On the axioms of topological electromagnetism

D. H. Delphenich[*]

Physics Department, Bethany College, Lindsborg, KS 67456, USA

The axioms of topological electromagnetism that were given by Hehl, Obukhov, and Rubilar are refined by the use of geometrical and topological notions that are found on orientable manifolds. The central problem of defining the spacetime electromagnetic constitutive law in terms of the geometrical and topological structure of the spacetime manifold is elaborated upon in the linear and nonlinear cases. The manner by which the spacetime metric might follow from the electromagnetic constitutive law is examined in the linear case. The possibility that the intersection form of the spacetime manifold might play a role in defining a topological basis for a nonlinear electromagnetic constitutive law is explored. The manner by which electromagnetic wave motion relates to the geometric structure is also discussed.

Contents

[*] E-mail: delphenichd@bethanylb.edu

1   Introduction

Although the interpretation of the electromagnetic field that is given by gauge field theory – viz., that the electromagnetic field strength 2-form F is the curvature 2-form associated with a connection 1-form A, which represents a choice of potential 1-form, on a $U(1)$-principal bundle that represents the freedom to assign the local gauge of F arbitrarily – is certainly geometrical in a formal sense, nevertheless, the $U(1)$-principal bundle and the connection itself seem to exist independently of the bundle of Lorentzian frames over spacetime and the Levi-Civita connection that is associated with the presence of gravitation. Hence, one of the long-standing problems of theoretical physics has been to discern whether the electromagnetic field can be described as a geometrical or topological object that exists naturally within the structure of the spacetime manifold, in the sense of reductions of the bundle of linear frames on spacetime and the various tensor and spinor bundles that are associated with them. Thus, such a representation of electromagnetism might involve the geometrical structure of spacetime, like the gravitational field, or its topological structure, or some combination of both.

Despite the many noble attempts at the unification of gravitation and electromagnetism (see Lichnerowicz [1] or Vizgin [2]), which usually took a purely geometric approach to the problem, only so much definitive progress was made. Einstein, in particular, felt that the unification of electromagnetism and gravitation would probably be closely related to two other long-standing problems: the causal interpretation of wave mechanics and the extension of electromagnetism to a nonlinear theory.

One of the consequences of the widespread interest in gauge field theories on the part of pure mathematics was the growing realization amongst physicists that topology, as well as geometry, played an important role in all such theories, including what is perhaps the simplest gauge field theory of physics, namely, classical electromagnetism. Furthermore, if one returns to the purely physical considerations, one must remember that, historically, the introduction of the methods of differential geometry into physics, by way of general relativity, actually originated in the consideration of the structure of the characteristic submanifolds for the linear wave equation that followed as a consequence of Maxwell's equations for electromagnetism. Consequently, there is something logically recursive about reintroducing the geometry of spacetime back into the foundations of electromagnetism.

A more logically straightforward approach would be to accept that the metric structure of spacetime is a *consequence* of the electromagnetic structure of spacetime and then try to be more precise about the nature of the phrase "electromagnetic structure." One might do well to accept the notion that the best structure to start with is a G-structure on spacetime [3], that is, a reduction of the bundle $GL(M)$ of linear frames on the spacetime manifold, since the metric structure of spacetime is a particular example of such a structure, namely, an $SO(3,1)$ reduction. In particular, we shall treat electromagnetism as something that relates to an $SL(4)$-structure on the spacetime manifold, i.e., a bundle of oriented unit-volume frames whose fundamental tensor field is the chosen unit volume element. However, the spacetime manifold is not assumed to be Lorentzian, *a priori*. The fundamental tensor for an $SL(4)$-structure on spacetime, viz., the unit volume element, then plays the central role in electromagnetism, just as the fundamental tensor for a $SO(3,1)$-structure, viz., a Lorentzian metric, plays the central role in gravitation.

Besides the axiomatization of electrodynamics, a second problem that one must confront is that of defining the reduction from the bundle of oriented unit-volume linear frames $SL(M)$ to the bundle of oriented Lorentzian frames $SO(3,1)(M)$ in terms of consequences that follow from electromagnetic field theory. We shall not pursue the details of the reduction here, except to indicate that the transition from electromagnetism to gravitation might follow from this spacetime phase transition and point out that the appearance of wavelike solutions to the field equations of electromagnetism is intimately related to it.

We mentioned that the spacetime metric should be the symbol of the wave operator that follows from the electrodynamics equations. This suggests that the appearance of a spacetime metric is associated with the structure of electromagnetic wave motion in spacetime. However, a wave-like solution to the Maxwell equations is represented by a particular type of 2-form F of rank 2, and such a 2-form defines a reduction *beyond* $SO(3,1)(M)$ to an $SO(2)$-structure on M, i.e., an $SO(2)$ reduction of the bundle of oriented

unit-volume linear frames. This is because a 2-form F of rank 2 defines a two-dimensional subspace of $T_x(M)$ at each $x \in M$ in the form of the associated system to F, in the sense of Cartan [4,5]. Hence, one might consider how the spacetime metric might appear for more general solutions to the Maxwell equations, which will be of rank four.

We see that there are really two distinct aspects to extending the theory of electromagnetism: the topological one, which is most intimately related to the nature of the sources of electromagnetic fields, and the geometrical one, which is concerned with exhibiting the presence of electromagnetism in Nature as a consequence of spacetime geometry, most likely at a more esoteric level of consideration than the metric one. The main focus of this article will be the topological aspect.

Our approach to this topological aspect is to introduce the analog of de Rham cohomology that one can define on multi-vectors – i.e., the exterior algebra over $T(M)$ – when M is orientable, oriented, and given a choice of unit volume element, and we call this analog *de Rham homology*. The axioms of topological – or *pre-metric* – electromagnetism can then be cast in the language of homology, except for the constitutive axiom. We also examine the possible role that the intersection form on the de Rham cohomology might play as a sort of "topological constitutive law."

The most promising ideas to start with in defining the geometric nature of the problem – viz., pre-metric electromagnetism – are the fact that the group $SL(4)$ takes on its most detailed interpretation in the role of projective transformations of homogeneous coordinates for $\mathbb{RP}^3$ and the fact that a linear electromagnetic constitutive law indirectly defines a complex structure for the vector space of bivectors on $\mathbb{R}^4$, hence, an almost-complex structure on the bundle of 2-forms over spacetime. This suggests that more attention could be given to the physical significance of projective geometry, especially complex projective geometry, in the geometry of spacetime. However, this research will follow in a later article, as the present one will be devoted solely to the topological problem.

2 De Rham homology on orientable manifolds

There are actually *two* ways by which the spacetime metric appears in the non-homogeneous Maxwell equations:

$$dF = 0, \quad d^*F = -\frac{4\pi}{c_0} * J. \tag{2.1}$$

As is commonly recognized [6], the obvious one is by way of the Hodge $*$-operator. The second one is by the introduction of the "constant" c_0, which is actually derivable from simplifying assumptions on the nature of the constitutive laws of the electromagnetic vacuum state. We shall discuss this latter aspect of Maxwell's equations in more detail later, but for the moment, we concentrate on the former one.

In effect, the Hodge $*$ isomorphism follows from a more general isomorphism # that one finds on orientable manifolds, in the form of Poincaré duality. In the absence of a metric, but the presence of a unit volume element, i.e., a globally non-zero $\mathcal{V} \in \Lambda^4(M)$, the only isomorphisms that we can define are:[1]

$$\# : \Lambda_k(M) \to \Lambda^{n-k}(M), \quad \mathbf{a} \mapsto i_{\mathbf{a}}\mathcal{V}. \tag{2.2}$$

In this expression, $\mathbf{a}$ is a k-vector field on M, so it can be expressed as a finite sum of expressions of the form $\mathbf{X} \wedge \mathbf{Y} \wedge \ldots \wedge \mathbf{Z}$, where $\mathbf{X}, \mathbf{Y}, \ldots, \mathbf{Z} \in \mathfrak{X}(M)$ are k vector fields. The interior product of a k-vector field and a p-form α is defined in general by assuming that it is k-linear and recursively defining:

$$i_{\mathbf{X} \wedge \mathbf{Y} \wedge \ldots \wedge \mathbf{Z}}\, \alpha = i_{\mathbf{X}}(i_{\mathbf{Y} \wedge \ldots \wedge \mathbf{Z}}\, \alpha). \tag{2.3}$$

---

[1] We temporarily revert to the general case of n-dimensional orientable differentiable manifolds.

The introduction of a metric then allows us to define a linear isomorphism between the tangent spaces to our manifold M and the cotangent spaces, and in so doing, an isomorphism of the exterior algebra $\Lambda^*(M)$ of k-vector fields on M with the exterior algebra $\Lambda^*(M)$ of exterior differential k-forms on M.

As is well known, the exterior derivative operator d on $\Lambda^*(M)$ makes $\Lambda^*(M)$ into a $\mathbb{Z}$-graded differential module (vector space, in fact). Since d is of degree $+1$, one can define the *de Rham cohomology modules* by $H^k(M;\mathbb{R}) = Z^k(M;\mathbb{R})/B^k(M;\mathbb{R})$, where $Z^k(M;\mathbb{R})$ consists of all of the closed k-forms and $B^k(M;\mathbb{R})$ consists of all exact k-forms. The exterior product of differential forms then gives rise to a ring structure on $H^*(M;\mathbb{R})$ whose product $[\alpha] \cup [\beta]$, which amounts to the usual "cup product" of algebraic topology [7–10], is derived from the exterior product:

$$[\alpha] \cup [\beta] = [\alpha \wedge \beta]. \tag{2.4}$$

As is less known, on an orientable manifold M one can use the isomorphisms (2.2) and the operator d to define a *codifferential* or "boundary" operator of degree -1 on $\Lambda_*(M)$ by way of:

$$\delta : \Lambda_k(M) \to \Lambda_{k-1}(M), \qquad \mathbf{a} \mapsto \delta\mathbf{a} = \#^{-1}d\#\mathbf{a}. \tag{2.5}$$

A simple local computation shows that if $\mathbf{X} \in \mathfrak{X}(M)$ then the components of $\delta\mathbf{X}$ with respect to a natural local frame field $\mathbf{e}_i = \partial/\partial x_i$ are going to agree with the usual divergence of $\mathbf{X}$ that one learns from vector calculus. An important point to emphasize is that, unlike the usual definition of the divergence operator in differential geometry, we have not needed to introduce a metric for ours.

It is not hard to see that the fact that $d^2 = 0$ implies that:

$$\delta^2 = 0. \tag{2.6}$$

Hence, δ makes $\Lambda_*(M)$ into a $\mathbb{Z}$-graded differential module. One then defines a homology by way of:

$$H_k(M;\mathbb{R}) = Z_k(M;\mathbb{R})/B_k(M;\mathbb{R}), \tag{2.7}$$

in which $Z_k(M;\mathbb{R})$ is the space of a *co-closed* k-vector fields ($\delta\mathbf{a} = 0$) and $B_k(M;\mathbb{R})$ consists of all *co-exact* k-vector fields ($\mathbf{a} = \delta\mathbf{b}$ for some $k + 1$-vector field $\mathbf{b}$).

A significant difference between the behavior of $H_*(M;\mathbb{R})$ and $H^*(M;\mathbb{R})$, besides the degree of the boundary map, is that the exterior product of a p-vector field $\mathbf{a}$ and a q-vector field $\mathbf{b}$ does not "descend to homology" since $\delta(\mathbf{a} \wedge \mathbf{b})$ does not generally equal $\delta\mathbf{a} \wedge \mathbf{b} + (-1)^p\mathbf{a} \wedge \delta\mathbf{b}$. One does have the following useful relation for the product of a smooth function f and a k-vector field $\mathbf{a}$:

$$\delta(f\mathbf{a}) = \#^{-1}(df \wedge \#\mathbf{a}) + f\delta\mathbf{a}. \tag{2.8}$$

When $\mathbf{X}$ is a vector field, this becomes:

$$\delta(f\mathbf{X}) = \mathbf{X}f + f\delta\mathbf{X}. \tag{2.9}$$

Before one suspects that something is missing in our homology, namely, a ring structure, keep in mind that one does not usually expect to find a ring structure on homology, at least in general. Nevertheless, one does find that the interior product of k-vector fields and p-forms ($p > k$) descends to (co)homology, where it takes the form of the usual "cap product" that one encounters in the topology of manifolds [7–10]:

$$H_k(M;\mathbb{R}) \times H^p(M;\mathbb{R}) \to H^{p-k}(M;\mathbb{R}), \qquad ([\mathbf{a}], [\beta]) \mapsto [\mathbf{a}] \cap [\beta] = [i_a\beta]. \tag{2.10}$$

In this definition, the [ ] brackets denote the homology or cohomology class that corresponds to the item inside.

When one descends to homology the cap product produces the isomorphisms of *Poincaré duality:*

$$H_k(M;\mathbb{R}) \cong H^{n-k}(M;\mathbb{R}), \quad [\mathbf{a}] \mapsto [\mathbf{a}] \cap [\mathcal{V}] = [\#\mathbf{a}]. \tag{2.11}$$

Since our coefficient ring is a field, it is correct to identify $H^k(M; \mathbb{R})$ as the dual of the vector space $H_k(M; \mathbb{R})$; i.e., $H^k(M; \mathbb{R}) \cong \mathrm{Hom}(H_k(M; \mathbb{R}); \mathbb{R}) = H_k(M; \mathbb{R})^*$. Hence, its elements can be regarded as linear functionals on homology classes. When M is compact, or we restrict ourselves to k-vector fields of compact support, we can represent the k-dimensional cohomology class $[\alpha]$ as a linear functional on k-dimensional homology classes $[\mathbf{b}]$ in integral form:

$$\alpha[\mathbf{b}] = \int_M \alpha \wedge \#\mathbf{b} = \int_{\tilde{\mathbf{b}}} \alpha. \tag{2.12}$$

However, in order for the last integral to make sense, we make use of the fact that de Rham's theorem for cohomology, which we consider to be an isomorphism of the de Rham cohomology of M with the singular cubic cohomology with real coefficients, also gives rise to a corresponding isomorphism of de Rham homology with singular cubic homology with real coefficients. Hence, the homology class $[\mathbf{b}]$ also corresponds to a k-dimensional singular cubic homology class $[\tilde{\mathbf{b}}]$ – which is represented by some closed singular cubic k-chain $\tilde{\mathbf{b}}$ – over which we can integrate. If the support of the k-vector field $\mathbf{b}$ is the image of the k-chain $\tilde{\mathbf{b}}$ then one can think of $\mathbf{b}$ as a k-vector field on $\tilde{\mathbf{b}}$.

This seems to be of immediate relevance to the investigation of the deeper nature of the source currents of physical fields. For instance, one associates real numbers (charges) with point sources and vector fields (currents) with line sources. Of course, one also associates real numbers with higher-dimensional k-chains, such as charge densities, but that might also relate to their contractibility. One might also note that stable currents only flow in conducting loops, which are closed 1-chains.

We also have the homological form of the *Poincaré Lemma:* every point of M has a neighborhood on which any co-closed k-vector field is co-exact. Indeed, one need only find the image of an open ball about the point in some coordinate chart.

In order to facilitate the physical interpretation, we refer to the homology class of a co-closed vector field $\mathbf{J}$ as a *conserved current* and the corresponding cohomology class in dimension $n-1$, $[\#\mathbf{J}] = [i_\mathbf{J}\mathcal{V}]$, as its *flux density*. For any $n-1$-dimensional submanifold S, the integral:

$$\mathbf{J}[S] = \int_S \#\mathbf{J} \tag{2.13}$$

is called the *total flux of* $\mathbf{J}$ *through* S. This allows us to give an integral form to our differential requirements on $\mathbf{J}$, namely:

$$\delta\mathbf{J} = 0 \quad \text{iff} \quad \mathbf{J}[S] = 0 \;\; \text{for all } S \text{ such that } \; S = \partial V \tag{2.14a}$$

$$\mathbf{J} = \delta\mathbf{B} \quad \text{iff} \quad \mathbf{J}[S] = 0 \;\; \text{for all } S \text{ such that } \; \partial S = 0. \tag{2.14b}$$

This can also be accounted for using Stokes's theorem (more precisely, *Gauss's* theorem) in the following form:

$$\delta\mathbf{J}[V] = \mathbf{J}[\partial S]. \tag{2.15}$$

One should observe that although we have sacrificed the ring structure of cohomology for the more limited structure of homology, nevertheless, in the present case, since our homology classes are represented by vector fields we also inherit the structure of a Lie algebra on the one-dimensional homology classes. This is because, as is straightforward to verify, the Lie bracket of conserved vector fields is conserved:

$$\text{if} \;\; \delta\mathbf{X} = 0 \;\; \text{and} \;\; \delta\mathbf{Y} = 0 \;\; \text{then} \;\; \delta[\mathbf{X}, \mathbf{Y}] = 0. \tag{2.16}$$

This is essentially the statement that the infinitesimal generators of volume-preserving diffeomorphisms on an orientable manifold define a Lie algebra. Furthermore, one can define an action of the conserved currents on the homology classes and cohomology classes by way of the Lie derivative operator:

$$L_{[X]}[\mathbf{a} \wedge \mathbf{b} \wedge \ldots \wedge \mathbf{c}] = [L_X(\mathbf{a} \wedge \mathbf{b} \wedge \ldots \wedge \mathbf{c})] \tag{2.17a}$$

where:

$$L_{\mathbf{X}}(\mathbf{a} \wedge \mathbf{b} \wedge \ldots \wedge \mathbf{c}) = [\mathbf{X}, \mathbf{a}] \wedge \mathbf{b} \wedge \ldots \wedge \mathbf{c} + \mathbf{a} \wedge [\mathbf{X}, \mathbf{b}] \wedge \ldots \wedge \mathbf{c} + \mathbf{a} \wedge \mathbf{b} \wedge \ldots \wedge [\mathbf{X}, \mathbf{c}] \tag{2.17b}$$

and

$$L_{[\mathbf{X}]}[\alpha] = [L_{\mathbf{X}}\alpha] \tag{2.17c}$$

in which:

$$L_{\mathbf{X}}\alpha = d i_{\mathbf{X}}\alpha + i_{\mathbf{X}}d\alpha. \tag{2.18}$$

3 Axiomatic topological electromagnetism [11–13]

We assume that we are given an orientable four-dimensional differentiable manifold M with a global unit volume element $\mathcal{V}$. The basic data of electromagnetism consists of a smooth vector field $\mathbf{J}$, which we call the *source current*, and a 2-form F, which we call the *electromagnetic field strength 2-form*.

Because we are trying to avoid all metric-related assumptions, we note that since we cannot distinguish statics from dynamics without the imposition of a rest frame for the motion in question, we must deal with field sources in the most general case, for which the field source is a conserved electric current.

One might easily argue that the vector field $\mathbf{J}$ itself defines a line field over its own support $\mathrm{supp}(\mathbf{J})$, which represents a partial Lorentzian structure on M. Physically, it represents the equivalence class of Lorentz frames at each point of $\mathrm{supp}(\mathbf{J})$ in which the components of $\mathbf{J}$ with respect to all but one leg are zero. In such a frame, $\mathbf{J}$ can be the source of only electrostatic fields, since it has zero spatial velocity in its own rest frame. The only way that a current $\mathbf{J}$ can feel a magnetic field in its own rest space is if $\mathbf{J}$ is not the source of that field. This fact will prove crucial in our discussion of the Lorentz force.

Note that any conserved electric current $\mathbf{J}$ can be decomposed into a product:

$$\mathbf{J} = \rho\mathbf{v} \tag{3.1}$$

in which ρ is a smooth non-negative function that represents a charge density and $\mathbf{v}$ is a smooth vector field that represents the velocity flow field of the charge-bearing matter, although not uniquely. The observation is really more in the nature of a physical interpretation of the conserved electric currents as being reducible to kinematical entities. Note that $\mathbf{v}$ can still be compressible, even though, by definition, $\mathbf{J}$ is not. This decomposition introduces a subtle physical equivalence in the form of the set of all $(\rho, \mathbf{v})$ such that $\mathbf{J} = \rho\mathbf{v}$ for a given incompressible vector field $\mathbf{J}$. This set, in turn, is clearly parameterized by the set of all smooth positive functions on the support of $\mathbf{J}$, which also forms a group under multiplication. We shall return to this group later in the context of dilatational gauge symmetries.

One should observe at this point that – at least in the eyes of electromagnetism – it is the *current* that is fundamental[2], not the decomposition into ρ and $\mathbf{v}$.

Since we cannot demand that $\mathbf{J}$ have compact spacelike support in the absence of a Lorentzian structure, we might also consider a class of $\mathbf{J}$ whose support is on "thickened curves." These would be 4-chains that have 1-chains as deformation retracts, such as four-dimensional cylinders of the form $(0,1) \times B^3$, or solid torii of the form $S^1 \times B^3$, where B^3 is an open 3-ball. Such a $\mathbf{J}$ could be called an *elementary current*. In the former case, one could go one step further and contract the curve to a point; in the latter case, whether one could contract S^1 to a point would depend upon whether M was assumed to be simply connected. In terms of homology, we are asking the question of whether a given 1-chain is a 1-cycle, and, if so, does it bound. This suggests a way of distinguishing static from dynamic currents, at least in the eyes of topology. Since we are dealing with differentiable 1-chains, we are also allowed the luxury of a naturally defined vector field along them, which could play the role of $\mathbf{v}$ above. Note that this also forces us to consider only $\mathbf{J}$'s that are obtained from $\mathbf{v}$ by a conformal factor.

---

[2] Here, we are making a thinly veiled attempt to endow our theory with a learned borrowing from the theory of current algebras.

As for the support of F, we make no assumptions at this point. However, one should keep in mind some of the considerations of the classical theory of electromagnetism, such as the way that the field strengths are not actually defined at source points, although they can be defined inside a spatially extended charge or current distribution. Similarly, causality considerations generally dictate that electromagnetic waves are only defined on light cones. However, we are trying to adhere to a strict regimen of relegating all metric considerations to a corollary status.

Ultimately, a complete theory of electromagnetism should describe the nature of the fields $\mathbf{J}$ and F and the way that they are coupled to each other, and possibly themselves. For instance, we need to know how the source current $\mathbf{J}$ produces the field F. Generally, this relationship takes the form of a partial differential equation, such as Poisson's equation. It can also take the form of an integral equation, as with Huygens's principle, which one can obtain as a fundamental solution to the differential equation. Of particular interest are the "radiative" solutions, which are presumably due to the possibility that $\mathbf{J}$ might be time varying, such as when $\mathbf{v}$ is also time varying. A key point to resolve is whether such radiative solutions are produced by a $\mathbf{J}$ that is associated with a constant acceleration in $\mathbf{v}$. If they are not, then we have another internal symmetry to the field theory, which relates to conformal symmetries that are more involved than homotheties, namely, the nonlinear inversion symmetries that correspond to the transformations from an inertial frame to a frame that moves with constant relative acceleration. Once again, we point out that if $\mathbf{J}$ is the source of F then it is entirely possible that a time variation in ρ can cancel a time variation in $\mathbf{v}$ in such a way that $\mathbf{J}$ has no time variation and produces no radiative F as a consequence. (However, since this only pertains to time variations within the line spanned by $\mathbf{v}$, this does not seem to suggest any immediate application to the problem of canceling the centripetal acceleration of atomic electrons.)

Conversely, if another electric current $\mathbf{J}'$ is in the field F that is produced by a source $\mathbf{J}$, one expects that there should be a coupling of F to $\mathbf{J}'$. In its simplest form, this is the Lorentz force law. However, the fact that accelerated charges might produce radiation in reaction to this force – i.e., a field F' – suggests that the situation is more involved, especially when one considers a nonlinear superposition of F' with F.

There is also the question of how the field F that is produced by one current $\mathbf{J}$ interacts with the field F' of another current $\mathbf{J}'$. Whether or not the interaction between the fields F and F' is one of simple superposition (tensor addition or what have you) depends on the degree of linearity in the field equation that couples currents to fields, since we could also say that the resulting field must be a solution of the same field equation when one uses the combination of $\mathbf{J}$ and $\mathbf{J}'$ as source. We shall not elevate this statement to the status of an axiom since it represents simply a physical interpretation of the solution to the field equations when the source is $\mathbf{J} + \mathbf{J}'$. In the following set of axioms, the degree of linearity in the field equation depends upon the nature of the constitutive law that one assumes for the medium in which the fields are defined.

Finally, if we are to leave open the possibility of nonlinear electromagnetism, we should seriously consider the possibility of self-interactions for $\mathbf{J}$ and F. The problem of defining a stable static charge distribution that is not pointlike seems to demand the introduction of a self-interaction of the charge distribution that is attractive at quantum-scale distances and dies away much faster than $1/r^2$. Similarly, the fact that an accelerating charge – even in the absence of an external field – will presumably radiate energy, which amounts to a decelerating force that acts on the charge, also represents a type of indirect self-interaction of $\mathbf{J}$ itself by way the intermediary of $\mathbf{J}$ producing a radiation field which then interacts with $\mathbf{J}$ as a Lorentz-type force.

The interaction of F with itself might take the form of a "saturation limit" term for the field strengths or something more elaborate that produces the phase transition of vacuum polarization past that limit. In fact, the usual statement of the Klein paradox amounts to asking how one is supposed to deal with self-interactions of the electromagnetic field under vacuum polarization without simply expanding one's scope to a nonlinear theory. The notion that there should be a saturation limit for electric field strengths as a result of vacuum polarization is at the basis for the Born-Infeld model [14, 15] for nonlinear electrodynamics.

The axioms of topological electromagnetism can be given the following form:

Axiom 1. *Conservation of electric charge:* $\mathbf{J}$ defines a homology class in $H_1(M;\mathbb{R})$, i.e.:

$$\delta J = 0. \tag{3.2}$$

Axiom 2. *Conservation of magnetic flux:* F defines a cohomology class in $H^2(M;\mathbb{R})$, i.e.:

$$dF = 0. \tag{3.3}$$

Axiom 3. Constitutive law: There exists a smooth map that takes one of the following forms:

$$f : \Lambda^2(M;\mathbb{R}) \to \Lambda_1(M;\mathbb{R}), \qquad\qquad F \mapsto \mathbf{J} \tag{3.4a}$$

$$f : \Lambda_1(M;\mathbb{R}) \times \Lambda^2(M;\mathbb{R}) \to C^\infty(M), \qquad f(F, \mathbf{J}) = 0. \tag{3.4b}$$

Furthermore, we make the following physical interpretations of certain geometric objects:

Physical interpretation 1. *Lorentz force:* The 1-form $i_{\mathbf{J}}F$ represents the force that is exerted by the electro-magnetic field F on the current $\mathbf{J}$. When $\mathbf{J}$ is the source of F this force is that of radiative reaction.

Physical interpretation 2. *Superposition:* When F is the field produced by a source $\mathbf{J}$ and F' is the field of a source $\mathbf{J}'$ then the field that is produced by the source $\mathbf{J} + \mathbf{J}'$[3] represents the (possibly nonlinear) superposition of the fields F and F'.

Physical interpretation 3. *Constitutive law.* If the operator $f = \delta \cdot \chi$ for some diffeomorphism $\chi : \Lambda^2(M;\mathbb{R}) \to \Lambda_2(M;\mathbb{R})$ then the equations:

$$\mathfrak{h} = \chi(F) \quad \text{and} \quad \delta\mathfrak{h} = \mathbf{J} \tag{3.5}$$

complete the usual Maxwell equations, in which any nonlinearity would originate in the constitutive law χ and reduces to linearity in the limiting case of weak field strengths.

In the paper of Hehl et al. [11], two other axioms were mentioned. In addition to the aforementioned three axioms, it was suggested that one would need to postulate the existence of a proper time simultaneity foliation [16] and raise the existence of a Lorentz force from the status of a physical interpretation to a mathematical axiom. In the present formulation, the latter axiom is unnecessary because if $\mathbf{J}'$ is another conserved current that is in the field F produced by the source current $\mathbf{J}$ then we can form the interior product:

$$\alpha = i_{\mathbf{J}'}F \tag{3.6}$$

with no further assumptions and interpret it as the force that F exerts on $\mathbf{J}'$. Hence, we simply include the Lorentz force as a physical interpretation, not an axiom.

As for the axiom concerning the existence of a proper time simultaneity foliation, which seems mostly directed towards facilitating the definition of the electric and magnetic 1-forms that one obtains from F when one chooses a rest frame, let us subdivide this issue into two sub-issues:

1. The existence of a Whitney sum decomposition of $T(M)$ into $L(M) \oplus \Sigma(M)$, where $L(M)$ is a (real) line bundle that represents a field of proper time directions and $\Sigma(M)$ is a rank-three vector sub-bundle that represents a distribution of spacelike tangent subspaces.

2. The integrability of the differential system on M that $\Sigma(M)$ represents into a codimension foliation of M by proper time simultaneity submanifolds.

---

[3]   The sum in this expression refers to the addition of 1-chains. Of course, the carrier of the sum is the union of the carriers and the coefficients at the points of their intersection is the vector sum, so this is still a vector addition, after all.

Although the subject at hand is discussed in more detail in [16], for now, we simply point out that such a Whitney sum decomposition is usually an artifact of defining a Lorentzian structure, by way of a choice of rest frame. Since we are trying to exhibit the reduction from the $SL(4)$ structure that facilitates electromagnetism to the bundle of oriented Lorentzian frames as a consequence of the axioms of electromagnetism, we shall try to avoid introducing the axiom of such a decomposition from the outset.

However, we can say this much about Lorentzian structures and rest frames, *a priori:* All it takes to define a Lorentzian structure, at least in the eyes of homotopy theory, is a line field, and one does have such a Lorentzian structure on the set of points, $NZ(\mathbf{J})$, at which $\mathbf{J}$ is non-zero; since this set is an open subset of M, it is also a manifold. The line $L_x(\mathrm{S})$ at every point $x \in NZ(\mathbf{J})$ is then the one that is spanned by $\mathbf{J}$ at that point. Furthermore, because the line field $L_x(NZ(\mathbf{J}))$ was defined by a non-zero vector field to begin with, it is an orientable line bundle. As for the matter of defining a Lorentzian pseudo-metric, it is important to note that the homotopy equivalence of a global line field on M with a Lorentzian pseudo-metric g does not define a *canonical* association of one with the other. In effect, any 3-plane in each tangent space to $x \in NZ(\mathbf{J})$ that does not contain the line $L_x(NZ(\mathbf{J}))$ could serve as an orthogonal complement.

The line bundle $L(NZ(\mathbf{J})) \to NZ(\mathbf{J})$ also defines a rest frame for the motion of the charge distribution that defines the current, namely the equivalence class of unit-volume linear frames at each point of $NZ(\mathbf{J})$ that have one leg in $L_x(NZ(\mathbf{J}))$. Since we have not specified which Lorentzian metric is associated with $L(NZ(\mathbf{J}))$ we can only define a reduction of $SL(M)$ to $SL(3)(M)$ as a result of this. Similarly, without a choice of g we have no unique choice of complementary sub-bundle to $L(NZ(\mathbf{J}))$ in $T(NZ(\mathbf{J}))$. Hence, the integrability of the $3+1$ decomposition of $T(NZ(\mathbf{J}))$ into a foliation of $NZ(\mathbf{J})$ by proper time simultaneity leaves becomes moot.

Notice that when we state the integral forms of the first two axioms, namely:

$$\mathbf{J}[V] = \int_V \#\mathbf{J} = 0 \qquad \text{for any bounding three-dimensional region } V \tag{3.7a}$$

$$F[S] = \int_S F = 0 \qquad \text{for any bounding two-dimensional region } S, \tag{3.7b}$$

we cannot make the usual stipulations about whether these regions are spacelike, since we are trying to relegate such purely Lorentzian concerns to *consequences* of the electromagnetic structure.

If we look at the first axiom in differential form, we observe that when $J = \rho\mathbf{v}$ it also takes the form of a continuity equation:

$$0 = \delta\mathbf{J} = d\rho(\mathbf{v}) + \rho\delta\mathbf{v} \tag{3.8}$$

or:

$$\mathbf{v}\rho = -\rho\delta\mathbf{v}, \tag{3.9}$$

which can also take form:

$$\delta\mathbf{v} = -\mathbf{v}(\ln\rho). \tag{3.10}$$

This shows that $\mathbf{v}$ is incompressible iff ρ is constant along the flow of $\mathbf{v}$. We shall return to the formal structure of equation (3.10) in a different context.

The integral form of the first axiom then becomes:

$$\int_N \mathbf{v}(\ln\rho)\mathcal{V} = -\int_N d\#\mathbf{v} = -\int_V \#\mathbf{v}. \tag{3.11}$$

whenever $V = \partial N$.

If one interprets the functional $F[S]$ as producing the *total magnetic flux* in an arbitrary compact orientable 2-dimensional submanifold S then one sees that if S bounds, so $S = \partial V$, then this makes physically intuitive sense and implies that the total magnetic charge in V is 0. However, when S does not bound a three-dimensional region, although the total magnetic flux in S might well be non-zero, nevertheless, one could not physically interpret the integral as the magnetic charge contained in a three-dimensional region since Gauss's theorem would no longer apply.

Of course, since we are defining a source current as a one-dimensional homology class and an electromagnetic field as a two-dimensional cohomology class one naturally must confront the question of whether there is only one such class or more than one, i.e., whether $H_1(M; \mathbb{R})$ or $H^2(M; \mathbb{R})$ vanishes. This naturally leads into the subject of our next section.

Finally, one must point out that there is still something formally incomplete about our system of axioms in that:

a) Ultimately it would be preferable to exhibit the fields F and $\mathbf{J}$ as having some intrinsic geometrical or topological origin, instead of introducing them as essentially logical primitives.

b) The constitutive law also seems to lack an immediate relationship to either the geometry or topology of the spacetime manifold.

Hence, the primary goals of refining the present axioms should probably be those of:

a) Exhibiting the spacetime metric and wave motion as special case corollaries to the constitutive laws.

b) Constructing the constitutive law from physical first principles and fundamental physical processes, as one might in the physics of condensed matter.

c) Illuminating a path into nonlinear electrodynamics that might define deeper foundations for quantum electrodynamics.

d) Suggesting a promising Ansatz for exhibiting the fields F and $\mathbf{J}$ as being naturally associated with geometrical or topological structures associated with the $SL(4)$ reduction of the bundle of linear frames on spacetime that is necessary in order to facilitate Poincaré duality.

4 Gauge symmetries and conservation laws

The reason for the use of the plural in the title of this section is that actually, we have *two* ways in which ambiguity can arise in the construction of our basic fields F and $\mathbf{J}$. The first is the usual one that relates to the possibility that the closed 2-form F is also exact; i.e., whether there exists a potential 1-form A such that:

$$F = dA. \tag{4.1}$$

There are two comments that must be made about this: First: Whether or not such an A exists globally depends upon the vanishing of $H^2(M; \mathbb{R})$. If that module does not vanish then one can find only local potential 1-forms. Second: If z is a closed 1-form then $A + z$ is also a potential 1-form for F. This defines an equivalence class of 1-forms by the equivalence relation:

$$A \sim A' \qquad \text{iff} \qquad A - A' \in Z^1(M; \mathbb{R}). \tag{4.2}$$

One refers to the equivalence relation as *gauge equivalence* and the transformation:

$$A \mapsto A + z \tag{4.3}$$

as a *gauge transformation of the second kind.* In order to describe what the gauge transformations of the first kind were, we need to account for the appearance of the usual $U(1)$ symmetry.

Customarily, one exhibits the closed form z as an exact form $d\lambda$, which means that zero-form λ represents a sort of infinitesimal generator for the gauge transformation. Naturally, the question of whether global z exist that are not exact depends upon the vanishing of $H^1(M; \mathbb{R})$, which is essentially the question of whether M is simply connected. However, since the customary physical arguments are usually local in scope, the Poincaré lemma allows us to find a suitable zero-form λ locally. This shows that the set of infinitesimal gauge transformations of the second kind can be effectively parameterized by the 0-forms. This means that we can regard the vector space of 0-forms as the (Abelian) Lie algebra of the group of gauge transformations. Since the Lie algebra over $\mathbb{R}$ can define the infinitesimal transformations of either the group $(\mathbb{R}^+, \times)$ or the circle group, which can also be considered as $SO(2)$ or $U(1)$, the group of gauge transformations can be either the Lie group of positive smooth functions on M under multiplication, or the Lie group of smooth functions from M into S^1. Since the usual physical discussions of gauge field theory are mostly concerned with infinitesimal transformations – those being the ones that define field variations – the choice of the Lie group $(\mathbb{R}^+, \times)$ or $U(1)$ for finite transformations is one of physical intuition, and this usually suggests that a gauge is like a phase; hence, the choice of $U(1)$ is customary.

The traditional reasoning behind this is that quantum mechanics introduces a $U(1)$ phase factor into the description of any particle by wave functions, whether they are photons, whose position distributions are obtained from complex wave functions, or electrons, whose position distributions are obtained from spinor wave functions. However, just as introducing a metric into the theory of electromagnetism seems premature when the metric should be a consequence of the propagation of electromagnetic waves, it also seems somewhat premature to introduce traditional quantum mechanical arguments into a theory that – one hopes – might shed new light on the deeper nature of quantum mechanics to begin with. Consequently, the author has been pursuing the possibility that the gauge structure for electromagnetism is not something one introduces axiomatically and independently from the geometry and topology of spacetime – as we already did with the fields F and $\mathbf{J}$ themselves – but something that follows naturally from the demands of physical interpretation. In particular, obtaining the gauge structure as actually an $SO(2)$-reduction of the bundle of linear frames on M seems promising (see [17]). However, that discussion would take us far from the immediate concerns, so we simply accept the traditional argument, for now, and proceed.

If we then give z the form (which is no loss of generality):

$$z = d(\ln g) = g^{-1}dg, \tag{4.4}$$

where $g: M \to U(1)$, then, if we introduce a slight redundancy, the transformation (4.3) takes the form:

$$A \mapsto g^{-1}Ag + g^{-1}dg \tag{4.5}$$

that one finds for the transformation of the local representative of a $U(1)$ connection 1-form. One then refers to the group $U(1)$ as a gauge group (of internal symmetries) for the field F and the choice of g (or z) as a choice of local $U(1)$ *gauge* for F, since F is locally equivalent to the pair (A, z). This set of associations also carries with it the interpretation of $F = dA$ as the curvature 2-form that is associated with the connection 1-form A.

In order to complete this geometric picture, we need to account for the appearance of a $U(1)$ principal bundle $P \to M$ on which the connection 1-form A would be defined; one generally refers to such a principal fibration as a *gauge structure* for the field theory. However, this gauge structure is usually constructed independently of the bundle of linear frames on spacetime as an essentially auxiliary object.

If one forms the electromagnetic field Lagrangian:

$$\mathcal{L} = \frac{1}{2}dA(\mathfrak{h}) + A(\mathbf{J}) \tag{4.6}$$

one sees that as long as one assumes that the constitutive equation gives $\mathfrak{h}$ as a function of F, and not A, then the kinetic energy term is gauge invariant. Under a gauge transformation, the interaction term $A(\mathbf{J})$

picks up a contribution of:

$$d\lambda(\mathbf{J}) = \mathbf{J}\lambda. \tag{4.7}$$

Since this does not always vanish, we see that the interaction term breaks the gauge invariance of $\mathcal{L}$. By Noether's theorem, the $U(1)$ internal symmetry of $\mathcal{L}$ is associated with a conserved current:

$$\mathbf{J} = \frac{\delta\mathcal{L}}{\delta A}, \tag{4.8}$$

which means that the $U(1)$ gauge symmetry is equivalent to charge conservation.

A less-discussed form of gauge equivalence is the one that pertains to $\mathbf{J}$ when one addresses the question of whether a *super-potential* $\mathfrak{h}$ exists for $\mathbf{J}$; i.e., an $\mathfrak{h} \in H_2(M;\mathbb{R})$:

$$\mathbf{J} = \delta\mathfrak{h}. \tag{4.9}$$

By the homological Poincaré lemma such an $\mathfrak{h}$ always exists locally. Whether this is also possible globally is a matter of whether $H_1(M;\mathbb{R})$ vanishes.

By the Hurewicz isomorphism theorem, this comes down to whether M is simply connected or not. If M is not simply connected, we could pass to the simply connected covering manifold $\tilde{M}$, where one would always have the existence of a globally defined $\mathfrak{h}$. However, since a global section of the covering map $s\colon M \to \tilde{M}$ would have to represent a global orientation, when M is not orientable s will necessarily be undefined at some points of M. The same will then be true of the image of $\mathbf{J}$ in $\mathfrak{X}(\tilde{M})$, namely $s_*\mathbf{J}$, as well as the pull back of $\mathfrak{h}$ to M by s, namely $s^*\mathfrak{h}$.

Once again, we see that $\mathfrak{h}$ is not uniquely defined. Since $\delta^2 = 0$, if $\mathbf{Z} \in Z_2(M;\mathbb{R})$ is a co-closed 2-form then the 2-form that one obtains by the transformation:

$$\mathfrak{h} \mapsto \mathfrak{h} + \mathbf{Z} \tag{4.10}$$

is also a superpotential for $\mathbf{J}$. This defines a different sort of gauge equivalence for superpotentials:

$$\mathfrak{h} \sim \mathfrak{h}' \qquad \text{iff} \qquad \mathfrak{h} - \mathfrak{h}' \in Z_2(M;\mathbb{R}). \tag{4.11}$$

Instead of looking for a 0-form to locally represent these transformations, we look for a 3-vector field $\mathbf{K}$ such that $\mathbf{Z} = \delta\mathbf{K}$. By Poincaré duality, $\mathbf{K}$ corresponds to a 1-form $\#\mathbf{K}$. Hence, the space of infinitesimal superpotential gauge transformations is parameterized by $\Lambda^1(M)$ this time. If we are only being local, we can exhibit these transformations by smooth functions from M to $\mathbb{R}^4$. If we wish to regard this as the Abelian Lie algebra of some Lie group of functions from M to an Abelian Lie group G then one should observe that $\mathbb{R}^4$ can cover any of the Lie groups $\mathbb{R}^4, S^1 \times \mathbb{R}^3, T^2 \times \mathbb{R}^2, T^3 \times \mathbb{R}, T^4$. This level of ambiguity suggests that we are perhaps following the wrong ansatz. Furthermore, there is a fundamental difference between the ambiguity in $\mathfrak{h}$ expressed by (4.11) and the gauge ambiguity in the potential 1-form for F: the components of the electromagnetic induction bivector are physically measurable, but the introduction of a potential 1-form has a somewhat mathematically formal character.

Hence, rather than pursue the analogous reasoning that we followed for the aforementioned gauge transformations, let us look at equation (3.10) again. This time, we shall put it into the form:

$$\delta\mathbf{v} = -i_{\mathbf{v}}(\rho^{-1}d\rho). \tag{4.12}$$

The 1-form $\rho^{-1}d\rho$ takes the same form as the $g^{-1}dg$ term that appeared in the context of the gauge transformations of A. We are then tempted to consider it to be a 1-form with values in the Lie algebra $\mathbb{R}$. However, since we are dealing with vector fields with non-zero divergence, we suspect that this time $\mathbb{R}$ is not the imaginary axis that generates the two-dimensional rotations, but the real axis that generates the dilatations, which is the multiplicative group $\mathbb{R}^+$.

If we recall that ρ originated when we chose to decompose $\mathbf{J}$, which is assumed to have zero divergence, into $\rho\mathbf{v}$ then one could also say that, in effect, since this choice of decomposition for $\mathbf{J}$ depends upon the choice of ρ, one must choose a *dilatational* gauge in order to define the vector field $\mathbf{v}$. It is intriguing that such a dilatational gauge is also associated with a choice of parameterization for an integral curve of $\mathbf{v}$, which also relates to mass and kinetic energy. At the moment, though, our concern is how it relates to charge conservation.

Now let us assume that $\rho^{-1}d\rho$ is the 1-form that appears when one makes a change of local gauge on an $\mathbb{R}^+$-principal bundle by way of ρ and transforms a local connection form σ from one gauge to the other:

$$\omega \mapsto \rho^{-1}\sigma\rho + \rho^{-1}d\rho. \tag{4.13}$$

This time, we have no problem identifying the $\mathbb{R}^+$-principal bundle in question, since it is undoubtedly the associated $\mathbb{R}^+$-principal bundle to $L(NZ(\mathbf{J}))$, which consists of all positive scalar multiples of the vector $\mathbf{J}_x$ at any point $x \in M$. We shall denote this $\mathbb{R}^+$-principal bundle by $SL(NZ(\mathbf{J}))$. Ordinarily, a line field on a manifold only defines an $\mathbb{R}^*(= \mathbb{R} - \{0\})$-principal bundle by way of the complement of its zero section. In order to reduce this to an $\mathbb{R}^+$-principal bundle, we would need to first show that the line field is orientable and then choose an orientation, which is simply a non-zero vector field on the support of the line field. However, this is also how we defined $L(NZ(\mathbf{J}))$ in the first place, so the orientation is automatic. One must note that the bundle $SL(NZ(\mathbf{J}))$ of unit volume 1-frames on $NZ(\mathbf{J})$ is topologically uninteresting, since the existence of a global section (over $NZ(\mathbf{J})$) makes it trivial: $SL(NZ(\mathbf{J})) = NZ(\mathbf{J}) \times \mathbb{R}^+$. Hence, an oriented one-frame on $NZ(\mathbf{J})$ is equivalent to smooth function on $NZ(\mathbf{J})$ with values in $\mathbb{R}^+$. In particular, if $\mathbf{v} = \lambda\mathbf{J}$ with $\lambda > 0$ then the function associated with the oriented 1-frame $\mathbf{v}$ is simply λ.

The question then becomes that of how we physically account for the appearance of a connection 1-form on $SL(NZ(\mathbf{J}))$. To address this issue, we first consider the role that it plays. Since ρ seems to be related to the scaling of the vector field $\mathbf{J}$ within the oriented line field that it spans, the issue would seem to be how one compares the unit of (one-dimensional) "volume" in the fiber $L_x(NZ(\mathbf{J}))$ at one point with the unit of volume at a neighboring point. (Of course, in dimension one a volume serves the same purpose as a norm or metric.) One must remember that in the absence of a metric there is nothing to distinguish any 1-frame of $L(NZ(\mathbf{J}))$ as having a "unit" volume, except when one makes such an association arbitrarily. Hence, if that choice is to be truly arbitrary then the scaling factor that takes any 1-frame to a rescaled one must represent an internal symmetry of the field theory. However, one notes that in the conventional theory of electromagnetism this homothety invariance is only true for the electromagnetic waves themselves and is broken by the introduction of a mass term.

The connection 1-form ω that we defined takes its values in $\mathbb{R}$, which is the Lie algebra of $(\mathbb{R}^+, \times)$. Like the electromagnetic potential 1-form, it too is an ordinary differential form. Consequently, we suspect that the only geometrical object that relates to it that also has an unambiguous physical meaning is its curvature, or field strength, 2-form:

$$\Omega = d\omega. \tag{4.14}$$

(The term $\omega \wedge \omega$ vanishes because the Lie algebra of $(\mathbb{R}^+, \times)$ is Abelian.)

Although it is tempting to speculate on whether Ω might relate, at least indirectly to the definition of the 2-form F, keep in mind that Ω is defined only over $NZ(\mathbf{J})$, whereas F is essentially defined on the complement of $NZ(\mathbf{J})$. Hence, one would have to replace the definition of F with a further "topological constitutive map":

$$\Lambda^2(NZ(\mathbf{J})) \rightarrow \Lambda^2(M - NZ(\mathbf{J})). \tag{4.15}$$

In effect, this is what we usually get from solving the field equations: a map from the source to the field. Of course, since we are defining Ω to be an exact form and F to be a closed form, if we pass to cohomology and the induced map is linear it would only take $[\Omega]$ to 0, so unless $[F] = 0$, as well – i.e., $F = dA$ – we should not expect this Ansatz to be productive in the context of topological electromagnetism unless the definition of Ω were weakened to something closed, but not exact.

5 The constitutive axiom

Naturally, the most open-ended axiom in the aforementioned set is the third one that concerns the existence of an electromagnetic constitutive map. The problem is to leave open the possibility that the constitutive law is nonlinear in such a way that it still relates to the geometrical or topological nature of the basic objects. One must unavoidably consider nonlinear theories of electromagnetism for both practical considerations, such as nonlinear optics, and more theoretical considerations, as when one is considering quantum electrodynamics. In either case, the domain of relevance tends to be the realm of large field strengths, such as the electric and magnetic fields of high-energy laser beams, or the electric field of an electron or an atomic nucleus when one is sufficiently close.

The existence of vacuum polarization in QED tends to suggest a nonlinear constitutive law for the spacetime vacuum itself. (For a lengthier discussion of such matters, see [18].) Since the constitutive laws of macroscopic media are usually obtained by considering the macrostates (in the thermodynamic sense) that follow from the microstates of more fundamental processes, such as the formation of electric or magnetic dipoles in charge distributions, this seems to be the most promising direction for trying to probe the nature of the constitutive law that governs the electromagnetic vacuum state.

In a topological theory of electromagnetism, one might also wish that the microstates in question contain fundamental topological information. For instance, Misner and Wheeler [19] once attributed the appearance of charge itself to the existence of the much-discussed "spacetime wormholes," which amount to topological objects that render $H^2(M; \mathbb{R})$ non-trivial. An unacceptable aspect of the attachment of wormholes in the eyes of pre-metric electromagnetism is that they are three-dimensional objects that look like $S^1 \times S^2$ and are attached (by connected sum) to the three-dimensional spacelike leaves of a chosen proper time simultaneity foliation. In fact, for the purposes of most spacetime models, if one regards spacetime as the Cauchy development of an initial spatial manifold Σ then the foliation is the "cylindrical" foliation $M = \mathbb{R} \times \Sigma$. In the absence of such a foliation, one can only attach four-dimensional objects to a four-dimensional manifold by connected sum and still produce a manifold. If one attaches, say, a point or a curve segment and wishes to still produce a four-dimensional manifold, one must make some further identifications and pass to a quotient, such as attaching the point at infinity to a plane by stereographic projection. Later, we shall discuss topological matters in the context of constitutive laws.

As we have defined the term, a constitutive law will generally be a differential equation in the two fields $\mathbf{J}$ and F. However, as pointed out above, this equation may factor through an essentially algebraic map $\mathfrak{h}$: $\Lambda^2(M) \to \Lambda_2(M)$ and the codifferential δ – i.e., $\delta\mathfrak{h} = \mathbf{J}$ – which is the traditional form.

One can appeal to the theory of gravitation for an analogy. The spirit of Sakharov's notion of gravitation as "metrical elasticity" is that if spacetime curvature is essentially related to the second covariant derivative of the strain that is associated with deforming the Minkowski metric $\eta_{\mu\nu}$ into the Lorentzian metric $g_{\mu\nu}$ – à la the Cauchy-Green conception of the strain tensor, which would amount to $g_{\mu\nu} - \eta_{\mu\nu}$ – and this is coupled to stress-energy-momentum tensor then the Einstein field equations take the form of a constitutive equation in differential form.

One can give the electromagnetic constitutive axiom different forms, depending upon whether one assumes that $F = dA$ for some $A \in \Lambda^1(M; \mathbb{R})$ or $\mathbf{J} = \delta\mathfrak{h}$ for some $\mathfrak{h} \in \Lambda_2(M; \mathbb{R})$. For instance, in the former case, one might consider laws of the form:

$$A = A(\mathbf{J}), \qquad \mathbf{J} = \mathbf{J}(A), \qquad f(A, \mathbf{J}) = \text{const.} \tag{5.1}$$

In the latter case, one might consider laws of the form:

$$F = F(\mathfrak{h}), \qquad \mathfrak{h} = \mathfrak{h}(F), \qquad f(F, \mathfrak{h}) = \text{const.} \tag{5.2}$$

In (5.1), the last expression represents a common form for the interaction term $A \wedge \#\mathbf{J} = (A_\mu J^\mu)\mathcal{V}$ in the electromagnetic Lagrangian that expresses how the source current gets coupled to the electromagnetic field. Similarly, the last expression of (5.2) is of the form of the field kinetic energy term $\frac{1}{2}F \wedge \#\mathfrak{h} = (\frac{1}{4}F_{\mu\nu}\mathfrak{h}^{\mu\nu})\mathcal{V}$ in such a Lagrangian.

Hence, we form the expressions:

$$\frac{1}{2} F^{\wedge} \# \mathfrak{h} = \frac{1}{2} F(\mathfrak{h}) \mathcal{V}, \quad A^{\wedge} \# \mathbf{J} = A(\mathbf{J}) \mathcal{V}, \tag{5.3}$$

which are actually equivalent to forming:

$$\frac{1}{2} F^{\wedge} H, \qquad A^{\wedge *} J, \tag{5.4}$$

when one deals with H as a 2-form and J as a 3-form. However, the latter form requires the introduction of a metric in order to associate the vector field $\mathbf{J}$ with the 1-form J and the bivector field $\mathfrak{h}$ with the 2-form H. Later, we shall examine the expression $F^{\wedge} H$ in the context of the intersection form of the spacetime manifold.

In local components, and neglecting the factor of $\mathcal{V}$, (5.3) gives the usual expressions:

$$\mathcal{L}_{\text{kin}} = \frac{1}{4} F_{\mu\nu} \mathfrak{h}^{\mu\nu}, \qquad \mathcal{L}_{\text{int}} = A_{\mu} J^{\mu}. \tag{5.5}$$

The 2-vector field $\mathfrak{h}$ of electromagnetic induction is the one that appears in the Maxwell equations, when we give them the form:

$$dF = 0, \qquad \delta\mathfrak{h} = \mathbf{J}, \qquad \mathfrak{h} = \mathfrak{h}(F), \tag{5.6}$$

The homological Poincaré lemma says that one can, at least locally, find a 2-vector field $\mathfrak{h}$ such that:

$$\mathbf{J} = \delta\mathfrak{h}. \tag{5.7}$$

Of course, in the present case, as opposed to the usual machinery of Maxwell's equations in terms of differential forms, $\mathfrak{h}$ is a 2-vector field, instead of a 2-form. However, this was always implicit when one wrote the Maxwell equations in the older local – but still metric-free – form:

$$\partial_{\mu} F_{\nu\lambda} + \partial_{\nu} F_{\lambda\mu} + \partial_{\lambda} F_{\mu\nu} = 0, \qquad \partial_{\mu} \mathfrak{h}^{\mu\nu} = J^{\nu}, \qquad \mathfrak{h}^{\mu\nu} = \mathfrak{h}^{\mu\nu}(F_{\mu\nu}). \tag{5.8}$$

The situation that is associated with the existence of a global $\mathfrak{h}$ on a manifold is simpler than the situation that relates to the existence of a global potential 1-form A such that $F = dA$. In the former case, where the issue is the vanishing of $H_1(M; \mathbb{R})$, one can define one's fundamental fields F, $\mathbf{J}$, $\mathfrak{h}$ on the simply connected covering manifold of M and pull them down to subsets of M by way of local orientations. In the latter scenario, the issue is the vanishing of $H^2(M; \mathbb{R})$, which cannot be sidestepped by passing to a covering manifold, and leads to the possible existence of magnetic monopoles and spacetime wormholes.

If one is given the Lagrangian $\mathcal{L} = \mathcal{L}(F, \mathbf{J})$ a *priori* then one can retrieve the constitutive law from:

$$\mathfrak{h} = \frac{\partial \mathcal{L}}{\partial F}. \tag{5.9}$$

If we sense that we are close to dealing with a spacetime metric then that is due to the fact that we are concerned with a kinetic energy, and, as one knows from relativistic point mechanics, kinetic energy is, in its simplest form, conformally related to the spacetime metric. Furthermore, the conformal factor is simply the rest mass of the point particle; of course, in the case of electromagnetic waves this conformal factor would be zero.

Another piece of the geometric puzzle that is associated with constitutive laws comes from the fact that the speed of propagation of electromagnetic waves in vacuo – i.e., c_0 – is actually a derived quantity that follows from the assumed form of the constitutive law for the classical electromagnetic vacuum. That law

amounts to the assumption that the electric and magnetic polarizability of that state is expressed by two constants, ε_0 and μ_0, which give:[4]

$$c_0^2 = \frac{1}{\varepsilon_0 \mu_0}. \tag{5.10}$$

As is well known, one of the most far-reaching consequences of quantum electrodynamics is the suggestion that this aforementioned assumption is an oversimplification of the nature of the electromagnetic vacuum state. In particular, vacuum polarization seems to contribute at every turn to the nature of elementary electromagnetic processes at the quantum level. Consequently, one expects that if classical electromagnetism leads to gravitation somehow then quantum electrodynamics might lead to quantum gravity (or at least an unambiguous definition of the concept). Alternatively, if that is not the case then perhaps it is only quantum electrodynamics that leads to gravitation in the first place; indeed, this is probably more likely.

An intriguing hint of how this might come to pass is given by the fact that (5.10) also suggests that one can rewrite the simplest spacetime metric – i.e., the Minkowski space metric – in the form:

$$\eta_{\mu\nu} = \operatorname{diag}\left(\frac{1}{\varepsilon_0 \mu_0}, -1, -1, -1\right) = \frac{1}{\varepsilon_0}\operatorname{diag}\left(\frac{1}{\mu_0}, -\varepsilon_0, -\varepsilon_0, -\varepsilon_0\right). \tag{5.11}$$

In this one elementary geometrical object, we can see Ansätze that might lead to the unification of gravitation, electromagnetism, vacuum polarization, and wave mechanics.

In order to gain more intuition about the nature of the problem of how the electromagnetic constitutive model leads to the spacetime metric as a consequence, we examine two special cases of an electromagnetic constitutive model: the linear case and the nonlinear case as it relates to nonlinear optics and the Born-Infeld model, which they constructed in order to account for vacuum polarization effects to some degree. We then make some observations and speculations on the problem of resolving the constitutive laws into geometrical and topological aspects of more fundamental processes.

5.1   Linear case

In conventional linear electrodynamics one ordinarily assumes the *a priori* existence of a Lorentzian metric g, so the constitutive relation takes the 2-form F to the 2-form H by way of:

$$H(F) = (\iota_g \times \iota_g) \cdot \chi(F), \tag{5.12}$$

where:

$$\iota_g : T(M) \to T^*(M), \qquad \mathbf{v} \mapsto i_{\mathbf{v}} g \tag{5.13}$$

is the linear isomorphism that takes each tangent vector to its metric-dual covector and:

$$\chi : \Lambda^2(M) \to \Lambda_2(M), \qquad F \mapsto \mathfrak{h} = \chi(F) \tag{5.14}$$

is a linear isomorphism of 2-forms with 2-vector fields. In this form, we see that in an unpolarized medium the correspondence (5.12) is simply the identity map and one should have simply:

$$\chi = (\iota_g \times \iota_g)^{-1}. \tag{5.15}$$

We then see that the effect of polarization is to deform the isomorphism of $\Lambda^2(M)$ with $\Lambda_2(M)$ that is defined when one has a Lorentzian metric g. One might even consider the possibility that this deformation originates in a deformation of the metric itself, $g \mapsto g' = g + \delta g$, which should make:

$$\iota_{g'} \times \iota_{g'} = \iota_g \times \iota_g + \iota_{\delta g} \times \iota_g + \iota_g \times \iota_{\delta g} + \iota_{\delta g} \times \iota_{\delta g}. \tag{5.16}$$

---

[4]   Although the units in which $c_0 = 1$ are sometimes called "God's units," the fact that they tend to obscure the deeper truth of this physical subtlety suggests that they are probably "Satan's units," since God does not play games with truth.

We could then define the polarization tensor field $\mathfrak{p}$ to be the difference between the deformed and undeformed states:

$$\mathfrak{p} = (\iota_{g'} \times \iota_{g'})^{-1} - (\iota_g \times \iota_g)^{-1}. \tag{5.17}$$

To relate this to the usual definition of the polarization tensor field we define:

$$P = (\iota_g \times \iota_g)\mathfrak{p} = (\iota_g \times \iota_g)(\iota_{g'} \times \iota_{g'})^{-1} - I, \tag{5.18}$$

which agrees with the usual conception of the polarization tensor field as $H - I$ if we set:[5]

$$H = (\iota_g \times \iota_g)(\iota_{g'} \times \iota_{g'})^{-1} = (\iota_g \times \iota_g)\chi. \tag{5.19}$$

Relative to a natural frame field, (5.12) takes the component form:

$$H_{\mu\nu} = g_{\mu\rho}g_{\nu\sigma}\mathfrak{h}^{\rho\sigma} = g_{\mu\rho}g_{\nu\sigma}\chi^{\rho\sigma\tau\upsilon}F_{\tau\upsilon}. \tag{5.20}$$

We can then represent the matrix of the map $g^\wedge g = (\iota_g \times \iota_g)$ by way of:

$$(g^\wedge g)_{\mu\rho\nu\sigma} = \frac{1}{2}(g_{\mu\rho}g_{\nu\sigma} - g_{\nu\rho}g_{\mu\sigma}). \tag{5.21}$$

When we are dealing with an oriented Lorentzian manifold, in addition to the isomorphism of $\Lambda^2(M)$ with itself that is defined by $(\iota_g \times \iota_g) \cdot \chi$, we also get another isomorphism from (the inverse of) Poincaré duality. If we compose the maps χ and # then we get the isomorphism:

$$*_\chi : \Lambda_2(M) \to \Lambda_2(M), \quad F \mapsto \#\chi(F). \tag{5.22}$$

We must be careful to distinguish the effect of this map from the previous one defined by $(\iota_g \times \iota_g) \cdot \chi$. In the limit of an unpolarized medium χ goes to $(\iota_g \times \iota_g)^{-1}$ so the map $(\iota_g \times \iota_g) \cdot \chi$ goes to I, but the map $*_\chi$ goes to the conventional form of the Hodge star isomorphism on a pseudo-Riemannian manifold. Hence, $*_\chi$ is a deformation of the Hodge star isomorphism due to polarization.

We now wish to eliminate the explicit reference to the metric in the aforementioned constructions. We first represent the electromagnetic induction by the 2-vector field $\mathfrak{h} = \chi(F)$, instead of the 2-form H. Furthermore, we can alternatively represent the tensor field χ by the fully contravariant fourth rank tensor field of the form $\chi \in \Lambda_2(M) \otimes \Lambda_2(M)$ that corresponds to the linear isomorphism (5.14).

The components of a tensor field of the form $\chi \in \Lambda_2(M) \otimes \Lambda_2(M)$ come with two symmetries already: First, there is antisymmetry in the first and last pair due to the fact that they pertain to 2-forms. Second, from Lagrangian considerations, one often assumes that there is also a transposition symmetry of the form:

$$\chi(\alpha, \bullet) = \chi(\bullet, \alpha), \quad \alpha \in \Lambda^2(M). \tag{5.23}$$

This corresponds to a symmetry in the components of χ:

$$\chi^{\mu\nu\tau\upsilon} = \chi^{\tau\upsilon\mu\nu}. \tag{5.24}$$

This latter symmetry, combined with the fact that χ is a bilinear and non-degenerate functional on the vector bundle $\Lambda_2(M)$, implies that we can also regard χ as a fiber metric on the vector bundle $\Lambda_2(M)$. The question then arises: under what circumstances can we resolve the fiber metric χ on $\Lambda_2(M)$ to a Lorentzian one, g, on $\Lambda_1(M) = T(M)$? From (5.21), which we rewrite as $\chi = g^\wedge g$, we see that we are essentially looking for the "exterior square root" of χ.

---

[5] In this expression, we intend that H be considered as an operator.

From the orientability of M, if we choose a unit-volume element $\mathcal{V} = \frac{1}{4!}\varepsilon^{\mu\nu\rho\sigma}\partial_\mu^\wedge\partial_\nu^\wedge\partial_\rho^\wedge\partial_\sigma$ in $T^*(M)$, i.e., a globally non-zero 4-vector field $\mathcal{V}$ on M, then one already has another scalar product defined for 2-forms on an orientable M:

$$\langle F, G \rangle = \mathcal{V}(F^\wedge G) = \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}G_{\rho\sigma}. \tag{5.25}$$

However, a simple comparison of symmetries shows that $\mathcal{V}$ cannot be represented by an exterior product of symmetric tensor fields, such as $g^\wedge g$. Hence, if we wish to take the exterior square root of χ then we must subtract off the contribution proportional to $\mathcal{V}$:

$$\chi_0 = \chi - \alpha\mathcal{V}, \tag{5.26}$$

for some appropriate scalar α.

In order to find an exterior square root of χ_0, we follow Hehl, et al. [11, 12] and define:[6]

$$* = \#\chi_0 = \#\chi + \alpha I, \tag{5.27}$$

which is a linear isomorphism of $\Lambda^2(M)$ with itself.

The square of # is:

$$*^2 = \#\chi_0\#\chi_0. \tag{5.28}$$

If we can find a form for χ_0 that satisfies the constraint:

$$*^2 = -1 \tag{5.29}$$

then we expect that we have reproduced the Hodge isomorphism, at least for 2-forms.

In order to solve (5.28) for χ_0, at least locally, we choose a local frame field $\mathbf{e}_\mu$ whose domain is an open subset $U \subset M$ and whose reciprocal coframe field is θ^μ. We enumerate the basis for $\Lambda^2(U)$ defined by all $\theta^\mu \wedge \theta^\nu$ with $\mu < \nu$ by E^I, $I = 1, 2, \ldots, 6$. We can express the matrices of # and χ_0 as block matrices:

$$[\#]_{IJ} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}, \qquad [\chi_0]^{IJ} = \begin{bmatrix} A & C \\ C^T & B \end{bmatrix}, \tag{5.30}$$

in which A and B are symmetric 3×3 real matrices. For the general solution of (5.28), A and C must take the form:

$$A = pB^{-1} - \frac{1}{\det B}N, \qquad C = B^{-1}K \tag{5.31}$$

in which K is an arbitrary antisymmetric 3×3 real matrix, which we express as $ad(k)$ for a $k \in \mathbb{R}^3$, where ad refers to the Lie algebra on $\mathbb{R}^3$ that is defined by cross product, and:

$$N = k \otimes k, \qquad p = \frac{tr(NB)}{\det B} - 1. \tag{5.32}$$

If one has a tensor field χ_0 that satisfies (5.28) then one can then decompose $\Lambda_2(M)$ into a direct sum $\Lambda_{+2}(M) \oplus \Lambda_{-2}(M)$ of rank-three sub-bundles by polarization:[7]

$$F = F_+ + F_- = \frac{1}{2}(F - {}^*F) + \frac{1}{2}(F + {}^*F). \tag{5.33}$$

---

[6]   In order to stay consistent with our notation, which is based in a common notation for Poincaré duality, we shall invert their use of # and *, though.

[7]   Note the non-intuitive sign convention.

One refers to the sub-bundles $\Lambda_{+2}(M)$ and $\Lambda_{-2}(M)$ as consisting of *self-dual* and *anti-self-dual* 2-forms, respectively, even though the eigenvalues are imaginary, not real, as in the Riemannian case.

Now, one can associate a conformal class of *Riemannian* metrics with every splitting of $\Lambda_2(M)$ into a direct sum $\Lambda_+(M) \oplus \Lambda_-(M)$ whose constituent sub-bundles have the same rank – viz., three. In fact, this corresponds to the splitting of Spin(4) into $SU(2) \times SU(2)$. It also relates to the fact the Hodge isomorphism is the same for all metrics in a conformal class.

Of course, the issue at hand is not to find a conformal class of Riemannian metrics, but a conformal class of *Lorentzian* metrics. As we have pointed out above, the existence of the former does not have to imply the existence of the latter. Similarly, the uniqueness of Riemannian metrics up to homotopy – they are all obtained by deformation retractions of the bundle of linear frames – does not imply the uniqueness of the homotopy class of Lorentzian metrics, which depends upon the vanishing of $H^3(M; \mathbb{Z}_2)$. Furthermore, the spin group that is associated with $SO(3, 1)$, namely, $SL(2; \mathbb{C})$, does not split the same way as Spin(4). However, if one regards a Lorentz boost B as a "Wick-rotated" Euclidian rotation R:

$$B = \mathfrak{w}^{-1} R \mathfrak{w}, \qquad \mathfrak{w} = \mathrm{diag}(1, i, i, i) \tag{5.34}$$

then, at the infinitesimal level, the decomposition of the Lie algebra $\mathfrak{so}(4) = \mathfrak{so}(3) \oplus \mathfrak{so}(3)$ becomes the *vector space* decomposition of the Lie algebra $\mathfrak{so}(3, 1) = \mathfrak{so}(3) \oplus \mathfrak{b}(3)$, where the vector subspace:

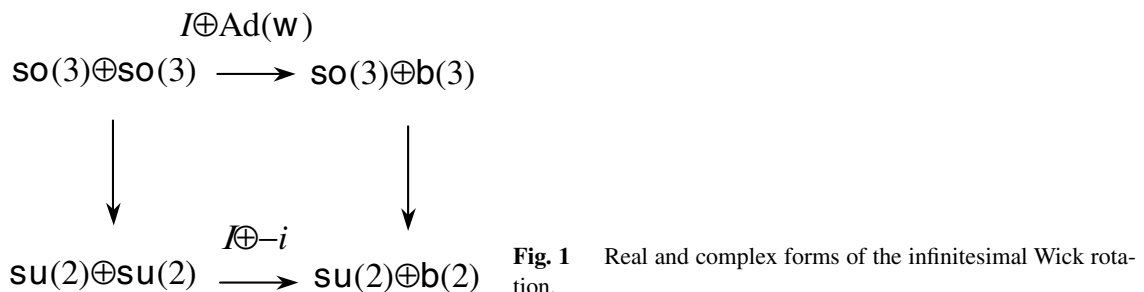$$\mathfrak{b}(3) = \mathfrak{w}^{-1} \mathfrak{so}(3) \mathfrak{w} \tag{5.35}$$

of infinitesimal boosts is not actually a Lie subalgebra of $\mathfrak{so}(3, 1)$. The next question is whether we also have:

$$\mathfrak{sl}(2; \mathbb{C}) \cong \mathfrak{su}(2) \oplus \mathfrak{b}'(2), \tag{5.36}$$

where $\mathfrak{b}'(2)$ is a three-dimensional subspace of infinitesimal boosts that one can obtain from $\mathfrak{su}(2)$ by a linear isomorphism. The answer is straightforward: By Hermitian polarization, we can decompose any $a \in \mathfrak{sl}(2; \mathbb{C})$ into a sum of a traceless Hermitian matrix and a traceless skew-Hermitian matrix:

$$a = h^+ + h^- = \frac{1}{2}(h + h^\dagger) + \frac{1}{2}(h - h^\dagger). \tag{5.37}$$

Since the traceless skew-Hermitian part is an element of the Lie algebra $\mathfrak{su}(2)$, we need only verify that the traceless Hermitian part represents an infinitesimal boost and can be obtained from a traceless skew-Hermitian matrix by a linear isomorphism. The first part follows from the isomorphism of $\mathfrak{sl}(2; \mathbb{C})$ with $\mathfrak{so}(3, 1) = \mathfrak{so}(3) \oplus \mathfrak{b}(3)$. As for the second part, the usual way to associate a Hermitian matrix with a skew-Hermitian one is to take h^- to $-ih^-$. We cannot actually represent this by a conjugation, so we simply regard the transformation defined by multiplying Hermitian matrices by i as the $\mathfrak{sl}(2; \mathbb{C})$ equivalent of the Wick rotation. We represent this situation by the commutative diagram:

$$
\begin{array}{ccc}
& I \oplus \mathrm{Ad}(\mathsf{w}) & \\
\mathsf{so}(3) \oplus \mathsf{so}(3) & \longrightarrow & \mathsf{so}(3) \oplus \mathsf{b}(3) \\
\Big\downarrow & & \Big\downarrow \\
& I \oplus -i & \\
\mathsf{su}(2) \oplus \mathsf{su}(2) & \longrightarrow & \mathsf{su}(2) \oplus \mathsf{b}(2)
\end{array}
$$

Fig. 1 Real and complex forms of the infinitesimal Wick rotation.

in which the vertical arrows are the complexifications.

Hence, we conclude that a splitting $\Lambda_2(M) = \Lambda_{2+}(M) \oplus \Lambda_{2-}(M)$ is associated with a conformal class $[g]$ of Lorentzian pseudo-metrics, up to topological obstructions. (For the actual construction of a representative g from B, N, K, see Hehl, et al. [11].) This is equivalent to a global field of light cones on M.

In summation, we see that in the case of linear constitutive laws the assumption of symmetry in the tensor χ, when regarded as a non-degenerate bilinear form on $\Lambda_2(M)$, defines χ as a fiber metric on that vector bundle. This, in turn, gives a splitting $\Lambda_2(M) = \Lambda_{2+}(M) \oplus \Lambda_{2-}(M)$ defined by the polarization of any 2-form by $* = \#\chi_0$, and if the existence of a global Lorentzian metric on M is topologically unobstructed then this leads to a conformal class $[g]$ of Lorentzian pseudo-metrics. Otherwise, when M is compact one will have to restrict one's consideration to an open submanifold of M that represents the complement of a finite set of points at which $[g]$ is undefined. (If M is not compact then there is no topological obstruction to a global Lorentzian metric.)

Before we leave the realm of linear constitutive laws, we pause to point out another mathematically intriguing aspect of constitutive laws as scalar products on $\Lambda_2(M)$: Because the linear automorphism on $\Lambda_2(M)$ that is defined by $* = \#\chi_0$ has the property that $*^2 = -1$, it defines an *almost-complex* structure on $\Lambda_2(M)$. In effect, the splitting $\Lambda_2(M) = \Lambda_{2+}(M) \oplus \Lambda_{2-}(M)$ into self-dual and anti-self-dual 2-forms is analogous to their splitting into real and imaginary parts. Since the vector bundle $\Lambda_2(M)$ is of rank six as a real vector bundle, it will be of rank three as a complex vector bundle. This implies that its structure group is $GL(3;\mathbb{C})$. If one introduces an orientation and a Hermitian structure on $\Lambda_2(M)$ then one can reduce its structure group to $SU(3)$. In fact, the orientation on $T(M)$ implies an orientation on $\Lambda_2(M)$, and the scalar product χ defines a Hermitian structure by way of $\chi(\alpha, \beta^{\dagger})$ where $\beta^{\dagger} = \mathrm{Re}\,(\beta) - \mathrm{Im}\,(\beta)$, in which $\beta = \mathrm{Re}\,(\beta) + \mathrm{Im}\,(\beta)$ is the decomposition of β into self-dual and anti-self-dual parts. All of this might suggest a possible geometrical representation for the color $SU(3)$ gauge structure of the strong interaction without the necessity of representing M as a complex three-manifold. However, our immediate concern is electromagnetism, so we let that pass.

5.2 Nonlinear constitutive laws

We might extend our constitutive law from a linear one to something that is homotopic to linear; i.e., something of the form:

$$\chi(F) = \chi^{(1)}(F) + N_s(F) \tag{5.38}$$

in which $\chi^{(1)} : \Lambda^2(M) \to \Lambda_2(M)$ is linear, $N_s : \Lambda^2(M) \to \Lambda_2(M)$ is nonlinear with $s \in [0, 1]$, and:

$$\lim_{s \to 0} N_s = 0. \tag{5.39}$$

In such a case, we are dealing with a constitutive law of the "generalized Taylor series" form: linear map + (nonlinear map that vanishes in the weak field limit). In nonlinear optics [20, 21], this level of approximation is referred to as "weak nonlinearity." The next levels of approximation are the quadratic and cubic ones:[8]

$$\chi(F) = \chi^{(0)} + \chi^{(1)}(F) + \chi_s^{(2)}(F \odot F) \tag{5.40a}$$

$$\chi(F) = \chi^{(0)} + \chi^{(1)}(F) + \chi_s^{(2)}(F \cdot F) + \chi_s^{(3)}(F \odot F \odot F). \tag{5.40b}$$

The leading constant term $\chi^{(0)}$ subsumes any residual polarization of the medium that remains when there is no applied electromagnetic field. Although that sounds physically uninteresting in the case of the classical spacetime electromagnetic vacuum, one should keep in mind that the Casimir effect is indicative of the reality of the zero-point field that one expects from quantum electrodynamic considerations. If one

---

[8] The notation $\odot$ refers to the symmetrized tensor product.

takes the position that quantum electrodynamics is really a phenomenological process for constructing the extension of linear electrodynamics into nonlinear electrodynamics in the realm of strong field strengths (see the "neoclassical" discussion in [18] or the quantum discussion in [22]) then one might still give this term serious consideration even for the electromagnetic vacuum.

The terms $\chi_s^{(2)}$ and $\chi_s^{(3)}$ are linear maps from the respective symmetrized tensor products of $\Lambda^2(M)$ with itself to $\Lambda_2(M)$. Hence, they can be represented in either form:

$$\chi_s^{(2)} : \Lambda^2(M) \odot \Lambda^2(M) \to \Lambda_2(M), \tag{5.41a}$$

or:

$$\chi_s^{(2)} \in \Lambda_2(M) \odot \Lambda_2(M) \otimes \Lambda_2(M), \tag{5.41b}$$

and:

$$\chi_s^{(3)} : \Lambda^2(M) \odot \Lambda^2(M) \odot \Lambda^2(M) \to \Lambda_2(M), \tag{5.42a}$$

or:

$$\chi_s^{(3)} \in \Lambda_2(M) \odot \Lambda_2(M) \odot \Lambda^2(M) \otimes \Lambda_2(M). \tag{5.42b}$$

Furthermore, $\chi_s^{(2)}$ and $\chi_s^{(3)}$ are expected to vanish in the limit $s \to 0$, which one intends to be the limit of weak field strength. Hence, s represents a sort of coupling constant for the nonlinear terms.

Although it would be mathematically natural to generalize all of this to the methodology of jet bundles or covariant derivatives in order to account for the successive terms as true higher-order derivatives of ℏ with respect to F, we shall take a more nonlinear optical route and treat the successive terms as simply phenomenologically defined tensor fields. Of course, the ultimate challenge to theoretical physics is to account for these tensor fields in terms of geometrically or topologically defined tensor fields; i.e., in terms of spacetime structure. Consequently, one should not keep the aforementioned generalities out of one's consideration completely.

In nonlinear optics, since:[9]

$$H(F) = F + P(F) = (I + P)(F), \tag{5.43}$$

one correspondingly regards the polarization 2-form $P(F)$ as the essential part in the eyes of nonlinear analysis and then decomposes the nonlinear operator P into linear, quadratic, etc., parts:

$$P(F) = P^{(0)} + \chi^{(1)}(F) + \chi^2(F \odot F) \tag{5.44a}$$

$$P(F) = P^{(0)} + \chi^{(1)}(F) + \chi^2(F \odot F) + \chi^3(F \odot F \odot F). \tag{5.44b}$$

However, since we are trying to avoid using the spacetime metric in the fundamental statements about electromagnetism, and P seems to relate mostly to deformations of the metric, we shall simply regard (5.41a,b) as the approach we shall take to weak nonlinearity.

Of all of the effective theories of nonlinear electrodynamics that are based in quantum electrodynamical considerations the one that seems to get the most attention is the Born-Infeld theory [14, 15]. It is based in the imposition of a maximal field strength that would precede the onset of vacuum polarization, as in QED.

---

[9] To be truly faithful to nonlinear optics, we should be talking about the electric field by itself, not the Minkowski 2-form, since the effects of large magnetic field strengths are usually treated as a separate class of phenomena. However, since our optical medium is the spacetime vacuum, and the $E - B$ decomposition follows from the imposition of a metric and a time orientation, we shall be more cautious from the outset.

The Born-Infeld Lagrangian for the electromagnetic field F takes the form:

$$\mathcal{L} = (\sqrt{E_c^2 + \mathcal{F} - \mathcal{G}} - E_c)\mathcal{V}, \tag{5.45}$$

in which E_c represents a critical electric field strength.

In order to define the Born-Infeld Lagrangian, one must first introduce the field invariants:

$$I_1 = F^{\wedge *}F = \mathcal{F}\mathcal{V}, \qquad I_2 = F^{\wedge}F = \mathcal{V}(F, F)\mathcal{V} = \mathcal{G}\mathcal{V}. \tag{5.46}$$

These correspond to the local expressions:

$$\mathcal{F} = \frac{1}{2}g^{\mu\nu}g^{\rho\sigma}F_{\mu\nu}F_{\rho\sigma}, \qquad \mathcal{G} = \varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}. \tag{5.47}$$

Ordinarily, one intends that the * isomorphism is due to Hodge duality, which, in turn, comes from the imposition of a Lorentzian structure on spacetime, which is why one usually constructs one's field Lagrangian from *Lorentz* invariant expressions in F – à la Mie [26]. In the pre-metric case, one must look for $SL(4)$-invariant expressions. Of the latter two expressions, only I_2 can be formed without introducing a metric, at least as stated. Indeed, it can be defined (as a 4-form) without recourse to $\mathcal{V}$, so it is actually $GL(4)$-invariant. The expression $\mathcal{F}$ takes the form of $||F||^2$ when one gives $\Lambda^2(M)$ the norm that comes from the scalar product $g^{\wedge}g$.

However, recall that if we are considering a χ such that $^{*2} = -1$, where $^* = \#\chi$ then, in effect, we have reconstructed the Hodge star, which is equivalent to a conformal class of metrics. Consequently, given such a χ we can still use the * isomorphism it defines in place of the Hodge star, and define a generalization of I_1 by:

$$I_1 = F^{\wedge *}F = \chi(F, F)\mathcal{V}. \tag{5.48}$$

However, since we have defined our replacement for the Hodge star only on 2-forms, none of the remaining three Lorentz invariants that Mie (with a correction from Weyl and Born) obtained, namely:

$$I_3 = A^{\wedge *}A \tag{5.49a}$$
$$I_4 = i_A F^{\wedge *}(i_A F) \tag{5.49b}$$
$$I_5 = i_A {}^*F^{\wedge *}(i_A {}^*F) \tag{5.49c}$$

can be defined without using a metric, since they all take the form of norm-squares of 1-forms.

The Lagrangians that involve norm-squares using the spacetime metric usually represent kinetic energy terms for the field, just as the elementary expression for the non-relativistic kinetic energy of a point mass involves using a conformal transform of the spatial metric tensor, whose conformal factor is $1/2m$. Similarly, the potential energy in a one-dimensional (or isotropic three-dimensional) Hookean material that has been displaced Δx from its equilibrium state involves a conformal transform of the spatial metric whose conformal factor is $1/2k$.

However, when one goes from point matter to extended matter, or isotropic to anisotropic media, or linear to nonlinear media, all of the aforementioned simplifications break down and the conformal transformation of the metric must be – in the linear case – replaced with a product of a conformal transformation and a volume-preserving shear, and with more elaborate expressions when one goes to the nonlinear realm.

In any event, the generalizations all amount to constitutive laws in one form or another. A metric tensor then emerges as essentially a special case of a process by which one associates tensor fields with their dual tensor fields. As we have seen, in the case of 2-forms on a four-manifold, a volume element gives another.

6 Topological constitutive laws

In order to make electromagnetism a completely topological phenomenon, in addition to giving a topological interpretation to the fundamental fields, we also need to find some topological origin for the constitutive properties of the medium in question. This is especially perplexing in the case of the spacetime vacuum itself, since, classically, the spacetime vacuum was assumed to have an electromagnetic structure that was expressed only by the constants ε_0 and μ_0. However, as pointed out above, quantum electrodynamics suggests that this picture of the spacetime vacuum is an oversimplification.

6.1 The intersection form

Since we are replacing our electromagnetic field strength 2-form F with its de Rham cohomology class $[F]$, we might also consider a constitutive axiom in the form of a bilinear functional on the second de Rham cohomology space itself, instead of $\Lambda_2(M)$:

$$H^2(M; \mathbb{R}) \times H^2(M; \mathbb{R}) \to \mathbb{R}. \tag{6.1}$$

Note that this time we are defining our bilinear functional on a *vector space*, not a $C^\infty(M)$-*module*, as we did for linear constitutive laws.

Now, any closed oriented orientable 4-manifold M already has one such functional defined by Poincaré duality:

$$Q([\alpha], [\beta]) = \int_M \alpha^\wedge \beta = Q([\beta], [\alpha]). \tag{6.2}$$

Inside the integral, α and β are any closed 2-forms that define each chosen de Rham class in dimension two. When M has no boundary, the integral will be independent of this choice of representative. If M has a boundary then, by Stokes's theorem, replacing α by $\alpha + d\lambda$ and β by $\beta + d\mu$ will produce a boundary term of $\alpha^\wedge \mu - \lambda^\wedge \beta + d\lambda^\wedge \mu$, which does not have to vanish, except for a restricted class of gauge transformations, such as λ, μ that vanish on ∂M.

What is not entirely obvious in (6.2) is how Poincaré duality and the choice of $\mathcal{V}$ affects the definition of Q. Hence, we rewrite it as:

$$Q([\alpha], [\beta]) = \int_M \alpha(^*\beta)\mathcal{V}, \tag{6.3}$$

in which the role of our choice of $\mathcal{V}$ and its associated Poincaré duality isomorphism * becomes clear.

Q is not only non-degenerate, but unimodular, as well, so its symmetry and bilinearity imply that Q defines a scalar product on the vector space $H^2(M; \mathbb{R})$. The bilinear functional Q is called the *intersection form* for a closed orientable M [10, 26–28]. As we shall see shortly, it plays a fundamental role in the topology of four-dimensional differentiable manifolds.

The relevance of the intersection form to electromagnetism should be unavoidable in our present formulation, since we are defining electromagnetic fields to be two-dimensional de Rham cohomology classes to begin with. Of course, the only way that Q is non-trivial is if we assume that the topology of M has a non-vanishing $H^2(M; \mathbb{R})$, i.e., that some F's do not admit global potential 1-forms, which gets one back into the realm of magnetic monopoles, wormholes, and other spacetime pathologies. Hence, we briefly summarize some of the fundamental features of the role that the intersection form plays in the topology of four-manifolds.

6.2 The topology of four-manifolds [26–28]

Algebraically, a scalar product can be classified by two integers: its *rank*, which equals the second Betti number for M in the present case:

$$b_2 = \dim(H^2(M; \mathbb{R})), \tag{6.4}$$

and its *signature*:

$$\tau = b^+ - b^-. \tag{6.5}$$

where b^+ is the maximal dimension of the subspaces of $H^2(M;\mathbb{R})$ on which Q is positive definite and b^- is the maximum dimension of the subspaces on which it is negative definite.

When one restricts the form Q to the free part of the integer cohomology, i.e., the free part of $H^2(M;\mathbb{Z})$, which defines an integer lattice of the form $\mathbb{Z}^{b_2}$, one can also define the *type* of Q according to whether it takes it values in the even integers or not. In the affirmative case one calls Q *even*, and in the contrary case one calls Q *odd* (although that does not have to imply that *all* of its values are odd integers).

For indefinite forms, the Hasse-Minkowski classification says:

$$Q = \begin{cases} I_l \oplus (-I_m) & \text{when } Q \text{ is odd} \\[2ex] l \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \oplus mE_8 & \text{when } Q \text{ is even.} \end{cases} \tag{6.6}$$

In this classification, the matrix E_8 is the Cartan matrix of the exceptional simple Lie algebra E_8.

For definite forms, the classification is not as straightforward. However, as we shall discuss later, Donaldson's theorem narrows down the field of definite forms that describe the intersection forms of simply connected closed four-manifolds considerably.

For simply connected four-dimensional manifolds, the intersection form is fundamental to the other characteristic classes, since they can all be derived from the information that is contained in Q:

Second Stiefel-Whitney: $Q(w_2[M], \alpha) = Q(\alpha, \alpha)(\bmod 2)$ for all $\alpha \in H^2(M;\mathbb{Z}_2)$

First Pontrjagin:[10] $p_1[M] = 3\tau[\mathcal{V}] = 3(b^+ - b^-)[\mathcal{V}]$

Euler: $e[M] = (b^+ b^-)[\mathcal{V}].$

(In these expressions, $\mathcal{V}$ is a choice of unit volume element.) The first and third Stiefel-Whitney classes vanish by simple connectedness and Poincaré duality; indeed this is actually true of w_3 for any compact orientable 4-manifold. The fourth Stiefel-Whitney class is the $\mathbb{Z}_2$ reduction of $e[M]$.

The second Stiefel-Whitney class plays an important role in the context of Spin(4) structures on M, namely it must vanish in order for an $SO(4)$-reduction of $GL(M)$ to admit a two-to-one covering by a Spin(4)-principal bundle. Moreover, the homotopy classes of such bundles are indexed by $H^1(M;\mathbb{Z}_2)$. In particular, if a simply connected four-manifold admits a spin structure then it must be unique. Moreover, a simply connected four-manifold admits a spin structure iff its intersection form, when restricted to integer cohomology classes, takes on only even values.

The mathematician's understanding of how the intersection form Q relates to the topology of simply connected four-manifolds is growing quite extensive by now. In 1949, Whitehead showed that if two closed simply connected four-manifolds have isomorphic intersection forms then they are homotopy equivalent. In fact, every unimodular symmetric bilinear form is the intersection form of some simply connected *topological* four-manifold. However, not all of them will be smoothable, or even piecewise-linear.

For the case of indefinite intersection forms, one has mostly examples to deal with. For instance, one can construct a compact orientable four-manifold with an intersection form $\operatorname{diag}(1,\ldots,1,-1,\ldots,-1) = l(1)\oplus m(-1)$ by taking the connected sum of l copies of $\mathbb{CP}^2$ (regarded as a real four-dimensional manifold instead of a complex two-dimensional one) with one orientation with m copies of $\overline{\mathbb{CP}^2}$, which is $\mathbb{CP}^2$ with the opposite orientation. This is because $H^2(\mathbb{CP}^2;\mathbb{Z}) = \mathbb{Z}$, reversing the orientation inverts the sign of the

---

[10] This result follows from either the Hirzebruch signature theorem [10, 29], or, more generally, the Atiyah-singer index theorem [29, 30].

intersection form, and the connected sum operation produces a direct sum in the cohomology. We point out that the standard generator for $H^2(\mathbb{C}P^2; \mathbb{Z})$ is essentially $\mathbb{C}P^1$, which is diffeomorphic to a 2-sphere.

For the case of definite intersection forms, one has Donaldson's celebrated theorem that the only positive definite intersection form on a smooth closed simply connected four-manifold is the standard one.

Now that we have discussed the notion of the intersection form of an orientable manifold, we can examine one possibility for rendering $H^2(M; \mathbb{R})$ non-trivial in a manner that seems closely related to the physical process of vacuum polarization in the presence of strong electric field strengths.

6.3 A possible topological mechanism for vacuum polarization

As a first attempt at giving a deeper physical basis for a topological constitutive axiom let us take the process of vacuum polarization literally and see what sort of expressions we can derive from it.

The basic process in vacuum polarization is that of the conversion of a photon into a virtual electron/positron pair when the field strengths of the photon exceed some threshold value. We first assume that the reason for using the word "virtual" is simply the fact that these processes and their inverses may be happening over such a short time interval that the measurement process cannot resolve them in experimental practice, even though their *collective* effect as a macroscopic ensemble is measurable; for instance, the Lamb shift. Hence, a virtual process might represent a real, but unobservable, process.

We assume that any sufficiently small region of the spacetime electromagnetic vacuum can exist in one of three phases:

a) Unpolarized; this is the *weak* field case.

b) Polarized by a bound electron/positron pair; this is the *critical* field case.

c) Stably charged, when the binding energy has been exceeded and the charges form stable disjoint distributions; we shall call this the *super-critical* field case.

Our main concern in this study is the transition between the first two phases, although we shall treat both phase transitions as if they were associated with the spontaneous breaking of the vacuum symmetry and the appearance of topological defects.

We next observe that when a photon resolves to a bound electron/positron pair, as long as the two charges are assumed to be separated by an actual displacement vector $\mathbf{d}$ there is also an electric dipole $2e\mathbf{d}$ that is associated with the pair. The vector $\mathbf{d}$ brings about a reduction in the symmetry of the vacuum state from the $SO(2)$ gauge symmetry of the electromagnetic Lagrangian that represents charge conservation to the $\{e\}$ symmetry of the vacuum state that a choice of non-zero vector brings about. The homogeneous space that is defined by this reduction is also $SO(2)$ – or S^1, if you prefer – whose first non-trivial homotopy group is $\pi_1(S^1) = \mathbb{Z}$. Hence, one should expect the topological obstruction to making such a reduction to be an element of $H^2(M; \mathbb{Z})$. In the language of defects, we are associating a "line defect" to some 2-cycles in M, which we regard as belonging to the singularity complex of F – i.e., its sources. In the usual Dirac construction of magnetic monopoles, if the 2-cycle is a 2-sphere then this line defect is its equator.

Consequently, let us imagine that topologically the formation of a charge pair is equivalent to the cavitation of a bubble in a fluid, i.e., a 2-cycle that does not bound. Furthermore, we regard this topology-changing process as described by the contraction of an open 2-ball to a point, while leaving a 2-sphere as a boundary to the remaining space. This is reminiscent of the fact that bubbles cannot form in real fluids unless there is point on which to nucleate. Note that the point that remains when one contracts a 2-ball is no longer in the same connected component of the spacetime manifold ([11]). Hence, we have added one generator to $H^2(M; \mathbb{Z})$ and one generator to $H^0(M; \mathbb{Z})$. We could also regard this process as one of attaching a copy of $\mathbb{C}P^1$ by connected sum and a point by disjoint union, although we shall disregard the contribution that the point makes and concentrate on the remaining component of the spacetime manifold. Hence, this sort

---

[11] Indeed, it has the wrong dimension to be included in the manifold that it left behind.

of topological modification produces an increase in the second Betti number of M, as well as the signature of M, so it should be manifest in the structure of Q.

We will assume, by symmetry, that the charge is effectively concentrated at the poles with $+e$, by definition, at the North pole and $-e$ at the South pole. Hence, the polar diameter, which we orient from the positive charge to the negative one, defines the displacement $\mathbf{d}$, as well as the line defect that gets associated with the 2-cycle that the bubble defines. As the electromagnetic field strength grows beyond the critical value for cavitation, one expects the 2-sphere to elongate into an ellipsoid of increasing polar diameter and decreasing equatorial diameter.

If one imagines this process of cavitation under high electric field strength as also happening in great numbers – which represents a somewhat different conception of "spacetime foam" – then one can pass to the continuum limit and define an overall density of electric dipoles; i.e., a polarization tensor field $\mathfrak{p}$, or simply the tensor field χ.

Note that the second phase transition that one can envision under vacuum polarization, viz., the actual separation of the charge pair, when it goes from a bound state to a scattering state, also represents a further topology-changing process. The single bubble that cavitated under the action of the surrounding field has become two disjoint bubbles. As for the point that the open 3-ball contracted to, although it sounds trivial, nevertheless, in the eyes of topology something must be done about it. In particular, if there is true mirror equivalence between matter and antimatter, one expects that if one of the free states has a disjoint point "inside" it then so should the other. However, the presence of such a point might also be an indicator of the instability of the topological state. When one removes the external field, one expects the vacuum polarization to disappear, and one imagines the spherical polarized boundary that previously formed collapsing back into the central point. Since electrons and positrons presumably exist in the absence of external fields as well, one might expect that the central point no longer exists in either.

A possible accounting of where the point went to is provided by the fact that since the process of separating one bubble into two bubbles – like the process of forming a bubble in the first place – cannot be effected by any continuous map there must be singularities for that process, i.e., points at which the map is discontinuous. If one imagines the bubble being stretched along the polar axis under increasing field strengths until the equator starts "necking down" then it is conceivable that the central point now serves as a point to which the equator contracts in order for the hemispheres to separate into disjoint bubbles. We schematically illustrate the process in Fig. 2–5, in which the parameter that changes between phases is the increasing electric field strength.

One could regard this process as a higher-dimensional analogue of the two-dimensional Lorentz cobordism that takes the form of the "trouser manifold." In that process, one facilitates the subdivision of one circle (the "waistline") into two disjoint circles (the two "cuffs") by means of a two-dimensional Lorentzian manifold whose boundary consists of the oriented disjoint union of the three circles and whose line field describes the dynamical system that effects this transition. Of course, a compact two-dimensional manifold M will admit a Lorentzian structure iff its Euler-Poincaré characteristic:

$$\chi[M] = 2 - 2g \tag{6.7}$$

vanishes, in which we have introduced the integer g that one calls the *genus* of M. Since this only vanishes for $g = 1$, we see that the trouser manifold ($g = 3$) does not admit a Lorentzian structure without singular points. But then, if one thinks of a Lorentzian structure as a one-dimensional differential system that describes the process of proper time evolution then one sees that the aforementioned process cannot be described by a flow of global diffeomorphisms anyway, or even a global flow of homeomorphisms.

In the case of bubble cavitation, we are looking at, first, the formation of a 2-sphere and a point, and then the separation of one 2-sphere into two disjoint 2-spheres in which the equator contracts to the point that was formed in the first transition and then gets absorbed into the surrounding manifold. If these three 2-spheres collectively define the boundary of a compact oriented 3-manifold then, since Poincaré duality makes $\chi[M]$ = 0 for any compact orientable 3-manifold, that oriented cobordism can also be given a Lorentzian structure.
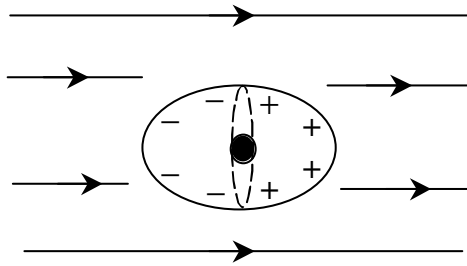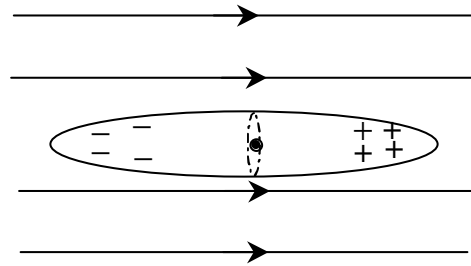
Fig. 2 Formation of bubble
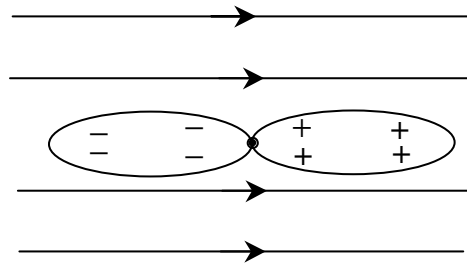

Fig. 3 Elongation.
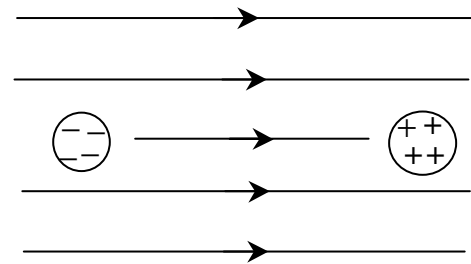

Fig. 4 Constriction of equator.

Fig. 5 Separation of new bubbles.

Since the super-critical phase transition has attached another copy of $\mathbb{C}P^2$ and thereby altered Q, a crucial issue is whether the change to the signature is reflected in the dimension of the maximal negative subspaces or that of the maximal positive subspaces. One suspects that an anti-particle should somehow balance a particle in a symmetrical sort of way, so perhaps the anti-particle has the opposite orientation and we are really attaching $\mathbb{C}P^2$ for the particle and $\overline{\mathbb{C}P^2}$ for the anti-particle. This would suggest that the symmetry that relates most fundamentally to the matter/anti-matter symmetry is the $\mathbb{Z}_2$ symmetry that one associates with the choice of orientation.

The possibility that past a critical electric field strength virtual or actual electron-positron pairs would form in such a way that the ultimate field strength would be reduced is usually referred to as the *Klein paradox*. If one regards such a process as simply another example of a more general class of phase transitions that are associated with spontaneous breaking of the ground state symmetry and the formation of topological defects then perhaps the phenomenon will not seem so paradoxical.

7 Electromagnetic waves

Since the appearance of the Minkowski scalar product in electromagnetism seems to be concerned with the appearance of wavelike solutions to the electromagnetic field equations, we should examine the way that one might use the structure of wave motion to induce the spacetime metric.

First, one might point out that the second order hyperbolic PDE that one usually considers to be the linear wave equation is not necessarily the most fundamental statement about the intrinsic nature of wave motion. For instance, one can also consider *first* order hyperbolic PDE's that one refers to as *conservation laws*. For instance, the two-dimensional linear wave operator can be regarded as the composition of two first order ones that define the equations:

$$\partial_t \phi - c^{-1}\partial_x \phi = 0, \qquad \partial_t \phi + c^{-1}\partial_x \phi = 0, \tag{7.1}$$

that have solutions of the form:

$$f(t, x) = f(t - c^{-1}x), \qquad g(t, x) = f(t + c^{-1}x). \tag{7.2}$$

Hence, just as the second order d'Alembertian operator is the product of two first order differential operators, similarly, the usual d'Alembert solution of the second order linear wave equation is simply the sum of solutions to two first order equations.

More generally, one can consider conservation laws of the form:

$$\partial_t \phi + c(t, x, \phi) \partial_x \phi = 0, \tag{7.3}$$

which include many of the popular nonlinear wave equations, as well. When one generalizes (7.3) to $n + 1$ dimensions, so that $\partial_x \phi$ becomes an n-dimensional vector, hyperbolicity becomes a property of the map χ. Of course, a drawback to this formulation of wave equations is that is presumes that space is distinct from time.

In the context of exterior differential forms, the equivalent statement to (7.1) is that when a k-form α on an orientable Riemannian (Lorentzian, resp.) manifold is both closed and co-closed:

$$d\alpha = 0, \qquad \delta\alpha = 0, \tag{7.4}$$

it is also harmonic:

$$\Delta\alpha = 0, \qquad (\Box\alpha = 0, \text{resp.}) \tag{7.5}$$

where:

$$\Delta(\Box, \text{resp.}) = \delta d + d\delta. \tag{7.6}$$

In the Riemannian case, the converse is also true. The fact that the converse does not have to be true in the Lorentzian case, along with the fact that many of the methods of Hodge theory in general break down for Lorentzian manifolds, is exactly why we choose to distinguish the Laplace operator on a Riemannian manifold from the d'Alembertian operator on a Lorentzian one.

In particular, the usual source-free Maxwell equations on a Lorentzian manifold:

$$dF = 0, \qquad \delta F = 0 \tag{7.7}$$

imply that F is harmonic in the d'Alembertian sense – i.e., wavelike. In fact, the inclusion of a source current 1-form J simply changes the homogeneous wave equation to a forced wave equation.

To bring the discussion back to the pre-metric context, we point out that if our manifold is at least orientable then we can make sense of the codifferential operator δ only when applied to k-vector fields. The compositions δd and $d\delta$ then become absurd. Consequently, we might look for a way around the use of the metric.

For instance, one might replace the metric with the constitutive law defined by χ, which still gives us a generalized Hodge duality – at least for 2-forms – by way of $^* = \#\chi$. Hence, we define the predictable codifferential operator on 2-forms:

$$\delta_\chi : \Lambda^2(M) \to \Lambda^2(M), \qquad \delta_\chi\alpha = -^*d^*\alpha. \tag{7.8}$$

However, $d^*\alpha$ is a 3-form. This would necessitate the extension of *, hence χ, to 3-forms. By Poincaré duality this would also imply an isomorphism of $\Lambda^1(M)$ with $\mathfrak{X}(M)$, or $T^*(M)$ with $T(M)$. Since such an isomorphism might then, by symmetrization, define a metric on M we suspect that perhaps the introduction of a metric is inevitable in order to define wave motion.

There are reasons to suspect that this situation is based in deeper considerations about the nature of wave motion. As has been discussed elsewhere [17], an essential construction that is associated with wave motion in a four-dimensional manifold is that of a pair of transverse foliations of M of codimension one. One of them defines the proper time simultaneity leaves relative to a choice of rest frame. The other defines the isophase hypersurfaces for the wave motion. Their intersections are two-dimensional instantaneous isophase surfaces.

Although we have been avoiding the introduction of the proper time simultaneity foliation as being essentially equivalent to the introduction of a metric, as it happens, we can still define the aforementioned two-dimensional foliation of instantaneous isophases purely on the basis of the algebraic properties of F.

The *rank* of a k-form α on a manifold M is defined to be the maximum dimension of its annihilating subspace A_x at each point $x \in M$. More precisely, a vector $\mathbf{v} \in T_x(M)$ is in A_x iff $i_\mathbf{v}\alpha = 0$. Another way of characterizing the rank of an exterior $k-$form α is that it is the smallest integer r such that $\alpha^r = 0$, in which we intend that the power refers to the exterior product.

For a 2-form, the rank must be even, so for a four-dimensional M, the rank of a 2-form can only be 0, 2, or 4. The first possibility corresponds the case in which the 2-form is simply 0 itself. Hence, if the 2-form in question is F then the physically interesting cases are when the 2-form in question has rank 2 or 4.

When F has rank 4, there are linearly independent 1-forms $\alpha, \beta, \gamma, \delta$ such that:

$$F = \alpha^\wedge\beta + \gamma^\wedge\delta. \tag{7.9}$$

Rank-four 2-forms are, in a sense, the "generic" non-zero 2-forms, since the rank-two 2-forms define a quadric hypersurface (viz., the Klein quadric) of lower dimension than six and the rank-four 2-forms are the complement of this hypersurface (plus zero, of course).

When F has rank 2, there are linearly independent 1-forms α and β such that:

$$F = \alpha^\wedge\beta. \tag{7.10}$$

One refers to the set $\{\alpha, \beta\}$ or $\{\alpha, \beta, \gamma, \delta\}$ as an *associated system* to F.

Alternately, if F has rank 2, one must have that:

$$F^\wedge F = 0. \tag{7.11}$$

When one gives F the usual $E - B$ decomposition, namely, $F = \theta^\wedge E + {}^*(\theta^\wedge B)$ for some choice of timeline unit 1-form θ, this last condition takes the form:

$$E^{\wedge *}B = 0. \tag{7.12}$$

Hence, if F has rank two then the E and B fields, which are spacelike in conventional electromagnetism, must be orthogonal or zero.

The 1-forms α and β collectively span a two-dimensional sub-bundle of $T^*(M)$ and annihilate a two-dimensional sub-bundle $\mathfrak{D}$ of $T(M)$. The sub-bundle $\mathfrak{D}$ defines a differential system on M, and, if it is integrable, the two-dimensional integral submanifolds define a codimension-two foliation of M that represents the instantaneous isophase surfaces of the electromagnetic wave motion defined by F. The integrability criterion is given by Frobenius:

$$\alpha^\wedge d\alpha = \beta^\wedge d\beta = 0. \tag{7.13}$$

An important subtlety to consider is that since α and β are globally linearly independent then they also define a global 2-frame field on M. Consequently, if this is true then M must be topologically inclined to admit such a field, i.e., it must have a degree of parallelizability of two. A necessary condition for this is that the top two Stiefel-Whitney classes of $T(M)$, w_3 and w_4, must vanish. Otherwise, α and β can only be defined on the complement of some singularity set.

Now that we have a codimension-two foliation the next issue to address is whether we can also define the complementary foliations whose intersections are the leaves of the latter foliation. However, the author admits that the answer to this question is not clear as of yet and defers that effort to further researches in order to prevent the present discussion from overstepping its scope.

8   Discussion

In conclusion, we shall summarize the most promising directions for further research into the topological nature of electromagnetism.

a) The representation of F and $\mathbf{J}$ by more fundamental geometrical or topological objects: It is not unreasonable to expect that the introduction of an $\mathfrak{sl}(4)$ connection on the $SL(4)$-principal bundle on M defined by a choice of $\mathcal{V}$ is unavoidable. In conventional electrodynamics when we use a potential 1-form A for F as the fundamental object the usual considerations of wanting the field theory to be invariant under local $U(1)$ gauge changes dictate that A must define a $U(1)$ connection 1-form. However, the cautious reader will note that the group that plays the fundamental role in topological electromagnetism seems to be $SL(4)$, not $SO(2)$. Resolving this source of confusion seems to be key to further progress.

b) Reduction from $SL(M)$ to $SO(3,1)(M)$: This is an essential step from the standpoint of introducing gravitation into the model, or rather, deducing gravitation from it. We have seen that such a reduction can follow from starting with a more general scalar product on 2-vector fields, but that only transfers the burden of significance to the choice of a constitutive law. One then must investigate the physical and mathematical bases for this choice in finer detail.

c) Role of the intersection form in constitutive laws: Although the constitutive law that we just mentioned certainly gives a simple and direct route for effecting the reduction from $SL(M)$ to $SO(3,1)(M)$, it nevertheless lacks an immediate topological construction. This either suggests that one cannot find a completely topological formulation for electromagnetism and must eventually resort to geometrical axioms or that we need to look for a more topological basis for the constitutive law. Since the intersection form seems to play a role vis-à-vis $H^2(M;\mathbb{R})$ that is analogous to the role played by the constitutive law in the context of $\Lambda^2(M)$, this seems to be a promising direction to investigate.

References

[1] A. Lichnerowicz, Théorie relativiste de la gravitational et de l'electromagnetisme (Masson and Co., Paris, 1955).
[2] V. P. Vizgin, Unified Field Theories in the First Third of the Twentieth Century (Birkhäuser, Boston, 1994).
[3] D. H. Delphenich, Spacetime G-structures I: topological defects, LANL Archives gr-qc/0304016; Spacetime G-structures II: geometry of the ground state, LANL Archives gr-qc/0401089.
[4] É. Cartan, Les systèmes différentielle extérieures et leur applications géométriques (Hermann, Paris, 1971).
[5] Y. Choquet-Bruhat, Géométrie differentielle et systèmes différentieles extérieures, (Dunod, Paris, 1964).
[6] W. Thirring, Classical Field Theory (Springer, Berlin, 1978).
[7] R. Bott and L. Tu, Differential Forms in Algebraic Topology (Springer, Berlin, 1982).
[8] M, Greenberg, Lectures on Algebraic Topology (Benjamin-Cummings, Boston, 1967).
[9] J. Vick, Homology Theory (Academic Press, New York, 1973).
[10] J. Milnor, J. Stasheff, Characteristic Classes (Princeton Univ. Press, Princeton, 1974).
[11] F. W. Hehl, et al., Ann. Phys. (Leipzig) **9** SI-71 (2000) [Special Issue for Journées Relativistes 99 in Weimar, edited by G. Neugebauer and R. Collier].

[12] F. W. Hehl and Y. N. Obukhov, Foundations of Classical Electrodynamics (Birkhäuser, Boston, 2003).

[13] E.J. Post, Formal Structure of Electromagnetics (Dover, New York, 1997).

[14] M. Born and L. Infeld, Proc. Roy. Soc. A **144**, 425 (1934).

[15] M. Born, Ann. Inst. Henri Poincaré **7**, 155 (1937).

[16] D. H. Delphenich, Proper Time Foliations of Lorentz Manifolds, LANL Archive gr-qc/0211066.

[17] D. H. Delphenich, Foliated Cobordism and Motion, LANL Archives, gr-qc/0209091.

[18] D. H. Delphenich, Nonlinear Electrodynamics and QED, LANL Archives hep-th/0309108.

[19] C. W. Misner and J. A. Wheeler, Ann. Phys. **2**, 525 (1957).

[20] D. L. Mills, Nonlinear Optics, 2nd ed. (Springer, Berlin, 1998).

[21] P. N. Butcher and D. Cotter, The Elements of Nonlinear Optics (Cambridge University Press, 1990).

[22] W. Dittrich and H. Gies, Probing the Quantum Vacuum (Springer, Berlin, 2000).

[23] H. Euler and W. Heisenberg, Z. Phys. **98**, 714 (1936).

[24] J. Schwinger, Phys. Rev. **82**, 664 (1951).

[25] G. Mie, Ann. Phys. **37**, 511 (1912); Ann. Phys. **39**, 40 (1912); **40**, 1 (1913).

[26] S. K. Donaldson and P. B. Kronheimer, The Geometry of Four-Manifolds (Oxford University Press, Oxford, 1990).

[27] M. H. Freedman and F. Luo, Selected Applications of Geometry to Low-dimensional Topology, Marker Lecture Series (American Mathematical Society, Providence, 1989).

[28] T. Petrie and J. Randall, Connections, Definite Forms, and Four-manifolds (Oxford Univ. Press, Oxford, 1990).

[29] F. Hirzebruch, Topological Methods in Algebraic Geometry (Springer, Berlin, 1978).

[30] P. Shanahan, The Atiyah-Singer Index Theorem, Lecture Notes in Mathematics Series, vol. 638 (Springer, Berlin, 1978).