	γ_{310}	γ_{120}		Pseudovectors	3-volumes
2	γ_{10}	γ_{20}	γ_{30}	γ_{23}	γ_{31}	γ_{12}	Bivectors	planes
1		γ_0	γ_1	γ_2	γ_3		Vectors	lines
0			1				Scalars	points

$$M = \langle M \rangle_0 + \langle M \rangle_1 + \langle M \rangle_2 + \langle M \rangle_3 + \langle M \rangle_4 \qquad \text{Multivectors}$$

Fig. 2. Graded basis for the spacetime algebra $\mathcal{C}_{1,3}$. Each multivector $M \in \mathcal{C}_{1,3}$ decomposes into a sum of distinct and independent grades $k = 0, 1, 2, 3, 4$, which can be extracted as grade-projections $\langle M \rangle_k$. The oriented basis elements of grade-1 $\{\gamma_\mu\}_{\mu=0}^3$ are an orthonormal basis $(\gamma_\mu \cdot \gamma_\nu = \eta_{\mu\nu})$ for the Minkowski 4-vectors $\mathcal{M}_{1,3}$. An oriented basis element of grade-k, such as $\gamma_{\mu\nu} \equiv \gamma_\mu \gamma_\nu = \gamma_\mu \wedge \gamma_\nu = -\gamma_\nu \wedge \gamma_\mu$ (with $\mu \neq \nu$), is constructed as a product of k of these orthonormal 4-vectors. Interchanging indices permute the wedge products, which only changes the sign of the basis element; hence, only the independent basis elements of each grade are shown. The color coding indicates the signature of each basis element (see Section 3.3.2), with blue being $+1$ and red being -1. The boxes and shading indicate useful dualities of the algebra: the solid and dashed boxes are (Hodge) dual under right multiplication of the pseudoscalar $I = \gamma_{0123}$ (see Section 3.5), while within each box the shaded region is dual to the unshaded region under right multiplication of the timelike basis vector γ_0 (see Section 3.6). These dualities are further detailed in Figs. 4 and 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

analogous to expressing a complex number as a sum of distinct real and imaginary parts. Just as with the study of complex numbers, it will be advantageous to consider these distinct parts as composing a unified whole, rather than separating them prematurely. We will explore this similarity more thoroughly in Section 3.5.

A significant benefit of combining both the dot and wedge products into a single associative product in this fashion is that an *inverse* may then be defined

$$a^{-1} = \frac{a}{a^2}, \tag{3.5}$$

provided a is not lightlike (i.e., $a^2 \neq 0$). Note that $a^2 = \epsilon_a |a|^2$ is a scalar by property (3.1d), so it trivially follows that $a^{-1}a = aa^{-1} = 1$. Importantly, neither the dot product nor the wedge product alone may be inverted; only their combination as the sum (3.2) retains enough information to define an inverse.

3.3. Multivectors

By iteratively appending all objects generated by the wedge product (3.4) to the initial vector space $\mathcal{M}_{1,3}$, we construct the full *spacetime algebra* $\mathcal{C}_{1,3}$. This notation indicates that the spacetime algebra is a *Clifford* algebra generated from the metric signature $(+, -, -, -)$. Importantly, all components in this Clifford algebra are purely *real*—we will not need any *ad hoc* addition of the complex scalar field in what follows.

The repeated wedge products produce 5 linearly independent subspaces of the total algebra, known as *grades*, which are illustrated in Fig. 2. Each grade is a distinct type of "directed number" [176]. The real scalars (pure numbers) $\alpha \in \mathbb{R}$ are grade 0, while the 4-vectors (line segments) $a \in \mathcal{M}_{1,3}$ are grade 1. The *bivectors* (plane segments) $(a \wedge b)$ appearing in (3.4) are grade 2. Successive wedge products will also produce *trivectors* (*pseudovectors*, 3-volume segments) $(a \wedge b \wedge c)$ of grade 3 and *quadvectors* (*pseudoscalars*, 4-volume segments) $(a \wedge b \wedge c \wedge d)$ of grade 4, which completes the algebra. We will refer to the elements of a grade-k subspace as k-blades in what follows to disambiguate them from the grade-1 vectors.

For concreteness, we systematically generate a complete graded basis for $\mathcal{C}_{1,3}$ as all independent products of the vectors $\{\gamma_\mu\}_{\mu=0}^3$ in a chosen basis of $\mathcal{M}_{1,3}$. The rich structure of the resulting graded basis is also detailed in Fig. 2. We choose the starting vector basis to be orthonormal $\gamma_\mu \cdot \gamma_\nu = \eta_{\mu\nu}$ in the sense of the Minkowski metric, so $\gamma_0^2 = 1$ and $\gamma_j^2 = -1$ for $j = 1, 2, 3$. The choice of notation for the basis is motivated by a deep connection to the Dirac γ-matrices that we will clarify in Section 3.8.

The basis of 0-blades is the real number 1. The basis of 1-blades is the chosen set of four orthonormal vectors γ_μ themselves. The $(16-4) = 12$ possible products of these vectors that produce bivectors (i.e., with $\mu \neq \nu$) $\gamma_\mu \gamma_\nu = \gamma_\mu \wedge \gamma_\nu \equiv \gamma_{\mu\nu} = -\gamma_{\nu\mu}$ produce only $(16-4)/2 = 6$ independent bivectors $\gamma_{10}, \gamma_{20}, \gamma_{30}, \gamma_{12}, \gamma_{23}, \gamma_{31}$ due to the antisymmetry of the wedge product; these independent elements form the oriented basis of 2-blades. Similarly, products of orthogonal vectors with these bivectors $\gamma_\mu \gamma_\nu \gamma_\delta = \gamma_\mu \wedge \gamma_\nu \wedge \gamma_\delta \equiv \gamma_{\mu\nu\delta}$ produce only 4 independent trivectors $\gamma_{123}, \gamma_{120}, \gamma_{230}, \gamma_{310}$ that form the oriented basis of 3-blades. Finally, the product of all four orthogonal vectors produces only a single independent element $\gamma_0 \gamma_1 \gamma_2 \gamma_3 = \gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3 \equiv \gamma_{0123}$ that serves as the basis for the 4-blades. Hence, there are $2^4 = 16$ independent basis elements for the spacetime algebra, partitioned into 5 grades of "4 choose k", $4!/k!(4-k)!$, independent basis k-blades.

The 4-blade γ_{0123} geometrically signifies the same unit 4-volume in any basis, and will be particularly important in what follows, so we give it a special notation

$$I \equiv \gamma_{0123}. \tag{3.6}$$

Table 2

Notational distinctions. Proper (frame-independent) quantities (top) have the different grades distinguished by font type, boldface, and capitalization for simplicity, while relative (frame-dependent) quantities (bottom) conform to traditional vector or index annotations for clarity.

Notation	Description
$\alpha,\ \beta,\ \omega$	Scalars (real numbers)
a, b, v, w	4-vectors
$\mathbf{F}$, $\mathbf{G}$	Bivectors
vI	Pseudovectors
αI	Pseudoscalars
M	Multivector
$\zeta = \alpha + \beta I$	Complex scalar
$z = v + wI$	Complex vector
∇	4-gradient, Dirac operator
$\underline{T}(b), \overline{T}(b)$	Tensor, adjoint tensor
γ_μ	Orthonormal basis (Lorentz frame) for 4-vectors
$\gamma^\mu = \gamma_\mu^{-1}$	Reciprocal basis for 4-vectors
$\vec{E}, \vec{B}$	Relative 3-vectors
$\vec{\sigma}_i = \gamma_i \gamma_0$	Orthonormal basis for 3-vectors
$\vec{\nabla} = \gamma_0 \wedge \nabla$	Relative 3-gradient
$\partial_0 = \gamma_0 \cdot \nabla$	Relative time derivative
$F^{\mu\nu},\ v^\mu$	Scalar components ($\mu = 0, 1, 2, 3$)
$F^{ij},\ v^k$	Spatial scalar components ($i, j, k = 1, 2, 3$)
$\eta_{\mu\nu}$	Minkowski metric tensor
$\delta_{\mu\nu}$	Euclidean metric tensor/Kronecker delta
$\epsilon_{ijk}, \epsilon_{\mu\nu\delta\kappa}$	Anti-symmetric Levi-Civita symbols

This algebraic element is the *pseudoscalar* for spacetime. That is, its scalar multiples αI act as scalars that flip sign under an inversion of the handedness of spacetime (i.e., flipping the orientation of any one basis direction). The notation of I is motivated by the fact that $I^2 = \gamma_0^2 \gamma_1^2 \gamma_2^2 \gamma_3^2 = (1)(-1)(-1)(-1) = -1$, so it will perform a similar function to the scalar imaginary i. Indeed, the element I provides spacetime with an intrinsic *complex structure* without any *ad hoc* introduction of the scalar complex numbers.[7] We will detail this complex structure in Section 3.5.

Any *multivector* $M \in \mathcal{C}_{1,3}$ in the spacetime algebra may be written as a sum of independent k-blades

$$M = \alpha + v + \mathbf{F} + \mathfrak{T} + \beta I \qquad (3.7)$$

where the Greek letters $\alpha,\ \beta$ are real numbers, the lowercase Roman letter v is a vector with 4 real components, the boldface Roman letter $\mathbf{F}$ is a bivector with 6 real components, and the Fraktur letter $\mathfrak{T}$ is a trivector with 4 real components. Each of these independent grades are distinct and proper geometric objects. We shall see in Section 3.5 that in practice we can dramatically simplify the description of multivectors by exploiting dualities of the algebra; in particular, we will be able to dispense with trivectors (and their elaborate notation) altogether by rewriting them as *pseudovectors* $\mathfrak{T} = vI$. To keep these distinctions conceptually clear, we shall make an effort to maintain useful notational conventions throughout this work that are summarized in Table 2 for reference.

Each k-blade in a multivector can be extracted by a suitable grade-projection, denoted as $\langle M \rangle_k$. For example, the bivector projection of (3.7) is $\langle M \rangle_2 = \mathbf{F}$. Notably, the scalar projection satisfies the cyclic property

$$\langle MN \rangle_0 = \langle NM \rangle_0 \qquad (3.8)$$

for any two multivectors M and N, so is an algebraic *trace* operation. In fact, this trace is entirely equivalent to the matrix trace operation, which can be verified by considering a (Dirac) matrix representation of the spacetime algebra $\mathcal{C}_{1,3}$ that simulates the noncommutative vector product using the standard matrix product. We will comment more on such a matrix representation in Section 3.8, but we shall not require it in what follows.

All proper physical quantities in spacetime must either correspond to pure k-blades, or multivectors that combine k-blades of differing grades. These are the only frame-independent objects permitted by the constraints of special relativity. Hence, the expansion (3.7) of a general multivector indicates the full mathematical arena in which relativistic physics must occur. The algebraic structure allows us to manipulate each of these proper objects with equal ease and finesse.

The added complication of multivectors $M \in \mathcal{C}_{1,3}$ over the original 4-vectors $v \in \mathcal{M}_{1,3}$ may seem excessive and intimidating at first glance; however, as we have been suggesting, the added structure contained in a full multivector is in fact necessary for a proper description of spacetime quantities. Indeed, as we will see in more detail in Sections 3.4 and 3.6, each distinct blade of a general multivector is in fact a familiar type of object that appears in the standard treatments of relativistic physics. We summarize several examples in Table 3 for reference. The multivector construction unifies all these

---

[7] It is worth emphasizing that there is no unique square root of -1. There are many algebraic elements in $\mathcal{C}_{1,3}$ that square to -1, each with distinct geometric significance. The scalar imaginary i of standard complex analysis does not specify any of this additional structure; hence, it is often enlightening to determine whether a particular $\sqrt{-1}$ is implied by a generic i that appears in a traditional physics expression. We will show in what follows that I is indeed the proper physical meaning of i throughout the electromagnetic theory.

Table 3
Examples of common physical quantities by grade. All these physical quantities appear in any proper description of relativistic physics, and fit naturally within the graded structure of spacetime algebra.

Grade	Blade type	Physical examples
0	Scalars	Charge q Mass m Proper time τ
1	Vectors	Coordinates x Proper velocity $w = dx/d\tau$ Energy–momentum p Force $dp/d\tau$ Electric potential a_e Electric current j_e
2	Bivectors	Electromagnetic field $\mathbf{F} = \vec{E} + \vec{B}I$ Angular momentum $\mathbf{M} = x \wedge p = -\vec{N} + \vec{L}I^{-1}$ Torque $d\mathbf{M}/d\tau$ Vorticity $\nabla \wedge w$
3	Pseudovectors	magnetic potential $a_m I$ Magnetic pseudocurrent $j_m I$ Helicity pseudocurrent XI
4	Pseudoscalars	Magnetic charge $q_m I$ Phase φI

quantities into a comprehensive whole in a principled way. Moreover, the Clifford product provides them with a wealth of additional structure, which we will exploit to make manipulations of multivectors simple in practice.

3.3.1. Bivectors: products with vectors

We will see in the following sections that several dualities of spacetime algebra allow a complete understanding in terms of scalars, vectors, and bivectors. Out of these three fundamental objects bivectors are the least familiar, so we shall make an effort to clarify their properties as we progress through this introduction.

We now consider how bivectors and vectors relate to one another in more detail, defining two useful operations (*contraction* and *inflation*) in the process. It is worth emphasizing that all operations in spacetime algebra are derived from the fundamental vector product and sum, so are introduced for calculational convenience, conceptual clarity, and for contact with existing formalisms. For reference, we list all of the auxiliary operations that will be introduced in this report in Table 4.

A product between a vector a and a bivector $\mathbf{F}$ can be split into two terms (vector and trivector) in an analogous way to the product of vectors in (3.2)

$$a\mathbf{F} = a \cdot \mathbf{F} + a \wedge \mathbf{F}. \tag{3.9}$$

The vector and trivector parts are constructed respectively as

$$a \cdot \mathbf{F} \equiv \langle a\mathbf{F} \rangle_1 = \frac{1}{2}(a\mathbf{F} - \mathbf{F}a) = -\mathbf{F} \cdot a, \tag{3.10}$$

$$a \wedge \mathbf{F} \equiv \langle a\mathbf{F} \rangle_3 = \frac{1}{2}(a\mathbf{F} + \mathbf{F}a) = \mathbf{F} \wedge a. \tag{3.11}$$

Notably, the grade-lowering dot product, or *contraction*, is antisymmetric while the grade-raising wedge product, or *inflation*, is symmetric, which is opposite to the vector product case (3.2). The use of the dot product to lower the grade and the wedge product to raise the grade has become conventional [5].

The geometric meaning of these products can be ascertained for spacelike $\mathbf{F}$ by decomposing the vector $a = a_\perp + a_\parallel$ into parts perpendicular ($a_\perp$) and parallel ($a_\parallel$) to the plane of $\mathbf{F}$. After factoring $\mathbf{F}$ into constituent orthogonal vectors (i.e., $b \cdot c = 0$), $\mathbf{F} = b \wedge c = bc$, we find

$$a\mathbf{F} = abc = \frac{1}{2}a_\parallel(bc - cb) + a_\perp \wedge b \wedge c, \tag{3.12}$$

$$\mathbf{F}a = bca = \frac{1}{2}(bc - cb)a_\parallel + a_\perp \wedge b \wedge c, \tag{3.13}$$

so the dot and wedge products have the simple forms

$$a \cdot \mathbf{F} = a_\parallel \cdot \mathbf{F} = (a_\parallel \cdot b)c - (a_\parallel \cdot c)b, \tag{3.14}$$

$$a \wedge \mathbf{F} = a_\perp \wedge \mathbf{F} = a_\perp \wedge b \wedge c. \tag{3.15}$$

For spacelike a, b, c, these expressions have intuitive geometric interpretations. The contraction $a \cdot \mathbf{F}$ produces a vector in the plane of $\mathbf{F}$ that is perpendicular to $a_\parallel$. Indeed, choosing $b = a_\parallel$ so $\mathbf{F} = a_\parallel \wedge c$ yields $a \cdot \mathbf{F} = c$. The wedge product
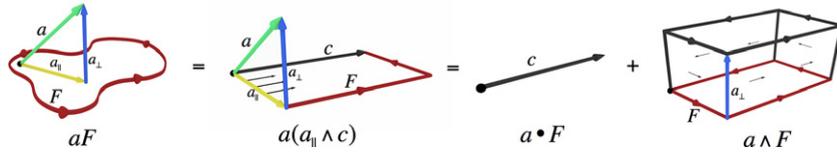
Fig. 3. The product $a\mathbf{F}$ between a vector a (green) and bivector $\mathbf{F}$ (red) decomposes into a vector $a \cdot \mathbf{F}$ and a trivector $a \wedge \mathbf{F}$ part. A spacelike $\mathbf{F}$ is an oriented plane segment that may be deformed into a parallelogram $\mathbf{F} = a_{\parallel} \wedge c$ that shares a side with the part of the spacelike vector $a = a_{\parallel} + a_{\perp}$ that is parallel to $\mathbf{F}$ (yellow). The contraction extracts the orthogonal vector $c = a \cdot \mathbf{F}$ (black), while the inflation drags the bivector $\mathbf{F}$ along the perpendicular vector $a_{\perp}$ (blue) to produce an oriented trivector volume $a_{\perp} \wedge \mathbf{F}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Algebraic operations. From the fundamental Clifford product we define a variety of convenient operations that are useful for calculations.

Description	Notation & Definition
Clifford product	ab
Vector dot product	$a \cdot b = (ab + ba)/2 = b \cdot a$
Vector wedge product	$a \wedge b = (ab - ba)/2 = -b \wedge a$
Projection to grade-k	$\langle M \rangle_k$
Algebraic trace	$\langle M \rangle_0$
Hodge duality	$M \mapsto MI^{-1}$
Relative frame duality	$M \mapsto M\gamma_0$
Bivector contraction	$a \cdot \mathbf{F} = (a\mathbf{F} - \mathbf{F}a)/2 = -\mathbf{F} \cdot a$
Bivector inflation	$a \wedge \mathbf{F} = (a\mathbf{F} + \mathbf{F}a)/2 = \mathbf{F} \wedge a$
Bivector dot product	$\mathbf{F} \cdot \mathbf{G} = (\mathbf{FG} + \mathbf{GF})/2 = \mathbf{G} \cdot \mathbf{F}$
Commutator bracket	$[\mathbf{F}, \mathbf{G}] = (\mathbf{FG} - \mathbf{GF})/2 = -[\mathbf{G}, \mathbf{F}]$
Bivector cross product	$\mathbf{F} \times \mathbf{G} = [\mathbf{F}, \mathbf{G}]I^{-1} = -\mathbf{G} \times \mathbf{F}$
Reversion (transpose)	$(ab)^{\sim} = ba$
	$(MN)^{\sim} = \widetilde{N}\widetilde{M}$
Positive magnitude	$|M|^2 = \lfloor\langle \widetilde{M}M \rangle_0|^2 + |\langle \widetilde{M}M \rangle_4 I^{-1}|^2$
Signature	$\epsilon_M = \langle \widetilde{M}M \rangle_0 / |M|^2$
Complex conjugation	$(\alpha + \beta I)^* = (\alpha - \beta I)$
	$(v + wI)^* = (v - wI)$
	$[\mathbf{f}\,e^{\varphi I}]^* = \mathbf{f}\,e^{-\varphi I}$
Relative reversion	$M^{\dagger} = \gamma_0 \widetilde{M} \gamma_0$
	$(\vec{E} + \vec{B}I)^{\dagger} = \vec{E} - \vec{B}I$

$a \wedge \mathbf{F}$ produces a trivector with magnitude equal to the volume of the parallelepiped constructed from dragging the plane segment $\mathbf{F}$ along a perpendicular vector $a_{\perp}$. These constructions are illustrated in Fig. 3.

3.3.2. Reversion and inversion

Before we continue, we make a brief diversion to define another useful operation on the spacetime algebra known as *reversion* [5], which reverses the order of all vector products: $(ab)^{\sim} = ba$. The reverse distributes across general multivector products recursively $(AB)^{\sim} = \widetilde{B}\widetilde{A}$, and is the algebraic equivalent of the *transpose* operation for matrices (as can be verified using a matrix representation). The reverse inverts itself $(\widetilde{M})^{\sim} = M$, which makes it an *involution* on the algebra. The reverse is also summarized in Table 4 for reference.

The reverse of a bivector $\mathbf{F}$ is its negation $\widetilde{\mathbf{F}} = -\mathbf{F}$, which can be seen by splitting each basis element of $\mathbf{F}$ into orthogonal factors $\gamma_{\mu\nu} = \gamma_{\mu}\gamma_{\nu} = \gamma_{\mu} \wedge \gamma_{\nu}$. The reversion of the factors then flips the wedge product $\widetilde{\gamma}_{\mu\nu} = \gamma_{\nu}\gamma_{\mu} = \gamma_{\nu} \wedge \gamma_{\mu} = -\gamma_{\mu} \wedge \gamma_{\nu} = -\gamma_{\mu\nu}$, which results in a negation. Each grade of the general multivector in (3.7) can be reversed in an analogous way according to

$$\widetilde{M} = \alpha + v - \mathbf{F} - \mathfrak{T} + \beta I. \tag{3.16}$$

Notably, the pseudoscalar reverses to itself $\widetilde{I} = \gamma_{3210} = \gamma_{0123} = I$.

As a useful application of the reversion, the product of a pure k-blade $M = \langle M \rangle_k$ with its reverse $\widetilde{M}$ produces a scalar. As an example, for $M = \gamma_{123}$ we have $\widetilde{M}M = \gamma_{321}\gamma_{123} = \gamma_{32}(\gamma_1)^2\gamma_{23} = -\gamma_3(\gamma_2)^2\gamma_3 = \gamma_3^2 = -1$. The resulting positive magnitude $|M|^2 \equiv |\widetilde{M}M| = 1$ is a product of the magnitudes of the factors of M, while the sign $\epsilon_M \equiv \widetilde{M}M/|M|^2 = -1$ is a product of the *signatures* of the factors. Hence, the reversed-square of a pure k-blade

$$\widetilde{M}M = M\widetilde{M} = \epsilon_M |M|^2 \tag{3.17}$$

produces a notion of *pseudonorm* for M with a net signature ϵ_M and positive magnitude $|M|^2$, in exact analogy with the definition (3.1d) for vectors.[8] It follows that if $|M|^2 \neq 0$, then

$$M^{-1} \equiv \frac{\widetilde{M}}{\widetilde{M}M} \tag{3.18}$$

is the inverse of a pure k-blade M that satisfies $M^{-1}M = MM^{-1} = 1$, in analogy with (3.5).

3.4. Reciprocal bases, components, and tensors

To connect with the standard tensor-analysis treatments of the electromagnetic field, where it is considered to be an antisymmetric rank-2 tensor $F^{\mu\nu}$, we now briefly consider the various ways of expanding proper multivector quantities in components. We shall see that we immediately recover and clarify the standard tensor analysis formulas, making the tensor analysis techniques available as a restricted consequence of spacetime algebra.

Given a reference basis γ_μ for $\mathcal{M}_{1,3}$, it is convenient to introduce a *reciprocal basis* of vectors

$$\gamma^\mu = (\gamma_\mu)^{-1}, \qquad \gamma^\mu \gamma_\nu = \gamma^\mu \cdot \gamma_\nu = \delta^\mu_\nu \tag{3.19}$$

defined by the algebraic inverse (3.5). This is an equally valid basis of vectors for $\mathcal{M}_{1,3}$ that satisfies the same metric relation $\gamma^\mu \cdot \gamma^\nu = \eta^{\mu\nu} = \eta_{\mu\nu}$. It also follows from the inverse that $\gamma^0 = \gamma_0$, while $\gamma^j = -\gamma_j$ for $j = 1, 2, 3$, so the reciprocal basis is the spatial inversion of the reference basis γ_μ.

Any multivector M is a geometric object that can be expanded in any basis, including the reciprocal basis. For example, a vector v has the expansions

$$v = \sum_\mu v^\mu \gamma_\mu = \sum_\mu v_\mu \gamma^\mu \tag{3.20}$$

in terms of the reference basis and the reciprocal basis. Traditionally, the components $v^\mu = \gamma^\mu \cdot v$ in the reference basis are called the *contravariant* components of v, while the components $v_\mu = \gamma_\mu \cdot v$ in the reciprocal basis are the *covariant* components of v; we see here that they can be interpreted as different ways of expressing the same geometric object. Evidently from (3.20) we have $v \cdot \gamma_\nu = v^\nu \eta_{\nu\nu} = v_\nu$, which shows how the two component representations are related by the metric (which effectively "raises" or "lowers" the index).

Since the graded basis elements of the spacetime algebra are constructed with the antisymmetric wedge product from the reference (or reciprocal) vector basis, then one can also expand a k-blade into redundant components that will be the same as the components of a rank-k antisymmetric tensor. For example, a bivector $\mathbf{F}$ may be expanded into components as

$$\mathbf{F} = \frac{1}{2} \sum_{\mu,\nu} F^{\mu\nu} \gamma_\mu \wedge \gamma_\nu = \frac{1}{2} \sum_{\mu,\nu} F_{\mu\nu} \gamma^\mu \wedge \gamma^\nu. \tag{3.21}$$

The components $F^{\mu\nu} = \gamma^\mu \cdot \mathbf{F} \cdot \gamma^\nu = \gamma^\nu \cdot (\gamma^\mu \cdot \mathbf{F}) = (\gamma^\nu \wedge \gamma^\mu) \cdot \mathbf{F} = \gamma^{\nu\mu} \cdot \mathbf{F} = (\gamma_{\mu\nu})^{-1} \cdot \mathbf{F}$ can be extracted using the contractions defined in Section 3.3.1, and are usually called the rank $(2,0)$ (contravariant) components of $\mathbf{F}$, while the components $F_{\mu\nu} = \gamma_\mu \cdot \mathbf{F} \cdot \gamma_\nu$ are the rank $(0,2)$ (covariant) components of $\mathbf{F}$. Similarly, the components $F^\mu{}_\nu = \gamma^\mu \cdot \mathbf{F} \cdot \gamma_\nu$ and $F_\mu{}^\nu = \gamma_\mu \cdot \mathbf{F} \cdot \gamma^\nu$ are rank $(1,1)$ (matrix) components for $\mathbf{F}$. There are overtly 16 real components in each case, of which only $(16-4)/2 = 6$ are nonzero and independent due to the antisymmetry; these six independent components correspond precisely to the components of $\mathbf{F}$ when expanded more naturally in a bivector basis. The factor of $1/2$ in (3.21) ensures that the redundant components (differing only by a sign) are not double-counted.

The various ranked tensor components corresponding to $\mathbf{F}$ all refer to the action of the same antisymmetric tensor $\underline{F}$, which is a multilinear *function*

$$\underline{F}(a, b) \equiv a \cdot \mathbf{F} \cdot b = b \cdot (a \cdot \mathbf{F}) = (b \wedge a) \cdot \mathbf{F}$$
$$= \sum_{\mu\nu} a^\mu F_{\mu\nu} b^\nu = \sum_{\mu\nu} a_\mu F^{\mu\nu} b_\nu = \sum_{\mu\nu} a_\mu F^\mu{}_\nu b^\nu = \sum_{\mu\nu} a^\mu F_\mu{}^\nu b_\nu \tag{3.22}$$

that takes two vector arguments $a = \sum_\mu a^\mu \gamma_\mu = \sum_\mu a_\mu \gamma^\mu$ and $b = \sum_\nu b^\nu \gamma_\nu = \sum_\nu b_\nu \gamma^\nu$ and produces a scalar through a total contraction with the bivector $\mathbf{F}$. We use the underbar notation $\underline{F}$ for functions to disambiguate them from products of multivectors. Note that the different ranks of components for the tensor $\underline{F}$ correspond to different ways of expanding the arguments a and b into different bases.

Importantly, the electromagnetic field is intrinsically a bivector $\mathbf{F}$ and *not* its associated antisymmetric tensor $\underline{F}$, which is the multilinear function that performs contractions with $\mathbf{F}$ to produce a scalar; the confusion between these two distinct

---

[8] It will be shown in Section 3.5.1 that bivectors $\mathbf{F}$ may not have a well-defined signature of ± 1. Nevertheless, these reversion equations will still be valid.
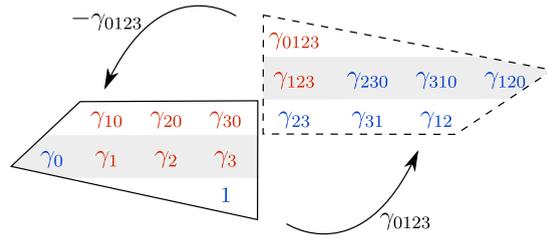
Fig. 4. Hodge duality illustrated using the same basis as Fig. 2. Right multiplication by the pseudoscalar $I = \gamma_{0123}$ converts an algebraic element to its dual, which is an orthogonal complement in the geometric sense. For example, the dual of a trivector γ_{123} is its unique normal vector $\gamma_{123}(-I) = \gamma_0$. Dual elements have opposite signatures, indicated here by the flip in color coding. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
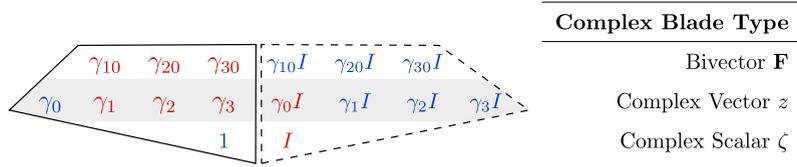
Fig. 5. Complex structure of the spacetime algebra. The complementary Hodge-dual halves of the algebra illustrated in Fig. 4 become intrinsically complex pairings by factoring out the pseudoscalar $I = \gamma_{0123}$ that satisfies $I^2 = -1$. The complex scalars and vectors have parts that are independent geometric objects. However, the bivectors are self-dual and irreducibly complex objects that cannot be further decomposed into frame-independent parts.

concepts arises because they have the same characteristic components $F^{\mu\nu}$. Component-based tensor analysis obscures this subtle conceptual distinction by neglecting the k-blades themselves in favor of the functions that can be defined by contractions with these k-blades, all while emphasizing component descriptions that depend on particular basis expansions. More distressingly, one cannot construct more general multivectors or the Clifford product using tensor notation. As a result, we regard component-based tensor analysis as a part of – but not a replacement for – the spacetime algebra used in this report.

3.5. The pseudoscalar I, Hodge duality, and complex structure

The expansion (3.7) of a multivector in terms of elements of differing grade is similar to the expansion of a complex number into its real and imaginary parts, as we emphasized in Section 3.3. In fact, this similarity is more than an analogy because the pseudoscalar I also satisfies $I^2 = -1$, which makes the subalgebra of the form $\zeta = \alpha + \beta I$ completely equivalent in practice to the complex scalar numbers. Hence, we do not need to additionally complexify the algebra in order to reap the benefits of complex analysis; the complex algebraic structure automatically appears within the spacetime algebra itself.

The pseudoscalar I has intrinsic geometric significance that goes well beyond the theory of scalar complex numbers, however, and has several additional interesting and useful properties. It commutes with elements of even grade (i.e., scalars, pseudoscalars, and bivectors), but *anticommutes* with elements of odd grade (i.e., vectors and trivectors). More surprisingly, a right product with I is a *duality transformation* from grade k to its orthogonal complement of grade $(4 - k)$. For example, the dual of a vector is its unique oriented orthogonal trivector $\gamma_0 I = \gamma_0 \gamma_0 \gamma_1 \gamma_2 \gamma_3 = \gamma_{123}$, while the dual of a trivector is its unique normal vector $\gamma_{123} I = (\gamma_0 I) I = -\gamma_0$. This duality is illustrated for the entire graded basis in Fig. 4. We will see in Section 5.2 that this duality is intimately connected to the field-exchange *dual symmetry* of the electromagnetic field.

The duality transformation induced by I is equivalent to the *Hodge-star* transformation in differential forms (though is arguably simpler to work with), and splits the spacetime algebra into two halves that are geometric complements of each other. Exploiting this duality, we can write any multivector M in an intrinsically complex form (in the sense of I) that pairs quantities with their duals

$$M = \zeta + z + \mathbf{F}. \tag{3.23}$$

We illustrate this decomposition in Fig. 5. The complex scalar ζ and complex vector z parts have the form

$$\zeta \equiv \alpha + \beta I, \qquad z \equiv v + wI = v - Iw, \tag{3.24}$$

where α, β are real scalars, and v, w are 4-vectors. As anticipated, the trivector $\mathfrak{T} = wI = -Iw$ has been expressed as the orthogonal complement of a vector (i.e., a *pseudovector*), which is a useful simplification.

Some care has to be taken with manipulations of complex vectors z, since I anti-commutes with vectors. As a result, complex scalars and vectors only quasi commute

$$\zeta z = (\alpha + \beta I)z = z(\alpha - \beta I) \neq z\zeta, \tag{3.25}$$

while complex scalars and bivectors commute normally: $\zeta \mathbf{F} = \mathbf{F}\zeta$. Thus, the pseudoscalar I is only algebraically equivalent to the usual notion of the scalar imaginary i when restricted to the *even* graded subalgebra of complex scalars and bivectors.

3.5.1. Bivectors: canonical form

A bivector like the electromagnetic field, $\mathbf{F}$, is an interesting special case for the complex decomposition (3.23) of a general multivector, since the full bivector basis is already *self-dual*. We can see this by computing the duality for a particular basis

$$
\begin{aligned}
&\gamma_{10}I = -\gamma_{23}, \qquad \gamma_{20}I = -\gamma_{31}, \qquad \gamma_{30}I = -\gamma_{12}, \\
&\gamma_{23}I = \gamma_{10}, \qquad \gamma_{31}I = \gamma_{20}, \qquad \gamma_{12}I = \gamma_{30}.
\end{aligned}
\tag{3.26}
$$

Geometrically, these duality relations express the fact that in 4-dimensions the orthogonal complement to any plane is another plane (not a normal vector as in 3-dimensions). We can express these relations more compactly by collecting cyclic permutations and writing $\gamma_{i0}I = -\epsilon_{ijk}\gamma_{jk}$ and $\gamma_{jk}I = \epsilon_{ijk}\gamma_{i0}$, where $i, j, k = 1, 2, 3$ and ϵ_{ijk} is the completely antisymmetric Levi-Civita symbol (no summation implied).

This self-duality of the bivector basis makes the signature of a general bivector *mixed*. That is, although each basis bivector has a well-defined signature, $\gamma_{i0}\widetilde{\gamma}_{i0} = (1)(-1) = -1$ or $\gamma_{jk}\widetilde{\gamma}_{jk} = (-1)(-1) = 1$, a general bivector $\mathbf{F}$ will be a mixture of these two different signatures. Indeed, consider the simple mixed bivector $\mathbf{F} = \gamma_{10} + \gamma_{23}$ as an example. Computing its reversed square produces

$$
\begin{aligned}
\widetilde{\mathbf{F}}\mathbf{F} &= [\gamma_{01} + \gamma_{32}][\gamma_{10} + \gamma_{23}], \\
&= \gamma_{01}\gamma_{10} + \gamma_{01}\gamma_{23} + \gamma_{32}\gamma_{10} + \gamma_{32}\gamma_{23}, \\
&= \gamma_1^2\gamma_0^2 + \gamma_{0123} + \gamma_{3210} + \gamma_2^2\gamma_3^2, \\
&= (-1)(1) + I + I + (-1)(-1), \\
&= 2I.
\end{aligned}
\tag{3.27}
$$

The cross-terms between the basis elements of different signature produce a *complex signature* $\epsilon_{\mathbf{F}} = I$ attached to its positive magnitude $|\mathbf{F}|^2 = 2$. Thus, mixing bases of differing signature is automatically handled by the algebra by making the total signature complex.

More generally, the reversed square of any bivector $\mathbf{F}$ will produce $\mathbf{F}\widetilde{\mathbf{F}} = \epsilon_F |\mathbf{F}|^2$, where $\epsilon_F = \alpha + \beta I$ is a complex signature with unit magnitude $\alpha^2 + \beta^2 = 1$. As such, this signature can be written using a *phase angle* by using the same algebraic techniques as standard complex analysis: $\epsilon_F = \exp(2\varphi I) = \cos 2\varphi + I \sin 2\varphi$. It follows that a general bivector $\mathbf{F}$ must have a canonical (i.e., frame-independent) form with an *intrinsic phase*

$$
\mathbf{F} = \mathbf{f} \exp(\varphi I) = \mathbf{f}(\cos\varphi + I \sin\varphi)
\tag{3.28}
$$

that mixes the bivector subspaces of different signature. Here we define $\mathbf{f}$ to be the *canonical bivector* for $\mathbf{F}$ that we choose to have a definite negative signature $\mathbf{f}\widetilde{\mathbf{f}} = -|\mathbf{F}|^2$ (for reasons that will become clear in Section 3.6). Its dual $\mathbf{f}I$ then has a definite positive signature, while the square produces the positive magnitude $\mathbf{f}^2 = |\mathbf{F}|^2$. It is worth emphasizing that the existence of this canonical complex form of a bivector is not at all apparent when working solely with its real components $F^{\mu\nu}$.

The phase and canonical bivector can be directly extracted from $\mathbf{F}$ according to

$$
\varphi = \frac{1}{2}\tan^{-1}\frac{\ell_2}{\ell_1}, \qquad \mathbf{f} = \mathbf{F}\exp(-\varphi I),
\tag{3.29}
$$

where $\ell_1 = \langle \mathbf{F}^2 \rangle_0 = |\mathbf{F}|^2 \cos 2\varphi$ and $\ell_2 = \langle \mathbf{F}^2 \rangle_4 I^{-1} = |\mathbf{F}|^2 \sin 2\varphi$ are the scalar parts of its square $\mathbf{F}^2 = \ell_1 + \ell_2 I$. Note that for null bivectors ($\mathbf{F}^2 = 0$) the decomposition (3.28) is not unique and the phase φ becomes degenerate (just as with standard complex analysis). This potentially degenerate intrinsic phase of a bivector will become very important in our discussion of electromagnetic waves in Section 5.4.

3.5.2. Bivectors: products with bivectors

A product between two bivectors $\mathbf{F}$ and $\mathbf{G}$ can be split into *three* distinct grades (scalar, bivector, and pseudoscalar), in contrast to the two grades in the vector–vector and vector–bivector products of (3.2) and (3.9),

$$
\mathbf{F}\mathbf{G} = \langle \mathbf{F} \cdot \mathbf{G} \rangle_0 + [\mathbf{F}, \mathbf{G}] + \langle \mathbf{F} \cdot \mathbf{G} \rangle_4.
\tag{3.30}
$$

We will find it instructive to consider the scalar and pseudoscalar parts together as the same complex scalar $\mathbf{F} \cdot \mathbf{G}$.

The symmetric part of the product produces this complex scalar, which we can better understand by making the canonical decompositions $\mathbf{F} = \mathbf{f}\exp(\vartheta I)$, $\mathbf{G} = \mathbf{g}\exp(\varphi I)$ according to (3.28),

$$
\mathbf{F} \cdot \mathbf{G} \equiv \frac{1}{2}(\mathbf{F}\mathbf{G} + \mathbf{G}\mathbf{F}) = (\mathbf{f} \cdot \mathbf{g}) \exp[(\vartheta + \varphi)I].
\tag{3.31}
$$

Since **f** and **g** have fixed signature, $\mathbf{f} \cdot \mathbf{g} = (\mathbf{fg} + \mathbf{gf})/2$ is a real scalar. In fact, it will become clear after we introduce the relative 3-vectors in Section 3.6.1 that this real scalar is completely equivalent to the usual (Euclidean) dot product from non-relativistic 3-vector analysis, which motivates this choice of notation.

It follows that the total dot product between arbitrary bivectors will have scalar and pseudoscalar parts of the form

$$\langle \mathbf{F} \cdot \mathbf{G} \rangle_0 = (\mathbf{f} \cdot \mathbf{g}) \cos(\vartheta + \varphi), \tag{3.32}$$

$$\langle \mathbf{F} \cdot \mathbf{G} \rangle_4 = (\mathbf{f} \cdot \mathbf{g}) \sin(\vartheta + \varphi) I. \tag{3.33}$$

Note that the intrinsic phases of **F** and **G** play a critical role in these products, and produce wave-like *interference* factors. It is worth emphasizing that these factors arise solely from the structure of spacetime itself, and not from any additional assumptions.

The antisymmetric part of **FG** is written as a *commutator bracket* that produces another bivector

$$[\mathbf{F}, \mathbf{G}] \equiv \langle \mathbf{FG} \rangle_2 = \frac{1}{2}(\mathbf{FG} - \mathbf{GF}). \tag{3.34}$$

We will explore the deep significance of this commutator bracket in Section 3.7, where it will play the role of a Lie bracket for the Lorentz group (motivating this choice of notation). For completeness here, we also define the *bivector cross product* as the (Hodge) dual of the commutator bracket

$$\mathbf{F} \times \mathbf{G} \equiv [\mathbf{F}, \mathbf{G}] I^{-1}. \tag{3.35}$$

After we introduce the relative 3-vectors in Section 3.6.1 it will become clear that this cross product is completely equivalent to the traditional 3-vector cross product (again, motivating this choice of notation).

3.5.3. Complex conjugation

At this point we define another useful operation: complex conjugation. It follows from (3.28) that the proper inverse (3.18) of a bivector **F** has the canonical form

$$\mathbf{F}^{-1} = \frac{\mathbf{f}}{|\mathbf{F}|^2} \exp(-\varphi I) \tag{3.36}$$

which must additionally invert the intrinsic phase of the complex scalar that appears in its signature.

To more compactly write this inverse, we introduce a complex conjugation of the spacetime algebra as a flip in sign of the pseudoscalar. For a multivector M, as in (3.23), this produces

$$M^* = M|_{I \mapsto -I} = (\alpha - \beta I) + (v - wI) + \mathbf{f} \exp(-\varphi I). \tag{3.37}$$

Geometrically, the flip in pseudoscalar sign corresponds to changing the orientation (handedness) of spacetime. Since $(M^*)^* = M$, conjugation is also an involution on the algebra.

For complex vectors, the conjugation is equal to reversion $\widetilde{z} = (v + wI)^{\sim} = v + Iw = v - wI = z^*$, since I anticommutes with vectors, which is a useful property. However, this anticommutation also forces us to be careful about applying conjugation to products, since it does *not* generally distribute across products that involve vectors, e.g., $(zM)^* \neq M^* z^*$, $(zM)^* \neq z^* M^*$. Nevertheless, conjugation does distribute over products involving only complex scalars and bivectors, just as in standard complex analysis. To aid memory about some of these subtler properties, we have collected a few commonly encountered identities in Table 5.

With this operation we can write the proper complex generalization of the blade inverse (3.18)

$$\zeta^{-1} = \frac{\zeta^*}{\zeta^* \zeta}, \qquad z^{-1} = (z^* z)^{-1} z^*, \qquad \mathbf{F}^{-1} = \frac{\mathbf{F}^*}{\mathbf{F}^* \mathbf{F}}, \tag{3.38}$$

since the quantities $\zeta^* \zeta = \alpha^2 + \beta^2$, $z^* z = \widetilde{z} z = v^2 + w^2 + 2(v \cdot w) I$, and $\mathbf{F}^* \mathbf{F} = \mathbf{f}^2 = |\mathbf{F}|^2$ are all invertible scalars.[9]

3.6. Relative frames and paravectors

To connect the spacetime algebra $\mathcal{C}_{1,3}$ with standard 3-vector analysis, we express the duality of the spacetime algebra in another way that selects a particular *inertial reference frame*. To do this, we decompose the bivector basis into the three elements that involve a specific timelike vector γ_0,

$$\vec{\sigma}_i \equiv \gamma_{i0} = \gamma_i \wedge \gamma_0, \tag{3.39}$$

---

[9] Note that $z^* z$ produces a complex scalar, so $(z^* z)^{-1}$ recursively follows the complex scalar rule in (3.38). Moreover, the inverse z^{-1} written here is both a left and a right inverse, even though $(z^* z)^{-1}$ and z^* do not commute in general. That is, the explicit right inverse is $z^{-1} = z^* (zz^*)^{-1} = [(zz^*)^{-1}]^* z^* = (z^* z)^{-1} z^*$, which is the same expression.

Table 5
Commonly encountered algebraic identities. All of these may derived in straightforward ways from the fundamental Clifford product, but we include them here for reference.

Algebraic identities	
Dot-wedge conversion:	$v \cdot (wI) = -(vI) \cdot w = (v \wedge w)I$
	$v \wedge (wI) = -(vI) \wedge w = (v \cdot w)I$
	$(vI) \cdot \mathbf{F} = v \wedge (\mathbf{F}I) = (v \cdot \mathbf{F})I$
	$(vI) \wedge \mathbf{F} = v \cdot (\mathbf{F}I) = (v \wedge \mathbf{F})I$
	$v \cdot \mathbf{F} \cdot w = w \cdot (v \cdot \mathbf{F}) = (w \wedge v) \cdot \mathbf{F}$
Trace cyclicity:	$\langle ABC \rangle_0 = \langle CAB \rangle_0$
	$\langle ABC \rangle_4 = \langle ABCI^{-1} \rangle_0 I$
Bivector–pseudoscalar commutation:	$\zeta \mathbf{F} = \mathbf{F}\zeta$
	$(\zeta \mathbf{F})^* = \zeta^* \mathbf{F}^* = \mathbf{F}^* \zeta^*$
Vector–pseudoscalar anti-commutation:	$(v + wI) = (v - Iw)$
	$z^* = \tilde{z}$
	$\zeta z = z\zeta^*$
	$(zM)^* \neq z^* M^*$
Relative frame splits:	$I = \gamma_{0123} = \vec{\sigma}_1 \vec{\sigma}_2 \vec{\sigma}_3$
	$v = (v_0 + \vec{v})\gamma_0 = \gamma_0(v_0 - \vec{v})$
	$\nabla = \gamma_0(\partial_0 + \vec{\nabla}) = (\partial_0 - \vec{\nabla})\gamma_0$
	$\mathbf{F} = \vec{E} + \vec{B}I$
3-vector products:	$\vec{E}\vec{B} = \vec{E} \cdot \vec{B} + \vec{E} \times \vec{B}I$
	$\langle \vec{A}\vec{B}\vec{C} \rangle_4 = \vec{A} \cdot (\vec{B} \times \vec{C})I$
	$\langle \vec{A}\vec{B}\vec{C} \rangle_2 = (\vec{A} \cdot \vec{B})\vec{C} - (\vec{A} \times \vec{B}) \times \vec{C}$
	$\quad = (\vec{A} \cdot \vec{B})\vec{C} + \vec{A}(\vec{B} \cdot \vec{C}) - (\vec{A} \cdot \vec{C})\vec{B}$
Relative reversion:	$\zeta^\dagger = \zeta^*$
	$\mathbf{F}^\dagger = \vec{E} - \vec{B}I$
	$\gamma_0 \widetilde{\mathbf{F}} = \mathbf{F}^\dagger \gamma_0$
	$(\vec{\sigma}_j \vec{\sigma}_k)^\dagger = \vec{\sigma}_k \vec{\sigma}_j$

with $i = 1, 2, 3$, and their duals $\vec{\sigma}_i I = -\epsilon_{ijk}\gamma_{jk}$. This *spacetime split* of the bivector basis into an inertial frame is illustrated in Fig. 6(a). The elements $\vec{\sigma}_i$ directly correspond to the three spatial directions γ_i orthogonal to γ_0. Importantly, they will act as a basis for the 3-vector quantities that are used in standard vector analysis treatments of electromagnetism. As such, we notate these basis elements using standard vector arrows and refer to them as *relative 3-vectors*. Geometrically, one can conceptualize the relative 3-vectors that are experienced by an inertial observer as plane segments in spacetime that are obtained by dragging the spatial frame γ_i along the timelike direction γ_0 of the observer's worldline. As we will make clear in Section 3.8, the 3-vectors $\vec{\sigma}_i$ also have a deep connection to the standard Pauli spin matrices, which motivates our choice of notation.

3.6.1. Bivectors: spacetime split and cross product

In terms of this relative 3-vector basis, a bivector takes the simple form

$$\mathbf{F} = E_1\gamma_{10} + E_2\gamma_{20} + E_3\gamma_{30} + B_1\gamma_{32} + B_2\gamma_{13} + B_3\gamma_{21} = \vec{E} + \vec{B}I \tag{3.40}$$

involving a relative 3-vector and the dual of a relative 3-vector

$$\vec{E} = E_1\vec{\sigma}_1 + E_2\vec{\sigma}_2 + E_3\vec{\sigma}_3, \qquad \vec{B}I = \left[B_1\vec{\sigma}_1 + B_2\vec{\sigma}_2 + B_3\vec{\sigma}_3 \right] I. \tag{3.41}$$

Unlike the canonical decomposition (3.28), splitting the bivector into dual relative pieces in this manner requires the specification of a particular timelike vector γ_0 in the basis (3.39), which spoils the frame-independence of each relative piece. Only the total bivector $\mathbf{F}$ produced in the sum (3.40) is a proper geometric object. We choose the suggestive labels $\vec{E}$ and $\vec{B}$ here to anticipate their correspondence to the electromagnetic field vectors in Section 5, but we can already see at this stage the inevitable equivalence in form of $\mathbf{F}$ to the Riemann–Silberstein complex vector (1.1). As with the complex canonical form of a bivector, the existence of this complex spacetime split is not at all apparent when working solely with its real components $F^{\mu\nu}$.

Briefly, we note that the product between a vector and a bivector that we defined in Section 3.3.1 can be used to compute the spacetime split (3.40) according to (noting $\gamma_0^{-1} = \gamma_0$)

$$\mathbf{F} = \mathbf{F}\gamma_0\gamma_0^{-1} = (\mathbf{F} \cdot \gamma_0)\gamma_0 + (\mathbf{F} \wedge \gamma_0)\gamma_0. \tag{3.42}$$

The first term contracts (i.e., lowers the grade of) $\mathbf{F}$ to a vector perpendicular to γ_0 and then multiplies by γ_0 to produce the relative 3-vector basis $\vec{\sigma}_i = \gamma_i\gamma_0$. The second term inflates (i.e., raises the grade of) $\mathbf{F}$ to a trivector including γ_0 and then contracts out the γ_0 part, leaving behind the dual basis $\vec{\sigma}_i I = -\epsilon_{ijk}\gamma_j\gamma_k$.
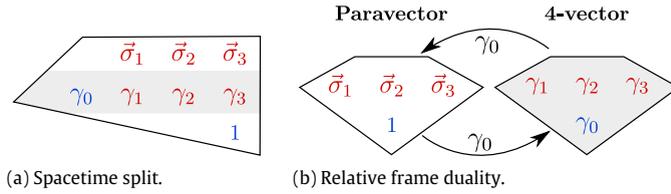
Fig. 6. Spacetime split and paravectors. (a) Performing a spacetime split of Fig. 5 selects a timelike vector γ_0 that corresponds to a particular inertial reference frame. This procedure partitions the proper bivectors into *relative* parts spanned by a 3-vector basis $\vec{\sigma}_i = \gamma_{i0}$ and its dual $\vec{\sigma}_i I = -\epsilon_{ijk}\gamma_{jk}$. (b) Each half of the split algebra then satisfies a secondary duality relation under right multiplication by γ_0. This duality maps proper 4-vectors to relative paravectors, and vice versa. As a result, the entire spacetime algebra may be expanded in terms of relative paravectors and duality factors, which provides an explicit embedding of traditional nonrelativistic 3-vector analysis inside spacetime algebra.

Similarly, we can now revisit the cross product between two bivectors, defined in Section 3.5.2 as the dual of the commutator bracket: $\mathbf{F} \times \mathbf{G} = [\mathbf{F}, \mathbf{G}]I^{-1}$. If we compute this cross product for the relative basis elements $\vec{\sigma}_i$ we find

$$\vec{\sigma}_1 \times \vec{\sigma}_2 = \vec{\sigma}_3, \qquad \vec{\sigma}_2 \times \vec{\sigma}_3 = \vec{\sigma}_1, \qquad \vec{\sigma}_3 \times \vec{\sigma}_1 = \vec{\sigma}_2. \tag{3.43}$$

These are precisely the defining relations for the familiar relative 3-vector cross product, which motivates its standard notation. Thus, the bivector cross product forms a closed nonassociative (Gibbs) subalgebra for relative 3-vectors that is entirely embedded within the spacetime algebra in a natural way. Hence, all the standard manipulations with the vector cross product still apply to the relative 3-vectors in the spacetime algebra. Moreover, the conceptual meaning of the cross product is enriched, since it is induced by the more general associative product of spacetime algebra.

3.6.2. Paravectors

Using the spacetime split (3.40) for the bivectors, we can rewrite a general multivector (3.23) in the following way

$$M = \left[(\alpha + \vec{E}) + v\right] + \left[(\beta + \vec{B}) + w\right]I, \tag{3.44}$$

where α, β are scalars, v, w are proper 4-vectors, and $\vec{E}$, $\vec{B}$ are relative 3-vectors.

We have further partitioned each half of the complex expansion (3.44) into even $(\alpha + \vec{E})$ and odd (v) grades. These partitions have four components each and are also dual to each other upon right multiplication by the chosen timelike unit vector γ_0. We illustrate this duality in Fig. 6(b). As an important example, the coordinate 4-vector x becomes

$$\begin{aligned} x\gamma_0 &= [ct\gamma_0 + x_1\gamma_1 + x_2\gamma_2 + x_3\gamma_3]\gamma_0, \\ &= ct + x_1\gamma_{10} + x_2\gamma_{20} + x_3\gamma_{30}, \\ &= ct + \vec{x}, \end{aligned} \tag{3.45}$$

which is the sum of a relative scalar time and a 3-vector of the relative spatial coordinates. This combination of a relative scalar and a relative 3-vector is commonly called a *paravector* [182,184] and is often written as a tuple $(ct, \vec{x})$ instead of a sum to indicate that the scalar and 3-vector parts of a 4-vector should be considered part of the same unified whole. Interestingly, the paravector that appears here as a sum of a scalar and a 3-vector is closely related to the original use of quaternions by Maxwell [134] to describe the essential unification of temporal and spatial parts of a 4-vector. We will comment more on the close relation between the 3-vectors that appear here and quaternions in Section 3.7.

We see that right-multiplication of a 4-vector by γ_0 isolates the *relative* quantities that correspond to the inertial frame of γ_0. Left multiplication flips the sign of the 3-vector part due to the wedge product in the 3-vector basis: $\gamma_0 x = ct - \vec{x}$. Thus, we can consistently recover the invariant coordinate interval in a direct way by using the paravector duality

$$x^2 = x\gamma_0\gamma_0 x = [ct + \vec{x}][ct - \vec{x}] = (ct)^2 - |\vec{x}|^2. \tag{3.46}$$

Observe that this style of derivation has completely bypassed the need to expand the expression in explicit components to obtain the final result. This is a general and convenient property of working with spacetime algebra: transitioning between proper and relative expressions is a component-free and straightforward exercise.

Using the vector–paravector duality, we can further expand the multivector (3.44) into a fully frame-dependent form relative to γ_0

$$M = \left[(\alpha + \vec{E}) + (\delta + \vec{p})\gamma_0\right] + \left[(\beta + \vec{B}) + (\omega + \vec{a})\gamma_0\right]I, \tag{3.47}$$

that involves only paravectors (with scalar parts α, δ, β, ω and vector parts $\vec{E}$, $\vec{p}$, $\vec{B}$, $\vec{a}$ discussed below) and the appropriate duality factors γ_0 and I. These paravectors are the quantities usually considered in frame-dependent treatments of relativistic systems that use traditional vector analysis. However, the geometric origins of the quantities involved become obscured when they are lifted outside the spacetime algebra.

Table 6
Transformation properties of the relative parts of a multivector M in a particular frame. Each checkmark indicates that the associated quantity changes sign under a spatial parity inversion, or is a proper scalar quantity that does not depend upon a specific Lorentz frame, or has components that change covariantly with the frame like parts of a proper 4-vector, or has components that change like parts of a proper bivector. These transformation properties are traditionally used to distinguish the distinct relative parts of a spacetime multivector.

Multivector part		Parity	Scalar	4-vector	Bivector
α	Proper scalar	×	✓	×	×
βI	Proper pseudoscalar	✓	✓	×	×
δ	Relative scalar	×	×	✓	×
ωI	Relative pseudoscalar	✓	×	✓	×
$\vec{p}$	Polar 3-vector	✓	×	✓	×
$\vec{E}$	Polar 3-vector	✓	×	×	✓
$\vec{a}$	Axial 3-vector	×	×	✓	×
$\vec{B}$	Axial 3-vector	×	×	×	✓

$$M = \left[(\alpha + \vec{E}) + (\delta + \vec{p})\gamma_0\right] + \left[(\beta + \vec{B}) + (\omega + \vec{a})\gamma_0\right] I$$

We can recover some distinguishing information about the scalars by examining their behavior under frame transformations and spatial reflections, which we summarize in Table 6. The scalar α is invariant under both frame transformations and spatial reflections, so is a proper scalar. The scalar β flips sign under spatial reflections but does not change under a frame transformation, so is a proper pseudoscalar. The scalar δ is invariant under spatial reflections but changes under a frame transformation, so is a relative scalar (i.e., temporal) part of a 4-vector. The scalar ω flips sign under a spatial reflection and changes under a frame transformation, so is a relative pseudoscalar part of a 4-pseudovector.

Similarly, we can distinguish the 3-vectors by looking at spatial reflections and frame transformation properties. The vectors $\vec{E}$, $\vec{p}$ flip sign under a spatial reflection, so are *polar* relative 3-vectors. The vectors $\vec{B}$, $\vec{a}$ do not flip sign under a spatial reflection, so are *axial* relative 3-vectors. However, while the combination of components $(\delta, \vec{p})$ transforms as a proper 4-vector and $(\omega, \vec{a})$ transforms as a proper 4-pseudovector, the combination of components in $(\vec{E}, \vec{B})$ transforms as what is usually considered as an antisymmetric rank-2 tensor (i.e., a bivector), as we indicated in Section 3.1 and further elucidated in Section 3.4.

For convenience, we also summarize the various expansions of a general multivector in Table 7. Comparing the proper expression of (3.7) to the frame-dependent expression of (3.47) makes the geometric origins of all these transformation properties clear. The proper combinations of relative quantities correspond to distinct geometric grades of a spacetime multivector. This geometric structure is left implicit in standard vector or tensor analysis methods, even though the transformation properties remain. The spacetime algebra makes this structure explicit, intuitive, and easily manipulated in calculations.

3.6.3. Relative reversion

To isolate the distinct quantities that are relevant for a particular frame, it is useful to define a *relative reversion* involution

$$\begin{aligned} M^\dagger &= \gamma_0 \widetilde{M} \gamma_0 \\ &= \left[(\alpha + \vec{E}) + (\delta - \vec{p})\gamma_0\right] + \left[-(\beta + \vec{B}) + (\omega - \vec{a})\gamma_0\right] I. \end{aligned} \tag{3.48}$$

Comparing with (3.47), note that this γ_0-dependent involution has the effect of the proper complex conjugation for the complex scalars

$$\zeta^\dagger = (\alpha + \beta I)^\dagger = \alpha - \beta I = \zeta^*, \tag{3.49}$$

but has the effect of conjugation *with respect to the spacetime split* for the complex bivectors

$$\mathbf{F}^\dagger = (\vec{E} + \vec{B}I)^\dagger = \vec{E} - \vec{B}I \neq \mathbf{F}^*. \tag{3.50}$$

Importantly, this type of conjugation is not frame-independent, but depends on the relative frame γ_0 due to the spacetime split.

In contrast, the relative reversion of a 4-vector only flips the sign of the relative 3-vector part

$$\begin{aligned} v^\dagger &= [(\delta + \vec{p})\gamma_0]^\dagger = (\delta - \vec{p})\gamma_0, \\ (wI)^\dagger &= [(\omega + \vec{a})\gamma_0 I]^\dagger = (\omega - \vec{a})\gamma_0 I, \end{aligned} \tag{3.51}$$

which is intrinsically different from complex conjugation.

Table 7
Multivector expansions into distinct geometric grades, intrinsically complex parts, and relative parts that depend upon a particular choice of Lorentz frame.

	Multivectors
Graded form	$M = \alpha + v + \mathbf{F} + wI + \beta I$
	α: scalar v: vector $\mathbf{F}$: bivector wI: pseudovector βI: pseudoscalar
Complex form	$M = \zeta + z + \mathbf{F}$
	$\zeta = \alpha + \beta I$: complex scalar $z = v + wI$: complex vector $\mathbf{F} = \mathbf{f}\exp(\varphi I)$: (irreducibly) complex bivector
Relative form	$M = \left[(\alpha + \vec{E}) + (\delta + \vec{p})\gamma_0 \right] + \left[(\beta + \vec{B}) + (\omega + \vec{a})\gamma_0 \right] I$
	α: proper scalar $\vec{E}$: relative polar 3-vector part of $\mathbf{F}$ δ: relative scalar part of v $\vec{p}$: relative polar 3-vector part of v βI: proper pseudoscalar $\vec{B}$: relative axial 3-vector part of $\mathbf{F}$ ωI: relative pseudoscalar part of wI $\vec{a}$: relative axial 3-vector part of wI

Using the appropriate relative reversions, we can then extract the relative parts of each geometric object in a systematic way

$$\alpha = \frac{\zeta + \zeta^\dagger}{2}, \qquad\qquad \beta = \frac{\zeta - \zeta^\dagger}{2I},$$
$$\vec{E} = \frac{\mathbf{F} + \mathbf{F}^\dagger}{2}, \qquad\qquad \vec{B} = \frac{\mathbf{F} - \mathbf{F}^\dagger}{2I},$$
$$\delta = \frac{v + v^\dagger}{2}\gamma_0^{-1}, \qquad\qquad \vec{p} = \frac{v - v^\dagger}{2}\gamma_0^{-1}, \qquad\qquad (3.52)$$
$$\omega = \frac{wI + (wI)^\dagger}{2}I^{-1}\gamma_0^{-1}, \qquad \vec{a} = \frac{wI - (wI)^\dagger}{2}I^{-1}\gamma_0^{-1}.$$

Note that we have to be careful about the order of the inversions in the case when elements do not commute.

3.7. Bivectors: commutator bracket and the Lorentz group

We now revisit the commutator bracket that we briefly defined for bivectors in Section 3.5.2. Surprisingly, this commutator bracket makes the bivectors a closed nonassociative *Lie algebra* [5]. Indeed, this feature has motivated the notation of the bracket. To see which Lie algebra the bivectors belong to, we compute the 15 following bracket relations between all six bivector basis elements:

$$\begin{aligned}
&[\gamma_{12}, \gamma_{23}] = \gamma_{31}, & &[\gamma_{23}, \gamma_{31}] = \gamma_{12}, & &[\gamma_{31}, \gamma_{12}] = \gamma_{23}, \\
&[\gamma_{10}, \gamma_{20}] = -\gamma_{12}, & &[\gamma_{20}, \gamma_{30}] = -\gamma_{23}, & &[\gamma_{30}, \gamma_{10}] = -\gamma_{31}, \\
&[\gamma_{31}, \gamma_{10}] = \gamma_{30}, & &[\gamma_{12}, \gamma_{20}] = \gamma_{10}, & &[\gamma_{23}, \gamma_{30}] = \gamma_{20}, \\
&[\gamma_{12}, \gamma_{10}] = \gamma_{20}, & &[\gamma_{23}, \gamma_{20}] = \gamma_{30}, & &[\gamma_{31}, \gamma_{30}] = \gamma_{10}, \\
&[\gamma_{23}, \gamma_{10}] = 0, & &[\gamma_{31}, \gamma_{20}] = 0, & &[\gamma_{12}, \gamma_{30}] = 0.
\end{aligned} \qquad (3.53)$$

We can simplify these bracket relations to a more familiar form by introducing the temporary (but standard) notations $\mathbf{S}_1 = \gamma_{23}$, $\mathbf{S}_2 = \gamma_{31}$, $\mathbf{S}_3 = \gamma_{12}$ and $\mathbf{K}_1 = \gamma_{10}$, $\mathbf{K}_2 = \gamma_{20}$, $\mathbf{K}_3 = \gamma_{30}$, and combining cyclic permutations

$$[\mathbf{S}_i, \mathbf{S}_j] = \epsilon_{ijk}\mathbf{S}_k, \qquad [\mathbf{K}_i, \mathbf{K}_j] = -\epsilon_{ijk}\mathbf{S}_k, \qquad [\mathbf{S}_i, \mathbf{K}_j] = \epsilon_{ijk}\mathbf{K}_k. \qquad (3.54)$$

These are precisely the Lie bracket relations that define the generators for the *Lorentz group*, where $\mathbf{S}_i$ generate spatial rotations, and $\mathbf{K}_i$ generate boosts. That is, the closed bivector subalgebra under the commutator bracket is precisely the Lie algebra of the Lorentz group. This connection between bivectors and the Lorentz group is extremely useful in practice for manipulating objects in spacetime.

As a consequence of this correspondence, exponentiating the commutator bracket of a spacetime bivector produces a restricted (proper, orthochronous) Lorentz transformation that is continuously connected to the identity. For example, to spatially rotate a bivector $\mathbf{F}$ by an angle θ in the plane $\mathbf{S}_3 = \vec{\sigma}_3 I^{-1}$ (i.e., around the relative axis $\vec{\sigma}_3$), then we would use the following exponentiation

$$
\begin{aligned}
\exp(\theta\,[\mathbf{S}_3, \cdot])\,\mathbf{F} &= \mathbf{F} + \theta[\mathbf{S}_3, \mathbf{F}] + \frac{\theta^2}{2}[\mathbf{S}_3, [\mathbf{S}_3, \mathbf{F}]] + \cdots, \\
&= \exp(\theta\,\mathbf{S}_3/2)\,\mathbf{F}\,\exp(-\theta\,\mathbf{S}_3/2), \\
&= \exp(-I\theta\,\vec{\sigma}_3/2)\,\mathbf{F}\,\exp(I\theta\,\vec{\sigma}_3/2).
\end{aligned}
\tag{3.55}
$$

The ability to resum an exponentiated commutator back into a double-sided product is an important identity that is commonly used in quantum-mechanics.[10] Indeed, the final form of the rotation in (3.55) is identical in form to the spatial rotation of a quantum mechanical spin, where the vector $\vec{\sigma}_3$ is usually interpreted as a Pauli matrix. We will return to this point in the next few sections.

As another important historical note, this double-sided product also appears as the proper form of a spatial rotation when using *quaternions*, where

$$
\mathbf{i} \equiv \mathbf{S}_1 = \vec{\sigma}_1 I^{-1}, \qquad \mathbf{j} \equiv -\mathbf{S}_2 = -\vec{\sigma}_2 I^{-1}, \qquad \mathbf{k} \equiv \mathbf{S}_3 = \vec{\sigma}_3 I^{-1}
\tag{3.56}
$$

are the usual "imaginary" quaternion elements[11] that all square to -1 and satisfy Hamilton's defining relation $\mathbf{ijk} = -1$ [135,136]. Note that the quaternion basis differs from the 3-vector basis by a factor of I^{-1} (i.e., the dual-basis), which is the fundamental reason that Maxwell's attempt to use them for describing relative spatial directions [134] was fraught with problems and eventually fell out of favor (despite leaving behind the legacy of using $\hat{\imath}, \hat{\jmath}$, and $\hat{k}$ to denote the spatial unit vectors). Quaternions are fundamentally related to spatial rotations, not spatial directions.

Another important and enlightening example of the Lorentz group is the boost from a frame with timelike vector γ_0 to a different frame moving at a relative velocity $\vec{v} = |v|\vec{\sigma}_3$ with respect to γ_0. Recall that a relative vector such as $\vec{v} = |v|\gamma_3 \wedge \gamma_0$ is really a bivector, so it can directly act as a generator for the appropriate Lorentz boost. Indeed, note that the relative unit vector $\vec{\sigma}_3 = \gamma_{30}$ is also equal to a boost generator $\vec{\sigma}_3 = \mathbf{K}_3$. The proper hyperbolic angle (i.e., *rapidity*) for the boost rotation is determined by the magnitude of the velocity $\alpha = \tanh^{-1}|v|/c$, while the proper plane of rotation is simply the relative unit vector itself $\vec{\sigma}_3$. Hence, the active Lorentz boost that rotates a bivector $\mathbf{F}$ from the $\vec{v}$ frame back into the reference γ_0 frame has the form

$$
\mathbf{F}' = \exp(-\alpha\,[\vec{\sigma}_3, \cdot])\,\mathbf{F} = \exp(-\alpha\,\vec{\sigma}_3/2)\,\mathbf{F}\,\exp(\alpha\,\vec{\sigma}_3/2).
\tag{3.57}
$$

Again, the conversion of the exponentiated commutator to a double-sided product makes the expression simpler. Note that the only difference in form between the boost (3.57) and the spatial rotation (3.55) is the factor of I in the exponent that changes the signature of the spacetime rotation.

To simplify (3.57), we split the bivector $\mathbf{F} = \mathbf{F}_\parallel + \mathbf{F}_\perp$ into pieces parallel $\mathbf{F}_\parallel$ and perpendicular $\mathbf{F}_\perp$ to $\vec{v}$, implying that $\vec{v}$ commutes with $\mathbf{F}_\parallel$ but anti-commutes with $\mathbf{F}_\perp$. The Lorentz transformation then becomes

$$
\begin{aligned}
\mathbf{F}' &= \exp(-\alpha\,\vec{\sigma}_3/2)\,\mathbf{F}\,\exp(\alpha\,\vec{\sigma}_3/2), \\
&= \mathbf{F}_\parallel + \exp(-\alpha\,\vec{\sigma}_3)\,\mathbf{F}_\perp, \\
&= \mathbf{F}_\parallel + \exp(-\tanh^{-1}|v|/c\,\vec{\sigma}_3)\,\mathbf{F}_\perp, \\
&= \mathbf{F}_\parallel + [\cosh(\tanh^{-1}|v|/c) - \sinh(\tanh^{-1}|v|/c)\,\vec{\sigma}_3]\,\mathbf{F}_\perp, \\
&= \mathbf{F}_\parallel + \frac{1 - \vec{v}/c}{\sqrt{1 - (v/c)^2}}\,\mathbf{F}_\perp,
\end{aligned}
\tag{3.58}
$$

where the boost dilation factor $\gamma = (1-(v/c)^2)^{-1/2}$ appears to scale only the perpendicular part of $\mathbf{F}$. If we make a spacetime split $\mathbf{F}_\perp = \vec{E}_\perp + \vec{B}_\perp I$ of the perpendicular bivector, then the transformed version is

$$
\begin{aligned}
\mathbf{F}'_\perp &= \gamma(1 - \vec{v}/c)\mathbf{F}_\perp = \gamma(1 - \vec{v}/c)(\vec{E}_\perp + \vec{B}_\perp I), \\
&= \gamma\left[\left(\vec{E}_\perp + \frac{\vec{v}}{c} \times \vec{B}_\perp\right) + \left(\vec{B}_\perp - \frac{\vec{v}}{c} \times \vec{E}_\perp\right)I\right],
\end{aligned}
\tag{3.59}
$$

---

[10] This equivalence and the resulting half-angle is a manifestation of the double-covering of the spatial rotation group SO(3) by the special unitary group SU(2), which is in turn isomorphic to the Spin(3) group generated by the rotation bivectors $\mathbf{S}_i = \vec{\sigma}_i I^{-1}$.

[11] Note the odd change in sign for $\mathbf{j}$: Hamilton accidentally defined a left-handed set, which has caused significant confusion about the proper application of quaternions in physics, and has obfuscated their equivalence to the rotation generators made apparent here.

where we have simplified the bivector product $\vec{v}\vec{E}_\perp = \vec{v} \times \vec{E}_\perp I$, since $\vec{v} \cdot \vec{E}_\perp = 0$ (and similarly for $\vec{B}$). Up to choices of units, this is precisely the transformation that we expect from a boost of the electromagnetic field [155,189].

3.7.1. Spinor representation

Recall that in (3.57) we converted the exponentiated commutator bracket into a double-sided product. Such a double-sided product is a more natural algebraic representation of a group transformation that can be extended from bivectors to the entire algebra. Specifically, if we notate the product factors as $U = \exp(-\alpha\,\vec{\sigma}_3/2)$, then the transformation in (3.57) has the form

$$\mathbf{F}' = U\,\mathbf{F}\,\widetilde{U}, \tag{3.60}$$

where we have used $\widetilde{\vec{\sigma}}_3 = -\vec{\sigma}_3$. The factors satisfy the normalization condition $U\widetilde{U} = \widetilde{U}U = 1$. Hence, this form of the transform makes it clear how the transformation preserves the product structure of the algebra,

$$\mathbf{GH} \mapsto U\mathbf{G}\widetilde{U}U\mathbf{H}\widetilde{U} = U(\mathbf{GH})\widetilde{U}, \tag{3.61}$$

precisely in the same manner that unitary representations of group transformations preserve product structure in quantum mechanics.

We call the quantity U that appears in (3.60) a *rotor* to emphasize that a Lorentz transformation has the geometric form of a *rotation* through spacetime. Specifically, the rotation generators $\mathbf{S}_i = \vec{\sigma}_i I^{-1}$ square to -1 and produce spatial rotations, while the boost generators $\mathbf{K}_i = \vec{\sigma}_i$ square to $+1$ and produce hyperbolic rotations in spacetime. Since the rotor representation of a Lorentz transformation produces a geometrically meaningful rotation, the bivector $\mathbf{F}$ can be replaced in (3.60) with any geometric object that can be rotated. Indeed, *any multivector M* can be Lorentz transformed according to $M' = UM\widetilde{U}$.

A rotor U is an example of a *spinor* ψ, which is an element of the even-graded subalgebra of the spacetime algebra [158] that may be decomposed as

$$\psi = \zeta + \mathbf{F}, \tag{3.62}$$

where $\zeta = \alpha + \beta I$ is a complex scalar and $\mathbf{F}$ is a bivector. This even-graded subalgebra is closed under the full Clifford product. As an example of this decomposition, the boost rotor $\exp(-\alpha\,\vec{\sigma}_3/2)$ can be expanded in terms of hyperbolic trigonometric functions as $\cosh(\alpha/2) - \vec{\sigma}_3\,\sinh(\alpha/2)$, which is a sum of a scalar and a bivector. Similarly, the spatial rotor $\exp(-\theta\,I\vec{\sigma}_3/2)$ can be expanded in terms of circular trigonometric functions as $\cos(\theta/2) - I\vec{\sigma}_3\,\sin(\theta/2)$.

Spinors also produce transformations of any multivector according to a double-sided product that preserves the product structure of the algebra. For example, if $\psi\widetilde{\psi} = \widetilde{\psi}\psi = 1$, then $M \mapsto \psi M\widetilde{\psi}$ is a *spinor representation* of a group transformation. Hence, the rotor form of (3.60) can also be understood as a spinor representation of the restricted Lorentz transformation. Similarly, if $\psi\psi^* = \psi^*\psi = 1$, then $M \mapsto \psi M\psi^*$ is another kind of group transformation that preserves the product structure of the algebra.

An example transformation of this latter type that we will find particularly useful in what follows is the *global phase rotation* characterized by the spinor $\psi = \exp(-\theta I/2)$. Interestingly, this phase rotation only affects complex vectors since I commutes with complex scalars and bivectors. That is, for a general multivector $M = \zeta + z + \mathbf{F}$ in complex form, we have

$$\psi M\psi^* = \zeta + z\,\exp(\theta I) + \mathbf{F}, \tag{3.63}$$

which is a proper group transformation since $\psi\psi^* = \psi^*\psi = 1$. This phase transformation is responsible for the field-exchange *dual symmetry* of the electromagnetic field, which we will detail in Section 5.2.

3.8. Pauli and Dirac matrices

A vast notational simplification of the Lorentz bracket relations (3.54) occurs when we use the duality relations to rewrite the bivectors $\mathbf{K}_i = \vec{\sigma}_i$ and their duals $\mathbf{S}_i = \vec{\sigma}_i I^{-1}$ in terms of a particular Lorentz frame of relative 3-vectors. The pseudoscalar I commutes with all bivectors and squares to -1, so all 15 bracket relations reduce to simple variations of the three fundamental relations

$$[\vec{\sigma}_i, \vec{\sigma}_j] = \epsilon_{ijk}\vec{\sigma}_k I, \tag{3.64}$$

which are simply another way of expressing the 3-vector cross products (3.43) in the relative frame. Observe that the relations (3.64) are also precisely the bracket relations that define the Pauli spin matrices used in quantum mechanics. Moreover, note that the relative 3-vector basis satisfies the relation

$$\vec{\sigma}_i\vec{\sigma}_j + \vec{\sigma}_j\vec{\sigma}_i = 2\delta_{ij}, \tag{3.65}$$

which indicates that the relative 3-vectors $\vec{\sigma}_i$ form a *Euclidean* Clifford subalgebra $\mathcal{C}_{3,0}$ in their own right.

In terms of this frame-dependent split, the proper pseudoscalar I can be written as $I = \vec{\sigma}_1\vec{\sigma}_2\vec{\sigma}_3$. That is, I is also the appropriate pseudoscalar (apparent unit 3-volume) for any relative frame, which is a nontrivial property. It also commutes with every element of this relative frame, making it completely equivalent in practice to a scalar imaginary i. Similarly, the

Table 8

The Clifford algebras in the nesting sequence of closed subalgebras $\mathcal{C}_{2,4} \supset \mathcal{C}_{4,1} \supset \mathcal{C}_{1,3} \supset \mathcal{C}_{3,0} \supset \mathcal{C}_{0,2} \supset \mathcal{C}_{0,1} \supset \mathcal{C}_{0,0}$, and their connection to common physics concepts, as well as the distinct types of spinors in quantum mechanics. Note how the $(+, -, -, -)$ signature of the spacetime algebra $\mathcal{C}_{1,3}$ appears naturally in the middle of this sequence.

Clifford algebras	Physics connections	Spinor types
$\mathcal{C}_{2,4}$	Conformal space	Penrose twistors
$\mathcal{C}_{4,1}$	Dirac electron	Dirac spinors
$\mathcal{C}_{1,3}$	**Spacetime/EM**	**Maxwell spinors**
$\mathcal{C}_{3,0}$	Relative 3-space	Pauli spinors
$\mathcal{C}_{0,2}$	Quaternion rotations	Hamilton spinors
$\mathcal{C}_{0,1}$	Complex (wave-like) phases	Schrödinger spinors
$\mathcal{C}_{0,0}$	Real numbers	Scalars

relative reversion operation $\psi^\dagger = \gamma_0 \widetilde{\psi} \gamma_0$ that we defined in (3.48) is the true reversion of the relative basis $\vec{\sigma}_i$ within this subalgebra: $(\vec{\sigma}_j \vec{\sigma}_k)^\dagger = \vec{\sigma}_k \vec{\sigma}_j$. As such, it also acts as the appropriate complex conjugation for the frame-dependent subalgebra according to (3.50). The fact that this relative complex conjugation is distinct from the global complex conjugation is not at all obvious when working solely within a relative frame. This subalgebra and relative conjugation is precisely what appears in the usual treatments of the Riemann–Silberstein vector (e.g., [40]).

Collecting these properties together, it is clear that the familiar Pauli matrices and complex scalar imaginary i can be viewed as a matrix representation of a particular spacetime split of the spinor subalgebra of spacetime. The matrix product simulates the noncommutative vector product. Indeed, this direct connection to the Pauli matrices has motivated our notation $\vec{\sigma}_i$ for the relative 3-vector basis [8]. Within the spacetime algebra, however, $\vec{\sigma}_i$ are not matrices—they are a bivector basis associated with the three spatial directions in a particular frame. Hence matrix constructions like $\sum_i a^i \vec{\sigma}_i$ that appear in quantum mechanics can be understood to define relative 3-vectors $\vec{a}$ in spacetime algebra (e.g., the reference spatial direction along which a spin can be aligned). There is nothing intrinsically quantum mechanical about these constructions.

Notably, the representation-free nature of spacetime algebra has exposed the deep structural relationship between the relative spatial directions, the cross product of relative 3-space, the Lie bracket relations of the Lorentz group, the quaternion algebra, the commutation relations of the Pauli spin matrices, and even the imaginary unit i that appears in these commutation relations. Using standard mathematical methods, these topics do not appear to have such an obvious connection. Indeed, the natural appearance of all these topics solely from a systematic construction of the geometry of spacetime highlights the deep geometric significance of their relation to one another.

As we briefly indicated at the beginning of this introduction, the embedding of Clifford algebras $\mathcal{C}_{3,0} \subset \mathcal{C}_{1,3}$ motivated our choice of metric signature for spacetime to be $(+, -, -, -)$. In fact, this embedding is part of a larger sequence of Clifford algebra embeddings $\mathcal{C}_{2,4} \supset \mathcal{C}_{4,1} \supset \mathcal{C}_{1,3} \supset \mathcal{C}_{3,0} \supset \mathcal{C}_{0,2} \supset \mathcal{C}_{0,1} \supset \mathcal{C}_{0,0}$ illustrated in Table 8, which corresponds to, in order: conformal space (Penrose twistors), relativistic electron (Dirac spinors), spacetime/electromagnetism (Maxwell spinors), relative 3-space (Pauli spinors), quaternions (Hamilton spinors), complex numbers (Schrödinger spinors), and the real numbers. The connection between this sequence and the successive approximations of quantum mechanical particles has been emphasized in [96,190–192]. The other choice of spacetime signature $(-, +, +, +)$ produces a distinct algebra $\mathcal{C}_{3,1}$ that does not belong to this sequence of algebras.

At this point we revisit the dot-product relation (3.3) for the 4-vectors γ_μ of the spacetime algebra $\mathcal{C}_{1,3}$

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\eta_{\mu\nu}. \tag{3.66}$$

This relation was historically used to define the Dirac γ-matrices that appear in the Dirac equation for the relativistic electron. We can thus view the Dirac matrices as a representation of the noncommutative spacetime algebra of 4-vectors γ_μ, where the matrix product simulates the noncommutative product. Indeed, our notation γ_μ is motivated by this connection [4]. As with the Pauli matrices, however, there is no need to construct such a matrix representation for the spacetime algebra to be well-defined. Matrix constructions like $\not{p} = \sum_\mu p^\mu \gamma_\mu$ that appear in the Dirac theory can be understood to define proper 4-vectors p in the spacetime algebra. Again, there is nothing intrinsically quantum mechanical about these constructions.

4. Spacetime calculus

Modern physical developments have required a mathematics that continually shifts its foundation and gets more abstract. Non-euclidean geometry and noncommutative algebra, which were at one time were considered to be purely fictions of the mind and pastimes of logical thinkers, have now been found to be very necessary for the description of general facts of the physical world. It seems likely that this process of increasing abstraction will continue in the future.

[Paul A.M. Dirac [193]]

With the spacetime algebra $\mathcal{C}_{1,3}$ now in hand, we consider how to construct *fields* on which we can perform calculus. This is the minimal extension of the algebra that is required to describe electromagnetism. Formally speaking, we replicate the

spacetime algebra $\mathcal{C}_{1,3}$ at every point x on a flat spacetime manifold so that it acts as a *tangent algebra* in which multivectors $M(x)$ can be locally defined. Collecting all the spacetime points x and their tangent algebras $\mathcal{C}_{1,3}(x)$ produces a *Clifford fiber bundle*. This bundle is the appropriate mathematical space to define *multivector fields* $M(x)$, such as the electromagnetic field, and to perform calculus on those fields.

The astonishing feature of the spacetime calculus that we will briefly construct over this bundle is that it permits the *directed integration* of multivector quantities by treating the points x themselves as vectors. This generalization recovers the usual scalar integration permitted by differential geometry as a special case, but also recovers the theorems of complex analysis and their generalizations. Moreover, it permits a coordinate-free definition of the vector derivative ∇ and a simple statement of the fully general fundamental (Stokes) theorem of calculus. We give a brief review here to give the essential background, but emphasize that more extensive and rigorous treatments may be found in [5,8,9,171,172,174].

4.1. Directed integration

Each vector point $x = \sum_\mu x^\mu \gamma_\mu$ on the oriented spacetime manifold corresponds to an oriented 4-volume $\mathrm{d}^4 x = \mathrm{d}x_0 \wedge \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3$ that can be decomposed into infinitesimal vector segments $\mathrm{d}x_\mu = |\mathrm{d}x|\gamma_\mu$ (i.e., differentials) that point along coordinate directions.[12]

One can then define a *directed integration* on the manifold as a Riemann summation of multivector products over a particular region S_k of dimension k. To do this, one encloses each point x of S_k in a finite oriented k-volume $\Delta^k x = |\Delta x|^k I_k$, where $I_k = \bigwedge_j \gamma_j$ is the unit k-volume element that encodes the orientation of the region S_k at the point x. One then computes a sum over all these points in the limit that the bounding intervals $|\Delta x|$ vanish and the number of points x increases to infinity in the standard way [171]

$$\int_{S_k} M_1(x)\,\mathrm{d}^k x\, M_2(x) = \lim_{|\Delta x| \to 0} \sum_{x \in S_k} M_1(x)\,\Delta^k x\, M_2(x). \tag{4.1}$$

The noncommutativity of the multivectors $M_1(x)$, $M_2(x)$ with the directed volume segment $\mathrm{d}^k x$ permits both right and left products in the integrand, which contrasts sharply with the usual definition of a scalar integral.

To understand the difference between scalar integration and directed integration, consider a k-volume S_k that has no boundary. The directed integral of this closed region must vanish $\int_{S_k} \mathrm{d}^k x = 0$. For example, if one sums small vectors around a curve, the net vector sum will be the vector connecting the start and end point of the curve; for a closed loop with no boundary points, this sum is simply zero. The volume of S_k, on the other hand, is given by the standard scalar integral of the undirected volume elements, which can be obtained from the directed integral by inverting the directed measure

$$\mathrm{Vol}(S_k) = \int_{S_k} |\mathrm{d}^k x| = \int_{S_k} \mathrm{d}^k x\, I_k^{-1} \tag{4.2}$$

using the unit k-volume element I_k for the measure $\mathrm{d}^k x = |\mathrm{d}^k x| I_k$. For a closed loop, this scalar integral will evaluate to the net length of the loop. As a side note, the implicit wedge products in $\mathrm{d}^k x$ that are used to compute the scalar measure $|\mathrm{d}^k x|$ in (4.2) are what produces the Jacobian determinant factor in the scalar integral when the coordinates are changed.

4.2. Vector derivative: the Dirac operator

If the spacetime points x are expanded into coordinates $x = \sum_\mu x^\mu \gamma_\mu$ using a standard basis γ_μ, then a proper *vector derivative* ∇ can be defined using the reciprocal basis γ^μ

$$\nabla = \sum_\mu \gamma^\mu \overleftrightarrow{\partial}_\mu, \tag{4.3}$$

where $\overleftrightarrow{\partial}_\mu = \overleftarrow{\partial}_\mu + \overrightarrow{\partial}_\mu$ is the usual partial derivative $\partial_\mu = \partial/\partial x^\mu$ along the coordinate $x^\mu = \gamma^\mu \cdot x$ that generally acts both to the left and to the right (one can think of this bidirectionality as a special case of the product rule for derivatives). The form of (4.3) is precisely the same as the *Dirac operator* that is usually constructed with gamma matrices and used in the Dirac equation. However, we see here that this operator has the natural interpretation as the proper vector derivative for a spacetime manifold *with no matrix representation required*.

---

[12] Formally speaking, the manifold can be constructed such that each oriented k-blade $\mathrm{d}^k x$ has an oriented boundary $\partial \mathrm{d}^k x = \sum_j (\mathrm{d}^{k-1}x_j^+ - \mathrm{d}^{k-1}x_j^-)$ of $(k-1)$-blades centered at pairs of boundary points $x_j^\pm$. For example, a vector segment $\mathrm{d}x$ has two ordered boundary points separated by the scalar interval $|\mathrm{d}x|$, a plane segment $\mathrm{d}^2 x$ has two ordered pairs of boundary line segments each separated by $|\mathrm{d}x|$, and a small cube of volume $\mathrm{d}^3 x$ has three pairs of boundary sides separated by $|\mathrm{d}x|$. These nested boundary relationships are formally constructed as an oriented chain complex via a limiting refinement procedure $\Delta^k x \to \mathrm{d}^k x$ that shrinks finite oriented volumes down to infinitesimal size [174]. The orientation information for each $\Delta^k x$ is encoded as a k-blade in the spacetime algebra at the corresponding point x.

4.2.1. Coordinate-free definition

For completeness, we emphasize that the directed integration permits a coordinate-free definition of the vector derivative in (4.3) as a generalized difference quotient [171]

$$M_1(x)\nabla M_2(x) = \lim_{\text{Vol}(S_4)\to 0} \frac{\int_{\partial S_4} M_1(x')[d^3x'I^{-1}(x)]M_2(x')}{\text{Vol}(S_4)}. \tag{4.4}$$

Here the numerator is the *directed* integral of the fields $M_1(x)$ and $M_2(x)$ over the *boundary* ∂S_4 of the spacetime region S_4, while the denominator is the scalar integral of S_4 that produces the volume $\text{Vol}(S_4)$ in (4.2). The local duality factor of the unit 4-volume $I^{-1}(x)$ in the numerator produces the normal vectors dual to the (trivector) boundary surfaces of ∂S_4.

The resulting derivative ∇ is thus a proper 4-vector algebraically that acts as a total derivative on both fields. When expanding it in local coordinates, one precisely obtains the form of the Dirac operator (4.3). In practice, one often needs to consider only the one-sided action of ∇ on a field, usually from the left, which amounts to setting $M_1(x) = 1$ in (4.4). Note that the algebraic and differential properties of ∇ can be treated independently according to (4.3). When it is not clear from context, it is conventional to use overdot notation to indicate which quantities are being differentiated in an expression, with the position of ∇ indicating its algebraic role as a vector (e.g., $\dot{\nabla}(a \cdot \dot{b})$ only differentiates b, but constructs a proper 4-vector) [9].

4.2.2. The relative gradient and d'Alembertian

The derivative ∇ can be restricted along any k-blade $M = \langle M \rangle_k$. Specifically, the *tangential derivative* along M is defined as

$$\nabla_M = M^{-1}(M \cdot \nabla), \tag{4.5}$$

which computes only the derivative along the surface M. When M is a unit vector a, then $a = a^{-1}$, so the tangential derivative recovers the usual directional derivative from vector calculus. Similarly, the complementary derivative is $\nabla_{MI^{-1}} = M^{-1}(M \wedge \nabla)$, which computes the derivative along the orthogonal surface MI^{-1} dual to M. The total derivative decomposes into a sum of these complementary derivatives $\nabla = \nabla_M + \nabla_{MI^{-1}}$ since $M^{-1}(M \cdot \nabla) + M^{-1}(M \wedge \nabla) = (M^{-1}M)\nabla = \nabla$.

Most usefully for our purposes here, we can decompose the derivative in terms of a frame relative to γ_0 by using tangential derivatives,

$$\nabla = \gamma_0^2\nabla = \gamma_0(\partial_0 + \vec{\nabla}) = (\partial_0 - \vec{\nabla})\gamma_0, \tag{4.6}$$

where $\nabla_{\gamma_0} = \gamma_0(\gamma_0 \cdot \nabla) = \gamma_0\partial_0$ is the tangential derivative along γ_0 and $\nabla_{\gamma_0 I^{-1}} = \gamma_0(\gamma_0 \wedge \nabla) = \gamma_0\vec{\nabla}$ is its dual. The frame-dependent 3-vector derivative

$$\vec{\nabla} = \gamma_0 \wedge \nabla = \sum_i \vec{\sigma}_i\partial_i \tag{4.7}$$

that appears in the decomposition (4.6) is precisely the relative 3-*gradient* operator used in standard vector analysis.

This decomposition of ∇ into a relative frame also makes it clear that ∇^2 is the d'Alembertian operator

$$\nabla^2 = (\partial_0 - \vec{\nabla})\gamma_0^2(\partial_0 + \vec{\nabla}) = \partial_0^2 - \vec{\nabla}^2 = \square. \tag{4.8}$$

In this sense, ∇ is the natural "square-root" of the d'Alembertian operator in spacetime, which was originally suggested by Dirac [194].

4.2.3. Example: field derivatives

The derivative of a scalar field $\alpha(x)$ produces the total gradient vector field

$$\nabla\alpha(x) = \gamma_0[\partial_0\alpha(x) + \vec{\nabla}\alpha(x)], \tag{4.9}$$

that includes the proper derivative along the timelike coordinate. Taking the dual of this equation via right multiplication by I immediately produces the derivative of a pseudoscalar field $\alpha(x)I$. It follows that the derivative of a complex scalar field $\zeta(x) = \alpha(x) + \beta(x)I$ is a complex vector field $z(x) = \nabla\zeta(x) = \nabla\alpha(x) + [\nabla\beta(x)]I$.

The derivative of a vector field $v(x) = [v_0(x) + \vec{v}(x)]\gamma_0$ produces a scalar field and a bivector field (i.e., a spinor field)

$$\nabla v(x) = \nabla \cdot v(x) + \nabla \wedge v(x),$$
$$= (\partial_0 - \vec{\nabla})\gamma_0^2(v_0 - \vec{v}),$$
$$= [\partial_0 v_0 + \vec{\nabla} \cdot \vec{v}] + [-(\partial_0\vec{v} + \vec{\nabla}v_0) + \vec{\nabla} \times \vec{v}I]. \tag{4.10}$$

The scalar part $\nabla \cdot v$ is the total *divergence* of $v(x)$, while the bivector part $\nabla \wedge v$ is the proper generalization of the *curl* of $v(x)$ to spacetime. Taking the dual of this equation produces the derivative of a trivector field $v(x)I$, which produces a pseudoscalar field and a bivector field (i.e., a spinor field). It follows that taking the derivative of a complex vector field

$z = a + gI$ generally produces a spinor field $\nabla z(x) = \psi(x) = \zeta(x) + \mathbf{F}(x)$ with a complex scalar part $\zeta = \nabla \cdot a + (\nabla \cdot g)I$ and a bivector part $\mathbf{F} = \nabla \wedge a + (\nabla \wedge g)I$.

The derivative of a bivector field $\mathbf{F}(x) = \vec{E}(x) + \vec{B}(x)I$ produces a vector field and a trivector field

$$
\begin{aligned}
\nabla \mathbf{F}(x) &= \nabla \cdot \mathbf{F}(x) + \nabla \wedge \mathbf{F}(x), \\
&= \gamma_0(\partial_0 + \vec{\nabla})(\vec{E} + \vec{B}I), \\
&= \gamma_0 \left[\vec{\nabla} \cdot \vec{E} + \partial_0 \vec{E} - \vec{\nabla} \times \vec{B} \right] + \gamma_0 \left[\vec{\nabla} \cdot \vec{B} + \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} \right] I,
\end{aligned}
\tag{4.11}
$$

which can be understood as a single complex vector field. Since derivatives of both complex scalars ζ and bivectors $\mathbf{F}$ produce complex vector fields z, it follows that spinors $\psi = \zeta + \mathbf{F}$ and complex vectors z are duals under the action of the derivative ∇.

Note that we can also write (4.11) in terms of the intrinsic phase of $\mathbf{F}(x)$ defined in Eq. (3.28)

$$
\begin{aligned}
\nabla \mathbf{F}(x) &= \nabla[\mathbf{f}(x) \exp([\varphi(x) + \varphi_0]I)], \\
&= [\nabla \mathbf{f}(x)] \exp([\varphi(x) + \varphi_0]I) + [\nabla \varphi(x)]\mathbf{F}(x)I
\end{aligned}
\tag{4.12}
$$

to show that the intrinsic bivector derivative $\nabla \mathbf{f}$ and the gradient of the *local* phase $\nabla \varphi$ contribute to the complex vector field. However any *global* intrinsic phase $\exp(\varphi_0 I)$ of $\mathbf{F}$ is a constant overall factor. This distinction between local and global phases holds for complex scalar and vector fields as well.

4.3. Fundamental theorem

The fundamental theorem of calculus readily follows from the integral definition of the derivative (4.4) and takes a particularly elegant form in terms of the tangential derivative along some surface S_k of dimension k [171]

$$
\int_{S_k} M_1(x) \mathrm{d}^k x \nabla_{S_k} M_2(x) = \int_{\partial S_k} M_1(x) \mathrm{d}^{k-1} x M_2(x).
\tag{4.13}
$$

This is the generalized *Stokes theorem* of calculus that states that the boundary integral is equal to the volume integral of the derivative. Computing this for various dimensional surfaces and k-blade fields reproduces all the standard variations of Stokes theorem, including the divergence theorem, the various Green's theorems, and the usual fundamental theorem of calculus in one dimension [172].

4.3.1. Example: Cauchy integral theorems

More surprisingly, if we consider the integration of appropriate Green's functions that invert the tangential derivative ∇_{S_k}, then we obtain and generalize the celebrated Cauchy integral theorems of complex analysis [9,172]. First, for any field $M(x)$ that satisfies $\nabla_{S_2} M(x) = 0$ everywhere in a closed planar region S_2, the following boundary integral vanishes according to (4.13)

$$
\oint_{\partial S_2} \mathrm{d}x\, M(x) = 0.
\tag{4.14}
$$

Second, Green's function that inverts the planar derivative is $\nabla_{S_2}[2\pi(x-x')]^{-1} = \delta(x-x')$. Hence, if $N(x) = \alpha_{-1}x^{-1} + M(x)$ is a modification of $M(x)$ that adds a pole, then we obtain the residue theorem

$$
\oint_{\partial S_2} \mathrm{d}x\, N(x) = \alpha_{-1} \int_{S_2} |\mathrm{d}^2 x| \mathbf{I}_2 \nabla_{S_2} x^{-1} = 2\pi \mathbf{I}_2 \alpha_{-1},
\tag{4.15}
$$

where $\mathbf{I}_2$ is the unit bivector for S_2 that squares to -1. Third, multiplying $M(x)$ with Green's function and then integrating produces the remaining Cauchy integral theorem

$$
\oint_{\partial S_2} \mathrm{d}x\, (x-x')^{-1} M(x) = \int_{S_2} \mathbf{I}_2 |\mathrm{d}^2 x| \nabla (x-x')^{-1} M(x) = 2\pi \mathbf{I}_2 M(x').
\tag{4.16}
$$

To rewrite these formulas in the standard form from complex analysis, we can expand the 2-vector $x = \alpha r + \beta y = \zeta r$, where r is a unit 2-vector that points along the "real axis", y is a unit 2-vector that points along the "imaginary axis", and $\zeta = \alpha + \beta \mathbf{I}_2$ is the "complex scalar" corresponding to x, with $\mathbf{I}_2 = yr$ being the unit planar area. Importantly, however, by considering different dimensional surfaces and using the fundamental theorem (4.13) with the correct Green's functions for derivatives along different k-volumes ∇_{S_k}, it is possible to generalize the Cauchy theorems above to the whole spacetime algebra [9,172].

5. Maxwell's equation in vacuum

One scientific epoch ended and another began with James Clerk Maxwell.

[Albert Einstein [195]]

The ultimate importance of the Maxwell theory is far greater than its immediate achievement in explaining and unifying the phenomena of electricity and magnetism. Its ultimate importance is to be the prototype for all the great triumphs of twentieth-century physics.

[Freeman J. Dyson [1]]

With the spacetime algebra developed in the previous two sections now in hand, we can explore the physical structure of electromagnetism as expressed with this proper geometric language. Wherever possible, we shall remind the reader of the previous mathematical results as they become relevant to the physics in what follows. As we shall see, all of electromagnetism essentially emerges without any additional effort directly from the geometric structure of spacetime algebra that we have already developed.

To proceed, we briefly examine the derivative of a bivector field that we calculated in Eq. (4.11) to make the following observation: the simplest differential equation that one can write for **F**

$$\nabla \mathbf{F} = 0 \tag{5.1}$$

is equivalent to all four of Maxwell's equations in vacuum, up to appropriate choices of constants that define units for the bivector components. This almost automatic appearance of Maxwell's vacuum equation is remarkable, since the deceptively simple differential equation (5.1) is trivially postulated solely from our careful development of the algebraic and differential *structure* of a spacetime manifold, with no reference to additional physical assumptions. We will return to this intriguing point later in Section 8.

5.1. Relative frame form

We can prove the equivalence of (5.1) to the usual form of Maxwell's equations in a relative frame with timelike vector γ_0 by writing **F** in terms of relative polar ($\vec{E}$) and axial ($\vec{B}$) parts obtained via the contractions in Eq. (3.42)

$$\mathbf{F} = \mathbf{F}\gamma_0\gamma_0^{-1} = (\mathbf{F} \cdot \gamma_0)\gamma_0 + (\mathbf{F} \wedge \gamma_0)\gamma_0 = \vec{E} + \vec{B}I. \tag{5.2}$$

Applying the identity (4.11) to (5.1) then yields

$$\nabla \mathbf{F}(x) = \gamma_0 \left[\vec{\nabla} \cdot \vec{E} + \partial_0 \vec{E} - \vec{\nabla} \times \vec{B} \right] + \gamma_0 \left[\vec{\nabla} \cdot \vec{B} + \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} \right] I = 0. \tag{5.3}$$

The independent vector and trivector parts of (5.1) can be separated, after which the duality factors γ_0 and I cancel, yielding the two paravector equations for the γ_0 frame[13]

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} + (\partial_0 \vec{E} - \vec{\nabla} \times \vec{B}) &= 0, \\ \vec{\nabla} \cdot \vec{B} + (\partial_0 \vec{B} + \vec{\nabla} \times \vec{E}) &= 0. \end{aligned} \tag{5.4}$$

Finally, the relative scalar and 3-vector parts of each paravector can be separated to produce the standard Gibbs 3-vector form for the four vacuum Maxwell equations [155,189]

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= 0, & \vec{\nabla} \times \vec{B} &= \partial_0 \vec{E}, \\ \vec{\nabla} \cdot \vec{B} &= 0, & \vec{\nabla} \times \vec{E} &= -\partial_0 \vec{B}. \end{aligned} \tag{5.5}$$

Rescaling these vectors with constant factors that set the (SI) units $\vec{E} \mapsto \vec{E}\sqrt{\epsilon_0}$, $\vec{B} \mapsto \vec{B}/\sqrt{\mu_0}$, and expanding $\partial_0 \mapsto c^{-1}\partial_t$ makes the equivalence apparent. Furthermore, with these replacements and the identity $c = 1/\sqrt{\epsilon_0\mu_0}$, it is clear that the proper bivector field

$$\mathbf{F} = \sqrt{\epsilon_0}\vec{E} + \vec{B}I/\sqrt{\mu_0} = \sqrt{\epsilon_0}(\vec{E} + c\vec{B}I) = (\vec{E}/c + \vec{B}I)/\sqrt{\mu_0} \tag{5.6}$$

is precisely equivalent to the Riemann–Silberstein complex electromagnetic field vector (1.1). Importantly, however, the pseudoscalar factor of I that appears here has rich physical meaning that is not apparent when using the scalar imaginary i appearing in other treatments (e.g., [40]). For brevity, we will omit the constants in the discussion to follow, since they may be easily restored when needed. The various formulations of the electromagnetic field and Maxwell's vacuum equation are compared in Tables 9 and 10 for reference, along with a reminder of the proper unit scalings for both SI and CGS conventions.

---

[13] These two paravector equations are in fact equivalent to Maxwell's preferred quaternion formulation of electromagnetism [134]. However, as mentioned in Section 3.7, in place of the basis 3-vector directions $\vec{\sigma}_i$ he used the (left-handed set of) basis quaternions $(\mathbf{i}, \mathbf{j}, \mathbf{k}) = (\vec{\sigma}_1 I^{-1}, -\vec{\sigma}_2 I^{-1}, \vec{\sigma}_3 I^{-1})$, which geometrically correspond to rotation planes and thus caused numerous subtle problems.

Table 9
The electromagnetic field **F** and its dual **G**, as expressed in various formalisms. In spacetime algebra **F** is a bivector, and its dual **G** is obtained by right-multiplication with the pseudoscalar I^{-1}. In differential forms **F** is a 2-form (antisymmetric rank-2 tensor) with components $F_{\mu\nu}$, and its dual **G** is obtained with the Hodge-star transformation $\star$. In component notation, this Hodge-star is the total contraction of $F_{\mu\nu}$ with the fully anti-symmetric Levi-Civita symbol $\epsilon_{\mu\nu\alpha\beta}$. The standard Gibbs 3-vector notation is reference-frame-dependent, but the 3-vectors of $\vec{E}$ and $\vec{B}$ can still be appropriately paired.

	Electromagnetic Field
Spacetime Algebra:	$\boxed{\mathbf{F} = \vec{E} + \vec{B}I}$ $\boxed{\mathbf{G} = \mathbf{F}I^{-1} = \vec{B} - \vec{E}I}$
Differential Forms:	$\mathbf{F} = \dfrac{1}{2}\sum_{\mu\nu} F_{\mu\nu}dx^{\mu} \wedge dx^{\nu}$ $\mathbf{G} = \star\mathbf{F}$
Tensor Components:	$G_{\mu\nu} = -\dfrac{1}{2}\sum_{\alpha\beta} \epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}$ $E_i = F_{i0}, \qquad B_i = G_{i0}$
Gibbs 3-Vectors:	$(\vec{E}, \vec{B})$ $(\vec{B}, -\vec{E})$
SI Units:	$\vec{E} \mapsto \sqrt{\epsilon_0}\,\vec{E}, \qquad \vec{B} \mapsto \vec{B}/\sqrt{\mu_0}, \qquad c \equiv 1/\sqrt{\epsilon_0\mu_0}$
CGS Units:	$\vec{E} \mapsto \vec{E}/\sqrt{4\pi}, \qquad \vec{B} \mapsto \vec{B}/\sqrt{4\pi}$

Eq. (5.1) is manifestly proper, but the four equations (5.5) are not so obviously invariant under changes of the inertial reference frame. We can see the effect of choosing a different frame by splitting $\mathbf{F} = \vec{E}' + \vec{B}'I$ along a different timelike vector γ_0' according to (3.42). This produces different relative 3-vector fields

$$\begin{aligned}
\vec{E}' &= (\mathbf{F}\cdot\gamma_0')\gamma_0' = (\vec{E}\cdot\gamma_0')\gamma_0' + (\vec{B}I\cdot\gamma_0')\gamma_0', \\
\vec{B}'I &= (\mathbf{F}\wedge\gamma_0')\gamma_0' = (\vec{E}\wedge\gamma_0')\gamma_0' + (\vec{B}I\wedge\gamma_0')\gamma_0',
\end{aligned} \tag{5.7}$$

which must also satisfy (5.5) (proved by the same derivation). These different relative fields accordingly mix the components of $\vec{E}$ and $\vec{B}$, even though they produce the *same* proper bivector field $\mathbf{F} = \vec{E} + \vec{B}I = \vec{E}' + \vec{B}'I$. The subtle and important role of the pseudoscalar I in managing the implicit relative bases $\vec{\sigma}_i = \gamma_i\gamma_0$ and $\vec{\sigma}_i' = \gamma_i'\gamma_0'$ becomes apparent when comparing these different frame expansions: using a scalar imaginary i would not preserve the frame invariance in the same way.

The mixing of components from such a change of frame will have precisely the same form as a Lorentz boost, such as the one derived in (3.59). To prove this, we note that we can always write the new frame vector γ_0' as a Lorentz boost of the old frame γ_0 with some relative velocity $\vec{v}$. This boost has the spinor representation $\gamma_0' = \psi\gamma_0\widetilde{\psi}$ with the rotation spinor $\psi = \exp(-\alpha\hat{v}/2)$, where the velocity unit 3-vector $\hat{v} = \vec{v}/|v|$ also acts as the plane of rotation for the boost, and $\alpha = \tanh^{-1}|v|/c$ is the appropriate rapidity (rotation angle). This Lorentz transformation produces the relation

$$\gamma_0' = \psi\gamma_0\widetilde{\psi} = \exp(-\alpha\hat{v})\gamma_0 = \frac{1 - \vec{v}/c}{\sqrt{1 - (|v|/c)^2}}\gamma_0. \tag{5.8}$$

After defining $\gamma = [1 - (|v|/c)^2]^{-1/2}$ to be the usual time-dilation factor for brevity, we can then directly expand **F** into the relative fields of the γ_0' frame using (5.2)

$$\begin{aligned}
\mathbf{F} &= \mathbf{F}\gamma_0'\gamma_0'^{-1} = \gamma^2\mathbf{F}\gamma_0\gamma_0(1 - \vec{v}/c)(1 + \vec{v}/c) \\
&= \gamma^2(\vec{E} + \vec{B}I)(1 - \vec{v}/c)(1 + \vec{v}/c), \\
&= \gamma^2[\vec{E} + \vec{B}I - \vec{E}\cdot(\vec{v}/c) - \vec{B}\cdot(\vec{v}/c)I + \vec{B}\times(\vec{v}/c) - \vec{E}\times(\vec{v}/c)I](1 + \vec{v}/c), \\
&= [\vec{E}_\parallel + \gamma(\vec{E}_\perp + \vec{v}\times\vec{B}_\perp/c)] + [\vec{B}_\parallel + \gamma(\vec{B}_\perp - \vec{v}\times\vec{E}_\perp/c)]I,
\end{aligned} \tag{5.9}$$

which is the same form as the direct Lorentz boost of **F** derived in (3.59).

Table 10
Maxwell's vacuum equation, as expressed in various formalisms. Notably, only spacetime algebra permits the formulation as the single boxed equation for the electromagnetic bivector field **F**, which can be algebraically expanded into various other forms as needed. Shown are successive expansions that are equivalent to the formalisms of differential forms, paravectors/quaternions, and Gibbs 3-vectors, respectively, to show how they are all naturally contained within spacetime algebra.

	Maxwell's Equation in Vacuum

Spacetime Algebra: $\boxed{\nabla \mathbf{F} = 0}$ $\quad$ $\boxed{\mathbf{F} = \vec{E} + \vec{B}I}$

$$\Downarrow$$

$$\nabla \cdot \mathbf{F} = 0, \qquad \nabla \wedge \mathbf{F} = 0$$

$$\Downarrow$$

$$\vec{\nabla} \cdot \vec{E} + (\partial_0 \vec{E} - \vec{\nabla} \times \vec{B}) = 0$$
$$\vec{\nabla} \cdot \vec{B} + (\partial_0 \vec{B} + \vec{\nabla} \times \vec{E}) = 0$$

$$\Downarrow$$

$$\vec{\nabla} \cdot \vec{E} = 0, \qquad \vec{\nabla} \times \vec{B} = \partial_0 \vec{E}$$
$$\vec{\nabla} \cdot \vec{B} = 0, \qquad \vec{\nabla} \times \vec{E} = -\partial_0 \vec{B}$$

Differential Forms: $\quad \mathrm{d}\mathbf{G} = \mathrm{d}\star\mathbf{F} = 0, \qquad \mathrm{d}\mathbf{F} = \mathrm{d}\star\mathbf{G} = 0$

Tensor Components: $\quad \partial_\nu F^{\mu\nu} = 0, \qquad \partial_\nu G^{\mu\nu} = 0$

Gibbs 3-Vectors: $\quad \begin{aligned} \vec{\nabla} \cdot \vec{E} &= 0, & \vec{\nabla} \times \vec{B} &= \partial_0 \vec{E} \\ \vec{\nabla} \cdot \vec{B} &= 0, & \vec{\nabla} \times \vec{E} &= -\partial_0 \vec{B} \end{aligned}$

SI Units: $\quad \vec{E} \mapsto \sqrt{\epsilon_0}\,\vec{E}, \qquad \vec{B} \mapsto \vec{B}/\sqrt{\mu_0}, \qquad \mathrm{c} \equiv 1/\sqrt{\epsilon_0 \mu_0}$

CGS Units: $\quad \vec{E} \mapsto \vec{E}/\sqrt{4\pi}, \qquad \vec{B} \mapsto \vec{B}/\sqrt{4\pi}$

5.2. Global phase degeneracy and dual symmetry

Notably, Maxwell's vacuum equation (5.1) is invariant under *global* phase transformations, such as those considered in (3.63). To understand the implications of this symmetry of Maxwell's equation, we will briefly revisit how this transformation acts on the electric field itself.

Specifically, we apply a phase rotation using the spinor representation $\psi = \exp(-\theta I/2)$ to Eq. (5.1) to find

$$\psi(\nabla \mathbf{F})\psi^* = e^{-\theta I/2}(\nabla \mathbf{F})e^{\theta I/2} = \nabla(e^{\theta I/2}\mathbf{F}e^{\theta I/2}) = \nabla \mathbf{F}\exp(\theta I) = 0, \tag{5.10}$$

which shows that the phase rotation cancels. Note that this phase transformation is equivalent to modifying the *intrinsic* phase of $\mathbf{F} \mapsto \mathbf{F}\exp(\theta I)$. This added phase mixes $\vec{E}$ and $\vec{B}$ at every point x in the same way[14] while leaving the inertial frame the same

$$\mathbf{F}\exp(\theta I) = (\cos\theta\vec{E} - \sin\theta\vec{B}) + (\sin\theta\vec{E} + \cos\theta\vec{B})I. \tag{5.11}$$

The interesting special case of the angle $\theta = -\pi/2$ exchanges the relative fields $\vec{E}$ and $\vec{B}$ up to a sign at every point x. The resulting bivector is dual to **F** and is typically given a special notation

$$\mathbf{G} = \mathbf{F}I^{-1} = \vec{B} - \vec{E}I, \tag{5.12}$$

---

[14] Note that we could consider converting this global gauge transformation of **F** into a local gauge transformation in the usual way by defining a covariant derivative using an auxiliary scalar field (see Sections 8.9.1 and 8.9.2), but we shall not do so here. However, see [196] for an interesting connection between performing this procedure here and the electroweak theory of particle physics.

which can also be written in tensor component notation as

$$G^{\mu\nu} = \star F^{\mu\nu} = -\frac{1}{2}\sum_{\alpha,\beta}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta} \tag{5.13}$$

using the Hodge-star operation from differential forms (or the completely antisymmetric Levi-Civita symbol $\epsilon^{\mu\nu\alpha\beta}$) as shown in Table 9 [68]. The dual bivector **G** has been particularly useful for studies in magnetic monopoles [101,102,193,197–213] and optical helicity [68,103–108,214], and will play an interesting role in the Lagrangian treatment described in Section 8.

This exchange freedom of the vacuum Maxwell equation [110,111] has historically been called the *dual symmetry* of the vacuum Maxwell fields [68,99,101,102,105–108], and has prompted considerable study into similar dualities in field and string theories beyond standard electromagnetism [112–125]. We see here that it appears as a structural feature of a bivector field on spacetime that exploits a global phase symmetry of the equation of motion. Note that the bivector magnitude $|\mathbf{F}|^2 \equiv \mathbf{F}^*\mathbf{F} = \mathbf{G}^*\mathbf{G}$ is always manifestly invariant under dual symmetry.

5.3. Canonical form

We now revisit the expansion of the electromagnetic field into its canonical form, introduced in Section 3.5.1. These expansions do not require Maxwell's equation, but will be useful in what follows.

Computing the square of $\mathbf{F} = \vec{E} + \vec{B}I$ produces two proper scalar fields as a single complex scalar

$$\mathbf{F}^2 = |\mathbf{F}|^2\exp(2\varphi I) = (|\vec{E}|^2 - |\vec{B}|^2) + 2(\vec{E}\cdot\vec{B})I. \tag{5.14}$$

These scalars are precisely those used to construct Lagrangian densities for the electromagnetic field [215]. The first scalar is precisely the usual electromagnetic Lagrangian, while the second term is the "axion" contribution that has been extensively discussed in particle physics [216]. We will return to the issue of electromagnetic Lagrangians in Section 8.

Denoting the two proper scalars as

$$\ell_1 = \langle\mathbf{F}^2\rangle_0 = |\vec{E}|^2 - |\vec{B}|^2, \tag{5.15}$$

$$\ell_2 = \langle\mathbf{F}^2\rangle_4 I^{-1} = 2(\vec{E}\cdot\vec{B}), \tag{5.16}$$

they determine the intrinsic local phase of **F** according to (3.29)

$$\varphi = \frac{1}{2}\tan^{-1}\frac{\ell_2}{\ell_1}, \tag{5.17}$$

which demonstrates that this phase is also a proper scalar field.

The canonical bivector corresponding to **F** can then be computed according to (3.29)

$$\mathbf{f} = \mathbf{F}\exp(-\varphi I) = (\cos\varphi\vec{E} + \sin\varphi\vec{B}) + (-\sin\varphi\vec{E} + \cos\varphi\vec{B})I, \tag{5.18}$$

where we can expand the trigonometric functions in terms of ℓ_1 and ℓ_2 as

$$\cos\varphi = \sqrt{\frac{1}{2}\left(1 + \frac{\ell_1}{\sqrt{\ell_1^2 + \ell_2^2}}\right)}, \qquad \sin\varphi = \sqrt{\frac{1}{2}\left(1 - \frac{\ell_1}{\sqrt{\ell_1^2 + \ell_2^2}}\right)}. \tag{5.19}$$

This transformation performs a rotation of the relative fields in the complex plane of the spacetime split at each local point x, while leaving the chosen frame of γ_0 invariant.

The proper complex conjugate of the field can be computed in a similar way according to (3.37),

$$\mathbf{F}^* = \mathbf{f}\exp(-\varphi I) = \mathbf{F}\exp(-2\varphi I) = \frac{(\ell_1\vec{E} + \ell_2\vec{B}) + (-\ell_2\vec{E} + \ell_1\vec{B})I}{\sqrt{\ell_1^2 + \ell_2^2}}. \tag{5.20}$$

It follows that the dual-symmetric field magnitude has the explicit and intuitive form

$$|\mathbf{F}|^2 = \mathbf{F}^*\mathbf{F} = \sqrt{\ell_1^2 + \ell_2^2}. \tag{5.21}$$

Note that when $\ell_2 \propto \vec{E}\cdot\vec{B} = 0$ the proper phase vanishes $\varphi = 0$ and the bivector field is purely canonical $\mathbf{F} = \mathbf{F}^* = \mathbf{f} = \vec{E} + \vec{B}I$. Indeed the pseudoscalar part in the pseudonorm (5.14) vanishes in this case. The magnitude $|\mathbf{F}|^2 = |\ell_1|$ becomes the pseudonorm involving only the traditional Lagrangian density $\ell_1 = |\vec{E}|^2 - |\vec{B}|^2$, albeit in a manifestly dual-symmetric form (due to the absolute value).

Similarly, note that when $\ell_1 = |\vec{E}|^2 - |\vec{B}|^2 = 0$ the proper phase is $\varphi = \pi/4$ and the canonical bivector field has the symmetrized form

$$\mathbf{f} = \frac{1}{\sqrt{2}}[(\vec{E} + \vec{B}) + (-\vec{E} + \vec{B})I]. \tag{5.22}$$

The conjugate becomes the dual $\mathbf{F}^* = \mathbf{F}I^{-1} = \vec{B} - \vec{E}I$ up to the sign of ℓ_2, and the magnitude becomes the other proper scalar $|\mathbf{F}|^2 = |\ell_2| = 2|\vec{E} \cdot \vec{B}|$ in a manifestly dual-symmetric form.

When $\ell_1 = \ell_2 = 0$ then $|\mathbf{F}|^2 = 0$, making $\mathbf{F}$ a null bivector. The phase $\varphi(x)$ correspondingly becomes degenerate. Similarly, the phase of $\mathbf{F}^*$ will also be degenerate, but will always be the negative of the degenerate phase of $\mathbf{F}$. For such a null field, the conditions $|\vec{E}| = |\vec{B}|$ and $\vec{E} \cdot \vec{B} = 0$ hold according to (5.14).

5.4. Electromagnetic waves

Although a null vacuum field $\mathbf{F}$ has a globally degenerate phase, it must still satisfy Maxwell's equation (5.1). The constraining derivative therefore breaks the *local* phase degeneracy by connecting the local values of $\varphi(x)$ at nearby points x. Only the global phase remains arbitrary in the form of the dual symmetry of the solution.

To see how Maxwell's equation (5.1) breaks the phase degeneracy of a null field for a specific and important solution, consider a field $\mathbf{F} = \mathbf{f}\exp[\varphi(x)I]$ with a *constant* canonical bivector $\mathbf{f}$. Maxwell's equation (5.1) then takes the simpler form

$$\nabla \mathbf{F} = [\nabla\varphi]\,\mathbf{f} = 0. \tag{5.23}$$

If $\nabla\varphi \neq 0$ and $\mathbf{f} \neq 0$ then both factors must be null factors that cannot be inverted, making $\mathbf{F}$ a null bivector. Since $\mathbf{f}$ is constant for all x, then $[\nabla\varphi]$ must not vary with x in order to satisfy (5.23); hence the phase must be linear $\varphi = \varphi_0 \pm x \cdot k$ in terms of a null wavevector $k = \nabla\varphi$ with units inverse to x. Furthermore, the constraint $k\mathbf{f} = 0$ implies that the constant bivector $\mathbf{f}$ can be written in terms of this null wavevector as

$$\mathbf{f} = sk = s \wedge k = -k \wedge s \tag{5.24}$$

for some constant spacelike vector s orthogonal to k: $s \cdot k = 0$. Since the global phase is arbitrary, we set $\varphi_0 = 0$, and thus obtain the (monochromatic) *plane wave solutions*

$$\mathbf{F}(x) = (sk)\,\exp[\pm(k \cdot x)I]. \tag{5.25}$$

Importantly, this complex form of the plane wave solution arises with no *ad hoc* introduction of the complex numbers [173], unlike the usual treatment of electromagnetic plane waves. The factor of I is the intrinsic pseudoscalar for spacetime, and both the "real" and "imaginary" parts of the expression produce meaningful and necessary parts of the same proper bivector. Note that the six components of $\mathbf{F}$ are now entirely contained in the specification of the null vector k and the spacelike orientation vector s orthogonal to k.

If we perform a spacetime split of $k = \gamma_0(\omega - \vec{k})$ to put (5.25) in more standard notation (setting $c = 1$ for convenience), we find that the null condition $k^2 = (\omega + \vec{k})(\omega - \vec{k}) = \omega^2 - |\vec{k}|^2 = 0$ implies the usual *dispersion relation* $|\vec{k}| = |\omega|$ in any frame.[15] Hence, we can factor out a frame-dependent frequency as a magnitude $k = \omega k_0$ to isolate a purely directional null factor $k_0 = \gamma_0(1 - \vec{\kappa})$ that contains the relative unit vector $\vec{\kappa} \equiv \vec{k}/|\vec{k}|$ for the chosen frame.

It follows from this decomposition that Maxwell's equation (5.23) can also be written as the eigenvalue constraint $\vec{\kappa}\mathbf{f} = \mathbf{f}$, which makes $\mathbf{f}$ an eigenbivector of the relative unit vector $\vec{\kappa}$ (in any frame). Performing a spacetime split $\mathbf{f} = sk = \vec{E} + \vec{B}I$ then reduces this constraint to two copies of the equation $\vec{\kappa}\vec{E} = \vec{B}I$, in accordance with the null condition $|\vec{E}| = |\vec{B}|$, which implies $\vec{\kappa} \times \vec{E} = \vec{B}$ since there is no scalar term. Therefore, $\vec{\kappa} \cdot \vec{E} = \vec{\kappa} \cdot \vec{B} = 0$ and the three relative vectors $\vec{\kappa}\vec{E}\vec{B} = +I|\vec{E}||\vec{B}|$ form a *right-handed* oriented set of basis vectors in any relative frame. Consequently, the appropriately scaled directional factor ωs for the null bivector $\mathbf{f} = sk = \omega sk_0$ is simply the purely spatial part of its spacetime split, $\omega s = \vec{E}\gamma_0$.

The canonical bivector for the plane wave (5.25) is therefore

$$\mathbf{f} = \omega sk_0 = (\vec{E}\gamma_0)k_0 = \vec{E}(1 - \vec{\kappa}) = (1 + \vec{\kappa})\vec{E}. \tag{5.26}$$

Hence, in a relative frame $\vec{E}\gamma_0$ determines a reference orientation and notion of amplitude $|\vec{E}|$ for $\mathbf{f}$, while the purely directional null factor $k_0 = \gamma_0(1 - \vec{\kappa})$ provides the proper wave orientation. The full relative-frame version of (5.25) thus has the familiar form

$$\begin{aligned}
\mathbf{F} &= \vec{E}\,(1 - \vec{\kappa})\,\exp[\pm(\omega t - \vec{k} \cdot \vec{x})I], \\
&= (\vec{E} + \vec{\kappa} \times \vec{E}I)\,\exp[\mp(\vec{k} \cdot \vec{x} - \omega t)I], \\
&= [\vec{E}\cos(\vec{k} \cdot \vec{x} - \omega t) \mp \vec{\kappa} \times \vec{E}\sin(\vec{k} \cdot \vec{x} - \omega t)] + [\vec{E}\sin(\vec{k} \cdot \vec{x} - \omega t) \pm \vec{\kappa} \times \vec{E}\cos(\vec{k} \cdot \vec{x} - \omega t)]I.
\end{aligned} \tag{5.27}$$

---

[15] Note that choosing a different frame will boost the frequency ω (i.e., red- or blue-shift it from the Doppler effect), but will leave the null vector k and the dispersion relation invariant.

We emphasize that an overall frequency scaling factor ω has been absorbed into the amplitude of $\vec{E}$ in the frame γ_0 by convention, so this relative amplitude will also change with the reference frame due to the Doppler effect. Indeed, we will see later in Section 7.5.1 that the energy-density of the field (which contains the square of this relative amplitude) is a frame-dependent quantity.

The relative form (5.27) makes it clear that the pure plane wave solution (5.25) with $-I$ is *right-hand circularly polarized* in the traditional sense. In each relative spatial plane orthogonal to the relative wavevector direction $\vec{\kappa}$ (i.e., $\vec{\kappa} \cdot \vec{x} = 0$), the relative vectors $\vec{E}$ and $\vec{B} = \vec{\kappa} \times \vec{E}$ rotate around the axis $\vec{\kappa}$ at a frequency ω with increasing t. This rotation shows that a circularly polarized wave has a definite *helicity* [45,46,48–85], which is a manifestation of the spin-1 nature of the electromagnetic field. Choosing the opposite sign of the pseudoscalar $+I$ in (5.25) correspondingly produces a *left-hand circularly polarized* plane wave by flipping the sign of the phase in (5.27), effectively flipping both the propagation direction $\vec{\kappa}$ and the direction of rotation with increasing t. It is now also clear that the dual-symmetric global phase freedom of the solution (5.25) corresponds to the arbitrary phase offset of the polarization rotation (5.27); that is, the relative vectors $\vec{E}$ and $\vec{B}$ may be arbitrarily rotated in unison around the propagation axis $\vec{\kappa}$ without changing the solution (in any frame).

To go beyond the monochromatic plane wave solution, we observe that in addition to flipping the sign of I in (5.25), we can also scale the magnitude ω of k arbitrarily. We can thus exploit the fact that Maxwell's equation is linear to construct new solutions as superpositions of all possible values of ω. It follows that for any given directional null vector k there exists an infinite number of corresponding (polychromatic) *wave-packet* solutions determined by scalar spectral weight functions $\alpha(\omega)$

$$
\begin{aligned}
\mathbf{F} &= sk \int_{-\infty}^{\infty} d\omega\, \alpha(\omega)\, \exp[-\omega\,(k \cdot x)I], \\
&= sk \int_{0}^{\infty} d\omega\, [\alpha_+(\omega)\, \exp[-\omega\,(k \cdot x)I] + \alpha_-(\omega)\, \exp[\omega\,(k \cdot x)I]],
\end{aligned}
\tag{5.28}
$$

where $\alpha_\pm(\omega) = \alpha(\pm|\omega|)$. The function $\alpha(\omega)$ can absorb the magnitude of s for each choice of ω, making $s^2 = -1$ a spacelike reference unit vector by convention. This total field $\mathbf{F}$ is a linear superposition of the right- and left-hand circularly polarized plane waves at each positive scalar frequency $|\omega|$, and still satisfies Maxwell's vacuum equation (5.1) by construction. Note that when $\alpha_+ = \alpha_-^*$ the integral becomes purely real and the resulting polarization is linear. In general, coefficients α_+ and α_- determine the Jones vector of the wave polarization in the basis of circular polarizations, which are attached to the coordinate frame specified by the vector s orthogonal to k.

The two signed solutions with $\pm|\omega|$ are traditionally called the *positive-frequency* and *negative-frequency* plane wave solutions of Maxwell's vacuum equation, respectively. Here we see that the positive factor $|\omega|$ arises from the degeneracy of the null vector k with respect to scaling. However, the sign of ± 1 directly corresponds to a choice of handedness for I that determines a particular circular polarization (i.e., *helicity*). These features are invariant in (5.28) because null vectors and the handedness of I do not depend on any particular reference frame.

The solution (5.28) can also be understood as a Fourier transform of the spectral function $\alpha(\omega)$, which can be simply evaluated to obtain

$$
\mathbf{F} = sk\, \widetilde{\alpha}(-k_0 \cdot x) = (\vec{E}_0 + \vec{\kappa} \times \vec{E}_0 I)\, \widetilde{\alpha}(\vec{\kappa} \cdot \vec{x} - ct),
\tag{5.29}
$$

in terms of the relative unit-vector direction $\vec{E}_0$. The (complex scalar) amplitude $\widetilde{\alpha}$ is the Fourier-transform of $\alpha(\omega)$ that completely determines the packet shape of the traveling wave. The bivector prefactor provides an overall linear polarization for a traveling wave packet by default, with electric and magnetic parts phase-offset by an angle of $\pi/2$. This polarization can be made elliptic or circular, however, when $\widetilde{\alpha}$ has additional complex structure (as demonstrated in (5.27) for the monochromatic circularly polarized case).

We can also go beyond a simple plane wave solution altogether to construct a more general (not necessarily null) solution from (5.25) as a superposition of plane waves in all null wavevector directions k. To do this we can rewrite the null factor as $k = \gamma_0(|\vec{k}| - \vec{k})$ in a particular reference frame γ_0, after which the components of $\vec{k}$ can be integrated as a 3-vector to obtain

$$
\mathbf{F} = \iiint_{-\infty}^{\infty} d^3\vec{k}\, s(\vec{k})\, k_0(\vec{k})\, [\alpha_+(\vec{k})\, e^{-(k \cdot x)I} + \alpha_-(\vec{k})\, e^{(k \cdot x)I}],
\tag{5.30}
$$

where $k_0 = k/|\vec{k}| = \gamma_0(1 - \vec{\kappa})$ is a purely directional factor in the frame γ_0, and where the reference directions $s(\vec{k}) = \vec{E}_0(\vec{k})\gamma_0$ are orthogonal to each $\vec{\kappa}$: $s(\vec{k}) \cdot k(\vec{k}) \propto \vec{E}_0(\vec{k}) \cdot \vec{\kappa} = 0$. This integral (5.30) can be understood as the construction of $\mathbf{F}$ through a spatial Fourier transform.

The functions $\alpha_\pm(\vec{k})$ encode the spectral decomposition and magnitude of the field. The unit vectors $s(\vec{k})$ encode the reference polarizations for each traveling mode of the field. Notably, the functions $\alpha_\pm(\vec{k})$ that appear here become the foundation for the canonical ("second") quantization of the field in quantum field theory (after normalization), where they become the raising and lowering operators for specific field modes up to appropriate scaling factors [39]. We will not further explore this quantization procedure here, but the interested reader can find excellent discussions of how to quantize the field using the complex Riemann–Silberstein vector in Refs. [26,40].

Table 11
Vector potential representations of the vacuum electromagnetic field, as expressed in various formalisms. Notably, spacetime algebra most naturally motivates an intrinsically complex vector potential z. Nevertheless, one can also manually simulate this structure using differential forms with a scalar imaginary i and the Hodge star $\star$ (e.g., [68,196]), after interpreting the independent electric (a_e) and magnetic (a_m) vector potentials as one-forms $\mathbf{a}_{e,m}$ with the same components. As we discuss in Section 8.3, imposing the additional constraint $\mathbf{F}_e = \mathbf{F}_m I = \mathbf{F}/\sqrt{2}$ can optionally symmetrize the contributions of the two vector potentials so that they describe the same (vacuum) field $\mathbf{F}$ [68,196].

<div align="center">

Vector Potentials

Spacetime Algebra: $\boxed{\mathbf{F} = \nabla z}$ $\boxed{z = a_e + a_m I}$

$\Downarrow$

$\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m I$
$\mathbf{F}_e = \nabla \wedge a_e, \quad \mathbf{F}_m = \nabla \wedge a_m$
$\nabla \cdot a_e = \nabla \cdot a_m = 0$

$\Downarrow$

$\mathbf{F} = \vec{E} + \vec{B}I, \quad a_e = (\phi_e + \vec{A})\gamma_0, \quad a_m = (\phi_m + \vec{C})\gamma_0$
$\vec{E} = -\vec{\nabla}\phi_e - \partial_0 \vec{A} - \vec{\nabla} \times \vec{C}$
$\vec{B} = -\vec{\nabla}\phi_m - \partial_0 \vec{C} + \vec{\nabla} \times \vec{A}$

Differential Forms: $\mathbf{F} = \mathbf{F}_e + i \star \mathbf{F}_m$
$\mathbf{F}_e = \mathrm{d}\mathbf{a}_e, \quad \mathbf{F}_m = \mathrm{d}\mathbf{a}_m$

Tensor Components: $(F_e)_{\mu\nu} = \frac{1}{2}(\partial_\mu(a_e)_\nu - \partial_\nu(a_e)_\mu)$
$(F_m)_{\mu\nu} = \frac{1}{2}(\partial_\mu(a_m)_\nu - \partial_\nu(a_m)_\mu)$

Gibbs 3-Vectors: $\vec{E} = -\vec{\nabla}\phi_e - \partial_0 \vec{A} - \vec{\nabla} \times \vec{C}$
$\vec{B} = -\vec{\nabla}\phi_m - \partial_0 \vec{C} + \vec{\nabla} \times \vec{A}$

</div>

6. Potential representations

> *By performing particular operations on symbols, we acquire the possibility of expressing the same thing in many different forms.*

[James Clerk Maxwell [217]]

For more general solutions of Maxwell's equation in vacuum (5.1), it can be useful to construct the bivector field $\mathbf{F}$ from auxiliary *potential* fields. There are three types of potential that are particularly convenient to use: (1) a vector potential, (2) a bivector potential, and (3) a scalar Hertz potential. We will consider each in turn, and show their explicit relationship.

6.1. Complex vector potential

The standard vector potential construction follows from the ansatz

$$\mathbf{F} = \nabla z, \tag{6.1}$$

where z is some initially unspecified multivector field. This ansatz is compared with other formalisms in Table 11. Due to the fundamental theorem (4.13), this ansatz can only describe *conservative* fields that vanish when integrated around the boundary of any closed surface, such as a loop.

Since $\mathbf{F}$ is a pure bivector field, ∇z must satisfy the following conditions. First, z must be a vector field to produce a bivector with a derivative. Second, the *dual symmetry* of $\mathbf{F}$ forces z to be correspondingly *complex*. Specifically, a phase-rotation of $\mathbf{F}$ induces the transformation

$$\mathbf{F} \mapsto \mathbf{F}\exp(\theta I) = (\nabla z)\exp(\theta I) = \nabla(ze^{\theta I}) = \nabla(e^{-\theta I/2}ze^{\theta I/2}), \tag{6.2}$$

which is a phase rotation $\psi z \psi^*$ of z with a spinor representation $\psi = \exp(-\theta I/2)$, exactly as used in (5.10) and (3.63).

It follows that z must be a complex vector field of the form

$$z = a_e + a_m I \tag{6.3}$$

with separate ("electric" a_e and "magnetic" a_m) parts,[16] that satisfies the equation

$$\nabla z = \nabla \wedge a_e + \nabla \cdot (a_m I) = \nabla \wedge a_e + (\nabla \wedge a_m)I = \mathbf{F}, \tag{6.4}$$

along with the additional *Lorenz–FitzGerald conditions*[17]

$$\nabla \cdot a_e = \nabla \cdot a_m = 0. \tag{6.5}$$

As a result, the potential fields must be manifestly *transverse* $\nabla a_e = \nabla \wedge a_e$ and $\nabla a_m = \nabla \wedge a_m$.

This transversality implies that there is an *additional gauge freedom* in the definition (6.1) of z due to the invariance under the replacement $z' = z + \nabla \zeta$. Provided that[18] $\nabla^2 \zeta = 0$, each new z' will produce the same field according to $\nabla z' = \nabla z + \nabla^2 \zeta = \nabla z = \mathbf{F}$. Moreover, each new z' will also satisfy the Lorenz–FitzGerald conditions since $\nabla \cdot \nabla \zeta = \nabla^2 \zeta = 0$.

6.1.1. Plane wave vector potential

If we revisit the null plane wave solution of (5.25), and use the relation $\mathbf{F} = \nabla z$, we can infer that the potential z for a polarized plane wave must have the simple form (being careful to remember that I anticommutes with 4-vectors)

$$z = \exp[\mp(k \cdot x)I](Is) = (Is) \exp[\pm(k \cdot x)I], \tag{6.6}$$

where the spacelike vector s determines a reference polarization direction and amplitude. To verify that this is the appropriate potential, we take its derivative

$$\nabla z = (\nabla e^{\mp(k \cdot x)I})Is = (kI)e^{\mp(k \cdot x)I}Is = -kse^{\pm(k \cdot e)I} = (sk)e^{\pm(k \cdot e)I}, \tag{6.7}$$

which reproduces (5.25). A second derivative will annihilate the null factor k to produce Maxwell's equation in the form $\nabla^2 z = 0$.

The complexity of the vector potential z is critical for obtaining (6.6). There are intrinsic phases in the traveling wave factor $\exp(\pm(k \cdot x)I)$. Furthermore, this is the vector potential for a null solution $\mathbf{F}^2 = (\nabla z)^2 = 0$, so the global intrinsic phase of the solution must be degenerate as a general structural property of null bivectors.

Moreover, we know from the derivation of the plane wave (5.27) that the directional vector $s = \vec{E}\gamma_0$ should be orthogonal to k in any frame. This constraint is another way of expressing the Lorenz–FitzGerald condition $\nabla \cdot z = 0$ for the gauge freedom since

$$\nabla \cdot z = \pm I(s \cdot \nabla) \exp[\pm(k \cdot x)I] = \pm(s \cdot k) \exp[\pm(k \cdot x)I] = 0. \tag{6.8}$$

We also know from the derivation of (5.27) that the vector s arises as a factor of a null bivector that can always be written in a relative frame as purely spatial vector $s = \vec{E}_0\gamma_0$. It then follows that $\vec{E}_0 \cdot \vec{k} = 0$, which makes the plane wave potential manifestly *transverse* in any relative frame: $\vec{\nabla} \cdot (z\gamma_0) = \pm(\vec{E}_0 \cdot \vec{k}) \exp[\mp(\vec{k} \cdot \vec{x} - \omega t)I] = 0$. This property will become important in Section 6.2.

It follows from (6.6) that the vector potential for the null wave packet in (5.28) has the form

$$z = Is \int_{-\infty}^{\infty} \frac{d\omega}{\omega} \alpha(\omega) e^{-\omega (k \cdot x)I},$$

$$= Is \int_0^{\infty} \frac{d\omega}{\omega} \left[\alpha_+(\omega) e^{-\omega (k \cdot x)I} - \alpha_-(\omega) e^{\omega (k \cdot x)I} \right], \tag{6.9}$$

where the derivative produces a sign flip between the spectral functions $\alpha_\pm$. The potential for the superposition of plane waves in (5.30) thus has the corresponding form

$$z = I \iiint_{-\infty}^{\infty} \frac{d^3\vec{k}}{|\vec{k}|} s(\vec{k}) [\alpha_+(\vec{k}) e^{-(k \cdot x)I} - \alpha_-(\vec{k}) e^{(k \cdot x)I}], \tag{6.10}$$

where the remaining factor of $|\vec{k}|$ can be absorbed into $\alpha_\pm$ as units and convention dictate. These potentials are also manifestly transverse in any frame since $s(\vec{k}) \cdot k = 0$ by construction for each null plane wave propagating in direction $\vec{k}$.

---

[16] Note that we will use the subscripts e and m throughout this report to distinguish the quantities that will become naturally associated with electric and magnetic *charges* in Section 7 (not to be confused with the relative electric and magnetic fields, which can both be produced by either type of charge). To recover traditional electromagnetism, all quantities with magnetic subscripts can be neglected, but we keep the discussion general.

[17] This condition is also known as the *Lorenz gauge*, which only partially fixes the gauge freedom to be more natural for manifestly relativistic treatments of the field.

[18] Note that if $\zeta = \alpha + \beta I$ such that $\nabla^2 \zeta \neq 0$, then we obtain $\nabla z' = \mathbf{F} + \nabla^2 \zeta$ with $\mathbf{F} = \nabla \wedge (a + \nabla \alpha) + [\nabla \wedge (g + \nabla \beta)]I = \langle \nabla z' \rangle_2$. Adding this projection in the field definition provides a more general (but less natural) gauge freedom that does not preserve the Lorenz–FitzGerald conditions.

6.1.2. Relative vector potentials

We perform a spacetime split of the vector potentials $a_e = \gamma_0(\phi_e - \vec{A})$ and $a_m = \gamma_0(\phi_m - \vec{C})$ into *relative scalar potentials* ϕ_e, ϕ_m, and *relative vector potentials* $\vec{A}, \vec{C}$. Note that for these relative quantities, we try to conform to existing notation conventions. The notation $\vec{C}$ for the magnetic vector potential has become standard in studies of dual symmetry in vacuum [68,99,101,102,105,106,108]. With these relative potentials, we can expand the field into a relative form

$$
\begin{aligned}
\mathbf{F} = \nabla z &= (\partial_0 - \vec{\nabla})[(\phi_e - \vec{A}) + (\phi_m - \vec{C})I], \\
&= [-\partial_0\vec{A} - \vec{\nabla}\phi_e - \vec{\nabla} \times \vec{C}] + [-\partial_0\vec{C} - \vec{\nabla}\phi_m + \vec{\nabla} \times \vec{A}]I,
\end{aligned}
\tag{6.11}
$$

where we have used the Lorenz–FitzGerald conditions (6.5) written in the relative frame,

$$
\partial_0\phi_e + \nabla \cdot \vec{A} = \partial_0\phi_m + \nabla \cdot \vec{C} = 0.
\tag{6.12}
$$

This split implies that the relative fields may be written as

$$
\vec{E} = -\vec{\nabla}\phi_e - \partial_0\vec{A} - \vec{\nabla} \times \vec{C}, \qquad \vec{B} = -\vec{\nabla}\phi_m - \partial_0\vec{C} + \vec{\nabla} \times \vec{A},
\tag{6.13}
$$

in terms of the relative potentials.

Under a gauge transform $z' = z + \nabla\zeta$, with $\zeta = \zeta_a + \zeta_g I$, these relative vector potentials become

$$
\begin{aligned}
z'\gamma_0 &= (\phi_e + \vec{A}) + (\phi_m - \vec{C})I + (\partial_0 - \vec{\nabla})(\zeta_a - \zeta_g I), \\
&= (\phi_e + \partial_0\zeta_a) + (\vec{A} - \vec{\nabla}\zeta_a) + (\phi_m - \partial_0\zeta_g)I - (\vec{C} - \vec{\nabla}\zeta_g)I.
\end{aligned}
\tag{6.14}
$$

Thus, the relative scalar and vector potentials can become mixed both by Lorentz transformations and gauge transformations in general. One must fix both the inertial reference frame and the gauge in order to completely fix the form of the potential.

6.1.3. Constituent fields

The decomposition of the field (6.4) can be interpreted as constructing the total field $\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m I$ from *two constituent bivector fields* that are derived from each vector potential separately. The "electric" part, $\mathbf{F}_e = \nabla \wedge a_e$, will become naturally associated with electric charge in Section 7, while the "magnetic" part, $\mathbf{F}_m = \nabla \wedge a_m$, will be correspondingly associated with magnetic charge. However, the dual symmetry of vacuum makes these constituent fields interchangeable. For example, we can exchange their roles by taking the dual of $\mathbf{F}$ as in (5.12)

$$
\mathbf{G} = \mathbf{F}I^{-1} = (\nabla \wedge a_m) - (\nabla \wedge a_e)I = \mathbf{F}_m - \mathbf{F}_e I = \vec{B} - \vec{E}I.
\tag{6.15}
$$

If we perform a spacetime split of each piece of the total field, $\mathbf{F}_e = \vec{E}_e + \vec{B}_e I$ and $\mathbf{F}_m I = \vec{E}_m + \vec{B}_m I$ (so $\mathbf{F}_m = \vec{B}_m - \vec{E}_m I$), then we find that the relative fields can be written as

$$
\begin{aligned}
\vec{E}_e &= -\vec{\nabla}\phi_e - \partial_0\vec{A}, \qquad \vec{B}_e = \vec{\nabla} \times \vec{A}, \\
\vec{E}_m &= -\vec{\nabla} \times \vec{C}, \qquad \vec{B}_m = -\vec{\nabla}\phi_m - \partial_0\vec{C}.
\end{aligned}
\tag{6.16}
$$

The total relative fields include contributions from both potentials, $\vec{E} = \vec{E}_e + \vec{E}_m, \vec{B} = \vec{B}_e + \vec{B}_m$.

Note that we will use the subscripts of e and m throughout this report to signify the eventual associations of physical quantities with either electric or magnetic *charges*. Traditionally, the electromagnetic field couples only to *electric* charges, so is described by $\mathbf{F}_e$ and its associated electric vector potential a_e. Therefore, to recover traditional electromagnetism, we can simply neglect all "magnetic" quantities with an m subscript in what follows. However, in the absence of charges (as in vacuum) there is no such constraint, so we keep the formulas dual-symmetric.

6.2. Bivector potential

We can make an important simplification of the gauge freedom of the complex vector potential z if we introduce the *locally transverse* $\vec{A}_\perp, \vec{C}_\perp$ and *longitudinal* $\vec{A}_\parallel, \vec{C}_\parallel$ parts of the relative vector potentials $\vec{A} = \vec{A}_\perp + \vec{A}_\parallel$ and $\vec{C} = \vec{C}_\perp + \vec{C}_\parallel$ in a particular frame γ_0. These pieces of the vector potentials are defined to satisfy

$$
\vec{\nabla} \cdot \vec{A}_\perp = \vec{\nabla} \cdot \vec{C}_\perp = 0,
\tag{6.17}
$$

$$
\vec{\nabla} \times \vec{A}_\parallel = \vec{\nabla} \times \vec{C}_\parallel = 0.
\tag{6.18}
$$

As a result, the gauge transform (6.14) produces

$$
z'_\perp \gamma_0 = \vec{A}_\perp - \vec{C}_\perp I = z_\perp \gamma_0,
\tag{6.19}
$$

$$
z'_\parallel \gamma_0 = (\phi_e + \partial_0\zeta_a) + (\vec{A}_\parallel - \vec{\nabla}\zeta_a) + (\phi_m - \partial_0\zeta_g)I - (\vec{C}_\parallel - \vec{\nabla}\zeta_g)I.
\tag{6.20}
$$

That is, in any local frame the *transverse* part of the vector potential is *gauge-invariant*.

Table 12

Various bivector potential representations of the vacuum electromagnetic field. All such representations fix a particular reference frame γ_0, and fix the gauge to make the longitudinal part of the vector potentials vanish $z_\parallel = 0$ (i.e., fixing the Coulomb gauge). With this simplification, the bivector potential $\mathbf{Z}$ contains the same information as the vacuum field $\mathbf{F}$ without added freedom. As such, expanding this fundamental bivector potential into various forms permits alternative strategies for solving Maxwell's vacuum equation.

Bivector potentials for vacuum fields	
Transverse bivector potential:	$\mathbf{Z} = z_\perp \gamma_0 = \vec{A}_\perp - \vec{C}_\perp I$
Field relation:	$\mathbf{F} = (\nabla \mathbf{Z})\gamma_0 \qquad (z_\parallel = 0)$
Vacuum equation:	$\nabla^2 \mathbf{Z} = 0$
Hertz bivector potential:	$\mathbf{Z} = -[\vec{\nabla}, \mathbf{\Pi}] = -\vec{\nabla} \times \mathbf{\Pi} I$
Field relation:	$\mathbf{F} = \nabla \gamma_0 [\vec{\nabla}, \mathbf{\Pi}] = (I\partial_0 + \vec{\nabla}\times)(\vec{\nabla} \times \mathbf{\Pi})$
Vacuum equation:	$\nabla^2(\vec{\nabla} \times \vec{\Pi}) = (\partial_0^2 - \vec{\nabla}^2)(\vec{\nabla} \times \mathbf{\Pi}) = 0$
Hertz scalar potential:	$\mathbf{\Pi} = \vec{\Pi}\Phi$
Field relation:	$\mathbf{F} = \nabla \gamma_0 [\vec{\nabla}, \vec{\Pi}]\Phi = (I\partial_0 + \vec{\nabla}\times)(\vec{\nabla} \times \vec{\Pi})\Phi$
Vacuum equation:	$\nabla^2\Phi = (\partial_0^2 - \vec{\nabla}^2)\Phi = 0$

This invariance leads to an interesting observation. After separating the vector potential z into a transverse and longitudinal piece in a particular frame, we can define a corresponding transverse *bivector potential* (e.g., as in [68])

$$\mathbf{Z} = z_\perp \gamma_0 = \vec{A}_\perp - \vec{C}_\perp I, \tag{6.21}$$

and longitudinal *spinor potential*

$$\psi_z = z_\parallel \gamma_0 = (\phi_e - \phi_m I) + (\vec{A}_\parallel - \vec{C}_\parallel I). \tag{6.22}$$

Both of these potentials can now be easily manipulated as proper geometric objects, even though their definitions involve a particular frame γ_0. Moreover, only the spinor potential ψ_z depends on the choice of gauge according to (6.19), which makes the bivector $\mathbf{Z}$ a *gauge-invariant* potential (restricted to the γ_0 frame). This separation of the total potential into a gauge-invariant part $\mathbf{Z}$ and a gauge-dependent part ψ_z in a particular frame is known as the *Chen decomposition* [88,218], and is important for the meaningful separation of the orbital and spin parts of the angular momentum (as we shall see in Section 8.5).

According to Maxwell's vacuum equation (5.1), we can then write

$$\nabla \mathbf{F} = \nabla^2 z = \nabla^2(\mathbf{Z} + \psi_z)\gamma_0 = 0. \tag{6.23}$$

In the absence of sources, the frame-dependent factor of γ_0 simply cancels to further simplify the analysis. Moreover, since ψ_z can be changed arbitrarily by the choice of gauge, according to (6.20), we can always fix the gauge to set the longitudinal potential ψ_z to zero in the local frame (i.e., choosing the Coulomb, or radiation gauge). After fixing the gauge in this manner, we can write Maxwell's vacuum equation (5.1) solely in terms of the transverse bivector potential

$$\nabla^2 \mathbf{Z} = \nabla^2(\vec{A}_\perp - \vec{C}_\perp I) = 0. \tag{6.24}$$

This bivector potential is summarized in Table 12 along with several of its derivatives (from the next section) for reference.

6.2.1. Plane wave bivector potential

As a notable special case, we showed that the plane wave potential (6.6) must be manifestly transverse in any relative frame in Section 6.1.1. Thus we can immediately write the corresponding gauge-invariant bivector potential for a plane wave

$$\mathbf{Z} = z_\perp \gamma_0 = (Is\gamma_0) \exp[\mp (k \cdot x)I] = I \exp[\mp (k \cdot x)I]\vec{E}_0. \tag{6.25}$$

Note that this bivector potential is already essentially equivalent to the total field bivector $\mathbf{F}$. There is no remaining arbitrariness in the potential, which emphasizes that the bivector potential is a gauge-invariant representation in the γ_0 frame.

The corresponding bivector potentials for the plane wave superpositions follow in a similar way from (6.9) and (6.10)

$$\mathbf{Z} = I\vec{E}_0 \int_{-\infty}^{\infty} \frac{d\omega}{\omega}\, \alpha(\omega)\, e^{\omega(k \cdot x)I}, \tag{6.26}$$

$$= I\vec{E}_0 \int_0^{\infty} \frac{d\omega}{\omega}\, [\alpha_+(\omega)\, e^{\omega(k \cdot x)I} - \alpha_-(\omega)\, e^{-\omega(k \cdot x)I}],$$

$$\mathbf{Z} = I \iiint_{-\infty}^{\infty} \frac{d^3\vec{k}}{|\vec{k}|}\, \vec{E}_0(\vec{k})\, [\alpha_+(\vec{k})\, e^{(k \cdot x)I} - \alpha_-(\vec{k})\, e^{-(k \cdot x)I}]. \tag{6.27}$$

Table 13

Maxwell's equation with both electric and magnetic sources, as expressed in various formalisms. As with the vacuum case in Tables 10 and 11, only spacetime algebra permits the formulation as the single boxed equation involving either the bivector field $\mathbf{F}$ or the complex vector potential field $z = a_e + a_m I$, while treating both electric sources j_e and magnetic sources j_m on equal footing as a single complex source $j = j_e + j_m I$. In contrast, the differential-form approach becomes strained with a nonzero magnetic source (one-form) $\mathbf{j}_m$ [198], but simplifies to a form similar to the vacuum case in Table 10 for purely electric sources $\mathbf{j}_e$. Note that for the SI unit scalings, we use the Ampere-meter convention for magnetic charges.

<div style="text-align:center">

Maxwell's Equation with Sources

</div>

Spacetime Algebra: $\boxed{\nabla \mathbf{F} = \nabla^2 z = j}$ $\qquad \boxed{j = j_e + j_m I}$

$$\Downarrow$$

$$\nabla \cdot \mathbf{F} = \nabla^2 a_e = j_e, \qquad \nabla \wedge \mathbf{F} = \nabla^2 a_m I = j_m I$$

$$\Downarrow$$

$$\mathbf{F} = \vec{E} + \vec{B}I, \qquad j_e = (c\rho_e + \vec{J}_e)\gamma_0, \qquad j_m = (c\rho_m + \vec{J}_m)\gamma_0$$
$$a_e = (\phi_e + \vec{A})\gamma_0, \qquad a_m = (\phi_m + \vec{C})\gamma_0$$

$$\vec{\nabla} \cdot \vec{E} + (\partial_0 \vec{E} - \vec{\nabla} \times \vec{B}) = \nabla^2(\phi_e - \vec{A}) = c\rho_e - \vec{J}_e$$
$$\vec{\nabla} \cdot \vec{B} + (\partial_0 \vec{B} + \vec{\nabla} \times \vec{E}) = \nabla^2(\phi_m - \vec{C}) = c\rho_m - \vec{J}_m$$

$$\Downarrow$$

$$\vec{\nabla} \cdot \vec{E} = \nabla^2 \phi_e = c\rho_e, \qquad \partial_0 \vec{E} - \vec{\nabla} \times \vec{B} = -\nabla^2 \vec{A} = -\vec{J}_e$$
$$\vec{\nabla} \cdot \vec{B} = \nabla^2 \phi_m = c\rho_m, \qquad \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} = -\nabla^2 \vec{C} = -\vec{J}_m$$

Differential Forms: $\quad \mathbf{F} = \mathbf{F}_e + i \star \mathbf{F}_m, \quad \star d \star \mathbf{F}_e = \star d \star d a_e = -\mathbf{j}_e$
$$\star d \star \mathbf{F}_m = \star d \star d a_m = -\mathbf{j}_m$$

Tensor Components: $\quad \partial_\nu F_e^{\mu\nu} = j_e^\mu, \quad \partial_\nu F_m^{\mu\nu} = j_m^\mu, \quad F^{\mu\nu} = F_e^{\mu\nu} + i \star F_m^{\mu\nu}$

Gibbs 3-Vectors: $\quad \vec{\nabla} \cdot \vec{E} = c\rho_e, \qquad \partial_0 \vec{E} - \vec{\nabla} \times \vec{B} = -\vec{J}_e$
$$\vec{\nabla} \cdot \vec{B} = c\rho_m, \qquad \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} = -\vec{J}_m$$

SI Units : $\quad (\vec{E}, \vec{C}, \phi_e) \mapsto \sqrt{\epsilon_0}\,(\vec{E}, \vec{C}, \phi_e), \quad (\vec{B}, \vec{A}, \phi_m) \mapsto (\vec{B}, \vec{A}, \phi_m)/\sqrt{\mu_0}$
$(c \equiv 1/\sqrt{\epsilon_0 \mu_0})$ $\qquad (\vec{J}_e, \rho_e) \mapsto \sqrt{\mu_0}\,(\vec{J}_e, \rho_e), \qquad (\vec{J}_m, \rho_m) \mapsto \sqrt{\mu_0}\,(\vec{J}_m, \rho_m)/c$

CGS Units: $\quad \vec{E} \mapsto \vec{E}/\sqrt{4\pi}, \qquad \vec{B} \mapsto \vec{B}/\sqrt{4\pi}, \qquad j \mapsto \sqrt{4\pi}\,j/c$

The only appreciable difference in structure from the vector potentials is the flip in sign of the phase factors associated with each $\alpha_\pm$, which occurs since γ_0 and I anticommute. Moreover, since the same functions $\alpha_\pm$ occur here as in the vector potentials in Section 6.1.1 and the plane wave fields in Section 5.4, the same procedure can be used to canonically quantize the bivector potential directly, if desired.

6.2.2. Field correspondence

The relative fields associated with $\mathbf{Z}$ still depend upon the defining frame γ_0

$$\mathbf{F} = (\nabla \mathbf{Z})\gamma_0 = \nabla z_\perp. \tag{6.28}$$

Thus, both the transversality properties of $\mathbf{Z}$ and its relation to the field $\mathbf{F}$ depend on the frame γ_0. Nevertheless, for any frame γ_0 it is possible to construct the transverse $\mathbf{Z}$ for that frame.

Using the correspondence (6.28), the relative fields have the form [98,103–105]

$$\vec{E} = -\partial_0 \vec{A}_\perp - \vec{\nabla} \times \vec{C}_\perp, \qquad \vec{B} = -\partial_0 \vec{C}_\perp + \vec{\nabla} \times \vec{A}_\perp, \tag{6.29}$$

which can be understood as the appropriate simplification of (6.13). Interestingly, Eqs. (6.29) plus the transversality conditions $\vec{\nabla} \cdot \vec{A}_\perp = \vec{\nabla} \cdot \vec{C}_\perp = 0$ resemble the four Maxwell's equations in vacuum (5.5) that we started with [105,178]. This resemblance stems from the fact that the derivative of the bivector potential **Z** has the same structural form (4.11) that also produced Maxwell's equations (5.5).

6.3. Hertz potentials

Solving equations with vector and bivector potentials can be simpler than solving the main bivector equation (5.1), but it would be simpler still to solve equations using a scalar potential. Hence, we are motivated to make another ansatz to further reduce the vector potential z or the bivector potential **Z** by using another derivative.

A naive attempt at performing such a reduction is to assume that $z = \nabla \phi$ for some complex scalar potential ϕ. However, this attempt fails since the resulting field $\mathbf{F} = \nabla^2 \phi$ would be a scalar. Hence, we instead consider simplifying the bivector potential **Z** further.

6.3.1. Hertz bivector potential

A suitable reduction was found by Hertz [219], who (after translating to the language of spacetime algebra) exploited the fact that the commutator bracket in Eq. (3.34) for bivectors produces another bivector. That is, if we write $z = \mathbf{Z}\gamma_0 = -\gamma_0 \mathbf{Z}^\dagger$, with a bivector potential $\mathbf{Z}^\dagger = \vec{A}_\perp + \vec{C}_\perp I$, then we can make the ansatz

$$\mathbf{Z}^\dagger = -[\vec{\nabla}, \mathbf{\Pi}] = -\vec{\nabla} \times \mathbf{\Pi} I, \tag{6.30}$$

making $\mathbf{Z}^\dagger$ the dual of the relative curl of a second-order[19] bivector potential **Π**. Note the restriction to a particular relative frame γ_0 implicit in the relative gradient $\vec{\nabla} = \gamma_0 \wedge \nabla$; this is the same fixed frame used to define **Z**.

The second-order bivector potential $\mathbf{\Pi} = \vec{\Pi}_e + \vec{\Pi}_m I$ is known as a *Hertz potential* for the field, and can be decomposed into electric $\vec{\Pi}_e$ and magnetic $\vec{\Pi}_m$ relative vector parts. The resulting field equation has the form [40]

$$\mathbf{F} = \nabla z = \nabla \mathbf{Z}\gamma_0 = -\nabla \gamma_0 \mathbf{Z}^\dagger,$$
$$= \nabla \gamma_0 [\vec{\nabla}, \mathbf{\Pi}] = (\partial_0 - \vec{\nabla})(\vec{\nabla} \times \mathbf{\Pi})I = (I\partial_0 + \vec{\nabla} \times)(\vec{\nabla} \times \mathbf{\Pi}), \tag{6.31}$$

where the last equality uses the fact that the relative divergence of a curl vanishes.

Maxwell's equation in vacuum then mandates

$$\nabla \mathbf{F} = \nabla^2 z = -\gamma_0 \nabla^2 \mathbf{Z}^\dagger = -\gamma_0 \nabla^2 (\vec{\nabla} \times \mathbf{\Pi})I = 0. \tag{6.32}$$

That is, the relative curl of the Hertz potential must satisfy the d'Alembert wave equation

$$\nabla^2(\vec{\nabla} \times \mathbf{\Pi}) = (\partial_0^2 - \vec{\nabla}^2)(\vec{\nabla} \times \mathbf{\Pi}) = 0. \tag{6.33}$$

This general construction by Hertz is quite useful for applications, such as describing dipole radiation fields or rectangular waveguide modes [220]. However, the construction still involves a bivector potential, so we will simplify it further.

6.3.2. Hertz complex scalar potential

We can avoid the complication of the Hertz bivector potential $\mathbf{\Pi} = \vec{\Pi} \Phi$ by factoring it into a *complex scalar potential* Φ and a fixed reference unit-3-vector $\vec{\Pi}$ in the relative frame. This reference direction is usually chosen to have physical significance, such as the propagation axis for a paraxial beam or the dipole-radiation axis [40,221]. The Maxwell constraint on the scalar function Φ can then be written

$$\nabla^2(\vec{\nabla} \times \vec{\Pi})\Phi = (\vec{\nabla} \times \vec{\Pi})\nabla^2 \Phi = 0, \tag{6.34}$$

making it sufficient for the potential Φ to be a complex scalar solution of the d'Alembert wave equation

$$\nabla^2 \Phi = (\partial_0^2 - \vec{\nabla}^2)\Phi = 0. \tag{6.35}$$

Since this relativistic scalar wave equation is well-understood, and may be systematically solved by separation of variables in a myriad of coordinate systems (e.g., [222]), constructing solutions of Maxwell's equation from solutions of this scalar wave equation is a powerful method for generating nontrivial solutions to Maxwell's equation in vacuum [40].

Given such a wave solution Φ, an associated electromagnetic field is

$$\mathbf{F} = \nabla \gamma_0 [\vec{\nabla}, \vec{\Pi}]\Phi = (I\partial_0 + \vec{\nabla} \times)(\vec{\nabla} \times \vec{\Pi})\Phi. \tag{6.36}$$

---

[19] Curiously, we can perform this trick as many times as we like by making each relative potential the curl of another generating potential. This ability to keep increasing the number of derivatives was also noticed in [105,178]. We could also use the form $\mathbf{Z}^\dagger = \nabla \mathbf{\Pi}\gamma_0$ to mirror the field definition (6.28), which would add an extra time derivative to the curl.

As we mentioned earlier, the frame-specific differential factor $\gamma_0[\vec{\nabla}, \vec{\Pi}]$ is a consequence of fixing the frame to define the transverse potential $\mathbf{Z}$. The resulting bivector field $\mathbf{F}$ will always be a proper (frame-independent) geometric object. This Hertz potential technique has been used to great effect by Hertz [219], Whittaker [223], Debye [224], and others [40,225,226].

The use of dual-symmetric fields and spacetime algebra makes the origin of the Hertz potentials particularly transparent. In particular, their construction relies upon further decomposing a *complex* 4-vector potential z into a transverse bivector potential $\mathbf{Z}$, after which that potential can be decomposed into a second-order bivector potential $\mathbf{\Pi}$, which in turn can be decomposed into a complex scalar potential Φ. Notably, the starting vector potential z is only complex when the dual symmetry of the vacuum Maxwell equation is preserved.

7. Maxwell's equation with sources

> *The theory I propose may therefore be called a theory of the Electromagnetic Field because it has to do with the space in the neighborhood of the electric or magnetic bodies, and it may be called a Dynamical Theory, because it assumes that in the space there is matter in motion, by which the observed electromagnetic phenomena are produced.*
>
> [James Clerk Maxwell [133]]

The simplest modification to Maxwell's equation in vacuum (5.1) that we can make is to add a source field (which we summarize in Table 13). Since the derivative of a bivector field can only produce a complex vector field, this modification produces

$$\nabla \mathbf{F} = j, \tag{7.1}$$

where the complex current $j = j_e + j_m I$ acts as the appropriate source field, and admits both vector (electric) and trivector (magnetic) parts. Note that this equation is no longer intrinsically invariant under the simple phase-rotation $\mathbf{F} \mapsto \mathbf{F} \exp(\theta I)$ due to the presence of the source current j; we will clarify this important point in Section 7.4.

Taking a second derivative decouples (7.1) into the set of equations

$$\nabla^2 \mathbf{F} = (\nabla \wedge j_e) + (\nabla \wedge j_m)I, \tag{7.2}$$
$$\nabla \cdot j_e = \nabla \cdot j_m = 0. \tag{7.3}$$

The first equation (for the bivector part) indicates that $\mathbf{F}$ satisfies a *wave equation* with a complex source equal to the curl of j. The second equation (for the scalar part) indicates the source satisfies a (charge-current) *continuity equation*.

7.1. Relative frame form

After breaking the source $j = j_e + j_m I$ into vector and trivector parts, Eq. (7.1) implies the two independent equations

$$\nabla \cdot \mathbf{F} = j_e, \qquad \nabla \wedge \mathbf{F} = j_m I, \tag{7.4}$$

which, under the spacetime splits $\mathbf{F} = \vec{E} + \vec{B}I$, $j_e = \gamma_0(c\rho_e - \vec{J}_e)$, and $j_m = \gamma_0(c\rho_m - \vec{J}_m)$, expand to the four equations

$$
\begin{aligned}
\vec{\nabla} \cdot \vec{E} &= c\rho_e, & \partial_0 \vec{E} - \vec{\nabla} \times \vec{B} &= -\vec{J}_e, \\
\vec{\nabla} \cdot \vec{B} &= c\rho_m, & \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} &= -\vec{J}_m,
\end{aligned}
\tag{7.5}
$$

according to (4.11). These are the familiar Maxwell's equations that include both electric and magnetic charges [101,102, 193,197–213]. Similarly, the continuity equation for the sources (7.3) produces the two equations

$$c\partial_0 \rho_e + \vec{\nabla} \cdot \vec{J}_e = 0, \qquad c\partial_0 \rho_m + \vec{\nabla} \cdot \vec{J}_m = 0, \tag{7.6}$$

which each have the familiar form [155].

7.2. Vector potentials revisited

If $\mathbf{F}$ is generated by a vector potential, $\mathbf{F} = \nabla z = \nabla(a_e + a_m I)$, then Maxwell's equation (7.1) is equivalent to a complex vector wave equation with a source

$$\nabla^2 z = j. \tag{7.7}$$

In this form it becomes clear that the electric potential a_e couples directly to the electric source j_e, while the magnetic potential a_m couples directly to the magnetic source j_m. Hence the total complex field $\mathbf{F} = (\nabla \wedge a_e) + (\nabla \wedge a_m)I = \mathbf{F}_e + \mathbf{F}_m I$ can be equivalently constructed by solving the two independent equations

$$\nabla \mathbf{F}_e = \nabla^2 a_e = j_e, \qquad \nabla \mathbf{F}_m = \nabla^2 a_m = j_m \tag{7.8}$$

that each have the same form as (7.1), but involve only non-complex vector fields and bivectors of fixed signature. Evidently the constituent fields $\mathbf{F}_e$ and $\mathbf{F}_m$ are *noninteracting, completely independent, and distinguished only by how they couple to distinct types of electromagnetic charge.* In particular, note how the e and m subscripts for all quantities are consistently matched; for traditional electromagnetism that considers only electric charges, the "magnetic" terms (with subscripts m) can thus be neglected.

In terms of a relative frame with $a_e = (\phi_e + \vec{A})\gamma_0$ and $j_e = (c\rho_e + \vec{J}_e)\gamma_0$, the wave equation $\nabla^2 a_e = j_e$ expands to

$$(\partial_0^2 - \vec{\nabla}^2)(\phi_e + \vec{A})\gamma_0 = (c\rho_e + \vec{J}_e)\gamma_0, \tag{7.9}$$

which further splits into the two wave equations [155]

$$(\partial_0^2 - \vec{\nabla}^2)\phi_e = c\rho_e, \qquad (\partial_0^2 - \vec{\nabla}^2)\vec{A} = \vec{J}_e. \tag{7.10}$$

These two equations are implicitly coupled by the Lorenz–FitzGerald condition $\nabla \cdot a_e = \partial_0 \phi_e + \vec{\nabla} \cdot \vec{A} = 0$, which can be applied to the first equation in order to produce the modified form

$$\vec{\nabla}^2 \phi_e + \partial_0(\vec{\nabla} \cdot \vec{A}) = -c\rho_e, \qquad (\partial_0^2 - \vec{\nabla}^2)\vec{A} = \vec{J}_e. \tag{7.11}$$

A similar split also applies to the equation $\nabla^2 a_m = j_m$ involving the magnetic potential.

7.3. Lorentz force

Similar to the appearance of Maxwell's equation (7.1), the corresponding Lorentz force law arises quite naturally from the structure of spacetime itself. To see this, consider what happens when the field $\mathbf{F} = \vec{E} + \vec{B}I$ is contracted with a 4-vector $w = (w_0 + \vec{w})\gamma_0$. According to (3.10), we have the general structure

$$\mathbf{F} \cdot w = \frac{1}{2}[\mathbf{F}w - w\mathbf{F}],$$
$$= \frac{1}{2}[(\vec{E} + \vec{B}I)(w_0 + \vec{w}) + (w_0 + \vec{w})(\vec{E} - \vec{B}I)]\gamma_0,$$
$$= (\vec{E} \cdot \vec{w})\gamma_0 + [w_0\vec{E} + \vec{w} \times \vec{B}]\gamma_0. \tag{7.12}$$

Now consider what happens if w is a proper velocity vector $w = dx/d\tau$, which implies $w^2 = c^2$. Inserting the spacetime split of x yields $dx/d\tau = d/d\tau[ct + \vec{x}]\gamma_0 = [d(ct)/d\tau + d\vec{x}/d\tau]\gamma_0$. Denoting $dt/d\tau = \gamma$ as the time dilation factor, we see that $d\vec{x}/d\tau = \gamma d\vec{x}/dt = \gamma \vec{v}$ is the scaled relative velocity $\vec{v}$. From $w^2 = c^2 = w_0^2 - |\vec{w}|^2 = \gamma^2(c^2 - |\vec{v}|^2)$ we find the standard formula $\gamma = (1 - |\vec{v}/c|^2)^{-1/2}$. Hence, we have the relative components

$$w_0 = \gamma c = \frac{c}{\sqrt{1 - |\vec{v}/c|^2}}, \qquad \vec{w} = \gamma \vec{v}, \tag{7.13}$$

as well as the elegant and useful relations [9]

$$\gamma = \frac{dt}{d\tau} = \frac{w}{c} \cdot \gamma_0, \qquad \frac{d\vec{x}}{d\tau} = w \wedge \gamma_0, \qquad \frac{\vec{v}}{c} = \frac{w \wedge \gamma_0}{w \cdot \gamma_0}. \tag{7.14}$$

Inserting (7.13) into (7.12) produces the suggestive relation

$$\mathbf{F} \cdot w = \gamma c \left[\vec{E} \cdot \vec{v}\right]\gamma_0 + \gamma \left[c\vec{E} + \vec{v} \times \vec{B}\right]\gamma_0. \tag{7.15}$$

Inverting factors and multiplying by a scalar charge q_0 then produces

$$[\mathbf{F} \cdot (q_0 w)]\frac{d\tau}{dt}\gamma_0 = q_0 c\, \vec{E} \cdot \vec{v} + q_0 \left[c\vec{E} + \vec{v} \times \vec{B}\right]. \tag{7.16}$$

The vector part on the right hand side is precisely[20] the *Lorentz force* $d\vec{p}/dt$ in a relative frame with $\vec{p} = m\vec{v}$, while the scalar part on the right hand side is precisely the corresponding rate of work $d\mathcal{E}/d(ct)$ with $\mathcal{E} = \gamma mc^2$. It follows that the simple contraction of $\mathbf{F}$ and a 4-vector current $q_0 w$ produces the proper Lorentz force

$$\frac{dp}{d\tau} = \mathbf{F} \cdot (q_0 w), \tag{7.17}$$

where $p = (\mathcal{E}/c + \vec{p})\gamma_0$ is the proper 4-vector energy–momentum.

We can expand upon this result by considering what happens if the vector current $q_0 w$ is allowed to be a *complex* vector current $j = q_0 w \exp(\theta I)$ similar to (7.1). Since $\mathbf{F} \cdot (wI) = (\mathbf{F} \wedge w)I$ is the proper contraction that produces a vector, we must

---

[20] The choice of units for $\vec{E}$, $\vec{B}$, and q_0 will induce appropriate scaling factors (see Table 14).

Table 14
The Lorentz Force Law, as expressed in spacetime algebra. We show the proper (frame-independent) formulation that supports both electric and magnetic charges, combined as a single complex charge $q = q_e + q_m I$. We also show the expansion of this force law in terms of a particular reference frame γ_0. Notably, the proper and relative velocities are related by elegant contraction formulas in spacetime algebra.

<div align="center">

Lorentz Force Law

</div>

Proper Force Law:	$\boxed{\dfrac{dp}{d\tau} = \langle \mathbf{F} j \rangle_1} = \frac{1}{2} [\mathbf{F} j + (\mathbf{F} j)^\sim] = q_e\, \mathbf{F} \cdot w + q_m\, (\mathbf{F} \wedge w) I$
	$p \equiv mw, \quad j \equiv qw, \quad q \equiv q_0 e^{\theta I} = q_e + q_m I$
Proper Velocity:	$w \equiv \dfrac{dx}{d\tau} = \gamma(\mathrm{c} + \vec{v})\gamma_0, \quad w^2 = \mathrm{c}^2$
	$\gamma\,\mathrm{c} \equiv \dfrac{dt}{d\tau} = w \cdot \gamma_0 = \mathrm{c}(1 - |\vec{v}/\mathrm{c}|^2)^{-1/2}$
	$\gamma\,\vec{v} \equiv \dfrac{d\vec{x}}{d\tau} = w \wedge \gamma_0$
Relative Force Law:	$\dfrac{dp}{d\tau}\dfrac{d\tau}{dt}\gamma_0 = \dfrac{d\mathcal{E}}{d(\mathrm{c}t)} + \dfrac{d\vec{p}}{dt}$
	$\mathcal{E} \equiv \gamma m \mathrm{c}^2, \quad \vec{p} \equiv m\vec{v}, \quad q_e = q_0 \cos\theta, \quad q_m = q_0 \sin\theta$
	$\dfrac{d\mathcal{E}}{d(\mathrm{c}t)} = q_e \mathrm{c}\, \vec{E} \cdot \vec{v} + q_m \mathrm{c}\, \vec{B} \cdot \vec{v}$
	$\dfrac{d\vec{p}}{dt} = q_e \mathrm{c}\, \vec{E} + q_m \mathrm{c}\, \vec{B} + q_e\, \vec{v} \times \vec{B} - q_m\, \vec{v} \times \vec{E}$
Relative Velocity:	$\dfrac{\vec{v}}{\mathrm{c}} \equiv \dfrac{w \wedge \gamma_0}{w \cdot \gamma_0}, \quad \gamma \equiv \dfrac{w}{\mathrm{c}} \cdot \gamma_0$
SI Units: $(\mathrm{c} \equiv 1/\sqrt{\epsilon_0 \mu_0})$	$\vec{E} \mapsto \sqrt{\epsilon_0}\, \vec{E}, \quad \vec{B} \mapsto \vec{B}/\sqrt{\mu_0}, \quad q_e \mapsto \sqrt{\mu_0}\, q_e, \quad q_m \mapsto \sqrt{\mu_0}\, q_m/\mathrm{c}$
CGS Units:	$\vec{E} \mapsto \vec{E}/\sqrt{4\pi}, \quad \vec{B} \mapsto \vec{B}/\sqrt{4\pi}, \quad q_0 \mapsto \sqrt{4\pi}\, q_0/\mathrm{c}$

also consider the wedge of the bivector field with the velocity w

$$\mathbf{F} \wedge w = \frac{1}{2}[\mathbf{F}w + w\mathbf{F}],$$
$$= \frac{1}{2}[(\vec{E} + \vec{B}I)(w_0 + \vec{w}) - (w_0 + \vec{w})(\vec{E} - \vec{B}I)]\gamma_0,$$
$$= [\vec{B} \cdot \vec{w}]\gamma_0 I^{-1} + [w_0 \vec{B} - \vec{w} \times \vec{E}]\gamma_0 I^{-1}. \tag{7.18}$$

The total contraction of $\mathbf{F}$ with $j = q_0 w \exp(\theta I) = j_e + j_m I$ then has the form

$$\frac{dp}{d\tau} = \langle \mathbf{F} j \rangle_1 = \mathbf{F} \cdot j_e + (\mathbf{F} \wedge j_m)I = \frac{1}{2}[\mathbf{F} j + (\mathbf{F} j)^\sim], \tag{7.19}$$

where the grade-1-projection makes the expression particularly elegant, while the last equality follows from the reversion properties $\widetilde{\mathbf{F}} = -\mathbf{F}$ and $\tilde{j} = j^*$ (recall Sections 3.3, 3.3.2 and 3.5.3). We summarize this proper Lorentz force that admits a complex source j in Table 14 for reference.

We must be careful to be fully consistent, however. That is, the physical momentum p is a vector quantity that can be decomposed into a mass-scaled velocity $p = mw$ (at least for classical point particles). Thus, if we make the current complex $q_0 w \mapsto q_0 w \exp(\theta I) = j$, as in (7.19), then the *charge* must become complex

$$q = q_0 \exp(\theta I) = q_e + q_m I, \tag{7.20}$$

leading to distinct electric and magnetic parts of the charge

$$q_e = q_0 \cos \theta, \qquad q_m = q_0 \sin \theta, \tag{7.21}$$

such that the observed charges have a well-defined ratio $q_m/q_e = \tan \theta$. (We will clarify the intimate connection between this ratio and the dual symmetry of the field in the next section.)

Inserting the relation $j = w[q_0 \exp(\theta I)]$ into (7.19) with $w = dx/d\tau$ as before yields the relative expression

$$\frac{dp}{d\tau} \frac{d\tau}{dt} \gamma_0 = \frac{d\mathcal{E}}{d(ct)} + \frac{d\vec{p}}{dt}, \tag{7.22}$$

where $\mathcal{E} = \gamma mc^2, \vec{p} = m\vec{v}$, and where

$$\frac{d\mathcal{E}}{d(ct)} = q_e c \vec{E} \cdot \vec{v} + q_m c \vec{B} \cdot \vec{v}, \tag{7.23a}$$

$$\frac{d\vec{p}}{dt} = q_e c \vec{E} + q_m c \vec{B} + q_e \vec{v} \times \vec{B} - q_m \vec{v} \times \vec{E}. \tag{7.23b}$$

After restoring units, these equations are precisely the correction to the Lorentz force law that properly accounts for the addition of magnetic monopoles [101,102,193,197–213]. This monopole description arises directly from the complexity of the current j and asserting that the total field **F** should couple to this complex current. We summarize these relative expressions in Table 14 for reference, along with both the SI and CGS unit conventions.

7.4. Dual symmetry revisited

We are now in a position to consider the meaning of dual symmetry in the presence of an electromagnetic source current j. The two equations that we must examine are Maxwell's equation (7.1) and the Lorentz force law (7.19), which we reproduce here for convenience in terms of both the field **F** and the vector potential z

$$\nabla \mathbf{F} = \nabla^2 z = j, \qquad \frac{dp}{d\tau} = \langle \mathbf{F} j \rangle_1 = \langle (\nabla z) j \rangle_1. \tag{7.24}$$

Keep in mind that $j = qw$ and $p = mw$ for classical point particles, where $w = dx/d\tau$ is the proper velocity.

It is now apparent that both these equations are invariant under a simultaneous phase rotation of both the vector potential z and the current j, using the spinor $\psi = \exp(-\theta I/2)$ from (5.10)

$$z \mapsto \psi z \psi^*, \qquad j \mapsto \psi j \psi^*. \tag{7.25}$$

Specifically, note that since $\psi z \psi^* = z \exp(\theta I)$ and $\psi j \psi^* = j \exp(\theta I)$, Maxwell's equation becomes

$$(\nabla^2 z) \exp(\theta I) = j \exp(\theta I), \tag{7.26}$$

so the phase factors cancel. Similarly, the Lorentz force law is also invariant

$$\langle (\nabla z e^{\theta I}) j e^{\theta I} \rangle_1 = \langle (\nabla z) j e^{-\theta I} e^{\theta I} \rangle_1 = \langle (\nabla z) j \rangle_1. \tag{7.27}$$

Thus, while it may initially appear that the dual symmetry of Maxwell's equation (7.1) is broken in the presence of a source current j, this is only strictly true if the description of the current is left unchanged. An equivalent physical picture is obtained if the current is also changed (via the same phase rotation of the coupling charge q). As such, choosing electric charges to be the primary type of charge is an arbitrary convention. *Regardless of how one describes the fundamental charges, the same forces (and thus the same physics) will still be described, as long as the fields are correspondingly transformed* [213].

In addition to this equivalence of representation, there is another important subtlety that appears when we consider a field solution of Maxwell's equation $\nabla \mathbf{F} = j$. Specifically, any solution to this equation will be the combination of a homogeneous solution $\mathbf{F}_0$ of the vacuum equation $\nabla \mathbf{F}_0 = 0$ and a particular solution that includes the source j. We can consider these two independent solutions to correspond to distinct vector potentials $z = z_0 + z_s$ that satisfy distinct equations of motion

$$\nabla^2 z_0 = 0, \qquad \nabla^2 z_s = j. \tag{7.28}$$

However, as noted above, we can always rotate the source j to become purely *electric* j_e using some phase rotation $\psi = \exp(-\theta I/2)$

$$\nabla^2 (\psi z_s \psi^*) = \psi j \psi^* = j_e. \tag{7.29}$$

After this rotation, there is no magnetic source, so the trivector part of the rotated potential $\langle \psi z_s \psi^* \rangle_3$ must satisfy the vacuum Maxwell equation, and therefore should be absorbed into the homogeneous equation as part of the vacuum potential
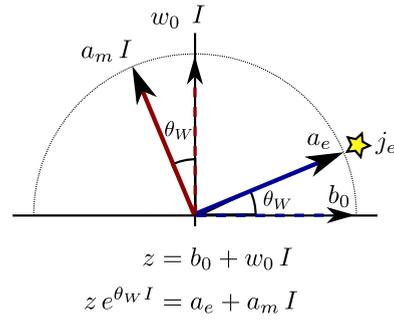
Fig. 7. Breaking dual symmetry. The neutral complex vector potential $z = b_0 + w_0 I$ that includes the neutral b_0 and w_0 bosons in the electroweak theory is dual-symmetric in vacuum, meaning that it is invariant under any global phase rotation $z \mapsto z \exp(\theta I)$. The Brout–Englert–Higgs–Guralnik–Hagen–Kibble (BEHGHK, or Higgs) mechanism fixes a particular (Weinberg) angle θ_W that couples the field to a purely electric source, breaking this dual symmetry. The vector part $a_e = \langle z \exp(\theta_W I) \rangle_1$ of the doublet becomes the electric vector potential for the photon (γ) that couples to vector electric source j_e. The pseudovector part $a_m I = \langle z \exp(\theta_W I) \rangle_3$ acquires mass from the Higgs mechanism to become the neutral (Z_0) boson of the weak force, which is the analogy to the magnetic vector potential that appears in the dual-symmetric formalism.

z_0. That is, the rotated source potential is a *pure (electric) vector* potential $a_{s,e}$, while the homogeneous vacuum potential will still be generally *complex*

$$\psi z_s \psi^* = a_{s,e}, \qquad \psi z_0 \psi^* = a_{0,e} + a_{0,m} I. \tag{7.30}$$

Thus, there is a profound difference between the independent source and vacuum parts of a field solution. The source piece must have a description as a pure vector potential matched with an electric source; this pair can then be jointly rotated into equivalent descriptions with both electric and magnetic sources as desired. In contrast, the vacuum piece is always manifestly dual-symmetric and thus admits an irreducibly complex vector potential description (such that no phase rotation exists that will make the vacuum potential a pure vector). Notably, this decomposition of the potential into distinct vacuum and matter parts is important in treatments of the quantized electromagnetic field [39]. We will return to this decomposition in Section 8.3, 8.9.2 and Fig. 7.

7.5. Fundamental conservation laws

Before turning our attention to macroscopic fields, we first provide a brief but enlightening derivation of the symmetric energy–momentum stress tensor,[21] as well as its associated angular momentum tensor [31,127]. These derivations will impose general constraints on the form of any local modifications to these tensors, which will become important in their Lagrangian derivation via Noether's theorem in Section 8.

7.5.1. Energy–momentum stress tensor

Conceptually, the energy–momentum tensor is a function $\underline{T}(b)$ that maps a constant vector direction b into the *flux* of energy–momentum that is passing through the hypersurface orthogonal to b. This flux of energy–momentum should satisfy a continuity relation

$$b \cdot \frac{dp}{d\tau} + \nabla \cdot \underline{T}(b) = 0, \tag{7.31}$$

stating that the rate of change of the energy–momentum (the Lorentz force (7.19)) projected along the direction of b should correspond to the divergence of the energy–momentum flux through the surface orthogonal to b. (This statement is analogous to how the current continuity equation $c\partial_0 \rho + \vec{\nabla} \cdot \vec{J} = 0$ from (7.3) relates the rate of change of the charge density to the divergence of the charge flux, or current.)

We can simplify (7.31) using the *adjoint* tensor[22] $\overline{T}$ that corresponds to the tensor $\underline{T}$ according to the vector dot product relation $d \cdot \underline{T}(b) = \overline{T}(d) \cdot b$. Applying this adjoint relation to (7.31) yields [9]

$$b \cdot \left[\frac{dp}{d\tau} + \overline{T}(\nabla) \right] = 0, \tag{7.32}$$

keeping in mind that ∇ still differentiates the adjoint tensor $\overline{T}$. Since b is arbitrary, we can dispense with it to find a direct constraint equation for the adjoint, in which we can substitute the Lorentz force law (7.19) to find

$$\overline{T}(\nabla) = -\frac{dp}{d\tau} = -\langle \mathbf{F}j \rangle_1. \tag{7.33}$$

---

[21] This symmetric tensor is also known as the Belinfante–Rosenfeld stress–energy tensor, or the kinetic stress–energy tensor [31,68,88,127].

[22] The underbar and overbar notation for a tensor and its adjoint is a useful notational device [9], so we adopt it here.

Table 15
The symmetric energy–momentum tensor and its associated angular momentum tensor, shown in both spacetime algebra and tensor component notation. Notably, spacetime algebra emphasizes that the energy–momentum tensor $\overline{T}_{\text{sym}}$ is a simple quadratic form of the field $\mathbf{F}$, which maps a vector direction b into the energy–momentum flux passing through a surface orthogonal to b, exactly as the energy–momentum current for the Dirac equation of the electron [9,87,96]. It is a symmetric tensor, meaning it equals its adjoint $\underline{T}_{\text{sym}}$, and is easily expanded algebraically into the usual expressions containing the energy-density of the field ε, the Poynting vector $\vec{P}$, and the Maxwell stress tensor. Similarly, the associated angular momentum tensor $\overline{M}_{\text{sym}}$ is a function that maps a vector direction b into the angular momentum flux through a surface orthogonal to b, and expands to the usual boost-momentum $\vec{N}$ and rotational-angular-momentum $\vec{J}$ vectors for the field. The adjoint of this tensor $\underline{M}_{\text{sym}}$ is a function that maps a plane of rotation (bivector) $\mathbf{R}$ into an energy–momentum flux.

Symmetric Energy-Momentum Tensor

Spacetime Algebra:
$$\boxed{\overline{T}_{\text{sym}}(b) = \underline{T}_{\text{sym}}(b) = \frac{1}{2}\mathbf{F}b\widetilde{\mathbf{F}}}$$

$$\Downarrow$$

$$\overline{T}_{\text{sym}}(\gamma_0) = \frac{1}{2}\mathbf{F}\gamma_0\widetilde{\mathbf{F}} = \frac{1}{2}\mathbf{F}\mathbf{F}^\dagger\gamma_0 = (\varepsilon + \vec{P})\gamma_0$$

$$\overline{T}_{\text{sym}}(\gamma_i) = \frac{1}{2}\mathbf{F}\vec{\sigma}_i\mathbf{F}^\dagger\gamma_0 = [\vec{P}\cdot\vec{\sigma}_i + \vec{E}(\vec{E}\cdot\vec{\sigma}_i) + \vec{B}(\vec{B}\cdot\vec{\sigma}_i) - \varepsilon\,\vec{\sigma}_i]\gamma_0$$

$$\varepsilon = \frac{1}{2}\left(|\vec{E}|^2 + |\vec{B}|^2\right), \quad \vec{P} = \vec{E}\times\vec{B}$$

Tensor Components:
$$T_{\text{sym}}^{\mu\nu} = T_{\text{sym}}^{\nu\mu} = \sum_\alpha F^{\mu\alpha}F_\alpha{}^\nu - \frac{1}{4}\eta^{\mu\nu}\sum_{\alpha,\beta}F^{\alpha\beta}F_{\beta\alpha}$$

Associated Angular Momentum Tensor

Spacetime Algebra:
$$\boxed{\overline{M}_{\text{sym}}(b) = x \wedge \overline{T}_{\text{sym}}(b)} \qquad \boxed{\underline{M}_{\text{sym}}(\mathbf{R}) = \underline{T}_{\text{sym}}(\mathbf{R}\cdot x)}$$

$$\Downarrow$$

$$\overline{M}_{\text{sym}}(\gamma_0) = -\vec{N} + \vec{J}I^{-1},$$

$$\vec{N} = (\text{c}t)\vec{P} - \varepsilon\,\vec{x}, \quad \vec{J} = \vec{x}\times\vec{P}$$

Tensor Components:
$$M_{\text{sym}}^{\alpha\beta\gamma} = x^\alpha\,T_{\text{sym}}^{\beta\gamma} - x^\beta\,T_{\text{sym}}^{\alpha\gamma}$$

After substituting Maxwell's equation (7.1) for the source j, we obtain

$$\overline{T}(\nabla) = -\frac{1}{2}\left[\mathbf{F}(\nabla\mathbf{F}) + (\widetilde{\mathbf{F}}\nabla)\widetilde{\mathbf{F}}\right] = -\frac{1}{2}\mathbf{F}\nabla\mathbf{F}. \tag{7.34}$$

Therefore, we can infer that the adjoint of the energy–momentum stress tensor must have the elegant quadratic form

$$\overline{T}_{\text{sym}}(b) = -\frac{1}{2}\mathbf{F}b\mathbf{F} = \frac{1}{2}\mathbf{F}b\widetilde{\mathbf{F}} = \frac{1}{2}\mathbf{f}b\widetilde{\mathbf{f}}, \tag{7.35}$$

which is the simplest dual-symmetric form that involves the field bivector $\mathbf{F} = \mathbf{f}\exp(\varphi I)$. This form is also *symmetric* (in the sense of the adjoint $\overline{T}_{\text{sym}} = \underline{T}_{\text{sym}}$) since

$$\overline{T}_{\text{sym}}(d) \cdot b = \langle \overline{T}_{\text{sym}}(d)b\rangle_0 = -\langle \mathbf{F}d\mathbf{F}b\rangle_0 = -\langle d\mathbf{F}b\mathbf{F}\rangle_0 = d\cdot\overline{T}_{\text{sym}}(b) \tag{7.36}$$

by the cyclic property of the scalar projection (3.8). Any proposed modification to this fundamental symmetric tensor must have a vanishing divergence, according to the general continuity equation (7.31). This symmetric tensor is summarized for reference in Table 15.

To verify that this symmetric energy–momentum stress tensor gives us the results we expect, we compute the energy–momentum flux through the surface perpendicular to γ_0 (restoring the SI unit scaling factors from (5.6)),

$$
\begin{aligned}
\overline{T}_{\mathrm{sym}}(\gamma_0) &= \frac{1}{2}\mathbf{F}\gamma_0\widetilde{\mathbf{F}} = \frac{1}{2}\mathbf{F}\mathbf{F}^\dagger\gamma_0, \\
&= \frac{1}{2}\left[\vec{E}+\vec{B}I\right]\left[\vec{E}-\vec{B}I\right]\gamma_0, \\
&= \left[\frac{1}{2}\left(|\vec{E}|^2+|\vec{B}|^2\right)+\vec{E}\times\vec{B}\right]\gamma_0 \equiv (\varepsilon + \vec{P})\gamma_0.
\end{aligned}
\tag{7.37}
$$

The first term is the energy density ε of the field, which is relative to a particular frame γ_0, as we anticipated in Section 5.4. The second term is the Poynting vector $\vec{P}$ that gives the corresponding relative field momentum [9]. The two together construct the proper field energy–momentum $\underline{T}(\gamma_0) = (\varepsilon + \vec{P})\gamma_0$ that will be measured by an observer traveling along a worldline pointing in the γ_0 direction.

As another example, we can compute the energy–momentum flux through a surface perpendicular to γ_i ($i = 1, 2, 3$) using $\gamma_i = \gamma_i\gamma_0^2 = \vec{\sigma}_i\gamma_0$,

$$
\begin{aligned}
\overline{T}_{\mathrm{sym}}(\gamma_i) &= \frac{1}{2}\mathbf{F}\gamma_i\widetilde{\mathbf{F}} = \frac{1}{2}\mathbf{F}\vec{\sigma}_i\mathbf{F}^\dagger\gamma_0, \\
&= \frac{1}{2}\left[\vec{E}+\vec{B}I\right]\vec{\sigma}_i\left[\vec{E}-\vec{B}I\right]\gamma_0, \\
&= \left[\frac{1}{2}\left(\vec{E}\vec{\sigma}_i\vec{E}+\vec{B}\vec{\sigma}_i\vec{B}\right)+\frac{1}{2}\left(\vec{E}\vec{\sigma}_i\vec{B}-\vec{B}\vec{\sigma}_i\vec{E}\right)I^{-1}\right]\gamma_0.
\end{aligned}
\tag{7.38}
$$

This form emphasizes the bilinear nature of the tensor with respect to the fields. The vector products can be expanded into a more familiar form by using the identity

$$
\vec{E}\vec{\sigma}_i\vec{B} = \vec{B}(\vec{E}\cdot\vec{\sigma}_i)+\vec{E}(\vec{B}\cdot\vec{\sigma}_i)-(\vec{E}\cdot\vec{B})\vec{\sigma}_i-(\vec{E}\times\vec{B})\cdot\vec{\sigma}_i I,
\tag{7.39}
$$

to produce

$$
\overline{T}_{\mathrm{sym}}(\gamma_i) = -\left[\vec{E}\times\vec{B}\cdot\vec{\sigma}_i\right]\gamma_0 + \left[\vec{E}(\vec{E}\cdot\vec{\sigma}_i)+\vec{B}(\vec{B}\cdot\vec{\sigma}_i)-\frac{1}{2}\left(|\vec{E}|^2+|\vec{B}|^2\right)\vec{\sigma}_i\right]\gamma_0.
\tag{7.40}
$$

The first term is the projection of the Poynting vector $\vec{P}$ onto the direction $\vec{\sigma}_i$. The second term is the contraction of the Maxwell stress tensor along the direction $\vec{\sigma}_i$.

7.5.2. Angular momentum tensor

Conceptually, angular momentum is a bivector that indicates a plane of rotation. The rotation in question is that of a coordinate vector as it rotates to track movement associated with a linear momentum. For a classical point particle, the plane segment containing this rotation is produced explicitly by a wedge product $\mathbf{M} = x \wedge p$ between a coordinate 4-vector x and a 4-momentum p [189]. Indeed, after introducing the spacetime splits $p = (\mathcal{E}/\mathrm{c}+\vec{p})\gamma_0$, $x = (\mathrm{c}t+\vec{x})\gamma_0$ we have

$$
x \wedge p = \frac{xp-px}{2} = [\mathcal{E}\vec{x}/\mathrm{c}-(\mathrm{c}t)\vec{p}] + \vec{x}\times\vec{p}I^{-1} \equiv -\vec{N}+\vec{L}I^{-1}.
\tag{7.41}
$$

The second term $\vec{L}I^{-1}$ of (7.41) involves I, and is part of the angular momentum that agrees with the usual notion of the relative angular momentum as the cross product $\vec{L} = \vec{x}\times\vec{p}$. Within spacetime algebra, however, we see that this cross product is the Hodge-dual of the proper angular momentum bivector $\vec{L}I^{-1}$ that generates *spatial rotations* as a Lorentz transformation, just as discussed in (3.55) of Section 3.7. In contrast, the first term $-\vec{N}$ of (7.41) is a bivector of opposite signature that generates *boost rotations*, as discussed in (3.57) of Section 3.7. The total invariant notion of angular momentum in spacetime includes both spatial rotation and boost components, since the 4-dimensional rotations in spacetime are Lorentz transformations that can be of either type.

With this intuition, the angular momentum tensor[23] for the electromagnetic field is straightforward to obtain directly from the adjoint symmetric energy–momentum tensor $\overline{T}_{\mathrm{sym}}(b)$. As seen in the last section, this latter tensor produces an energy–momentum density associated with a particular direction b. Hence, taking the wedge product of that energy–momentum density will produce its associated angular momentum density

$$
\overline{M}_{\mathrm{sym}}(b) = x \wedge \overline{T}_{\mathrm{sym}}(b).
\tag{7.42}
$$

---

[23] Specifically, this is the Belinfante–Rosenfeld angular momentum tensor [31,68,88,127].

As with the energy–momentum tensor, the angular momentum tensor $\overline{M}_{\mathrm{sym}}$ is a function that maps a constant vector direction b into the angular momentum flux passing through the hypersurface orthogonal to b. This tensor is also summarized for reference in Table 15.

Applying the angular momentum tensor to the timelike direction γ_0 produces the intuitive result

$$\overline{M}_{\mathrm{sym}}(\gamma_0) = x \wedge \overline{T}_{\mathrm{sym}}(\gamma_0) = [\varepsilon \vec{x} - (ct)\vec{P}] + \vec{x} \times \vec{P}I^{-1} \equiv -\vec{N} + \vec{J}I^{-1}, \tag{7.43}$$

with the energy-density and Poynting vector as in (7.41), and where $\vec{N}$ is the *boost angular momentum* and $\vec{J}$ is the *rotational angular momentum* of the electromagnetic field. According to commonly accepted notations, we denote the rotational angular momentum of the field as $\vec{J}$ (not to be confused with the charge current!), because it includes both the orbital ($\vec{L}$) and spin ($\vec{S}$) parts [48–51] (see Section 8.7).

Just as the energy–momentum tensor $\overline{T}_{\mathrm{sym}}$ has an adjoint tensor $\underline{T}_{\mathrm{sym}}$, the angular momentum tensor $\overline{M}_{\mathrm{sym}}$ also has an adjoint. We can compute this adjoint by contracting the tensor with any pure bivector $\mathbf{R}$, which should be understood as the plane segment for a rotation (i.e., generating a Lorentz transformation as in Section 3.7)

$$\mathbf{R} \cdot \overline{M}_{\mathrm{sym}}(b) = \langle \mathbf{R}x\overline{T}_{\mathrm{sym}}(b)\rangle_0 = \langle \underline{T}_{\mathrm{sym}}(\langle \mathbf{R}x\rangle_1)b\rangle_0 = \underline{T}_{\mathrm{sym}}(\mathbf{R} \cdot x) \cdot b, \tag{7.44}$$

yielding the adjoint tensor

$$\underline{M}_{\mathrm{sym}}(\mathbf{R}) = \underline{T}_{\mathrm{sym}}(\mathbf{R} \cdot x). \tag{7.45}$$

This adjoint tensor is a function that takes the plane of rotation $\mathbf{R}$, contracts it with the coordinate direction x to find the orthogonal direction b such that $\mathbf{R} = b \wedge x_{\parallel} = bx_{\parallel}$ (see the discussion in Section 3.3.1), and then returns the energy–momentum flux $\underline{T}_{\mathrm{sym}}(b)$ through the hypersurface perpendicular to that direction.

As an example, applying $\underline{M}_{\mathrm{sym}}$ to a bivector $\mathbf{R} = \gamma_i\gamma_0 = \vec{\sigma}_i$ that generates boost rotations produces

$$\underline{M}_{\mathrm{sym}}(\vec{\sigma}_i) = \underline{T}_{\mathrm{sym}}(\vec{\sigma}_i \cdot x) = (ct)\,\underline{T}_{\mathrm{sym}}(\gamma_i) - x_i\,\underline{T}_{\mathrm{sym}}(\gamma_0), \tag{7.46}$$

which is the correct energy–momentum flux through the hyperplane perpendicular to the part of the coordinate vector x that lies in the boost rotation plane $\mathbf{R} = \vec{\sigma}_i$.

7.6. Macroscopic fields

The preceding discussion has assumed the free propagation of a (*micro*scopic) field $\mathbf{F}$ through vacuum. In the presence of a medium, however, it is useful to partition the total field into a *macro*scopic distinction between a *free* part $\mathbf{F}_f$ that is coupled to freely moving charges in the conduction bands of the medium, and a *bound* part $\mathbf{F}_b$ that is coupled to the bound charges held rigid by the lattice structure of the medium [155,227]. Such a partitioning of the fields and sources must have the general form

$$\mathbf{F} = \mathbf{F}_f + \mathbf{F}_b, \qquad j = j_f + j_b. \tag{7.47}$$

To see how the macroscopic fields in (7.47) are expressed in a relative frame, we consider the spacetime splits (using conventional (SI) units)

$$\mathbf{F}_f = \vec{D}/\sqrt{\epsilon_0} + \sqrt{\mu_0}\vec{H}I, \qquad \mathbf{F}_b = -\vec{\mathcal{P}}/\sqrt{\epsilon_0} + \sqrt{\mu_0}\vec{\mathcal{M}}I. \tag{7.48}$$

Here $\vec{D}$ is the usual electric *displacement*, $\vec{H}$ is the macroscopic magnetic field, $\vec{\mathcal{P}}$ is the *electric polarization* field of the medium, and $\vec{\mathcal{M}}$ is the *magnetization* field of the medium.[24] From the definition (7.47) we see that the total electromagnetic field $\mathbf{F} = \sqrt{\epsilon_0}\vec{E} + \vec{B}I/\sqrt{\mu_0}$ is related to the macroscopic fields in the following way

$$\epsilon_0\vec{E} = \vec{D} - \vec{\mathcal{P}}, \qquad \frac{\vec{B}}{\mu_0} = \vec{H} + \vec{\mathcal{M}}, \tag{7.49}$$

which agrees with the standard definition of the electric polarization and magnetization vectors [155,227].

Imposing Maxwell's equation (7.1) on the partitions in (7.47) then couples the fields and sources in a nontrivial way

$$\nabla\mathbf{F}_f = j_f - (\nabla\mathbf{F}_b - j_b). \tag{7.50}$$

We can then *define* the split (7.47) to force the final bracketed term of (7.50) to vanish, which purposefully couples the free macroscopic field with only the *free* charge-currents. This definition forces the separation of the total field into two equations

$$\nabla\mathbf{F}_f = j_f, \qquad \nabla\mathbf{F}_b = j_b. \tag{7.51}$$

---

[24] Note that we use calligraphic script here for the polarization and magnetization to disambiguate them from momentum and angular momentum in other sections. They will only be used within this section.

The first of these split equations is Maxwell's *macroscopic* equation for free charges. The second is a constraint equation that defines the *electromagnetic polarization field* that is bound to the medium [155,227]. Note that the bound charges j_b will generally respond to the total field **F**, so will shift around while staying confined to the lattice structure, creating an indirect coupling between the two split equations in (7.51). We will return to this important point later.

In terms of a relative frame, the free fields $\vec{D}$ and $\vec{H}$ satisfy the macroscopic Maxwell's equations in (7.51)

$$\begin{aligned}
\vec{\nabla} \cdot \vec{D} &= c\rho_{e,f}, & \partial_0 \vec{D} - \vec{\nabla} \times \vec{H} &= -\vec{J}_{e,f}, \\
\vec{\nabla} \cdot \vec{H} &= c\rho_{m,f}, & \partial_0 \vec{H} + \vec{\nabla} \times \vec{D} &= -\vec{J}_{m,f},
\end{aligned} \tag{7.52}$$

that involve only the *free* charge-currents. Similarly, the bound fields $\vec{\mathcal{P}}$ and $\vec{\mathcal{M}}$ satisfy the constraint equations from (7.51)

$$\begin{aligned}
-\vec{\nabla} \cdot \vec{\mathcal{P}} &= c\rho_{e,b}, & -\partial_0 \vec{\mathcal{P}} - \vec{\nabla} \times \vec{\mathcal{M}} &= -\vec{J}_{e,b}, \\
\vec{\nabla} \cdot \vec{\mathcal{M}} &= c\rho_{m,b}, & \partial_0 \vec{\mathcal{M}} - \vec{\nabla} \times \vec{\mathcal{P}} &= -\vec{J}_{m,b},
\end{aligned} \tag{7.53}$$

that involve only the *bound* charge-currents.

The subtlety that arises from this separation of the total source into free and bound parts is that the partition could implicitly contain complementary *induced* charge-currents j_i that emerge from complicated interactions with the medium. That is, one can generally arrange for the relation

$$j_f = j_{f,0} + j_i, \qquad j_b = j_{b,0} - j_i, \tag{7.54}$$

such that the total source is $j = j_f + j_b = j_{f,0} + j_{b,0}$, so any induced part cancels and is only visible to the separation of the macroscopic fields. Interestingly, the induced charge-current j_i could change the apparent nature of the separated charge-current. For example, even if one considers a purely electric (vector) total charge-current, $j = j_e$, an exotic material (e.g., [228–232]) could induce an effective *magnetic* (trivector) part, $j_i = j_{m,i}I$, to the total free charge-current, $j_f = j_{e,f} + j_{m,i}I$. The free macroscopic field $\mathbf{F}_f$ will then be sensitive to this induced magnetic source, which will appear as apparent magnetic monopoles embedded in the medium [213].

To see this behavior more clearly in the relative frame, consider the case when the total current $j = j_e = (c\rho_e + \vec{J}_e)\gamma_0$ is purely electric (as is traditional and expected), but a magnetic part $j_iI = (c\rho_{m,i} + \vec{J}_{m,i})\gamma_0I$ is then induced by an exotic medium (e.g., [228–232]). In this case, according to (7.1) and (7.47), the following macroscopic and microscopic equations will be satisfied

$$\begin{aligned}
\vec{\nabla} \cdot \vec{D} &= c\rho_{e,f}, & \partial_0 \vec{D} - \vec{\nabla} \times \vec{H} &= -\vec{J}_{e,f}, \\
\vec{\nabla} \cdot \vec{H} &= c\rho_{m,i}, & \partial_0 \vec{H} + \vec{\nabla} \times \vec{D} &= -\vec{J}_{m,i}, \\
\epsilon_0 \vec{\nabla} \cdot \vec{E} &= c\rho_e, & \epsilon_0 \partial_0 \vec{E} - \vec{\nabla} \times \vec{B}/\mu_0 &= -\vec{J}_e, \\
\vec{\nabla} \cdot \vec{B} &= 0, & \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} &= 0.
\end{aligned} \tag{7.55}$$

Note that the divergence equation for $\vec{H}$ and the curl equation for $\vec{D}$ acquire induced *magnetic source* terms, even though the corresponding equations for the total fields $\vec{B}$ and $\vec{E}$ show no evidence of such a magnetic source.

The standard practice [155,227] to simplify the large set of equations (7.55) is to drop the middle four equations and consider only the remaining four equations that are hybridized between the macro- and microscopic fields and involve only the free electric charge

$$\begin{aligned}
\vec{\nabla} \cdot \vec{D} &= c\rho_{e,f}, & \partial_0 \vec{D} - \vec{\nabla} \times \vec{H} &= -\vec{J}_{e,f}, \\
\vec{\nabla} \cdot \vec{B} &= 0, & \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} &= 0.
\end{aligned} \tag{7.56}$$

These equations are then supplemented by the first constraint equation in (7.53) that involves the bound electric charge

$$-\vec{\nabla} \cdot \vec{\mathcal{P}} = c\rho_{e,b}, \qquad -\partial_0 \vec{\mathcal{P}} - \vec{\nabla} \times \vec{\mathcal{M}} = -\vec{J}_{e,b}. \tag{7.57}$$

All six of this reduced set of equations then involve only the *electric* part of the charge, regardless of any magnetic charge that may have been induced by the medium. This simplification avoids any consideration of the induced magnetic source in favor of a purely electric description.

Note that the six standard equations (7.56) and (7.57) for macroscopic fields can be equivalently written in a frame-independent way as the triplet of equations [9]

$$\nabla \cdot \mathbf{F}_f = j_f, \qquad \nabla \wedge \mathbf{F} = j_m I, \qquad \nabla \cdot \mathbf{F}_b = j_b. \tag{7.58}$$

Table 16

Macroscopic Maxwell's equations. The total microscopic field $\mathbf{F}$ is partitioned into a free part $\mathbf{F}_f$ and a bound part $\mathbf{F}_b$, each coupled to free j_f and bound j_b charge-currents in a medium, respectively. The bound field contains the electric polarization $\vec{\mathcal{P}}$ and magnetization $\vec{\mathcal{M}}$ densities of the medium, while the free field contains the electric displacement $\vec{D}$ and magnetic field strength $\vec{H}$. Out of the total set of equations implied by Maxwell's microscopic equation, a reduced set of the indicated three equations is typically considered, augmented by phenomenological constitutive relations between $\mathbf{F}$ and $\mathbf{F}_f$.

<hr>

<div align="center">Macroscopic Maxwell's Equations</div>

<hr>

Microscopic: $\boxed{\nabla \mathbf{F} = j}$ $\boxed{\mathbf{F} = \vec{E} + \vec{B}I}$

$$\vec{\nabla} \cdot \vec{E} = c\rho_e, \qquad \partial_0 \vec{E} - \vec{\nabla} \times \vec{B} = -\vec{J}_e$$
$$\vec{\nabla} \cdot \vec{B} = c\rho_m, \qquad \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} = -\vec{J}_m$$

Macroscopic: $\boxed{\mathbf{F} = \mathbf{F}_f + \mathbf{F}_b}$ $\boxed{\mathbf{F}_f = \vec{D} + \vec{H}I}$ $\boxed{\mathbf{F}_b = -\vec{\mathcal{P}} + \vec{\mathcal{M}}I}$

$$\vec{E} = \vec{D} - \vec{\mathcal{P}} \qquad \vec{B} = \vec{H} + \vec{\mathcal{M}} \qquad j = j_f + j_b$$

Free fields: $\boxed{\nabla \mathbf{F}_f = j_f}$

$$\vec{\nabla} \cdot \vec{D} = c\rho_{e,f}, \qquad \partial_0 \vec{D} - \vec{\nabla} \times \vec{H} = -\vec{J}_{e,f}$$
$$\vec{\nabla} \cdot \vec{H} = c\rho_{m,f}, \qquad \partial_0 \vec{H} + \vec{\nabla} \times \vec{D} = -\vec{J}_{m,f}$$

Bound fields: $\boxed{\nabla \mathbf{F}_b = j_b}$

$$-\vec{\nabla} \cdot \vec{\mathcal{P}} = c\rho_{e,b}, \qquad -\partial_0 \vec{\mathcal{P}} - \vec{\nabla} \times \vec{\mathcal{M}} = -\vec{J}_{e,b}$$
$$\vec{\nabla} \cdot \vec{\mathcal{M}} = c\rho_{m,b}, \qquad \partial_0 \vec{\mathcal{M}} - \vec{\nabla} \times \vec{\mathcal{P}} = -\vec{J}_{m,b}$$

Reduced set: $\boxed{\nabla \cdot \mathbf{F}_f = j_f}$ $\boxed{\nabla \wedge \mathbf{F} = j_m I}$ $\boxed{\nabla \cdot \mathbf{F}_b = j_b}$

$$\vec{\nabla} \cdot \vec{D} = c\rho_{e,f}, \qquad \partial_0 \vec{D} - \vec{\nabla} \times \vec{H} = -\vec{J}_{e,f}$$
$$\vec{\nabla} \cdot \vec{B} = c\rho_m, \qquad \partial_0 \vec{B} + \vec{\nabla} \times \vec{E} = -\vec{J}_m$$
$$-\vec{\nabla} \cdot \vec{\mathcal{P}} = c\rho_{e,b}, \qquad -\partial_0 \vec{\mathcal{P}} - \vec{\nabla} \times \vec{\mathcal{M}} = -\vec{J}_{e,b}$$

<hr>

SI Units : $\vec{E} \mapsto \sqrt{\epsilon_0}\,\vec{E}, \quad \vec{B} \mapsto \vec{B}/\sqrt{\mu_0}, \quad j_e \mapsto \sqrt{\mu_0}\,j_e, \quad j_m \mapsto \sqrt{\mu_0}\,j_m/c$
$(\mathrm{c} \equiv 1/\sqrt{\epsilon_0 \mu_0})$ $\vec{D} \mapsto \vec{D}/\sqrt{\epsilon_0}, \quad \vec{H} \mapsto \sqrt{\mu_0}\vec{H}, \quad \vec{\mathcal{P}} \mapsto \vec{\mathcal{P}}/\sqrt{\epsilon_0}, \quad \vec{\mathcal{M}} \mapsto \sqrt{\mu_0}\,\vec{\mathcal{M}}$

CGS Units: $\vec{E} \mapsto \vec{E}/\sqrt{4\pi}, \quad \vec{B} \mapsto \vec{B}/\sqrt{4\pi}, \quad j \mapsto \sqrt{4\pi}\,j/\mathrm{c}$
$\vec{D} \mapsto \vec{D}/\sqrt{4\pi}, \quad \vec{H} \mapsto \vec{H}/\sqrt{4\pi}, \quad \vec{\mathcal{P}} \mapsto \sqrt{4\pi}\,\vec{\mathcal{P}}, \quad \vec{\mathcal{M}} \mapsto \sqrt{4\pi}\,\vec{\mathcal{M}}$

<hr>

The first equation involves the free electric charge. The second becomes the Bianchi identity for the total field when the total magnetic source j_m is zero. The third is the constraint for the bound electric charge. It is important to remember, however, that these three equations arise as only part of the fully separated equations in (7.51), which are parts of Maxwell's single equation for the total (microscopic) field (7.1). All these relationships are summarized for reference in Table 16.

The information that is lost by neglected the middle equations in (7.55) can be phenomenologically restored by determining a *constitutive relation* between the macroscopic fields $\vec{D}$ and $\vec{H}$ and the total microscopic fields $\vec{E}$ and $\vec{B}$. As mentioned previously, such a relation should exist because the configuration of the bound charge-current j_b will generally depend upon the total field $\mathbf{F}$ in the medium, which in turn alters the constraint for the electromagnetic polarization in (7.47) that determines the macroscopic fields. Determining the effective functional dependence of this polarization field is a nontrivial and material-specific phenomenological problem [155,227]. In the simplest case there is no medium (and thus no electromagnetic polarization), so $\vec{D} = \epsilon_0 \vec{E}$ and $\vec{H} = \vec{B}/\mu_0$ according to (7.49). In the next simplest case the relation is linear and one can define new constants ϵ and μ such that $\vec{D} = \epsilon\vec{E}$ and $\vec{H} = \vec{B}/\mu$. More generally, the appropriate constitutive relations depend on the total field $\mathbf{F}$, and are frequency-dependent (i.e., dispersive) [155,227,233].

It is also worth remarking that there is a considerable interest to artificial media and materials that restore the dual symmetry of vacuum on the level of macroscopic medium parameters (e.g., $\epsilon = \mu$) [19,107,108,214,234–236]. The

conservation of the optical helicity associated with such symmetry (see Section 8.8) provides a new insight into the light propagation and scattering dynamics.

8. Lagrangian formalism

> *It is impossible to study this remarkable theory without experiencing at times the strange feeling that the equations and formulas somehow have a proper life, that they are smarter than we, smarter than the author himself, and that we somehow obtain from them more than was originally put into them.*
>
> [Heinrich Hertz [237]]

> *You can recognize truth by its beauty and simplicity. When you get it right, it is obvious that it is right – at least if you have any experience – because usually what happens is that more comes out than goes in.*
>
> [Richard P. Feynman [238]]

It is worth emphasizing that up to this point we have introduced *no physical postulates* other than the need for spacetime itself (and thus its associated algebra). We made the casual observation that a proper spacetime bivector **F** naturally splits into polar and axial 3-vector parts $\mathbf{F} = \vec{E} + \vec{B}I$ with respect to a particular reference frame in Eq. (3.40), and noted that these parts transform under a Lorentz boost exactly as we expect the electromagnetic field to transform in Eqs. (3.59) and (5.9). We then found that the structure of the simplest bivector differential equation $\nabla\mathbf{F} = 0$ in Eq (5.1) is precisely Maxwell's equation in vacuum, while the next simplest modification $\nabla\mathbf{F} = j$ in Eq. (7.1) is Maxwell's equation that properly includes both electric and magnetic sources, albeit in a compact form as a single complex vector current j. We also made the casual observation that contracting the bivector **F** with such a complex current j gives the correct Lorentz force equation $dp/d\tau = \langle \mathbf{F}j \rangle_1$ that supports both electric and magnetic charges in Eq. (7.19). Combining the Lorentz force law and Maxwell's equation using a general conservation constraint in Eq. (7.31) then produces the correct symmetric energy–momentum stress tensor $\underline{T}(\gamma_0) = \mathbf{F}\gamma_0\widetilde{\mathbf{F}}/2$ in Eq. (7.35) as a simple quadratic form of the field (which is the same functional form as the energy–momentum current of the Dirac electron [9,87,96]).

Despite the richness of these emergent characteristics of the spacetime geometry, all suggestive notation, terminology, and constant unit factors have been window-dressing for features that are essentially inevitable when considering the simplest dynamics one can ascribe to any bivector field in spacetime, bereft of physical interpretation *a priori*. The fact that these features exactly correspond to the electromagnetic theory of both electric and magnetic charges is remarkable unto itself. More importantly, however, this fact prompts a careful reassessment of the physical foundations and postulates of classical electromagnetic field theory.

The more complete physical foundation of electromagnetism comes from a Lagrangian field-theory perspective. Fundamentally, the electromagnetic field **F** emerges as a necessary consequence of preserving a local U(1) gauge symmetry (i.e., a local phase rotation) in the Lagrangian for a complex (charged) matter field [168], as we will detail in Section 8.9.1. However, with such a Lagrangian approach it becomes clear that the *vector potential z* acts as the dynamical gauge field that preserves the charge symmetry, with the usual electromagnetic field $\mathbf{F} = \nabla z$ arising only as a derivative quantity. Thus, the vector potential z will be our new starting point, from which we will systematically derive the corresponding conserved physical quantities of the field.

8.1. Traditional vacuum-field Lagrangian

As we have shown in Sections 5.2 and 6, the dual symmetry of the vacuum Maxwell's equation $\nabla^2 z = 0$ implies that the vector potential should be invariant under the global phase symmetry $z \mapsto z \exp(\theta I)$ in vacuum, and thus must be intrinsically *complex*. This symmetry poses a problem for the traditional electromagnetic Lagrangian, which can be written in various equivalent forms:

$$\mathcal{L}_{\text{trad}}(x) = \frac{\langle (\nabla z)^2 \rangle_0}{2} = \frac{\langle \mathbf{F}^2 \rangle_0}{2} = \frac{\ell_1}{2} = \frac{|\vec{E}|^2 - |\vec{B}|^2}{2} = \frac{1}{4} \sum_{\mu\nu} F_{\mu\nu} F^{\nu\mu}. \tag{8.1}$$

We will motivate why this Lagrangian appears from a gauge theory perspective in Section 8.9.1. Traditionally $z = a_e$ is considered to be a pure (electric) vector potential that generates a single associated constituent field $\mathbf{F}_e = \nabla \wedge a_e$ [29,127,155]. Indeed, recall that the subscripts e and m throughout this report refer to the electric and magnetic types of *charge*, so in the traditional electromagnetism (with only electric charges) all "magnetic" (subscript m) quantities can be neglected. However, we keep $z = a_e + a_m I$ complex here, since no electric or magnetic charge exists in *vacuum* to select either a_e or a_m as a preferred vector potential.

The traditional Lagrangian (8.1) is *not dual-symmetric*, and, hence, is problematic for the description of dual-symmetric vacuum fields that require such a complex potential. To see this explicitly, we expand the complex vector potential z in (8.1) into its constituent parts, impose the Lorenz–FitzGerald conditions $\nabla \cdot a_e = \nabla \cdot a_m = 0$ discussed in Section 6, and define the two constituent fields $\mathbf{F}_e = \nabla \wedge a_e = \vec{E}_e + \vec{B}_e I$ and $\mathbf{F}_m = \nabla \wedge a_m = \vec{B}_m - \vec{E}_m I$ such that $\mathbf{F} = \nabla z = \mathbf{F}_e + \mathbf{F}_m I$, as in

Section 6.1.3, then the traditional Lagrangian expands into the form

$$\mathcal{L}_{\text{trad}}(x) = \frac{\langle \mathbf{F}_e^2 \rangle_0 - \langle \mathbf{F}_m^2 \rangle_0}{2} + \langle \mathbf{F}_e \cdot \mathbf{F}_m I \rangle_0,$$

$$= \frac{1}{2}[|\vec{E}_e|^2 - |\vec{B}_e|^2 + |\vec{E}_m|^2 - |\vec{B}_m|^2] - [\vec{E}_e \cdot \vec{B}_m + \vec{B}_e \cdot \vec{E}_m], \tag{8.2}$$

which shows an intrinsic interplay between the constituent fields[25] that alters the original structure of (8.1).

This additional structure appears because the traditional Lagrangian (8.1) is not invariant under the dual symmetry phase rotation $z \mapsto z \exp(\theta I)$. Instead, it transforms as

$$\mathcal{L}_{\text{trad}}(x) \mapsto \frac{\ell_1}{2} \cos 2\theta - \frac{\ell_2}{2} \sin 2\theta, \tag{8.3}$$

and oscillates between the two proper scalars $\ell_1 = |\vec{E}|^2 - |\vec{B}|^2$ and $\ell_2 = 2\vec{E} \cdot \vec{B}$ of the total field as the phase is changed [68]. (As we noted in Section 5.3, the second term that appears here is the "axion" term proposed in [216] using precisely this dual-symmetric phase-rotation.) The lack of global phase invariance of the Lagrangian (8.1) is problematic for maintaining the dual-symmetric complexity of the vector potential z. However, one can always avoid this problem by explicitly *breaking* the dual-symmetry to keep $z = a_e$ a pure (electric) vector potential, as in the traditional approach. In this case the constituent fields associated with the magnetic vector potential a_m vanish ($\vec{E}_m = \vec{B}_m = 0$) in (8.2).

We are thus motivated to consider a simple correction to the traditional Lagrangian that addresses these shortcomings for the description of a vacuum field. We wish to modify the Lagrangian to make it manifestly dual-symmetric under a global phase rotation of z.

8.2. Dual-symmetric Lagrangian

We can correct the traditional Lagrangian in a simple way to make it properly dual-symmetric by adding a complex conjugate to the traditional expression

$$\mathcal{L}_{\text{dual}}(x) = \frac{1}{2} \langle (\nabla z)(\nabla z^*) \rangle_0. \tag{8.4}$$

Notably, this modification makes the Lagrangian a simple quadratic form for the complex vector field z, which is the expected form for the kinetic energy of such a field. This form is thus a logical choice for a properly dual-symmetric Lagrangian. We must now verify that this modification produces sensible results.

In terms of constituent fields, (8.4) expands into the following equivalent forms:

$$\mathcal{L}_{\text{dual}}(x) = \frac{\langle (\nabla a_e)^2 \rangle_0}{2} + \frac{\langle (\nabla a_m)^2 \rangle_0}{2} = \frac{\langle \mathbf{F}_e^2 \rangle_0}{2} + \frac{\langle \mathbf{F}_m^2 \rangle_0}{2},$$

$$= \frac{1}{2}(|\vec{E}_e|^2 - |\vec{B}_e|^2) + \frac{1}{2}(|\vec{B}_m|^2 - |\vec{E}_m|^2),$$

$$= \frac{1}{4} \sum_{\mu\nu} [(F_e)_{\mu\nu}(F_e)^{\nu\mu} + (F_m)_{\mu\nu}(F_m)^{\nu\mu}]. \tag{8.5}$$

Perhaps surprisingly, the dual-symmetric form of (8.4) is simply the sum of *two copies of the traditional Lagrangian* (8.2), one for each independent constituent field a_e and a_m in the complex vector potential $z = a_e + a_m I$. Notably, a dual-symmetric Lagrangian of precisely this form was previously considered in [196,209,210,212] from a different starting point. The traditional and dual-symmetric Lagrangians are compared for reference in Table 17.

8.3. Constrained dual-symmetric Lagrangian

The fact that the dual-symmetrized Lagrangian (8.4) explicitly involves the second constituent field $\mathbf{F}_m$ as an independent but necessary field prompts the question whether this second field is really independent in vacuum, or should be further constrained. In particular, if one believes the *total* dual-symmetric field $\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m I$ should be the true electromagnetic field, then one can postulate that the different constituent fields $\mathbf{F}_e = \nabla \wedge a_e$ and $\mathbf{F}_m = \nabla \wedge a_m$ actually describe the *same* total field. We can thus consider imposing an additional nontrivial *constraint* to force these two constituent fields to give equal contributions to the measurable vacuum field

$$\mathbf{F}_e = \mathbf{F}_m I = \frac{\mathbf{F}_{\text{meas}}}{\sqrt{2}}, \tag{8.6}$$

---

[25] Again, the subscripts of e and m on the constituent relative fields indicate that they would normally be associated with either electric or magnetic *charges*, which do not exist in vacuum.

Table 17

Electromagnetic Lagrangians in the traditional and dual-symmetric field theories. The traditional Lagrangian is a kinetic term for only the constituent vector potential a_e associated with *electric* charge. The dual-symmetric Lagrangian is a kinetic term for the full complex vector potential z, and preserves the dual symmetry of vacuum. The electric a_e and magnetic a_m vector potentials contribute to this Lagrangian as *independent* fields that each satisfy a copy of the traditional Lagrangian [196,209,210,212]. In the constrained formalism, the two vector potentials of the dual-symmetric Lagrangian describe the *same* (measurable) vacuum field $\mathbf{F}_{\text{meas}}$ in a symmetrized way [68,106]. Note that with this constraint the dual-symmetric Lagrangian formally vanishes; however, the constraint can be implemented using a Lagrange multiplier, or as a final step.

<div align="center">

Electromagnetic Lagrangian Densities

Traditional: $\quad z = a_e \equiv a \quad \mathbf{F} = \nabla a_e = \mathbf{F}_e$

$$\boxed{\mathcal{L}_{\text{trad}}(x) = \frac{\langle (\nabla a)^2 \rangle_0}{2}} = \frac{\langle \mathbf{F}^2 \rangle_0}{2} = \frac{1}{4} \sum_{\mu\nu} F_{\mu\nu} F^{\nu\mu}$$

Dual-symmetric: $\quad z = a_e + a_m I \quad \mathbf{F} = \nabla z = \mathbf{F}_e + \mathbf{F}_m I$

$$\boxed{\mathcal{L}_{\text{dual}}(x) = \frac{\langle (\nabla z)(\nabla z^*) \rangle_0}{2}} = \frac{\langle \mathbf{F}_e^2 \rangle_0}{2} + \frac{\langle \mathbf{F}_m^2 \rangle_0}{2}$$

$$= \frac{1}{4} \sum_{\mu\nu} [(F_e)_{\mu\nu} (F_e)^{\nu\mu} + (F_m)_{\mu\nu} (F_m)^{\nu\mu}]$$

Constrained: $\quad \boxed{\mathbf{F}_e \to \mathbf{F}_m I \to \mathbf{F}_{\text{meas}}/\sqrt{2} = \mathbf{G}_{\text{meas}} I/\sqrt{2}}$

$$\mathcal{L}_{\text{dual}}(x) \to \frac{\langle \mathbf{F}_{\text{meas}}^2 \rangle_0}{4} + \frac{\langle \mathbf{G}_{\text{meas}}^2 \rangle_0}{4} = \frac{1}{8} \sum_{\mu\nu} [F_{\mu\nu} F^{\nu\mu} + G_{\mu\nu} G^{\nu\mu}]$$

</div>

so $\vec{E}_e = \vec{E}_m = \vec{E}_{\text{meas}}/\sqrt{2}$ and $\vec{B}_e = \vec{B}_m = \vec{B}_{\text{meas}}/\sqrt{2}$, and the total field magnitude is preserved:

$$|\mathbf{F}|^2 = \mathbf{F}^* \mathbf{F} = \mathbf{F}_e^* \mathbf{F}_e + \mathbf{F}_m^* \mathbf{F}_m \to \mathbf{F}_{\text{meas}}^* \mathbf{F}_{\text{meas}} = |\mathbf{F}_{\text{meas}}|^2. \tag{8.7}$$

With this constraint we find that $\mathbf{F}_m = \mathbf{F}_{\text{meas}} I^{-1}/\sqrt{2} = \mathbf{G}_{\text{meas}}/\sqrt{2}$, so the two constituent fields become precisely proportional to the measurable field $\mathbf{F}_{\text{meas}}$ and its Hodge-dual $\mathbf{G}_{\text{meas}} = \mathbf{F}_{\text{meas}} I^{-1}$.

Inserting the constraint (8.6) into (8.5) produces

$$\mathcal{L}_{\text{dual}}(x) \to \frac{\langle \mathbf{F}_{\text{meas}}^2 \rangle_0 + \langle \mathbf{G}_{\text{meas}}^2 \rangle_0}{4} = \frac{1}{8} \sum_{\mu\nu} (F_{\mu\nu} F^{\nu\mu} + G_{\mu\nu} G^{\nu\mu}). \tag{8.8}$$

The functional form of this constrained Lagrangian density is then precisely equal to the dual-symmetric form recently postulated in [68,106]. Curiously, this constrained form identically *vanishes* since $\mathbf{G}_{\text{meas}}^2 = (\mathbf{F}_{\text{meas}} I^{-1})^2 = -\mathbf{F}_{\text{meas}}^2$. Thus, the constraint (8.6) should be imposed either as a Lagrange multiplier constraint in the dual-symmetric Lagrangian (8.4), or as a final step in derivations that retain both constituent fields $\mathbf{F}_e$ and $\mathbf{F}_m$ [68,106].

Notably, the constraint (8.6) also forces the vector potential z to be *irreducibly complex*. Indeed, if one could rotate z to a pure vector potential a_0 with some phase rotation $z = a_0 \exp(\theta I)$, then the constraint (8.6) would force the measurable field $\mathbf{F}_{\text{meas}}$ to identically vanish. To see this, note that $z = a_e + a_m I = a_0 \exp(\theta I)$, so $a_e = a_0 \cos\theta$ and $a_m = a_0 \sin\theta$. Therefore, $\mathbf{F}_e = \cos\theta \, \mathbf{F}_0$ and $\mathbf{F}_m = \sin\theta \, \mathbf{F}_0$ where $\mathbf{F}_0 = \nabla \wedge a_0$. The constraint (8.6) then implies: $\cos\theta \, \mathbf{F}_0 = \sin\theta \, \mathbf{F}_0 I = \mathbf{F}_{\text{meas}}/\sqrt{2}$. Due to the factor of I, this relation is only satisfiable if $\mathbf{F}_0 = 0$ for any θ, making $\mathbf{F}_{\text{meas}} = 0$. Therefore, to obtain nontrivial vacuum fields that satisfy the constraint (8.6), the vector potential z must remain complex under *all* phase rotations.

However, the irreducible complexity of z forced by the constraint (8.6) presents a significant problem for *coupling to matter*. The equation $\nabla z^* = \nabla(a_e - a_m I) = \mathbf{F}_e - \mathbf{F}_m I \to 0$ vanishes with this constraint, which is inconsistent with Maxwell's source equation (7.1), written in its conjugated form $\nabla^2 z^* = j^*$. Specifically, $\nabla^2 z^* = \nabla(\nabla z^*) \to 0$ with the constraint, so any source j must vanish. Thus, the constraint (8.6) can only describe *vacuum* fields that are not coupled to matter, which corroborates our general discussion about the fundamental difference between the irreducibly complex vacuum and reducibly complex source parts of the complex vector potential z in Section 7.4.

Since fields must couple to matter (charges) to be physically measured, the constraint (8.6) in its current form is problematic. Either we must modify the constraint (8.6) to properly accommodate sources (see, e.g., [239] for ideas regarding

how this modification could be accomplished), or the constrained Lagrangian becomes inconsistent with the presence of sources. At the same time, the independent "electric" and "magnetic" parts of the unconstrained dual-symmetric Lagrangian (8.5) can be coupled to the corresponding charges. In this case one can speculate that the presence of "electric matter" explicitly breaks the dual-symmetry by coupling to only one of the constituent fields $\mathbf{F}_e = \nabla \wedge a_e$. In the absence of other fields in the model, the remaining constituent field $\mathbf{F}_m$ in the dual-symmetric Lagrangian (8.4) would then be undetectable and only contribute to gravity as an invisible and noninteracting energy density (e.g., a *dark light* contribution to dark matter [196]). We will revisit this issue of the second constituent field more carefully in Section 8.9.2.

8.4. Maxwell's equation of motion

We can obtain the equation of motion for z implied by all three candidate Lagrangians (Table 17) by treating the constituent vector potentials a_e and a_m independently, and applying the appropriate Euler–Lagrange equation [9,127] to each one. For example,

$$\sum_\mu \frac{\partial}{\partial x^\mu} \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu a_e)} \right) - \frac{\partial \mathcal{L}}{\partial a_e} = \nabla^2 a_e = 0, \tag{8.9}$$

is the proper equation of motion for a_e, where $x = \sum_\mu x^\mu \gamma_\mu$. The equation for a_m can be obtained similarly, yielding $\nabla^2 a_m = 0$. It follows that the equation of motion for $z = a_e + a_m I$ is Maxwell's vacuum equation (5.1) in a manifestly dual-symmetric form,

$$\nabla^2 z = \nabla \mathbf{F} = 0, \tag{8.10}$$

for any of the three candidate Lagrangians. This means, that the above three Lagrangian theories are indistinguishable from each other on the level of vacuum equations of motion. However, they do differ from each other in other field-theory aspects [68].

8.5. Field invariants: Noether currents

According to Noether's theorem [109], any continuous symmetry of a Lagrangian leads to a conservation law. We can use this theorem to systematically find the *canonical* conserved currents for the vacuum electromagnetic field, using each of the three Lagrangian densities (8.1), (8.4) and (8.8) in Table 17.

Recall the general procedure for obtaining the conserved currents [9,127]:

1. Suppose one transforms each field a (and possibly the coordinates x) in a Lagrangian density $\mathcal{L}[a(x), \nabla a(x)]$ according to a group transformation $\underline{U}_{\alpha,g}$ that depends on a continuous scalar parameter α, as well as a generating element g

$$a'(x') = \underline{U}_{\alpha,g}[a(x')], \tag{8.11a}$$

$$x' = \underline{U}_{\alpha,g}^{-1}(x) = \overline{U}_{\alpha,g}(x), \tag{8.11b}$$

$$\nabla' = \underline{U}_{\alpha,g}(\nabla). \tag{8.11c}$$

If $\mathcal{L}[a'(x'), \nabla' a'(x')] = \mathcal{L}[a(x), \nabla a(x)]$ then $\underline{U}_{\alpha,g}$ is a symmetry of the Lagrangian density.[26]

2. The infinitesimal field and coordinate increments are then

$$\delta a(g) \equiv \partial_\alpha a'(x')|_{\alpha=0} = \partial_\alpha a'(x)|_{\alpha=0} + (\delta x(g) \cdot \nabla) a(x), \tag{8.12a}$$

$$\delta x(g) \equiv \partial_\alpha x'|_{\alpha=0}, \tag{8.12b}$$

and depend on the chosen generator g.

3. The change in the transformed Lagrangian $\mathcal{L}'$ is therefore

$$\partial_\alpha \mathcal{L}'(x')|_{\alpha=0} = \left\langle \frac{\partial \mathcal{L}}{\partial a} \delta a(g) \right\rangle_0 + \sum_\mu \left\langle \frac{\partial \mathcal{L}}{\partial (\partial_\mu a)} \partial_\mu \delta a(g) \right\rangle_0 + (\delta x(g) \cdot \nabla) \mathcal{L}, \tag{8.13}$$

which must vanish from the symmetry. In the presence of multiple independent fields (such as a_e and a_m), the right hand side contains corresponding terms for each independent field.

---

[26] Note that if both the coordinates x and the fields a are transformed, then one should typically transform ∇ in the same way as a so that the kinetic terms ∇a transform similarly to a. Since ∇ and x are inversely related, this implies that x and a should transform oppositely with the group transformation.

4. After applying the Euler–Lagrange equation (8.9) for each a in the first term (i.e., essentially using Maxwell's equation), (8.13) simplifies to a divergence

$$\partial_\alpha \mathcal{L}'|_{\alpha=0} = \nabla \cdot \underline{J}_N(g) = 0, \qquad (8.14)$$

of a *conserved Noether current* (tensor) that depends on the generator g

$$\underline{J}_N(g) \equiv \sum_\mu \gamma_\mu \left\langle \frac{\partial \mathcal{L}}{\partial(\partial_\mu a)} \delta a(g) \right\rangle_0 + \mathcal{L} \, \delta x(g). \qquad (8.15)$$

5. In any reference frame $\underline{J}_N(g) = [\xi_N(g) + \vec{J}_N(g)]\gamma_0$, it follows from this conservation law that $\partial_0 \xi_N(g) - \vec{\nabla} \cdot \vec{J}_N(g) = 0$. Hence, we can integrate over the spatial coordinates (assuming $\vec{J}_N(g)$ vanishes at infinity) to find a conserved *Noether charge*

$$\underline{Q}_N(g) = \int d^3x \, \xi_N(g), \qquad \partial_0 \underline{Q}_N(g) = 0. \qquad (8.16)$$

We now use this procedure to find the Noether currents for the various electromagnetic Lagrangians. Note that it will be sufficient to compute everything using the dual-symmetric Lagrangian (8.4), and then constrain the final results appropriately. For the traditional Lagrangian (8.1), we will simply neglect all terms related to the constituent field $\mathbf{F}_m$ associated with magnetic charges. For the constrained Lagrangian, we will enforce the additional constraint $\mathbf{F}_e = \mathbf{F}_m = \mathbf{F}/\sqrt{2}$, Eq. (8.6), which makes both vector potentials a_e and a_m correspond to the same electromagnetic field. We will distinguish the expressions for the constrained Lagrangian using an arrow symbol "$\rightarrow$" for clarity in what follows.

8.6. Canonical energy–momentum stress tensor

We first consider translations of spacetime, which will produce the conservation of the energy–momentum of the field. Suppose we translate the position vectors x in the Lagrangian density by a small amount α in an arbitrary direction b, so $x' = x - \alpha b = \exp[-\alpha(b \cdot \nabla)]x$. This translation leaves the Lagrangian (8.4) invariant since $\nabla' = \nabla$. It follows from (8.12) that the increment for x is simply b, while the increment for a_e is the corresponding directional derivative

$$\delta x[b] = -b, \qquad (8.17a)$$
$$\delta a_e[b] = -(b \cdot \nabla)a_e, \qquad (8.17b)$$

and similarly for a_m. Computing the change from this translation according to (8.13) thus produces the conserved Noether current

$$
\begin{aligned}
\underline{T}(b) &= \sum_\mu \gamma_\mu \left\langle \frac{\partial \mathcal{L}}{\partial(\partial_\mu a_e)}(-b \cdot \nabla)a_e + \frac{\partial \mathcal{L}}{\partial(\partial_\mu a_m)}(-b \cdot \nabla)a_m \right\rangle_0 - b\mathcal{L}, \\
&= \sum_\mu \gamma_\mu \left\langle (\partial_\mu a_e)(-b \cdot \nabla)a_e + (\partial_\mu a_m)(-b \cdot \nabla)a_m \right\rangle_0 - b\mathcal{L}, \\
&= \nabla a_e \cdot [(-b \cdot \nabla)a_e] + \nabla a_m \cdot [(-b \cdot \nabla)a_m] - b\mathcal{L}, \\
&= (b \cdot \dot{\nabla})\langle \dot{a}_e \mathbf{F}_e + \dot{a}_m \mathbf{F}_m \rangle_1 - \frac{b}{2}(\langle \mathbf{F}_e^2 \rangle_0 + \langle \mathbf{F}_m^2 \rangle_0), \\
&\rightarrow \frac{1}{2}(b \cdot \dot{\nabla})\langle \dot{a}_e \mathbf{F} + \dot{a}_m \mathbf{G} \rangle_1.
\end{aligned}
\qquad (8.18)
$$

Recall that the overdots indicate that only the potentials are being differentiated. This conserved tensor $\underline{T}(b)$ is the *canonical energy–momentum stress tensor* [68,127], which was first derived while preserving dual symmetry by Gaillard and Zumino [115].

The canonical energy–momentum tensor is *not* symmetric with respect to its adjoint: $\underline{T} \neq \overline{T}$, which can be obtained as

$$\overline{T}(b) = \partial_d \langle \underline{T}(d)b \rangle_0 = \dot{\nabla}\langle (\dot{a}_e \mathbf{F}_e + \dot{a}_m \mathbf{F}_m)b \rangle_0 - \frac{b}{2}(\langle \mathbf{F}_e^2 \rangle_0 + \langle \mathbf{F}_m^2 \rangle_0),$$

$$\rightarrow \frac{1}{2}\dot{\nabla}\langle (\dot{a}_e \mathbf{F} + \dot{a}_m \mathbf{G})b \rangle_0. \qquad (8.19)$$

This adjoint form involves the scalar (real) parts of total bivector contractions.

Importantly, in contrast to the symmetric energy–momentum tensor in (7.35), the canonical tensor in (8.18) and (8.19) explicitly involves the vector-potentials a_e and a_m, making it *gauge-dependent*. Note also that the first summand (involving $\mathbf{F}_e$ and a_e) in the non-constrained tensor is equivalent to the conserved canonical tensor for the traditional Lagrangian (8.1).

Table 18

Canonical energy–momentum stress tensors, conserved by translation symmetry. Unlike the symmetric tensor in Table 15, the canonical tensor naturally separates the Poynting vector $\vec{P}$ into an *orbital* part $\vec{P}_o$ and a separate *spin* part $\vec{P}_s$ that does not contribute to energy–momentum transport (for all three Lagrangians) [64,68,243,244]. The traditional Lagrangian produces field-asymmetric parts that are *electric*-biased [68,127]. The dual-symmetric Lagrangian in Table 17 adds the complementary constituent field that is *magnetic*-biased. Constraining this Lagrangian makes both contributions describe the same total field [68].

<div align="center">

Canonical Energy-Momentum Tensor

</div>

$$\underline{T}(\gamma_0) = (\varepsilon + \vec{P})\gamma_0, \quad \overline{T}(\gamma_0) = (\varepsilon + \boxed{\vec{P}_o})\gamma_0, \quad \vec{P} = \vec{P}_o + \boxed{\vec{P}_s}$$

Traditional:

$$\boxed{\underline{T}(b) = (b \cdot \dot{\nabla})\langle \dot{a}\mathbf{F}\rangle_1 - \frac{b}{2}\langle \mathbf{F}^2\rangle_0} \quad \boxed{\overline{T}(b) = \dot{\nabla}\langle \dot{a}\mathbf{F}b\rangle_0 - \frac{b}{2}\langle \mathbf{F}^2\rangle_0}$$

$z = a_e \equiv a$

$$\vec{P}_o = \dot{\vec{\nabla}}(\vec{E} \cdot \dot{\vec{A}}), \quad \vec{P}_s = -(\vec{E} \cdot \vec{\nabla})\vec{A}$$

$\mathbf{F} = \nabla a_e = \mathbf{F}_e$

$$T^{\mu\nu} = \sum_\alpha F^{\alpha\mu}\partial^\nu a_\alpha - \frac{\eta^{\mu\nu}}{4}\sum_{\alpha\beta} F_{\alpha\beta}F^{\beta\alpha}$$

Dual-symmetric:

$$\boxed{\underline{T}(b) = (b \cdot \dot{\nabla})\langle \dot{a}_e\mathbf{F}_e + \dot{a}_m\mathbf{F}_m\rangle_1 - \frac{b}{2}\left[\langle\mathbf{F}_e^2\rangle_0 + \langle\mathbf{F}_m^2\rangle_0\right]}$$

$$\boxed{\overline{T}(b) = \dot{\nabla}\langle(\dot{a}_e\mathbf{F}_e + \dot{a}_m\mathbf{F}_m)b\rangle_0 - \frac{b}{2}\left[\langle\mathbf{F}_e^2\rangle_0 + \langle\mathbf{F}_m^2\rangle_0\right]}$$

$z = a_e + a_m I$

$$\vec{P}_o = \dot{\vec{\nabla}}\left[\vec{E}_e \cdot \dot{\vec{A}} + \vec{B}_m \cdot \dot{\vec{C}}\right] \quad \vec{P}_s = -[(\vec{E}_e \cdot \vec{\nabla})\vec{A} + (\vec{B}_m \cdot \vec{\nabla})\vec{C}]$$

$\mathbf{F} = \nabla z$
$= \mathbf{F}_e + \mathbf{F}_m I$

$$T^{\mu\nu} = \sum_\alpha ((F_e)^{\alpha\mu}\partial^\nu(a_e)_\alpha + (F_m)^{\alpha\mu}\partial^\nu(a_m)_\alpha)$$
$$- \frac{\eta^{\mu\nu}}{4}\sum_{\alpha\beta}\left[(F_e)_{\alpha\beta}(F_e)^{\beta\alpha} + (F_m)_{\alpha\beta}(F_m)^{\beta\alpha}\right]$$

Constrained:

$$\boxed{\underline{T}(b) \to \frac{(b \cdot \dot{\nabla})}{2}\langle \dot{a}_e\mathbf{F} + \dot{a}_m\mathbf{G}\rangle_1} \quad \boxed{\overline{T}(b) \to \frac{\dot{\nabla}}{2}\langle(\dot{a}_e\mathbf{F} + \dot{a}_m\mathbf{G})b\rangle_0}$$

$\mathbf{F}_e = \mathbf{F}_m I \to \dfrac{\mathbf{F}}{\sqrt{2}}$

$$\vec{P}_o \to \frac{1}{2}\dot{\vec{\nabla}}\left[\vec{E} \cdot \dot{\vec{A}} + \vec{B} \cdot \dot{\vec{C}}\right] \quad \vec{P}_s \to -\frac{1}{2}[(\vec{E} \cdot \vec{\nabla})\vec{A} + (\vec{B} \cdot \vec{\nabla})\vec{C}]$$

$\mathbf{F} = \mathbf{G}I$

$$T^{\mu\nu} \to \frac{1}{2}\sum_\alpha (F^{\alpha\mu}\partial^\nu(a_e)_\alpha + G^{\alpha\mu}\partial^\nu(a_m)_\alpha)$$

Hence, the two summands in the unconstrained tensor are *separately conserved* energy–momentum tensors for the two constituent fields $\mathbf{F}_e$ and $\mathbf{F}_m$. We detail the canonical energy–momentum tensor in Table 18 for reference.

For comparison, we also compute the *symmetric* energy–momentum tensor from (7.35) explicitly in terms of the constituent fields to find

$$\underline{T}_{\text{sym}}(b) = \overline{T}_{\text{sym}}(b) = \frac{1}{2}\mathbf{F}b\widetilde{\mathbf{F}} = -\frac{1}{2}(\mathbf{F}_e + \mathbf{F}_m I)b(\mathbf{F}_e + \mathbf{F}_m I),$$
$$= \frac{1}{2}\mathbf{F}_e b\widetilde{\mathbf{F}}_e + \frac{1}{2}\mathbf{F}_m b\widetilde{\mathbf{F}}_m. \tag{8.20}$$

The cross-term involves a vector minus its reverse, so it vanishes to leave the sum of the symmetric tensors for each constituent field. This form as a sum is consistent with the canonical tensors derived from Noether's theorem. In particular the identity (8.20) implies that the energy density $\varepsilon = \varepsilon_e + \varepsilon_m$ and the Poynting vector $\vec{P} = \vec{P}_e + \vec{P}_m$ decompose into sums of the corresponding quantities for the constituent fields.

It is instructive to find the difference between the symmetric (Section 7.5.1) and canonical energy–momentum tensors:

$$\underline{K} \equiv \underline{T}_{\text{sym}} - \underline{T}, \qquad \overline{K} \equiv \overline{T}_{\text{sym}} - \overline{T}. \tag{8.21}$$

Using the identity $(b \cdot \nabla)a = b \cdot (\nabla a) + \nabla(b \cdot a)$, as well as the identity $2[b \cdot (\nabla a)] \cdot (\nabla a) - b\langle(\nabla a)^2\rangle_0 = -(\nabla a)b(\nabla a)$, we find the explicit forms

$$
\underline{K}(b) = (\mathbf{F}_e \cdot \nabla)(a_e \cdot b) + (\mathbf{F}_m \cdot \nabla)(a_m \cdot b),
$$
$$
\rightarrow \frac{1}{2}[(\mathbf{F} \cdot \nabla)(a_e \cdot b) + (\mathbf{G} \cdot \nabla)(a_m \cdot b)], \tag{8.22}
$$
$$
\overline{K}(b) = \langle b\mathbf{F}_e \dot{\nabla}\rangle_0 \dot{a}_e + \langle b\mathbf{F}_m \dot{\nabla}\rangle_0 \dot{a}_m,
$$
$$
\rightarrow \frac{1}{2}\left[\langle b\mathbf{F}\dot{\nabla}\rangle_0 \dot{a}_e + \langle b\mathbf{G}\dot{\nabla}\rangle_0 \dot{a}_m\right]. \tag{8.23}
$$

Since both the symmetric and canonical energy–momentum tensors satisfy the continuity equations, the difference term $\underline{K}(b)$ must vanish under a divergence, i.e., it does not contribute to the energy–momentum transport. Using the vacuum Maxwell's equation (5.1), we can make this property apparent by writing (8.22) as a total derivative:

$$
\underline{K}(b) = -\nabla[(b \cdot a_e)\mathbf{F}_e + (b \cdot a_m)\mathbf{F}_m],
$$
$$
\rightarrow -\frac{1}{2}\nabla[(b \cdot a_e)\mathbf{F} + (b \cdot a_m)\mathbf{G}]. \tag{8.24}
$$

Historically, the canonical energy–momentum tensor $\underline{T}$ that is produced by Noether's theorem was considered problematic due to its non-symmetric and gauge-dependent form (as well as the enigmatic appearance of the spin in the angular momentum tensor, as we shall see in the next section). Then, in 1940, Belinfante and Rosenfeld [240–242] suggested a general symmetrization procedure by adding a "virtual" contribution $\underline{K}$ (i.e., a total divergence) to the canonical tensor, which resulted in the familiar symmetric tensor $\underline{T}_{\text{sym}}$. Usually, the symmetric Belinfante–Rosenfeld energy–momentum tensor is considered to be the physically meaningful tensor, because it is gauge-invariant and is naturally coupled to gravity [127]. However, recent studies of the *local* momentum and angular momentum densities of optical fields have shown that it is actually the *canonical* momentum density from $\underline{T}$ (in the Coulomb gauge) that determines the optical pressure on small dipole particles or atoms that is observed in experiments [52–54,59–61,66,68–72,74,75,85]. Furthermore, as we will see in the next section, the observable separation of the spin and orbital angular momenta of light [46,48–56] is also described by the *canonical* tensors rather than the symmetrized Belinfante–Rosenfeld tensors [58,68]. Notably, this need for the canonical energy–momentum and corresponding angular momentum tensors has also been actively discussed in relation to the separation of the spin and orbital angular momenta of gluon fields in quantum chromodynamics (QCD) [88].

To obtain more familiar expressions for the canonical energy–momentum tensor, we expand Eqs. (8.18) and (8.19) in terms of a relative frame, using the spacetime splits $\mathbf{F}_{e,m} = \vec{E}_{e,m} + \vec{B}_{e,m}I$ (that become constrained to $\mathbf{F} = \vec{E} + \vec{B}I$, with $\mathbf{G} = \mathbf{F}I^{-1} = \vec{B} - \vec{E}I$), as well as the (transverse) Coulomb gauge $a_e = \vec{A}\gamma_0$ and $a_m = \vec{C}\gamma_0$ for the potentials[27]:

$$
\underline{T}(\gamma_0)\gamma_0 = \varepsilon + \vec{P},
$$
$$
\underline{T}(\gamma_i)\gamma_0 = -(\vec{\sigma}_i \cdot \dot{\vec{\nabla}})\left[\vec{E}_e \cdot \dot{\vec{A}} + \dot{\vec{A}} \times \vec{B}_e + \vec{B}_m \cdot \dot{\vec{C}} - \dot{\vec{C}} \times \vec{E}_m\right],
$$
$$
\rightarrow -\frac{1}{2}(\vec{\sigma}_i \cdot \dot{\vec{\nabla}})\left[\vec{E} \cdot \dot{\vec{A}} + \vec{B} \cdot \dot{\vec{C}} + \dot{\vec{A}} \times \vec{B} - \dot{\vec{C}} \times \vec{E}\right], \tag{8.25}
$$

$$
\overline{T}(\gamma_0)\gamma_0 = \varepsilon + \dot{\vec{\nabla}}\left[\vec{E}_e \cdot \dot{\vec{A}} + \vec{B}_m \cdot \dot{\vec{C}}\right],
$$
$$
\rightarrow \varepsilon + \frac{1}{2}\dot{\vec{\nabla}}\left[\vec{E} \cdot \dot{\vec{A}} + \vec{B} \cdot \dot{\vec{C}}\right],
$$
$$
\overline{T}(\gamma_i)\gamma_0 = -\vec{P} \cdot \vec{\sigma}_i - \dot{\vec{\nabla}}\left[(\dot{\vec{A}} \times \vec{B}_e - \dot{\vec{C}} \times \vec{E}_m) \cdot \vec{\sigma}_i\right],
$$
$$
\rightarrow -\vec{P} \cdot \vec{\sigma}_i - \frac{1}{2}\dot{\vec{\nabla}}\left[(\dot{\vec{A}} \times \vec{B} - \dot{\vec{C}} \times \vec{E}) \cdot \vec{\sigma}_i\right]. \tag{8.26}
$$

Here $\varepsilon = \frac{1}{2}(|\vec{E}|^2 + |\vec{B}|^2)$ and $\vec{P} = \vec{E} \times \vec{B}$ are the energy density and Poynting vector derived in Section 7.5.1, and we have used the fact that the energy densities and Poynting vectors for the constituent fields, $\varepsilon_e, \varepsilon_m, \vec{P}_e, \vec{P}_m$, sum to the total field quantities according to (8.20): $\varepsilon = \varepsilon_e + \varepsilon_m$, $\vec{P} = \vec{P}_e + \vec{P}_m$. The vector part of $\overline{T}(\gamma_0)\gamma_0$ represents the canonical (orbital) momentum density $\vec{P}_o$, i.e., $\overline{T}(\gamma_0)\gamma_0 = \varepsilon + \vec{P}_o$ (we describe this quantity below).

---

[27] Recall from Section 6.2 that this choice of gauge is equivalent to using the gauge-invariant transverse bivector potential $\mathbf{Z} = z_\perp \gamma_0$ in the frame γ_0, and then fixing the gauge to make the gauge-dependent longitudinal spinor potential $\psi_z = z_\parallel \gamma_0$ vanish, since it is arbitrary and thus unmeasurable.

The associated Noether charges from (8.16) are the integrals of the energy and momentum densities over all space, i.e., the total energy and momentum of the field:

$$\int \varepsilon \, d^3x = \text{const}, \qquad \int \vec{P} \, d^3x = \int \vec{P}_o \, d^3x = \text{const}. \tag{8.27}$$

These 4 conserved charges are derived from (8.16) and (8.18) by choosing the generator b to be the 4 orthogonal vector directions γ_μ. Note that these equations are valid only for sufficiently localized square-integrable fields. Importantly, the second equation shows that the difference between the canonical and Poynting momentum densities (i.e., the difference between the canonical and symmetric energy–momentum tensors) does not affect the integral momentum value.

We also expand the correction Belinfante tensor (8.22) and (8.23) in a similar way:

$$\underline{K}(\gamma_0)\gamma_0 = 0, \tag{8.28}$$

$$\begin{aligned}
\underline{K}(\gamma_i)\gamma_0 &= -\left[(\vec{E}_e \cdot \vec{\nabla})(\vec{A} \cdot \vec{\sigma}_i) + (\vec{B}_m \cdot \vec{\nabla})(\vec{C} \cdot \vec{\sigma}_i)\right] - \frac{1}{2}\left[\vec{E}_e(\vec{E}_e \cdot \vec{\sigma}_i) + \vec{B}_e(\vec{B}_e \cdot \vec{\sigma}_i) + \vec{E}_m(\vec{E}_m \cdot \vec{\sigma}_i) + \vec{B}_m(\vec{B}_m \cdot \vec{\sigma}_i)\right] \\
&\quad + \left[\vec{B}_e \times \vec{\nabla}(\vec{A} \cdot \vec{\sigma}_i) - \vec{E}_m \times \vec{\nabla}(\vec{C} \cdot \vec{\sigma}_i)\right], \\
&\rightarrow -\frac{1}{2}\left[(\vec{E} \cdot \vec{\nabla})(\vec{A} \cdot \vec{\sigma}_i) + (\vec{B} \cdot \vec{\nabla})(\vec{C} \cdot \vec{\sigma}_i)\right] - \frac{1}{2}\left[\vec{E}(\vec{E} \cdot \vec{\sigma}_i) + \vec{B}(\vec{B} \cdot \vec{\sigma}_i) - \vec{B} \times \vec{\nabla}(\vec{A} \cdot \vec{\sigma}_i) + \vec{E} \times \vec{\nabla}(\vec{C} \cdot \vec{\sigma}_i)\right],
\end{aligned}$$

$$\begin{aligned}
\overline{K}(\gamma_0)\gamma_0 &= -\left[(\vec{E}_e \cdot \vec{\nabla})\vec{A} + (\vec{B}_m \cdot \vec{\nabla})\vec{C}\right], \\
&\rightarrow -\frac{1}{2}\left[(\vec{E} \cdot \vec{\nabla})\vec{A} + (\vec{B} \cdot \vec{\nabla})\vec{C}\right],
\end{aligned} \tag{8.29}$$

$$\begin{aligned}
\overline{K}(\gamma_i)\gamma_0 &= \frac{1}{2}\left[(\vec{\sigma}_i \cdot \vec{E}_e)\vec{E}_e + (\vec{\sigma}_i \cdot \vec{B}_e)\vec{B}_e + (\vec{\sigma}_i \cdot \vec{E}_m)\vec{E}_m + (\vec{\sigma}_i \cdot \vec{B}_m)\vec{B}_m\right] - \left[\vec{\sigma}_i \cdot (\vec{B}_e \times \vec{\nabla})\vec{A} - \vec{\sigma}_i \cdot (\vec{E}_m \times \vec{\nabla})\vec{C}\right], \\
&\rightarrow \frac{1}{2}\left[(\vec{\sigma}_i \cdot \vec{E})\vec{E} + (\vec{\sigma}_i \cdot \vec{B})\vec{B}\right] - \frac{1}{2}\left[\vec{\sigma}_i \cdot (\vec{B} \times \vec{\nabla})\vec{A} - \vec{\sigma}_i \cdot (\vec{E} \times \vec{\nabla})\vec{C}\right].
\end{aligned}$$

For all the above tensors, the terms involving only the "electric" parts (i.e., $\mathbf{F}_e$, a_e, $\vec{E}_e$, $\vec{B}_e$, and $\vec{A}$) are equivalent to the corresponding tensors for traditional electromagnetism [68,127]. The dual-symmetric Lagrangian adds complementary "magnetic" terms to these traditional tensors. In turn, the constrained dual-symmetric Lagrangian makes both these halves correspond to the same field, producing dual-symmetrized canonical tensors that agree with those obtained very recently in [68].

8.6.1. Orbital and spin momentum densities

In the above expressions for the energy–momentum tensors, it is important to analyze the energy–momentum densities, which are given by the vector terms in $\underline{T}_{\text{sym}}(\gamma_0)\gamma_0$, $\underline{K}_{\text{sym}}(\gamma_0)\gamma_0$, and their adjoints. As we shall see in the next section, these densities will be responsible for the generation of the angular momentum of the field in a way similar to (7.43), but with a clear separation of the spin and orbital contributions.

First, observe in (8.28) that the correction to the timelike component vanishes: $\underline{K}(\gamma_0) = 0$, so the canonical tensor $\underline{T}(\gamma_0)$ yields precisely the same energy–momentum density as the symmetric tensor $\underline{T}_{\text{sym}}(\gamma_0)$. That is, both the correct energy density $\varepsilon = (|\vec{E}|^2 + |\vec{B}|^2)/2$ and Poynting vector $\vec{P} = \vec{E} \times \vec{B}$ are obtained in (8.25). These quantities satisfy a continuity equation, i.e., the Poynting theorem, or energy-transport equation [155].

In contrast, the adjoint correction in (8.29) is nonzero. Remarkably, the vector terms in $\overline{T}(\gamma_0)\gamma_0$ and $\overline{K}(\gamma_0)\gamma_0$ are the so-called *canonical (orbital) and spin momentum densities* [64,65,68,73,74]:

$$\begin{aligned}
\vec{P}_o &= \overline{T}(\gamma_0)\gamma_0 - \varepsilon, \\
&= \dot{\vec{\nabla}}[\vec{E}_e \cdot \dot{\vec{A}} + \vec{B}_m \cdot \dot{\vec{C}}], \\
&\rightarrow \frac{1}{2}\dot{\vec{\nabla}}[\vec{E} \cdot \dot{\vec{A}} + \vec{B} \cdot \dot{\vec{C}}],
\end{aligned} \tag{8.30}$$

$$\begin{aligned}
\vec{P}_s &= \overline{K}(\gamma_0)\gamma_0, \\
&= -[(\vec{E}_e \cdot \vec{\nabla})\vec{A} + (\vec{B}_m \cdot \vec{\nabla})\vec{C}], \\
&\rightarrow -\frac{1}{2}[(\vec{E} \cdot \vec{\nabla})\vec{A} + (\vec{B} \cdot \vec{\nabla})\vec{C}].
\end{aligned} \tag{8.31}$$

Together, the orbital and spin momentum densities form the Poynting vector in the corresponding sector of the symmetric tensor $\overline{T}_{\text{sym}}(\gamma_0)\gamma_0$: $\vec{P}_o + \vec{P}_s = \vec{P}$. Furthermore, the wedge products of these momentum densities generate, respectively, the orbital and spin angular momenta of the field [65,68,74,127,243,244], as we shall see in the next section. It follows from

Eq. (8.27) that the spin momentum density does not contribute to the integral momentum of the localized field:

$$\int \vec{P}_s \, d^3x = 0. \tag{8.32}$$

Importantly, although it is the total Poynting vector that is gauge-invariant and coupled to gravity [127], the spin and orbital degrees of freedom manifest themselves drastically differently in local interactions with matter. Optical and quantum measurements of the momentum density of light reveal the canonical (orbital) momentum (in the Coulomb gauge) (8.30), while the spin momentum (8.31) remains largely unobservable and virtual [68,69,74,85]. The appearance of the canonical momentum density in optics can be explained by the fact that it has a clear physical interpretation. Namely, for monochromatic optical fields, it is given by the local expectation value of the quantum momentum operator $\hat{\vec{p}} \propto -i\vec{\nabla}$, which makes it proportional to the local gradient of the phase, i.e., the *local wavevector* of the field [64,69]. In contrast, the Poynting vector does not have such intuitively-clear wave interpretation.

8.7. Canonical angular momentum tensor

We are now in a position to consider rotational symmetries of spacetime, which produces the conservation of the relativistic angular momentum of the field. Suppose that we perform a general restricted Lorentz transformation (i.e., a spatial rotation, boost, or combination thereof). As we discussed in Section 3.7, the bivectors themselves are the generators of this group, so the appropriate transformation is $\underline{U}(a, \alpha) = U_\alpha a \widetilde{U}_\alpha$, where $U_\alpha = \exp(\alpha \mathbf{R}/2)$ and $\mathbf{R}$ is a unit bivector corresponding to the plane of rotation. A symmetry of the Lagrangian follows if we first transform the fields $a' = \underline{U}(a, \alpha) = U_\alpha a \widetilde{U}_\alpha$ then counter-transform the coordinates $x' = \overline{U}(x, \alpha) = \widetilde{U}_\alpha x U_\alpha$, as discussed in (8.11).

From this symmetry transformation, the corresponding increments (8.12) have the form

$$\delta x(\mathbf{R}) = -\mathbf{R} \cdot x, \tag{8.33a}$$
$$\delta a(\mathbf{R}) = \mathbf{R} \cdot a + [(-\mathbf{R} \cdot x) \cdot \nabla]a. \tag{8.33b}$$

The change (8.13) of the Lagrangian thus produces the conserved Noether current

$$
\begin{aligned}
\underline{M}(\mathbf{R}) &= \sum_\mu \gamma_\mu \left\langle (\partial_\mu a_e) \delta a_e(\mathbf{R}) + (\partial_\mu a_m) \delta a_m(\mathbf{R}) \right\rangle_0 - \delta x(\mathbf{R}) \mathcal{L}, \\
&= (a_e \cdot \mathbf{R} + [(\mathbf{R} \cdot x) \cdot \nabla]a_e) \cdot \mathbf{F}_e + (a_m \cdot \mathbf{R} + [(\mathbf{R} \cdot x) \cdot \nabla]a_m) \cdot \mathbf{F}_m - (\mathbf{R} \cdot x) \cdot \nabla \mathcal{L}, \\
&\rightarrow \frac{1}{2} \left[(a_e \cdot \mathbf{R} + [(\mathbf{R} \cdot x) \cdot \nabla]a_e) \cdot \mathbf{F} + (a_m \cdot \mathbf{R} + [(\mathbf{R} \cdot x) \cdot \nabla]a_m) \cdot \mathbf{G} \right].
\end{aligned}
\tag{8.34}
$$

This conserved current is the *canonical angular momentum tensor*, which transforms bivectors (i.e., Lorentz transformation generators) into vectors (linear momenta), as we discussed in Section 7.5.2.

After comparing (8.34) to the form of $\underline{T}$ in (8.18), we can make the simplification

$$\underline{M}(\mathbf{R}) = \underline{L}(\mathbf{R}) + \underline{S}(\mathbf{R}), \qquad \underline{L}(\mathbf{R}) = \underline{T}(\mathbf{R} \cdot x), \tag{8.35}$$

where

$$
\begin{aligned}
\underline{S}(\mathbf{R}) &\equiv (a_e \cdot \mathbf{R}) \cdot \mathbf{F}_e + (a_m \cdot \mathbf{R}) \cdot \mathbf{F}_m, \\
&\rightarrow \frac{1}{2}[(a_e \cdot \mathbf{R}) \cdot \mathbf{F} + (a_m \cdot \mathbf{R}) \cdot \mathbf{G}].
\end{aligned}
\tag{8.36}
$$

Here, $\underline{L}(\mathbf{R})$ is the *orbital angular momentum* of the field (including the boost momentum), which is related to the canonical energy–momentum tensor in the same way as in Eq. (7.45) for the symmetric (Belinfante–Rosenfeld) tensors. The additional term $\underline{S}(\mathbf{R})$ that appears is the *spin angular momentum* tensor. Importantly, this spin tensor describes *intrinsic* (i.e., independent of the radius-vector x) angular momentum, and it *cannot* be obtained from the energy–momentum tensor.

The corresponding adjoint angular momentum tensors transform unit vectors b (linear momentum directions) into bivectors (associated angular momenta), so can be easier to interpret physically. The symmetric and canonical adjoint angular momentum tensors have the intuitive forms

$$\overline{M}_{\text{sym}}(b) = x \wedge \overline{T}_{\text{sym}}(b), \qquad \overline{M}(b) = x \wedge \overline{T}(b) + \overline{S}(b) \equiv \overline{L}(b) + \overline{S}(b), \tag{8.37}$$

where the adjoint spin angular momentum tensor is

$$
\begin{aligned}
\overline{S}(b) &= a_e \wedge (b \cdot \mathbf{F}_e) + a_m \wedge (b \cdot \mathbf{F}_m), \\
&\rightarrow \frac{1}{2}[a_e \wedge (b \cdot \mathbf{F}) + a_m \wedge (b \cdot \mathbf{G})].
\end{aligned}
\tag{8.38}
$$

We detail these angular momentum tensors in Table 19 for reference.

Table 19
Canonical angular momentum tensors, conserved by Lorentz transformations. Unlike the symmetric tensor in Table 15, the canonical tensor naturally separates into an extrinsic *orbital* part $\underline{L}$ that contains the wedge product of the orbital part $\vec{P}_o$ of the Poynting vector $\vec{P}$ in Table 18, as well as a separate *intrinsic spin* part $\underline{S}$ containing the spin vector $\vec{S}$. As with the energy–momentum tensor, the traditional Lagrangian produces electric-biased quantities, while the dual-symmetric Lagrangian adds complementary magnetic-biased terms. Constraining the Lagrangian symmetrizes the field contributions.

<div align="center">

Canonical Angular Momentum Tensor

</div>

$$\boxed{\underline{M}(\mathbf{R}) = \underline{L}(\mathbf{R}) + \underline{S}(\mathbf{R})} \qquad \boxed{\overline{M}(b) = \overline{L}(b) + \overline{S}(b)}$$
$$\underline{L}(\mathbf{R}) = \underline{T}(\mathbf{R} \cdot x) \qquad \overline{L}(b) = x \wedge \overline{T}(b)$$

$$\overline{S}(\gamma_0) = \boxed{\vec{S}}\,I^{-1} \qquad \overline{L}(\gamma_0) = [\varepsilon\vec{x} - (ct)\boxed{\vec{P}_o}] + \vec{x} \times \boxed{\vec{P}_o}\,I^{-1}$$
$$M^{\alpha\beta\gamma} = L^{\alpha\beta\gamma} + S^{\alpha\beta\gamma} \qquad L^{\alpha\beta\gamma} = x^\alpha T^{\beta\gamma} - x^\beta T^{\alpha\gamma}$$

Traditional:
$$\boxed{\underline{S}(\mathbf{R}) = (a \cdot \mathbf{R}) \cdot \mathbf{F}} \qquad \boxed{\overline{S}(b) = a \wedge (b \cdot \mathbf{F})}$$
$$z = a_e \equiv a \qquad \vec{S} = \vec{E} \times \vec{A}$$
$$\mathbf{F} = \nabla a_e = \mathbf{F}_e \qquad S^{\alpha\beta\gamma} = F^{\gamma\alpha}a^\beta - F^{\gamma\beta}a^\alpha$$

Dual-symmetric:
$$\boxed{\underline{S}(\mathbf{R}) = (a_e \cdot \mathbf{R}) \cdot \mathbf{F}_e + (a_m \cdot \mathbf{R}) \cdot \mathbf{F}_m}$$
$$\boxed{\overline{S}(b) = a_e \wedge (b \cdot \mathbf{F}_e) + a_m \wedge (b \cdot \mathbf{F}_m)}$$
$$z = a_e + a_m I \qquad \vec{S} = \vec{E}_e \times \vec{A} + \vec{B}_m \times \vec{C}$$
$$\mathbf{F} = \nabla z = \mathbf{F}_e + \mathbf{F}_m I \qquad S^{\alpha\beta\gamma} = F_e^{\gamma\alpha}a_e^\beta - F_e^{\gamma\beta}a_e^\alpha + F_m^{\gamma\alpha}a_m^\beta - F_m^{\gamma\beta}a_m^\alpha$$

Constrained:
$$\boxed{\underline{S}(\mathbf{R}) = \frac{1}{2}[(a_e \cdot \mathbf{R}) \cdot \mathbf{F} + (a_m \cdot \mathbf{R}) \cdot \mathbf{G}]}$$
$$\boxed{\overline{S}(b) = \frac{1}{2}[a_e \wedge (b \cdot \mathbf{F}) + a_m \wedge (b \cdot \mathbf{G})]}$$
$$\mathbf{F}_e = \mathbf{F}_m \to \frac{\mathbf{F}}{\sqrt{2}} \qquad \vec{S} \to \frac{1}{2}[\vec{E} \times \vec{A} + \vec{B} \times \vec{C}]$$
$$\mathbf{F} = \mathbf{G}I \qquad S^{\alpha\beta\gamma} \to \frac{1}{2}\left[F^{\gamma\alpha}a_e^\beta - F^{\gamma\beta}a_e^\alpha + G^{\gamma\alpha}a_m^\beta - G^{\gamma\beta}a_m^\alpha\right]$$

The appearance of the spin angular momentum in (8.36) and (8.38) indicates the principal difference between the canonical and symmetrized pictures of the angular momentum in field theory. Namely, the Belinfante–Rosenfeld angular momentum tensor $\overline{M}_{\text{sym}}$ in (8.37) (and Section 7.5.2) is simply the wedge product of the symmetric energy–momentum tensor $\overline{T}_{\text{sym}}$ with a coordinate radius x, so does not contain any new information about local properties of the field. In contrast, the canonical angular momentum tensor $\overline{M}$ has both an extrinsic orbital part $\overline{L}$ and an independent intrinsic spin part $\overline{S}$, which signifies additional degrees of freedom for the field. Critically, the *spin and orbital degrees of freedom are separated only in the canonical picture*. As we mentioned before, these degrees of freedom have been historically considered to have no separate meaning in both orthodox quantum electrodynamics and classical electromagnetic field theory, but nonetheless they clearly correspond to separate observable effects in local light–matter interactions in optical experiments [48–56]. Furthermore, the integral values (Noether charges) of the spin and orbital angular momenta are *separately conserved* quantities [47], and the corresponding Noether currents $\overline{L}$ and $\overline{S}$ can also be brought to meaningful separately conserved forms [58]. Thus, akin to the local momentum densities, the optical angular momentum properties are described by the *canonical* Noether currents rather than the symmetrized Belinfante–Rosenfeld currents [58,68,69,74] (see also [88] for a discussion of this same issue in QCD).

For simplicity, we only compute the timelike components of these tensors in a relative frame (in the Coulomb gauge):

$$\overline{M}_{\text{sym}}(\gamma_0) = [\varepsilon \vec{x} - (ct)\vec{P}] + \vec{x} \times \vec{P} I^{-1} \equiv \vec{N} + \vec{J} I^{-1}, \tag{8.39}$$

$$\overline{L}(\gamma_0) = [\varepsilon \vec{x} - (ct)\vec{P}_o] + \vec{x} \times \vec{P}_o I^{-1} \equiv \vec{N}_o + \vec{L} I^{-1}, \tag{8.40}$$

$$\overline{S}(\gamma_0) = \left[\vec{E}_e \times \vec{A} + \vec{B}_m \times \vec{C} \right] I^{-1} \equiv \vec{S} I^{-1},$$
$$\rightarrow \frac{1}{2} [\vec{E} \times \vec{A} + \vec{B} \times \vec{C}] I^{-1}. \tag{8.41}$$

For the orbital angular momentum $\overline{L}$, the associated orbital linear momentum $\vec{P}_o$ is given by (8.30), while the energy density ε and full Poynting vector $\vec{P}$ are the usual expressions. Recall from our discussion in Section 7.5.2 that each bivector term in the angular momentum that involves I generates spatial rotations, while the remaining bivectors generate boosts. One can then clearly see that the intrinsic spin angular momentum (8.41) is a purely spatial rotation (around the axis $\vec{S}$) that appears explicitly only in the canonical angular momentum tensor [68].

The Noether charges (integral conserved quantities) associated with the symmetric and canonical angular-momentum tensors are as follows:

$$\int \vec{J} \, d^3x = \int (\vec{L} + \vec{S}) \, d^3x = \text{const}, \qquad \int \vec{N} \, d^3x = \int \vec{N}_o \, d^3x = \text{const}. \tag{8.42}$$

These 6 conserved charges represent the integral angular momentum and boost momentum of the field, and are derived from (8.16) and (8.35) by choosing the generator **R** to be the 6 orthogonal bivector directions $\gamma_{\mu\nu}$. Together with the 4 energy–momentum integrals (8.27) they form 10 Poincaré invariants of the field associated with the 10-parameter Poincaré group of spacetime symmetries [19,31,68,127].

The seeming absence of the intrinsic spin contribution in the Belinfante–Rosenfeld angular momentum (8.39) causes the so-called "spin-of-a-plane-wave paradox" [74,127,243–246]. Namely, the local density of the total angular momentum $\overline{M}_{\text{sym}}(\gamma_0)$ vanishes in a circularly polarized plane electromagnetic wave, while the spin density $\overline{S}(\gamma_0)$ is non-zero, as it should be. The explanation of this paradox lies in the fact that, in contrast to the extrinsic orbital angular momentum (8.40), the relation between the spin angular momentum density (8.41) and the spin part of the momentum density (8.31) is essentially *nonlocal*:

$$\int \vec{S} \, d^3x = \int \vec{x} \times \vec{P}_s \, d^3x, \tag{8.43}$$

so applies only for sufficiently localized fields, such as wave packets.

We can see that the concepts of the canonical (orbital) momentum and independent spin density is more physically meaningful and intuitive than the concept of the Poynting vector and its corresponding Belinfante–Rosenfeld angular momentum. Akin to the canonical momentum density, the spin angular momentum density has a clear physical meaning in the case of monochromatic optical fields. Namely, it is proportional to the ellipticity of the local polarization of the field and is directed along its normal axis. In other words, the rotation of the electric and magnetic field vectors in optical fields generates a well-defined spin angular momentum density.

Most importantly for our study, the separation of the spin and orbital degrees of freedom is closely related to the *dual symmetry* between the electric and magnetic contributions. The constrained forms of the spin momentum density in (8.41), as well as the orbital momentum density in (8.30), are dual-symmetric in the fields (as recently emphasized in [68]), while the traditional Lagrangian produces dual-asymmetric spin and orbital characteristics, containing only the electric part (e.g., $\mathbf{F}_e$ and a_e) [68,127]. This asymmetry is closely related to the fact that the spin and orbital degrees of freedom of electromagnetic waves are coupled to dual-asymmetric matter via the *electric*-dipole coupling. Nonetheless, for vacuum fields, the dual-symmetric canonical Noether currents provide a more natural and self-consistent picture [68].

8.8. Helicity pseudocurrent

In addition to the Poincaré symmetries of spacetime (translations and rotations) considered above, the Lagrangian $\mathcal{L}_{\text{dual}}$ in (8.4) exhibits one more continuous symmetry, namely, the *dual symmetry*. Hence, we can determine the proper conserved quantity that should be associated with the Lagrangian invariance with respect to the dual "electric–magnetic" rotations $\psi = \exp(\theta I/2)$. Importantly, this symmetry does *not* exist in the traditional Lagrangian (8.1) since only the *electric* vector potential a_e is usually considered, which has no complex phase.

The proper field increment for the transformations $a' = \psi a \psi^*$ is

$$\delta a(I) = aI^{-1}. \tag{8.44}$$

Notably, this field increment is simply the Hodge-dual of a. Noether's theorem (8.15) thus produces a conserved *helicity pseudovector current*

$$
\begin{aligned}
\underline{X}(I) &= \sum_\mu \gamma_\mu \left\langle (\partial_\mu a_e) a_e I^{-1} + (\partial_\mu a_m) a_m I^{-1} \right\rangle_4 , \\
&= \sum_\mu \gamma_\mu [(\partial_\mu a_e) \cdot a_e + (\partial_\mu a_m) \cdot a_m] I^{-1}, \\
&= a_e \wedge (\nabla a_e) + a_m \wedge (\nabla a_m), \\
&= a_e \wedge \mathbf{F}_e + a_m \wedge \mathbf{F}_m, \\
&= [a_e \cdot (\mathbf{F}_e I^{-1}) + a_m \wedge (\mathbf{F}_m I^{-1})] I, \\
&\rightarrow \frac{1}{2} [a_e \cdot \mathbf{G} + a_m \cdot \mathbf{F}] I.
\end{aligned}
\tag{8.45}
$$

Interestingly, after factoring out I from the pseudovector, each constituent potential becomes coupled to the *Hodge-dual* of its associated field. If the additional constraint in the last line is imposed, then the vector potentials effectively cross-couple through the total field $\mathbf{F} = \mathbf{G}I$. We detail this pseudovector current in Table 20 for reference. It is also worth noting that the Weinberg–Witten theorem [22,26,38] that forbids conserved vector currents for massless spin-1 electromagnetic fields is not violated here, for two reasons: first, the conserved helicity pseudovector current here is not a grade-1 vector, but is rather a grade-3 pseudovector; second, as noted by Gaillard and Zumino [115], this conserved current is not gauge-invariant.

Writing this result in a relative frame using the Coulomb gauge yields

$$
\underline{X}(I) = (\chi I + \vec{S} I^{-1}) \gamma_0, \qquad \chi = \vec{A} \cdot \vec{B}_e - \vec{C} \cdot \vec{E}_m,
$$
$$
\rightarrow \frac{1}{2} (\vec{A} \cdot \vec{B} - \vec{C} \cdot \vec{E}).
\tag{8.46}
$$

The pseudoscalar term is the *helicity density* χ of the electromagnetic field. When integrated over the spatial coordinates, it produces the conserved integral helicity, i.e., the corresponding Noether charge (8.16):

$$
\int \chi \, d^3 x = \text{const.}
\tag{8.47}
$$

This integral helicity is equal to the difference between the numbers of right-hand and left-hand circularly polarized photons [78,98–100,103–105].

The pseudovector term in the helicity pseudocurrent (8.46) is the helicity flux density $\overline{S}(\gamma_0) = \vec{S} I^{-1}$, which is precisely equal to the *spin density* (8.41) [68,105,106]. Intriguingly, the constrained form of the helicity χ matches the known expression for the conserved optical helicity in vacuum [68,78,99,103–106]. Indeed, the field-asymmetric electric part (or magnetic part) is not independently conserved since the corresponding symmetry in the Lagrangian does not exist unless the dual-symmetric form (8.4) is used. This fact gives strong indirect support for the requirement of the additional field $\mathbf{F}_m$ in the Lagrangian: without its complementary contribution, the conserved optical helicity χ is not obtainable as a Noether current from the Lagrangian.[28]

8.9. Sources and symmetry-breaking

Thus far we have focused on what can be said by constructing the general dual-symmetric Lagrangian in (8.4) for a pure vacuum field that corresponds to an intrinsically complex vector potential z with the appropriate gauge freedom. We found that we could reproduce (and generalize) all the traditional results that stem from the Lagrangian (8.1) using this ansatz. In particular, we derived dual-symmetrized formulas for the various canonical Noether currents (i.e., the energy–momentum, angular-momentum, spin, and helicity currents). However, an isolated vacuum field is an incomplete picture, since such a field can only be measured through coupling to charged probes.

We now detail how the dual symmetry of the vacuum field is broken by the gauge mechanism that couples it to charge. We will find that a charge essentially introduces a preferred angle of rotation for the dual symmetry, and thus will couple only to a (conventional) electric part of the field that is governed by a pure electric vector potential a_e. The magnetic potential a_m can remain after breaking the symmetry, but it is uncoupled to the electric charge and is thus unmeasurable (except by its effect on gravity, in principle [196]). However, one can speculate that this extra potential is still required for the complete description of the vacuum field, and may be measurable by its coupling to other fields not considered in isolated electromagnetism (e.g., gravity).

---

[28] We note, however, that by carefully considering the symmetries of the integrated *action* it is still possible to derive the conserved integral helicity (8.47) using the traditional Lagrangian [99,102,204]. Curiously, this alternative method reproduces the *constrained* form of the dual-symmetric helicity density that we derive here, even though the form of the spin vector $\vec{S}$ for the traditional Lagrangian is not the same as for the constrained dual-symmetric Lagrangian [68].

Table 20
Canonical helicity pseudocurrent, conserved by the continuous *dual symmetry* of the vacuum Lagrangian. This symmetry does not exist for the traditional Lagrangian, for which the conserved *integral* helicity can only be obtained from the symmetry of the *action* [99,102,204]. The pseudoscalar part of the helicity pseudocurrent is the helicity density χ, while the relative pseudovector part is precisely the spin density $\vec{S}$ from Table 19. Curiously, only the *constrained* Lagrangian produces the field-symmetrized expression for the helicity density that is known in the literature [68,99,102–106,214]; this Lagrangian also produces the corresponding dual-symmetric spin density $\vec{S}$, which differs from the dual-asymmetric spin produced by the traditional Lagrangian [68] (see Table 19).

Helicity Pseudocurrent
$\underline{X}(I) = [\chi I + \vec{S}I^{-1}]\gamma_0$

Traditional:	$z = a_e \equiv a, \quad \mathbf{F} = \nabla a_e = \mathbf{F}_e$
	No dual symmetry of the Lagrangian.
Dual-symmetric:	$z = a_e + a_m I, \quad \mathbf{F} = \nabla z = \mathbf{F}_e + \mathbf{F}_m I$
	$\boxed{\underline{X}(I) = [a_e \cdot (\mathbf{F}_e I^{-1}) + a_m \cdot (\mathbf{F}_m I^{-1})]I}$
	$\chi = \vec{A} \cdot \vec{B}_e - \vec{C} \cdot \vec{E}_m$
Constrained:	$\mathbf{F}_e \to \mathbf{F}_m \to \mathbf{F}/\sqrt{2} = \mathbf{G}I/\sqrt{2}$
	$\boxed{\underline{X}(I) \to \frac{1}{2}[a_e \cdot \mathbf{G} + a_m \cdot \mathbf{F}]I}$
	$\chi \to \frac{1}{2}(\vec{A} \cdot \vec{B} - \vec{C} \cdot \vec{E})$

8.9.1. Gauge mechanism

Physically, a pure vector potential a_e acts as a *connection field* for the proper parallel transport of a coupled charged field across spacetime, which provides spacetime with an effective *curvature* that corresponds to the electromagnetic field $\mathbf{F} = \nabla \wedge a_e$ [159–163,165,167–170]. This curvature alters the geodesics for the charged field from its typically straight lines, which can be understood as the effect of the electromagnetic force on the charged field. The gauge freedom of a_e does not affect this induced curvature, so is not directly observable in terms of this physical force. However, the influence of a_e as a dynamical field can be indirectly observed in other topologically-induced field dynamics, such as the Aharonov–Bohm effect [164,166,247].

From this point of view, a gauge field like a_e can be understood as a dynamical mechanism for effectively *warping the spacetime metric* by using a simpler flat spacetime metric as an static embedding space. The curvature of the effective metric then depends on the charge configuration, in an entirely analogous manner to how the curvature of the actual spacetime metric depends on the configuration of the energy–momentum density in the gravitational theory of General Relativity [127,189]. Indeed, it has been shown that all currently measured predictions of General Relativity may be reproduced by a simpler dynamical gauge field reformulation over a flat and static spacetime metric [8,9,175]; this gauge theory of gravity only appreciably differs in its (as-yet-unmeasured) predictions for spin-fields and torsion, and in its inability to handle global changes to the topology of spacetime (which may be unmeasurable or unphysical anyway).

To make the gauge origins of electromagnetism evident, we consider a complex scalar field $\phi(x)$ as the simplest prototype for a charged matter spinor field [168]. The complexity of the field is the origin of its charge; indeed, the two possible signs of the charge correspond to the two possible orientations of the pseudoscalar I of the spacetime 4-volume.

The field Lagrangian density for this complex field ϕ must be a proper scalar field so it must involve only quadratic terms like

$$\mathcal{L}(x) = \nabla\phi^* \cdot \nabla\phi - m^2\phi^*\phi. \tag{8.48}$$

According to the Euler–Lagrange equations $\nabla \cdot (\delta\mathcal{L}/\delta\nabla\phi^*) = \delta\mathcal{L}/\delta\phi^*$, this bare Lagrangian produces the relativistic Klein–Gordon equation for the scalar field: $(\nabla^2 + m^2)\phi = 0$, as well as its complex conjugate. This Lagrangian is invariant

under the *global* phase transformation $\phi(x) \mapsto \phi(x)\exp(\theta I)$ (which is in fact identical to the dual-symmetry phase transformation for the electromagnetic field from Section 5.2).

The idea for a gauge field then arises from the observation that the mass term involving $\phi^*\phi = |\phi|^2$ is also invariant under a *local* phase rotation of the complex field $\phi \mapsto \exp[q_e \theta(x)I]\phi$, but the kinetic term that involves derivatives $\nabla\phi$ is not. Here $q_e = \pm|q_e|$ is an appropriate coupling strength (i.e., charge) that also indicates the orientation for the pseudoscalar I. Performing the phase rotation with the derivative yields $\nabla\phi \mapsto \exp(q_e\theta I)(\nabla\phi + q_e[\nabla\theta]\phi I)$. The extra term involving the gradient of the local phase spoils the invariance of the Lagrangian.

To fix this problem, the phase gradient can be absorbed into the definition of an auxiliary vector field a_e that must be the same type of blade as the phase gradient $\nabla\theta$ (i.e., a pure vector field). Specifically, we can modify the derivative in the Lagrangian to a *covariant derivative* that includes this field

$$D \equiv \nabla - q_e a_e I. \tag{8.49}$$

The extra term from the phase rotation of ϕ can then be canceled by a corresponding modification of the vector field

$$a_e \mapsto a_e + \nabla\theta. \tag{8.50}$$

Since this shift of the vector field must not change anything fundamental, the field a_e must be defined only up to the addition of an arbitrary gradient term, which is the origin of the gauge freedom of a_e. After the introduction of this gauge field, the phase rotation commutes with the covariant derivative

$$D\phi \mapsto (D - q_e[\nabla\theta]I)(\exp[q_e\theta I]\phi) = \exp(q_e\theta I)D\phi. \tag{8.51}$$

We then rewrite the Lagrangian (8.48) in a properly invariant form by replacing all bare derivatives with covariant derivatives

$$\mathcal{L}(x) = D\phi^* \cdot D\phi - m^2\phi^*\phi \tag{8.52}$$
$$= \nabla\phi^* \cdot \nabla\phi + q_e w \cdot a_e I + (|q_e a_e|^2 - m^2)|\phi|^2,$$
$$w \equiv (\nabla\phi^*)\phi - (\nabla\phi)\phi^* = 2[\phi_1(\nabla\phi_2) - \phi_2(\nabla\phi_1)]. \tag{8.53}$$

Note that the addition of the gauge field results in a linear coupling between a_e and the proper 4-current $q_e w$, which is the charge q_e times the proper 4-velocity w of the complex field $\phi = \phi_1 + \phi_2 I$. The gauge field also renormalizes the mass term; however, this term is an artifact of the quadratic kinetic term for the scalar field. Spinor fields with linear kinetic terms (such as the Dirac electron field) do not acquire such a term.

The linear coupling in (8.52) implies that a_e must *itself* be a dynamical field with its own kinetic term. However, the form of this kinetic term is initially unspecified. Adding the simplest quadratic contraction of ∇a_e as a kinetic term produces the gauge-invariant *scalar electrodynamics* Lagrangian

$$\mathcal{L}(x) = D\phi^* \cdot D\phi + \frac{1}{2}\langle(\nabla a_e)^2\rangle_0 - m^2\phi^*\phi$$
$$= \nabla\phi^* \cdot \nabla\phi + \frac{1}{2}\langle(\nabla a_e)^2\rangle_0 + q_e w \cdot a_e I + (|q_e a_e|^2 - m^2)|\phi|^2. \tag{8.54}$$

If we define the bivector field $\mathbf{F}_e = \nabla a_e = \nabla \wedge a_e$ as before, then we find that the postulated kinetic term for a_e in (8.54) is precisely the traditional Lagrangian (8.1).

Applying the Euler–Lagrange equation $\nabla \cdot (\delta\mathcal{L}/\delta\nabla a_e) = \delta\mathcal{L}/\delta a_e$ to the gauge-invariant Lagrangian (8.54) produces Maxwell's equation (7.1) as the equation of motion for the dynamical gauge field a_e

$$\nabla^2 a_e = \nabla\mathbf{F}_e = j_e, \tag{8.55}$$

where the pure vector current (for scalar electrodynamics) has the explicit form

$$j_e = q_e w + j_a, \qquad j_a = [q_e|\phi|^2][q_e a_e]. \tag{8.56}$$

The charge-current of the field $q_e w$ is modified in this case by an extra nonlinear term j_a that couples the charge density $q_e|\phi|^2$ back to the gauge field itself. Again, this term does not appear for fields with linear kinetic terms (like the Dirac electron).

Importantly, the extra term that is being canceled by the gauge approach is the (vector) phase gradient $\nabla\theta$ arising from the transformation $\exp(q_e\theta I)$. As such, only a pure vector potential a_e can be modified as in (8.50) to cancel this extra term, in accordance with the standard treatment of electromagnetism. An additional trivector potential $a_m I$ is thus not motivated by the usual gauge transformation $\exp(q_e\theta I)$.

8.9.2. Breaking dual symmetry

Notably, there is no overt restriction from the gauge mechanism itself that forbids adding a second kinetic term to the Lagrangian that corresponds to the second potential a_m (as in (8.4)) and then treating $z = a_e + a_m I$ as an intrinsically complex vector field. The gauge mechanism only motivates the necessary introduction of a_e, but it does not specify whether

a_e is the *complete* gauge field. The gauge mechanism only implies that if additional symmetry for a_e exists in vacuum prior to coupling, then it must be *broken* by the coupling to the charged field.

In fact, we already know that the electromagnetic theory does in fact arise from precisely such a symmetry-breaking process. In the Glashow–Weinberg–Salam electroweak theory [248–250] the vector potential a_e is itself a hybridization of two massless vector boson fields w_0 and b_0 that each obey Lagrangians $\langle(\nabla w_0)^2\rangle_0/2$ and $\langle(\nabla b_0)^2\rangle_0/2$ that are identical in form to the traditional electromagnetic Lagrangian (8.1). After constructing a complex boson field for this electroweak doublet $z = b_0 + w_0 I$, it then behaves identically to the (unconstrained) dual-symmetric formalism explored in this report. In particular, the sum of Lagrangians for b_0 and w_0 can be rewritten as the single complex field Lagrangian $\langle(\nabla z)(\nabla z^*)\rangle_0/2$, which is precisely our dual-symmetric form (8.4) motivated by the intrinsic complex structure of spacetime. Furthermore, there is evidence that performing a gauge-procedure (as in the previous section) to convert the global dual symmetry of this doublet to a local symmetry in its own right recovers additional structure of the electroweak theory [196].

For the electroweak theory, the Brout–Englert–Higgs–Guralnik–Hagen–Kibble (BEHGHK, or Higgs) mechanism [251–253] is responsible for breaking the dual symmetry of this doublet, as schematically shown in Fig. 7. In effect, the gauge coupling to the scalar BEHGHK boson (Higgs) field fixes a specific dual symmetry (Weinberg) rotation angle θ_W that fixes $a_e = \langle z \exp(\theta_W I)\rangle_1$ and $a_m = \langle z \exp(\theta_W I)\rangle_3 I^{-1}$ as the appropriate matter-coupled vector fields. The field a_e becomes the massless electromagnetic field vector potential for the photon (γ), while the field a_m acquires mass and corresponds to the neutral (Z_0) boson of the weak force. The details of this emergent coupling are much more involved than the simple electromagnetic model considered here, but the essential character of the broken dual symmetry is identical.

We can therefore understand the dual-symmetric formulation of electromagnetism as a simpler model of the neutral vacuum vector boson fields before the symmetry is broken by coupling to matter. In this dual-symmetric model, the gauge field a_m is not coupled to matter at all, but is still left over as a massless part of the total vacuum field. It is needed to preserve the symmetry of the vacuum, which *a priori* has no preferred reference axis for the dual-symmetric rotation $\exp(\theta I)$. The introduction of charged matter breaks this symmetry to force a particular reference axis for the dual symmetry, exactly as discussed in Section 7.4. It follows that the only quantities that can be measured (according to this model) from all the calculated conserved quantities will be the electric quantities (involving the potential a_e, its field $\mathbf{F}_e$, and the source j_e) just as with the standard formulation of electromagnetism. This does not, however, prevent the magnetic parts ($a_m, \mathbf{F}_m, j_m$) from being measurable in principle using other interactions that are not considered in this simple model, such as the additional isospin interaction that is acquired by the neutral (Z_0) boson that mediates the weak force.

The more remarkable note for the purposes of this report is that this dual-symmetric structure appears almost entirely from the assumption of spacetime. It is the geometry of spacetime itself that produced the intrinsically dual-symmetric bivector field $\mathbf{F}$ and associated complex vector potential z, as well as the especially simple forms of Maxwell's equation $\nabla \mathbf{F} = \nabla^2 z = j$ and the Lorentz force law $dp/d\tau = \langle\mathbf{F}j\rangle_1$. The use of spacetime algebra makes the geometric origin of all these quantities especially apparent.

9. Conclusion

> *It requires a much higher degree of imagination to understand the electromagnetic field than to understand invisible angels. … I speak of the $\vec{E}$ and $\vec{B}$ fields and wave my arms and you may imagine that I can see them … [but] I cannot really make a picture that is even nearly like the true waves.*
>
> [Richard P. Feynman [254]]

In this report we have presented a detailed introduction to the natural (Clifford) algebra of the Minkowski spacetime geometry, and have demonstrated its practicality and power as a tool for studying electromagnetism. Spacetime algebra not only preserves the existing electromagnetic formalisms, but also generalizes and unifies them into one comprehensive whole. This unification produces tremendous insight about the electromagnetic theory in a comparatively simple way. Below, we briefly review a few of the most important features that have emerged from the spacetime algebra approach, noting that we have also summarized a more complete list in the introductory Section 1.2, which we encourage the reader to review.

In particular, we emphasize that spacetime has an *intrinsic complex structure*, where the *ad hoc* scalar imaginary i is replaced with the physically meaningful unit 4-volume (pseudoscalar) I. Within this algebra, the electromagnetic field is an intrinsically complex bivector field $\mathbf{F} = \vec{E} + \vec{B}I$ on spacetime, which is a reference-frame-independent generalization of the Riemann–Silberstein complex field vector. Similarly, the vector potential for this field becomes complex, with associated complex bivector and scalar (Hertz) potentials. In all these representations, Maxwell's equations reduce to a single, and inexorable, equation. Similarly, the proper Lorentz force reduces to the simplest contraction of a current with the field, while the symmetric energy–momentum tensor is the simplest quadratic function of the field (identical in form to the energy–momentum current of the relativistic Dirac electron) and produces the associated angular momentum tensor as a wedge product.

The global phase of each complex field representation is arbitrary in vacuum (just like the global phase of a quantum wavefunction), and is responsible for the continuous dual (electric–magnetic field-exchange) symmetry of the vacuum electromagnetic field. Adding a source breaks this dual symmetry of the vacuum; however, making the charge of the source similarly complex (with separate electric and magnetic parts) preserves a similar global phase symmetry of the total system.

Choosing the source to be purely electric fixes this arbitrary global phase by convention, but the same physics will be described for any choice of global phase, which will produce equivalent descriptions that use both electric and magnetic charges. The dual symmetry of the complex field can be broken, however, by coupling to matter or other fields (as is the case in the electroweak theory).

In the Lagrangian treatment, the dual symmetry of the vacuum motivates the complex vector potential to be the primary dynamical field. Its constituent "electric" and "magnetic" parts effectively satisfy two separate traditional electromagnetic Lagrangians. As such, the traditionally conserved Noether currents (i.e., the canonical energy–momentum and angular-momentum tensors) are reproduced, but in an appropriately field-symmetrized form. The additional dual symmetry also immediately produces a conserved helicity current that cannot be easily derived in the traditional electromagnetic theory. Notably, the canonical Noether currents explicitly separate the orbital and spin parts of the local linear and angular momentum densities of the field, which are currently under active discussion in relation to both theoretical and experimental aspects of optics, QED, and QCD.

We also emphasize that the representation-free nature of spacetime algebra exposes deep structural relationships between mathematical topics that are seemingly disconnected in standard approaches. A detailed list of the remarkable and beneficial features of the spacetime algebra approach is provided in Section 1.2. Some of the most poignant examples of this unification include:

- The Hodge-star duality operation from differential forms is equivalent to multiplication by I^{-1}.
- The 4 Dirac matrices γ_μ from the relativistic electron theory are representations of unit 4-vectors in spacetime algebra (which require no such matrix representation).
- The 3 Pauli matrices $\vec{\sigma}_i = \gamma_i\gamma_0$ are representations of relative spatial directions, which are unit plane segments (bivectors) that drag the spatial coordinates γ_i along the proper-time direction γ_0.
- Their duals $\vec{\sigma}_i I^{-1}$ are spacelike planes that are equivalent to the quaternions. (There is thus nothing intrinsically quantum mechanical about any of these objects.)
- The commutator brackets of these 6 bivectors are precisely the Lie algebra of the Lorentz group, and reduce to the Pauli spin commutation relations in a relative frame.
- The spin commutation relations are simply the Hodge-duals of the usual 3-vector cross product relations.
- Exponentiating the bivector for a spatial plane produces a rotation spinor, while exponentiating a timelike plane produces a Lorentz boost spinor.
- These spinors can Lorentz-transform any geometric object with a double-sided product, in precisely the same way as the unitary transformations familiar from quantum mechanics, or the quaternionic formulation of spatial rotations.

It is our hope that this review will prompt continued research into the applications of spacetime algebra for both electromagnetic theory and other relativistic field theories. We feel that the language of spacetime algebra is practical, mature, and provides tremendous physical insight that can be more difficult to obtain when using other mathematical formulations of spacetime. Moreover, we feel that it is a natural extension of the existing electromagnetic formalisms in use, which gives it the distinct advantage of preserving (and augmenting) prior intuition and experience gained from working with those formalisms.

Acknowledgments

References

[1] F.J. Dyson, Why is Maxwell's theory so hard to understand?, in: James Clerk Maxwell Commemorative Booklet, James Clerk Maxwell Foundation, Edinburgh, 1999.
[2] E.P. Wigner, The unreasonable effectiveness of mathematics in the natural sciences, Comm. Pure Appl. Math. 13 (1) (1960) 1–14.
[3] S. Perkowitz, Empire of Light, Joseph Henry Press, 1999.
[4] D. Hestenes, Space-Time Algebra, Routledge, 1966.
[5] D. Hestenes, G. Sobczyk, Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics, D. Riedel Publishing, Dordrecht, 1987.
[6] D. Hestenes, Oersted medal lecture 2002: Reforming the mathematical language of physics, Amer. J. Phys. 71 (2) (2003) 104–121.
[7] D. Hestenes, Spacetime physics with geometric algebra, Amer. J. Phys. 71 (6) (2003) 1–24.
[8] D. Hestenes, Gauge theory gravity with geometric calculus, Found. Phys. 35 (6) (2005) 903–970.
[9] C. Doran, A. Lasenby, Geometric Algebra for Physicists, Cambridge University Press, Cambridge, 2007.
[10] O. Laporte, G. Uhlenbeck, Application of spinor analysis to the Maxwell and Dirac equations, Phys. Rev. 37 (1931) 1380–1397.
[11] J.R. Oppenheimer, Note on light quanta and the electromagnetic field, Phys. Rev. 38 (1931) 725–746.
[12] B. Mignani, E. Ricami, M. Baldo, About a Dirac-like equation for the photon according to Ettore Majorana, Nuovo Cimento Lett. 11 (1974) 568.

[13] W.J. Archibald, Field equations from particle equations, Can. J. Phys. 33 (1955) 565–572.
[14] R.H. Good, Particle aspect of the electromagnetic field equations, Phys. Rev. 105 (1957) 1914–1918.
[15] E. Giannetto, A Majorana-Oppenheimer formulation of quantum electrodynamics, Lett. Nuovo Cimento 44 (1985) 140–144.
[16] W.I. Fushchich, A.G. Nikitin, Symmetries of Maxwell's Equations, D. Riedel Publishing, Dordrecht, 1987.
[17] M.V. Berry, in: U.M. Bregola, G. Marmo, G. Morandi (Eds.), Anomalies, Phases, Defects, Bibliopolis, Napoli, 1990.
[18] I. Bialynicki-Birula, On the photon wave function, Acta Phys. Pol. A 86 (1994) 97.
[19] I. Bialynicki-Birula, Photon wave function, Prog. Opt. 36 (1996) 245–294.
[20] J.E. Sipe, Photon wave functions, Phys. Rev. A 52 (1995) 1875–1883.
[21] S. Esposito, Covariant majorana formulation of electrodynamics, Found. Phys. 28 (1998) 231–244.
[22] Y.P. Stepanovsky, From Maxwell equation to Berry's phase and sonoluminescence: Problems of theory of electromagnetic and other Massless fields, Electromagnetic Phenomena 1 (1998) 180–218.
[23] D.H. Kobe, A relativistic Schrödinger-like equation for a photon and its second quantization, Found. Phys. 19 (1999) 1203–1231.
[24] A. Gersten, Maxwell equations—the one-photon quantum equation, Found. Phys. 31 (8) (2001) 1211.
[25] B. Smith, M. Raymer, Two-photon wave mechanics, Phys. Rev. A 74 (6) (2006) 062104.
[26] B.J. Smith, M.G. Raymer, Photon wave functions, wave-packet quantization of light, and coherence theory, New J. Phys. 9 (11) (2007) 414.
[27] D. Dragoman, Photon states from propagating complex electromagnetic fields, J. Opt. Soc Am. B 24 (2007) 922–927.
[28] F. Tamburini, D. Vicino, Photon wave function: A covariant formulation and equivalence with QED, Phys. Rev. A 78 (5) (2008) 052116.
[29] B. Thidé, Electromagnetic Field Theory, Dover Publications, 2014.
[30] S.M. Barnett, Optical Dirac equation, New J. Phys. 16 (2014) 093008.
[31] B. Thidé, The physics of angular momentum radio, arXiv:1410.4268.
[32] M.H.L. Pryce, The mass-centre in the restricted theory of relativity and its connexion with the quntum theory of elementary particles, Proc. R. Soc. Lond. Ser. A 195 (1948) 62–81.
[33] T.D. Newton, W.E. P, Localized states for elementary systems, Rev. Modern Phys. 21 (1949) 400–406.
[34] A.S. Wightman, On the localizability of quantum mehcanical systems, Rev. Modern Phys. 34 (1962) 845–872.
[35] B.-S. Skagerstam, Localization of massless spinning particles and the Berry phase, arXiv:hep-th/9210054.
[36] O. Keller, On the theory of spatial localization of photons, Phys. Rep. 411 (2005) 1.
[37] H. Bacry, Localizability and Space in Quantum Physics, Springer, Berlin, 1988.
[38] S. Weinberg, E. Witten, Limits on massless particles, Phys. Lett. 96B (1980) 59–62.
[39] C. Cohen-Tannoudji, J. Dupont-Roc, G. Grynberg, Photons and Atoms, Wiley-Interscience, New York, 1997.
[40] I. Bialynicki-Birula, Z. Bialynicka-Birula, The role of the Riemann–Silberstein vector in classical and quantum theories of electromagnetism, J. Phys. A: Math. Theor. 46 (2013) 053001.
[41] H. Weber, Die Partiellen Differential-Gleichungen Der Mathematischen Physik: Nach Riemann's Vorlesungen Bearbeitet, Friedrich Vieweg und Sohn, Braunschweig, 1901.
[42] L. Silberstein, Elektromagnetische Grundgleichungen in bivectorieller Behandlung, Ann. Phys. 327 (1907) 579.
[43] L. Silberstein, Nachtrag zur Abhandlung über 'Elektromagnetische Grundgleichungen in bivectorieller Behandlung', Ann. Phys. 329 (1907) 783.
[44] B.E. Saleh, M.C. Teich, A.V. Sergienko, Wolf equations for two-photon light, Phys. Rev. Lett. 94 (22) (2005) 223601.
[45] J.F. Nye, M.V. Berry, Dislocations in wave trains, Proc. R. Soc. Lond. Ser. A 336 (1974) 165–190.
[46] L. Allen, M.W. Beijersbergen, R.J.C. Spreeuw, J.P. Woerdman, Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes, Phys. Rev. A 45 (1992) 8185–8189.
[47] S. van Enk, G. Nienhuis, Commutation rules and eigenvalues of spin and orbital angular momentum of radiation fields, J. Mod. Phys. 41 (1994) 963–977.
[48] L. Allen, M.J. Padgett, M. Babiker, The orbital angular momentum of light, Prog. Opt. 39 (1999) 291–372.
[49] D.L. Andrews, Structured Light and Its Applications, Academic Press, Amsterdam, 2008.
[50] J.P. Torres, L. Torner, Twisted Photons, Wiley-VCH, 2011.
[51] D.L. Andrews, M. Babiker, The Angular Momentum of Light, Cambridge University Press, Cambridge, 2013.
[52] A.T. O'Neil, I. MacVicar, A. Allen, M.J. Padgett, Intrinsic and extrinsic nature of the orbital angular momentum of a light beam, Phys. Rev. Lett. 88 (2002) 053601.
[53] V. Garcés-Chávez, D. McGloin, M.J. Padgett, W. Dultz, H. Schmitzer, K. Dholakia, Observation of the transfer of the local angular momentum density of a multiringed light beam to an optically trapped particle, Phys. Rev. Lett. 91 (2003) 093602.
[54] J.E. Curtis, D.G. Grier, Structure of optical vortices, Phys. Rev. Lett. 90 (2003) 133901.
[55] Y. Zhao, J.S. Edgar, G.D.M. Jeffries, D. McGloin, D.T. Chiu, Spin-to-orbital angular momentum conversion in a strongly focused optical beam, Phys. Rev. Lett. 99 (2007) 073901.
[56] H. Adachi, S. Akahoshi, K. Miyakawa, Orbital motion of spherical microparticles trapped in diffraction patterns of circularly polarized light, Phys. Rev. A 75 (2009) 063409.
[57] B. Roy, N. Ghosh, A. Banerjee, S.D. Gupta, S. Roy, Manifestations of geometric phase and enhanced spin Hall shifts in an optical trap, New J. Phys. 16 (2014) 083037.
[58] K.Y. Bliokh, J. Dressel, F. Nori, Conservation of the spin and orbital angular momenta in electromagnetism, New J. Phys. 16 (2014) 093037.
[59] J. Arlt, Handedness and azimuthal energy flow of optical vortex beams, J. Modern Opt. 50 (2003) 1573–1580.
[60] J. Leach, S. Keen, M.J. Padgett, C. Saunter, G.D. Love, Direct measurement of the skew angle of the Poynting vector in a helically phased beam, Opt. Express 14 (2006) 11919–11924.
[61] M. Yeganeh, S. Rasouli, M. Dashti, S. Slussarenko, E. Santamato, E. Karimi, Reconstructing the poynting vector skew angle and wavefront of optical vortex beams via two-channel moiré deflectometry, Opt. Lett. 38 (2013) 887–889.
[62] R.A. Terborg, K. Volke-Sepúlveda, Quantitative characterization of the energy circulation in helical beams by means of near-field diffraction, Opt. Express 21 (11) (2013) 3379–3387.
[63] M.R. Dennis, A.C. Hamilton, J. Courtial, Superoscillation in speckle patterns, Opt. Lett. 33 (2008) 2976–2978.
[64] M.V. Berry, Optical currents, J. Opt. A.: Pure Appl. Opt. 11 (2009) 094001.
[65] A. Bekshaev, K.Y. Bliokh, M. Soskin, Internal flows and energy circulation in light beams, J. Opt. 13 (2011) 053001.
[66] S. Kocsis, B. Braverman, S. Ravets, M.J. Stevens, R.P. Mirin, L.K. Shalm, A.M. Steinberg, Observing the average trajectories of single photons in a two-slit interferometer, Science 332 (6034) (2011) 1170–1173.
[67] M.R. Dennis, J.B. Götte, Beam shifts for pairs of plane waves, J. Opt. 15 (2013) 014015.
[68] K.Y. Bliokh, A.Y. Bekshaev, F. Nori, Dual electromagnetism: helicity, spin, momentum, and angular momentum, New J. Phys. 15 (3) (2013) 033026.
[69] K.Y. Bliokh, A.Y. Bekshaev, A.G. Kofman, F. Nori, Photon trajectories, anomalous velocities, and weak measurements: A classical interpretation, New J. Phys. 15 (2013) 073022.
[70] S.M. Barnett, M.V. Berry, Superweak momentum transfer near optical vortices, J. Opt. 15 (2013) 125701.
[71] S. Huard, C. Imbert, Measure de l'impulsion echangee au cours de l'interaction onde evanescente-atome, Opt. Commun. 24 (1978) 185–189.
[72] S. Huard, Exchange momentum between moving atoms and a surface wave: theory and experiment, Can J. Phys. 57 (1979) 612–618.
[73] K.Y. Bliokh, F. Nori, Transverse spin of a surface polariton, Phys. Rev. A 85 (2012) 061801(R).
[74] K.Y. Bliokh, A.Y. Bekshaev, F. Nori, Extraordinary momentum and spin in evanescent waves, Nature Comm. 5 (2014) 3300.
[75] A.Y. Bekshaev, K.Y. Bliokh, F. Nori, Transverse spin and momentum in two-wave interference, arXiv:1407.6786.
[76] Y. Tang, A.E. Cohen, Optical chirality and its interaction with matter, Phys. Rev. Lett. 104 (2010) 163901.

[77] E. Hendry, T. Carpy, J. Johnston, M. Popland, R.V. Mikhaylovskiy, A.J. Lapthorn, S.M. Kelly, L.D. Barron, N. Gadegaard, M. Kadodwala, Ultrasensitive detection and characterization of biomolecules using superchiral fields, Nature Nanotechnol. 5 (2010) 783–787.
[78] K.Y. Bliokh, F. Nori, Characterizing optical chirality, Phys. Rev. A 83 (2011) 021803(R).
[79] Y. Tang, A.E. Cohen, Enhanced enantioselectivity in excitation of chiral molecules by superchiral light, Science 332 (2011) 333–336.
[80] E. Hendry, R.V. Mikhaylovskiy, L.D. Barron, M. Kadodwala, T.J. Davis, Chiral electromagnetic fields generated by arrays of nanoslits, Nano. Lett. 12 (2012) 3640–3644.
[81] M. Schäferling, D. Dregely, M. Hentschel, H. Giessen, Tailoring enhanced optical chirality: Design principles for chiral plasmonic nanostructures, Phys. Rev. X 2 (2012) 031010.
[82] A. Canaguier-Durand, J.A. Hutchinson, C. Genet, T.W. Ebbesen, Mechanical separation of chiral dipoles by chiral light, New J. Phys. 15 (2013) 123037.
[83] R.P. Cameron, S.M. Barnett, A.M. Yao, Discriminatory optical force for chiral molecules, New J. Phys. 16 (2012) 013020.
[84] G. Tkachenko, E. Brasselet, Optofluidic sorting of material along optical chiral light, Nature Commun. 5 (2014) 3577.
[85] K.Y. Bliokh, Y.S. Kivshar, F. Nori, Magnetoelectric effects in local light–matter interactions, Phys. Rev. Lett. 113 (2014) 033601.
[86] A.I. Akhiezer, V. Berestetskii, Quantum Electrodynamics, Interscience, New York, 1965.
[87] V.B. Berestetskii, E.M. Lifshitz, L.P. Pitaevskii, Quantum Electrodynamics, Pergamon, Oxford, 1982.
[88] E. Leader, C. Lorcé, The angular momentum controversy: What's it all about and does it matter?, Phys. Rep. 541 (2014) 163–248.
[89] H.M. Wiseman, Grounding Bohmian mechanics in weak values and Bayesianism, New J. Phys. 9 (2007) 165.
[90] F.L. Traversa, G. Albareda, M. Di Ventra, X. Oriols, Robust weak-measurement protocol for Bohmian velocities, Phys. Rev. A 87 (2013) 052124.
[91] J. Dressel, K.Y. Bliokh, F. Nori, Classical field approach to quantum weak measurements, Phys. Rev. Lett. 112 (2014) 110407.
[92] E. Madelung, Eine anschauliche Deutung der Gleichung von Schrödinger, Naturwissenschaften 14 (45) (1926) 1004.
[93] E. Madelung, Quantentheorie in hydrodynamischer form, Z. Phys. 40 (3–4) (1927) 322–326.
[94] D. Bohm, A suggested interpretation of the quantum theory in terms of hidden variables. I, Phys. Rev. 85 (1952) 166–179.
[95] D. Bohm, A suggested interpretation of the quantum theory in terms of hidden variables. II, Phys. Rev. 85 (1952) 180–193.
[96] B.J. Hiley, R.E. Callaghan, Clifford algebras and the Dirac-Bohm quantum-Hamilton-Jacobi equation, Found. Phys. 42 (1) (2012) 192–208.
[97] D.L. Andrews, M.M. Coles, Measures of chirality and angular momentum in the electromagnetic field, Opt. Lett. 37 (2012) 3009–3011.
[98] D.J. Candlin, Analysis of the new conservation law in electromagnetic theory, Neural Comput. 37 (1965) 1390–1395.
[99] M.G. Calkin, An invariance property of the free electromagnetic field, Am. J. Phys 33 (1965) 958–960.
[100] R.F. O'Connell, D.R. Tompkins, Physical interpretation of generalized conservation laws, Neural Comput. 39 (1965) 391–394.
[101] D. Zwanziger, Quantum field theory of particles with both electric and magnetic charges, Phys. Rev. 176 (1968) 1489–1495.
[102] S. Deser, C. Teitelboim, Duality transformations of Abelian and non-Abelian gauge fields, Phys. Rev. D 13 (1976) 1592–1597.
[103] G.N. Afanasiev, Y.P. Stepanovsky, The helicity of the free electromagnetic field and its physical meaning, Neural Comput. 109 A (1996) 271–279.
[104] J.L. Trueba, A.F. Rañada, The electromagnetic helicity, Eur. J. Phys. 17 (1996) 141.
[105] R.P. Cameron, S.M. Barnett, A.M. Yao, Optical helicity, optical spin and related quantities in electromagnetic theory, New J. Phys. 14 (2012) 053050.
[106] R.P. Cameron, S.M. Barnett, Electric-magnetic symmetry and Noether's theorem, New J. Phys. 14 (2012) 123019.
[107] I. Fernandez-Corbaton, X. Zambrana-Puyalto, G. Molina-Terriza, Helicity and angular momentum: A symmetry-based framework for the study of light–matter interactions, Phys. Rev. A 86 (2012) 042103.
[108] I. Fernandez-Corbaton, X. Zambrana-Puyalto, N. Tischier, X. Vidal, M.L. Juan, G. Molina-Terriza, Electromagnetic duality symmetry and helicity conservation for the Macroscopic Maxwell's equations, Phys. Rev. Lett. 111 (2013) 060401.
[109] E. Noether, Invariante variationsprobleme, Nachr. Akad. Wiss. Gottingen. Math.-Phys. Kl. 2 (1918) 235–257.
[110] O. Heaviside, On the forces, stresses, and fluxes of energy in the electromagnetic field, Phil. Trans. R. Soc. A 183 (1892) 423.
[111] J. Larmor, A dynamical theory of the electric and luminiferous medium. Part III. Relations with material media, Phil. Trans. R. Soc. A 190 (1897) 205.
[112] M. Born, L. Infeld, Foundations of the new field theory, Proc. Roy. Soc. (London) A144 (1934) 425.
[113] E. Schrödinger, Contributions to Born's new theory of the electromagnetic field, Proc. Roy. Soc. (London) A 150 (1935) 465.
[114] S. Ferrara, J. Scherk, B. Zumino, Algebraic properties of extended supergravity theories, Nuclear Phys. B 121 (1977) 393.
[115] M.K. Gaillard, B. Zumino, Duality rotations for interacting fields, Nuclear Phys. B 193 (1981) 221–244.
[116] A. Sen, Electric-magnetic duality in string theory, Nuclear Phys. B 404 (1993) 109–126.
[117] J.H. Schwarz, A. Sen, Duality symmetric actions, Nuclear Phys. B 411 (1994) 35–63.
[118] P. Pasti, D. Sorokin, M. Tonin, Note on manifest Lorentz and general coordinate invariance in duality symmetric models, Phys. Lett. B 352 (1995) 59–63.
[119] G.W. Gibbons, D.A. Rasheed, Electric-magnetic duality rotatoins in non-linear electrodynamics, Nuclear Phys. B 454 (1995) 185–206.
[120] C.M. Hull, P.K. Townsend, Unity of superstring dualities, Nuclear Phys. B 438 (1995) 109–137.
[121] S. Deser, Ö. Sarioğlu, Hamiltonian electric/magnetic duality and Lorentz invariance, Phys. Let. B 423 (1998) 369–372.
[122] E. Cremmer, B. Julia, H. Lü, C.N. Pope, Dualisation of dualities, Nuclear Phys. B 523 (1998) 73–144.
[123] J.M. Figueroa-O'Farrill, Electromagnetic Duality for Children: Lecture notes 1995-6, 1998, http://www.maths.ed.ac.uk/~jmf/Teaching/EDC.html.
[124] P. Aschieri, S. Ferrara, B. Zumino, Duality rotations in nonlinear electrodynamics and in extended supergravity, Riv. Nuovo Cimento 31 (2008) 625–707.
[125] P. Aschieri, S. Ferrara, S. Theisen, Constitutive relations, off shell duality rotations and the hypergeometric form of born-infeld theory, in: S. Bellucci (Ed.), Breaking of Supersymmetry and Ultraviolet Divergences in Extended Supergravity, in: Proc. Phys., vol. 153, Springer, 2014, pp. 23–41.
[126] S.M. Barnett, Rotation of electromagnetic fields and the nature of optical angular momentum, J. Modern Opt. 57 (2010) 1339–1343.
[127] D.E. Soper, Classical Field Theory, Wiley, New York, 1976.
[128] R.J. Scully, The Demon and the Quantum, Wiley-VCH, Weinheim, 2007.
[129] J.C. Maxwell, On physical lines of force, I. The theory of molecular vortices applied to magnetic phenomena, Phil. Mag. 21 (1861) 161–175.
[130] J.C. Maxwell, On physical lines of force, II. The theory of molecular vortices applied to electric currents, Phil. Mag. 21 (1861) 281–291. 388–348.
[131] J.C. Maxwell, On physical lines of force, III. The theory of molecular vortices applied to statical electricity, Phil. Mag. 23 (1862) 12–24.
[132] J.C. Maxwell, On physical lines of force, IV. The theory of molecular vortices applied to the action of magnetism on polarised light, Phil. Mag. 23 (1862) 85–95.
[133] J.C. Maxwell, A dynamical theory of the electromagnetic field, Roy. Soc. Trans. 155 (1865) 459.
[134] J.C. Maxwell, A Treatise on Electricity and Magnetism, second ed., Oxford University Press, 1881.
[135] W.R. Hamilton, Lectures on Quaternions, Hodges and Smith, Dublin, 1853.
[136] W.R. Hamilton, Elements of Quaternions, Longmans, Green, and Co., New York, 1901.
[137] O. Heaviside, Electro-Magnetic Theory, Electrician Printing and Publishing Company, Ltd., London, 1893.
[138] E.B. Wilson, Vector Analysis: A Text-Book for the Use of Students of Mathematics and Physics, Founded Upon the Lectures of J. Willard Gibbs Ph.D. LL.D., Yale University Press, 1901.
[139] H. Grassmann, Die Lineale Ausdehnungslehre — Ein neuer Zweig der Mathematik, Verlag von Otto Wigand, Leipzig, 1844.
[140] W.K. Clifford, Elements of Dynamic, MacMillan and Co., London, 1878.
[141] H.A. Lorentz, Simplified theory of electrical and optical phenomena in moving systems, Proc. Roy. Neth. Acad. Arts Sci. 1 (1899) 427–442.
[142] H.A. Lorentz, Weiterbildung der Maxwellshen Theorie. Elektronentheorie., Ency. Math. Wiss. 5 (1904) 145–288.
[143] H.A. Lorentz, Electromagnetic phenomena in a system moving with any velocity smaller than that of light, Proc. Roy. Neth. Acad. Arts Sci. 6 (1904) 809–831.
[144] H. Poincaré, Sur la dynamique de l'électron, C. R. Acad. Sci. 140 (1905) 1504–1508.
[145] H. Poincaré, Sur la dynamique de l'électron, Rend. Circ. Mat. Pal. 21 (1906) 129–176.
[146] A. Einstein, Zur elektrodynamik bewegter Körper, Ann. Phys. 322 (1905) 891–921.

[147] H. Minkowski, Das Relativitätsprinzip, Ann. Phys. 352 (1907) 927–938.
[148] A. Einstein, Über die elektromagnetischen Grundgleichungen für bewegte Körper, Ann. Phys. 331 (1908) 532–540.
[149] A. Einstein, Relativität und gravitation. Erwiderung auf eine Bemerkung von M. Abraham, Ann. Phys. 38 (1912) 1059–1064.
[150] A. Sommerfeld, Zur Relativitätztheorie I: Vierdimensionale Vektoralgebra, Ann. Phys. 337 (1910) 749–776.
[151] A. Sommerfeld, Zur Relativitätztheorie II: Vierdimensionale, Ann. Phys. 337 (1910) 749–776.
[152] H. Minkowski, Raum und Zeit, Phys. Zeit. 10 (1908) 75–88.
[153] E. Cartan, Sur certaines expressions différentielles et le probème de Pfaff, Ann. Sci. l'École Norm. Sup. 16 (1899) 239–332.
[154] E.B. Wilson, The space-time manifold of relativity. The non-Euclidean geometry of mechanics and electromagnetics, Proc. Amer. Acad. Arts Sci. 48 (1911) 387–507.
[155] J.D. Jackson, Classical Electromagnetism, Wiley, New York, 1999.
[156] M. Göckeler, T. Schücker, Differential Geometry, Gauge Theories, and Gravity, Cambridge University Press, 1989.
[157] A. Einstein, Die Grundlage der allgemeinen Relativitätstheorie, Ann. Phys. 49 (1916) 769–822.
[158] A. Crumeyrolle, Orthogonal and Symplectic Clifford Algebras, Kluwer Academic Publishers, Dordrecht, 1990.
[159] C.N. Yang, R. Mills, Conservation of isotopic spin and isotopic gauge invariance, Phys. Rev. 96 (1954) 191–195.
[160] C.N. Yang, Einstein's impact on theoretical physics, Lecture given at the Second Marcel Grossmann Meeting held in honor of the 100th anniversary of the birth of Albert Einstein, Phys. Today 33 (1980) 42.
[161] C.N. Yang, Topology and gauge theory in physics, in: In Memory of Akira Tonomura, Physicist and Electron Microscopist, World Scientific, Singapore, 2014, pp. 41–49.
[162] C.N. Yang, The conceptual origins of Maxwell's equations and gauge theory, Phys. Today 67 (2014) 45.
[163] S.-S. Chern, J. Simons, Characteristic forms and geometric invariants, Ann. of Math. 99 (1974) 48–69.
[164] Y. Aharonov, D. Bohm, Significance of electromagnetic potentials in the quantum theory, Phys. Rev. 115 (3) (1959) 485–491.
[165] G. 't Hooft, M. Veltman, Regularization and renormalization of gauge fields, Nuclear Phys. B 44 (1972) 189.
[166] A. Tonomura, N. Osakabe, T. Matsuda, T. Kawasaki, J. Endo, S. Yano, H. Yamada, Evidence for Aharonov-Bohm effect with magnetic field completely shielded from electron wave, Phys. Rev. Lett. 56 (1986) 792–795.
[167] E. Witten, Topological quantum field theory, Comm. Math. Phys. 117 (1988) 353.
[168] S. Weinberg, The Quantum Theory of Fields I, Cambridge University Press, Cambridge, 1995.
[169] M. Nakahara, Geometry, Topology and Physics, second ed., IOP Publishing, London, 2003.
[170] A.C.T. Wu, C.N. Yang, Evolution of the concept of the vector potential in description of fundamental interactions, Internat. J. Modern Phys. A 21 (2006) 3235–3277.
[171] D. Hestenes, Multivector calculus, J. Math. Anal. and Appl. 24 (2) (1968) 313–325.
[172] D. Hestenes, Multivector functions, J. Math. Anal. and Appl. 24 (3) (1968) 467–473.
[173] D. Hestenes, Vectors, spinors, and complex numbers in classical and Quantum physics, Amer. J. Phys. 39 (9) (1971) 1013.
[174] G. Sobczyk, Simplicial calculus with geometric algebra, in: A. Micali, R. Boudet, J. Helmstetter (Eds.), Fundamental Theories of Physics, vol. 47, Springer, Netherlands, 1989, pp. 279–292.
[175] A. Lasenby, C. Doran, S. Gull, Gravity, gauge theories and geometric algebra, Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci. 356 (1737) (1998) 487–582.
[176] R. Abłamowicz, G. Sobczyk, Lectures on Clifford (Geometric) Algebras and Applications, Birkhäuser, Basel, 2004.
[177] V. de Sabbatta, B.K. Datta, Geometric Algebra and Applications to Physics, Taylor & Francis Group, 2007.
[178] S.J. van Enk, The covariant description of electric and magnetic field lines of null fields: application to Hopf-Rañada solutions, J. Phys. A: Math. Theor. 46 (2013) 175204.
[179] F. Colombo, I. Sabadini, F. Sommen, D.C. Struppa, Analysis of Dirac Systems and Computational Algebra, Birkhäuser, Boston, 2004.
[180] S. Delanghe, F. Sommen, V. Soucek, Clifford Algebra and Spinor-Valued Functions: A Function Theory for the Dirac Operator, Springer, Netherlands, 2012.
[181] W. Baylis, Clifford (Geometric) Algebras with Applications in Physics, Mathematics, and Engineering vol 1, Birkhäuser, Boston, MA, 1996.
[182] D. Hestenes, New Foundations for Classical Mechanics, Kluwer Academic Publishers, Dordrecht, 1999.
[183] E.B. Corrochano, G. Sobczyk, Geometric Algebra with Applications in Science and Engineering, Birkhäuser, Boston, 2001.
[184] W. Baylis, Electrodynamics: A Modern Geometric Approach, Birkhäuser, 2002.
[185] C. Perwass, Geometric Algebra with Applications in Engineering, Springer, New York, 2010.
[186] L. Dorst, J. Lasenby, Guide to Geometric Algebra in Practice, Springer, New York, 2011.
[187] A. Macdonald, Linear and Geometric Algebra, CreateSpace Independent Publishing Platform, 2011.
[188] A. Macdonald, Vector and Geometric Calculus, CreateSpace Independent Publishing Platform, 2012.
[189] L.D. Landau, E.M. Lifshitz, The Classical Theory of Fields, Butterworth-Heinemann, 1975.
[190] B.J. Hiley, A note on the role of idempotents in the extended Heisenberg algebra, Proc. of the Int. Meeting of ANPA 22 (2001) 107–121.
[191] B.J. Hiley, R.E. Callaghan, The Clifford Algebra approach to Quantum Mechanics A: The Schrödinger and Pauli Particles, 2010, http://arxiv.org/abs/1011.4031.
[192] B.J. Hiley, R.E. Callaghan, The Clifford Algebra approach to Quantum Mechanics B: The Dirac Particle and its relation to the Bohm Approach, 2010, http://arxiv.org/abs/1011.4033.
[193] P.A.M. Dirac, Quantised singularities in the electromagnetic field, Proc. R. Soc. Lond. Ser. A 133 (1931) 60.
[194] P.A.M. Dirac, The quantum theory of the electron, Proc. R. Soc. Lond. A 117 (778) (1928) 610–624.
[195] R. Arianrhod, Einstein's Heroes: Imagining the World Through the Language of Mathematics, Oxford University Press, Oxford, 2006.
[196] C.A.Z. Vasconcellos, D. Hadjimichef, CP Violation in Dual Dark Matter, 2014, http://arxiv.org/abs/arXiv:1404.0409.
[197] P.A.M. Dirac, The theory of magnetic poles, Phys. Rev. 74 (1948) 817–830.
[198] N. Cabibbo, E. Ferrari, Quantum electrodynamics with Dirac monopoles, Neural Comput. 23 (1962) 1147.
[199] J. Schwinger, Magnetic charge and quantum field theory, Phys. Rev. 144 (1966) 1087–1093.
[200] F. Rohrlich, Classical theory of magnetic monopoles, Phys. Rev. 150 (1966) 1104–1111.
[201] D. Zwanziger, Local-Lagrangian quantum field theory of electric and magnetic charges, Phys. Rev. D 3 (1971) 880–891.
[202] M.Y. Han, L.C. Biedenharn, Manifest dyality invariance in electrodynamics and the Cabibbo-Ferrari theory of magnetic monopoles, Neural Comput. 2 (1971) 544.
[203] B. Mignani, E. Ricami, Complex electromagnetic four-potential and the Cabibbo-Ferrari relation for magnetic monopoles, Neural Comput. 30 (1975) 533.
[204] S. Deser, Off-shell electromagnetic duality invariance, J. Phys. A: Math. Gen. 15 (1982) 1053.
[205] R. Gambini, S. Salamó, Null-plane analysis of electromagnetism with magnetic charges, Neural Comput. 49 (1979) 276.
[206] R. Gambini, S. Salamó, A. Trias, Self-dual covariant Lagrangian formulation of electromagnetism with magnetic charges, Lett. Nuovo Cimento 49 (1979) 276.
[207] J.H. Schwartz, A. Sen, Duality symmetric actions, Nuclear Phys. B 411 (1994) 35.
[208] P. Pasti, D. Sorokin, M. Tonin, Duality symmetric actions with manifest space-time symmetries, Phys. Rev. D 52 (1995) R4277–R4281.
[209] D. Singleton, Magnetic charge as a 'hidden' gauge symmetry, Internat. J. Theoret. Phys. 34 (1995) 37–46.
[210] D. Singleton, Electromagnetism with magnetic charge and two photons, Amer. J. Phys. 64 (1996) 452–458.
[211] A. Khoudeir, N.R. Pantoja, Covariant duality symmetric actions, Phys. Rev. D 53 (1996) 5974–5977.
[212] A. Kato, D. Singleton, Gauging dual symmetry, Internat. J. Theoret. Phys. 41 (2002) 1563–1572.
[213] Y.M. Shnir, Magnetic Monopoles, Springer, Berlin, 2005.

[214] T.G. Philbin, Lipkin's conservation law, Noether's theorem, and the relation to optical helicity, Phys. Rev. A 87 (2013) 043843.
[215] R. Battesti, C. Rizzo, Magnetic and electric properties of a quantum vacuum, Rep. Progr. Phys. 76 (2013) 016401.
[216] S. Weinberg, A new light boson?, Phys. Rev. Lett. 40 (1978) 223.
[217] J.C. Maxwell, Address to the Mathematical and Physical Sections of the British Associations, Liverpool, 1865.
[218] X.-S. Chen, X.-F. Lü, W.-M. Sun, F. Wang, T. Goldman, Spin and orbital angular momentum in gauge theories: Nucleon spin structure and multipole radiation revisited, Phys. Rev. Lett. 100 (2008) 232002.
[219] H. Hertz, Die Kräfte electrischer Schwingungen, behandelt nach der Maxwell'schen Theorie, Ann. Phys. 272 (1889) 1.
[220] E.A. Essex, Hertz vector potentials of electromagnetic theory, Amer. J. Phys. 45 (1977) 1099.
[221] M.V. Berry, Lateral and transverse shifts in reflected dipole radiation, Proc. R. Soc. A 467 (2133) (2011) 2500–2519.
[222] E. Kalnins, W.J. Miller, Lie theory and the wave equation in space-time: R-separable solutions of the wave equation $\Psi_{tt} - \Delta_3 \Psi = 0$, J. Math. Phys. 19 (1978) 1247.
[223] E.T. Whittaker, On an expression of the electromagnetic field due to electrons by means of two scalar potential functions, Proc. Lond. Math. Soc. 1 (1904) 367.
[224] P. Debye, Der lichtdruck auf kugeln von beliebigem material, Ann. Phys. 335 (1909) 57.
[225] A. Nisbet, Hertzian electromagnetic potentials and associated gauge transformations, Proc. R. Soc. A 231 (1955) 250.
[226] J.L. Synge, Relativity: The Special Theory, North-Holland, Amsterdam, 1965.
[227] L.D. Landau, E.M. Lifshitz, L.P. Pitaevskii, Electrodynamics of Continuous Media, Butterworth-Heinemann, 1984.
[228] C. Castelnovo, R. Moessner, S.L. Sondhi, Magnetic monopoles in spin ice, Nature 451 (2008) 42–45.
[229] S.R. Giblin, S.T. Bramwell, P.C.W. Holdsworth, D. Prabhakaran, I. Terri, Creation and measurement of long-lived magnetic monopole currents in spin ice, Nat. Phys. 7 (2011) 252–258.
[230] H.-B. Braun, Magnetic monopoles: Quenching the fire in spin ice, Nat. Phys. 10 (2014) 88–89.
[231] C. Paulsen, M.J. Jackson, E. Lhotel, B. Canals, D. Prabhakaran, K. Matsuhira, S.R. Giblin, S.T. Bramwell, Far-from-equilibrium monopole dynamics in spin ice, Nat. Phys. 10 (2014) 137–139.
[232] M.W. Ray, E. Ruokokowski, S. Kandel, M. Möttönen, D.S. Hall, Observation of Dirac monopoles in a synthetic magnetic field, Nature 505 (2014) 657–660.
[233] P. Milonni, Fast Light, Slow Light and Left-Handed Light, Taylor and Francis, New York, 2004.
[234] A.Y. Savchenko, B.Y. Zel'dovich, Birefringence by a smoothly inhomogeneous locally isotropic medium: Three-dimensional case, Phys. Rev. E 50 (1994) 2287–2292.
[235] X. Zambrana-Puyalto, I. Fernandez-Corbaton, M.L. Juan, X. Vidal, G. Molina-Terriza, Duality symmetry and Kerker conditions, Opt. Lett. 38 (2013) 1857–1859.
[236] N.e.a. Tischler, Experimental control of optical helicity in nanophotonics, Light Sci. Appl. 3 (2014) e183.
[237] H. Hertz, Über die Beziehungen Zwischen Licht und Elektrizität. Eind Vortrag gehalten bei der 62, Versammlund dentscher Naturforscher und Ärzte in Heidelberg, Bonn, 1889.
[238] K.C. Cole, Sympathetic Vibrations: Reflections on Physics as a Way of Life, Bantam, 1985.
[239] R.P. Cameron, On the 'second potential' in electrodynamics, J. Opt. 16 (2014) 015708.
[240] F. Belinfante, Physica 6 (1939) 887–898.
[241] F. Belinfante, Physica 7 (1940) 449–474.
[242] L. Rosenfeld, Mém. Acad. Roy. Belg. Sci. 18 (1940) 1–30.
[243] H. Ohanian, What is spin?, Amer. J. Phys. 54 (1986) 500–505.
[244] K. Mita, Virtual probability current associated with the spin, Amer. J. Phys. 68 (2000) 259–264.
[245] R. Zambrini, S. Barnett, Local transfer of optical angular momentum to matter, J. Modern Opt. 52 (2005) 1045–1052.
[246] L. Allen, M.J. Padgett, Response to question no. 79. Does a plane wave carry spin angular momentum?, Amer. J. Phys. 70 (2002) 567–568.
[247] E.J. Konopinski, What the electromagnetic vector potential describes, Amer. J. Phys. 46 (1978) 499–502.
[248] S.L. Glashow, Partial-symmetries of weak interactions, Nuclear Phys. 22 (1961) 579.
[249] S. Weinberg, A model of leptons, Phys. Rev. Lett. 19 (1961) 1264.
[250] A. Salam, Elementary particle physics, in: N. Svartholm (Ed.), Proceedings of the 8th Nobel Symposium, Almqvist and Wiksell, Stockholm, 1968, p. 367.
[251] F. Englert, R. Brout, Broken symmetry and the mass of gauge vector mesons, Phys. Rev. Lett. 13 (1964) 321.
[252] P.W. Higgs, Broken symmetries, massless particles and gauge fields, Phys. Lett. 12 (1964) 132.
[253] G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, Global conservation laws and massless particles, Phys. Rev. Lett. 13 (1964) 585.
[254] R.P. Feynman, The Feynman Lactures on Physics, vol. II, Addison-Wesley, Reading, Massachusetts, 1964.