The quantum electromagnetic field in multiply connected space

1979 J. Phys. A: Math. Gen. 12 403

(http://iopscience.iop.org/0305-4470/12/3/016)

J. Phys. A: Math. Gen., Vol. 12, No. 3, 1979. Printed in Great Britain

The quantum electromagnetic field in multiply connected space

Rafael Sorkin†

Department of Applied Mathematics and Astronomy, University College, PO Box 78, Cardiff, UK

Received 24 January 1978

Abstract. A gauge-invariant and microscopically causal theory of the sourceless electromagnetic field in a *topologically arbitrary* globally hyperbolic asymptotically flat *background metric* is proposed. The topology manifests itself in an algebra of superselected quantities, one of which is the net (apparent) charge. Relative to a particular Cauchy hypersurface, the field resolves into a 'Coulombic' part generating the above algebra and a 'radiative' part expressible in terms of photon creation and destruction operators. An appendix extends 'Hodge theory' to non-compact, but asymptotically-flat three-manifolds.

1. Introduction

In an earlier paper (Sorkin 1977) I described a 'quasi-local' model of electric charge based on the sourceless Maxwell equations in a space–time with suitable topology. More specifically, the metric, which was treated as a background, was assumed to be temporally but not necessarily spatially orientable, with all topological complications confined to a spatially compact region, K. It followed then—with $F_{\mu\nu}$ being identified contrary to custom as an *axial* rather than a polar tensor—that K can never display net magnetic charge but *can* display net *electric* charge in the non-orientable case—at least for the 'non-orientable handle' I described.

In order to bring the above model closer to reality it would be necessary to incorporate the metric as a dynamical field and then to quantise the entire system. The present paper undertakes only a small part of this task—quantisation of the electromagnetic field in an unquantised background metric. As a by-product, we will get in § 4 a comprehensive answer to the question, posed implicitly in Sorkin (1977), of the relation between the topology of a space–time and the character of the purely classical electromagnetic fields it admits. (See also lemma 2 of the appendix.)

The terminology of the present paper (some of which is described in more detail in the appendix) agrees with that of Sorkin (1977); units are chosen so that $c = \hbar = 4\pi\epsilon_0 = 1$; '$A \natural B$' means $AB = BA$; '$A := B$' means that the equation $A = B$ *defines* A.

2. Maxwell's equations in '3+1 form'

While it might be preferable to work directly with the four-dimensional field $F_{\mu\nu}$, one of the mathematical tools we will need—Hodge theory— seems to have been forged only

† Present address: Box 14, Enrico Fermi Institute, 5630 Ellis Avenue, Chicago, Illinois 60637, USA.

in the context of manifolds with positive-definite metric. In addition, the notion of canonical quantisation seems to depend inevitably on the existence of Cauchy hypersurfaces. Let us therefore consider a foliation of the space–time M into suitable space-like hypersurfaces $t=$ constant and describe $F_{\mu\nu}$ relative to a particular one of these, $\mathscr{H}$, using a local system of coordinates in which $x^0 \equiv t$.

If γ_{jk}, with inverse γ^{jk}, is the induced metric on $\mathscr{H}$ then in our chosen coordinates $\gamma_{jk}=g_{jk}$ while, as is then easily checked,

$$\gamma^{jk}=g^{jk}-g^{0j}g^{0k}/g^{00}. \tag{2.1}$$

The field $F_{\mu\nu}$ is most naturally resolved into magnetic and electric parts by defining

$$B_{jk}=F_{jk}, \tag{2.2}$$

which is the 'pull-back' of F to $\mathscr{H}$, and

$$\mathscr{E}^k=\mathscr{F}^{0k} \tag{2.3}$$

(where $\mathscr{F}^{\mu\nu} := \sqrt{-g}\, g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}$), which might be called the 'intersection' of $\mathscr{F}^{\mu\nu}$ with $\mathscr{H}$. It is well known that the former, and can also be shown† that the latter, is a tensorial quantity in $\mathscr{H}$; in particular the components B_{jk} $(=-B_{kj})$ and $\mathscr{E}^k$ depend only on the choice of coordinates in $\mathscr{H}$ itself. Moreover, since $\mathscr{H}$ is by assumption externally oriented (as defined by Sorkin (1977)) an (internal) orientation of $\mathscr{H}$ at $x\in\mathscr{H}$ is equivalent to one of M there. Therefore the postulated axial character of $F_{\mu\nu}$ (hence also of $\mathscr{F}^{\mu\nu}$) is inherited by B_{jk} and $\mathscr{E}^k$ which are accordingly an axial two-form and an axial vector density‡.

Let us write Maxwell's equations as

$$\partial\wedge F := \partial_\alpha F_{\beta\gamma}+\partial_\beta F_{\gamma\alpha}+\partial_\gamma F_{\alpha\beta}=0 \tag{2.4}$$

and dually

$$\partial \lrcorner \mathscr{F} := \partial_\mu \mathscr{F}^{\mu\nu}=0. \tag{2.5}$$

When time and space indices are distinguished explicitly each of (2.4) and (2.5) gives rise to one 'initial-value constraint' and one 'evolution equation'. The constraint pair comes already expressed in terms of B and $\mathscr{E}$:

$$\partial\wedge B := \partial_i B_{jk}+\partial_j B_{ki}+\partial_k B_{ij}=0 \tag{2.6}$$

$$\partial \lrcorner \mathscr{E} := \partial_k \mathscr{E}^k=0; \tag{2.7}$$

but the evolution pair involves further components of $F_{\mu\nu}$ and $\mathscr{F}^{\mu\nu}$:

$$\partial_0 B_{jk}+\partial_j F_{k0}-\partial_k F_{j0}=0 \tag{2.8}$$

$$\partial_0 \mathscr{E}^k+\partial_j \mathscr{F}^{jk}=0. \tag{2.9}$$

We therefore have to express F_{k0} and $\mathscr{F}^{jk}$ in terms of B and $\mathscr{E}$.

† The most direct way to see this is through the easy-to-verify equality

$$\mathscr{F}^{0k}=\tfrac{1}{2}\epsilon^{kab}\tfrac{1}{2}\epsilon_{ab\mu\nu}\mathscr{F}^{\mu\nu}$$

which displays $\mathscr{E}^k$ as the dual (in $\mathscr{H}$) of the pull-back to $\mathscr{H}$ of the dual (in M) of $\mathscr{F}^{\mu\nu}$. Here, as throughout, Greek indices range from 0 to 3, italic from 1 to 3 and $\epsilon_{A\ldots B}$ has components ±1 or 0 in any coordinate system.

‡ Eschewing axial quantities, one could work with ${}^*F_{\mu\nu} := \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\mathscr{F}^{\alpha\beta}$, $E_{mn} := \epsilon_{mnk}\mathscr{E}^k$ and $\mathscr{B}^k := \tfrac{1}{2}\epsilon^{kab}B_{ab}$. Alternatively, one could highlight the symmetry between electric and magnetic fields by using either of the pairs E_{mn}, B_{mn} or $\mathscr{E}^k$, $\mathscr{B}^k$.

In preparation for this, note that the determinants of γ_{jk} and $g_{\mu\nu}$ are related by

$$g^{00}g = \gamma \tag{2.10}$$

and define

$$\nu = |g^{00}|^{-1/2} = |g|^{1/2}|\gamma|^{-1/2} \tag{2.11}$$

$$N^k = g^{0k}/|g^{00}| = -g^{0k}/g^{00} \tag{2.12}$$

which are the customary 'lapse' and 'shift' functions conforming to our particular choice of coordinates. (Geometrically ν and N^k give respectively the length and the orthogonal projection into $\mathscr{H}$ of the vector $\partial/\partial t$ which joins one hypersurface $x^0 = \text{constant}$ to the 'next'.) We will, of course, raise and lower italic indices with γ_{jk} and use $\sqrt{\gamma}$ to convert between $\mathscr{H}$ tensors and $\mathscr{H}$-tensor densities.

Now

$$F^0{}_k = F^0{}_k$$

$$F^{0m}g_{mk} = g^{00}F_{0k} + g^{0j}F_{jk},$$

which in the light of (2.2), (2.3) and (2.10)–(2.12) becomes

$$F_{k0} = \nu E_k - N^j B_{jk}. \tag{2.13}$$

Similarly (using $\mathfrak{S}$ to denote skewing in j and k)

$$\mathscr{F}_{jk} = g_{j\mu}g_{k\nu}\mathscr{F}^{\mu\nu}$$

$$= g_{jk}g_{kn}\mathscr{F}^{mn} + \mathfrak{S}g_{j0}g_{kn}\mathscr{F}^{0n}.$$

But

$$g_{j0}g^{00} + g_{jm}g^{n0} = g_{j\mu}g^{\mu 0} = 0 \Rightarrow$$

$$g_{j0} = -(g^{00})^{-1}g_{jm}g^{m0} = \gamma_{jm}N^m = N_j \tag{2.14}$$

which can be substituted into the above to give (using also (2.11))

$$\nu\mathscr{B}_{jk} = \gamma_{jm}\gamma_{kn}\mathscr{F}^{mn} + \mathfrak{S}N_j\gamma_{kn}\mathscr{E}^n$$

or

$$\mathscr{F}^{jk} = \nu\mathscr{B}^{jk} - N^j \wedge \mathscr{E}^k \tag{2.15}$$

where

$$N^j \wedge \mathscr{E}^k := N^j\mathscr{E}^k - N^k\mathscr{E}^j.$$

Substituting these results into (2.8) and (2.9) yields the evolutionary Maxwell's equations in the form†

$$\partial B_{jk}/\partial t + \partial_j \wedge (\nu E_k - N^a B_{ak}) = 0 \tag{2.16}$$

† Notice the well known symmetry of these equations, as well as of (2.6) and (2.7), with respect to the 'dualisation'

$$B \to *\mathscr{E} = \epsilon_{mnk}\mathscr{E}^k$$

$$\mathscr{E} \to -*B = -\tfrac{1}{2}\epsilon^{kmn}B_{mn}.$$

In fact, explicit use of this symmetry, which derives from that of (2.4) and (2.5), would allow one to pass directly from (2.16) to (2.17) without ever having evaluated $\mathscr{F}^{jk}$, and similarly to avoid half of the verifications in § 3 below. (Note however that dualisation changes axial to polar and therefore does not in general make sense globally.)

$$\partial \mathscr{E}^k/\partial t + \partial_j(\nu \mathscr{B}^{jk} - N^j \wedge \mathscr{E}^k) = 0. \tag{2.17}$$

Here, of course, $\partial_j \wedge X_k$ means $\partial_j X_k - \partial_k X_j$.

In this form the equations look coordinate dependent since ν and N^k involve particular components of $g^{\mu\nu}$. But by introducing explicitly the vector field $\partial/\partial t$ with components $\xi^\mu = \delta^\mu_0$ we can interpret them in a fully 'geometrical' manner as governing the change in $\mathscr{E}$ and B when one moves along ξ^μ from $\mathscr{H}(0) = \mathscr{H}$ to $\mathscr{H}(\epsilon)$ where $\mathscr{H}(s) := \{x \in M \mid t(x) = s\}$. For under the interpretation of $\partial/\partial t$ as the 'Lie derivative'†, $\pounds_\xi$, of N^k as the orthogonal projection of ξ^μ into $\mathscr{H}$, and of ν as the distance from $\mathscr{H}(0)$ to $\mathscr{H}(\epsilon)$ each term of (2.16) and (2.17) acquires a covariant meaning. By defining ν to be negative when ξ^μ points backward in time with respect to $\mathscr{H}$ we can even free ξ^μ altogether from being tied to any local coordinate system whatever. Written in an index-free notation the thus reinterpreted equations read

$$\underset{\xi}{\pounds} B + \partial \wedge (\nu E - N \lrcorner B) = 0 \tag{2.18}$$

$$\underset{\xi}{\pounds} \mathscr{E} + \partial \lrcorner (\nu \mathscr{B} - N \wedge \mathscr{E}) = 0 \tag{2.19}$$

in which the densities $\mathscr{E}$ and $\mathscr{B}$ are implicitly contravariant while E and B are implicitly covariant, and where ν, N^k are derived from ξ^μ as just described.

We get a final simplification by using a connecting vector ξ^μ which is perpendicular to $\mathscr{H}$. In this case $N^k = 0$, $\nu = \mathrm{sgn}(\xi^0)\|\xi\|$ and with the notation $\dot{X} := \pounds_\xi X$ the evolution equations approach very closely to their flat-space appearance:

$$\dot{B} + \partial \wedge (\nu E) = 0 \tag{2.20}$$

$$\dot{\mathscr{E}} + \partial \lrcorner (\nu \mathscr{B}) = 0. \tag{2.21}$$

For future reference let us evaluate the 'ξ momentum' $P[\xi, \mathscr{H}]$ associated with an arbitrary vector field ξ^μ and the hypersurface $\mathscr{H}$:

$$\begin{aligned} \mathfrak{T}^\mu_\nu &= \mathscr{F}^{\mu\alpha} F_{\nu\alpha} - \tfrac{1}{4}\delta^\mu_\nu \mathscr{F}^{\alpha\beta} F_{\alpha\beta} \\ P[\xi, \mathscr{H}] &= \int_{\mathscr{H}} \mathfrak{T}^\mu_\lambda \xi^\lambda \, \mathrm{d}\sigma_\mu. \end{aligned} \tag{2.22}$$

Working for convenience in a local system in which, as before, $\xi^\mu = \delta^\mu_0$ and $x^0 \equiv t$, we find

$$\mathfrak{T}^\mu_\lambda \xi^\lambda \, \mathrm{d}\sigma_\mu = \mathfrak{T}^0_0 \, \mathrm{d}^3 x$$

and

$$\begin{aligned} \mathfrak{T}^0_0 &= \mathscr{F}^{0k} F_{0k} - \tfrac{1}{4}(2\mathscr{F}^{0k} F_{0k} + \mathscr{F}^{jk} F_{jk}) \\ &= -\tfrac{1}{2}(\mathscr{E}^k F_{k0} + \tfrac{1}{2}\mathscr{F}^{jk} B_{jk}) \\ &= -\tfrac{1}{2}[\mathscr{E}^k(\nu E_k - N^j B_{jk}) + \tfrac{1}{2}(\nu \mathscr{B}^{jk} - N^j \wedge \mathscr{E}^k) B_{jk}] \end{aligned}$$

† Strictly speaking, '$\pounds_\xi B$', for example, does not make sense since B is not a two-form in M but only one in $\mathscr{H}(s)$ for each s. What $\partial B/\partial t$ really represents is

$$(\partial T(s)^* B(s)/\partial s)|_{s=0}$$

where $\mathscr{H} = \mathscr{H}(0)$, $B(s)$ is $F_{\mu\nu}$ restricted to $\mathscr{H}(s)$, and $T(s)^* B(s)$ is $B(s)$ pulled back from $\mathscr{H}(s)$ to $\mathscr{H}$ via the identification of their points set up by ξ^μ. Thus '$\pounds_\xi$' as used here is a generalisation—but an obvious one—of the usual notation.

with the help of (2.13) and (2.15). Since the result is again covariant, we can perform the integration to obtain

$$P[\xi, \mathscr{H}] = \int -\tfrac{1}{2}(\mathscr{E}^m E_m + \tfrac{1}{2}\mathscr{B}^{mn} B_{mn})\nu + N^m B_{mn}\mathscr{E}^n \, \mathrm{d}^3 x \tag{2.23}$$

where—as always—ν and N are related to ξ^μ as before.

3. Formal quantisation

As throughout this paper, let M be a time-oriented space–time, suitably regular at spatial infinity, and such that through each event there passes an asymptotically flat Cauchy hypersurface. We seek to specify commutation relations for $F_{\mu\nu}(x)$ which vanish at space-like separations ('microscopic causality'), are compatible with Maxwell's equations, and for which the (formal) operator $P[\xi, \mathscr{H}]$ of equation (2.23) generates a deformation along ξ.

To this end, let us pick a Cauchy hypersurface and introduce therein the invariant bi-tensor

$$\delta_j(x, y)^k$$

in which j is a covariant index of weight 0 at x and k is a contravariant index of weight 1 at y, and which is defined by

$$\iint \mathfrak{A}^j(x)\delta_j(x, y)^k A_k(y) \, \mathrm{d}^3x \, \mathrm{d}^3y = \int \mathfrak{A}^j(x) A_j(x) \, \mathrm{d}^3x$$

for all test vector fields $A_j(x)$ and test vector densities $\mathfrak{A}^k(x)$. In any local coordinate system δ has the components

$$\delta_j(x, y)^k = \delta_j^k \delta(x - y), \tag{3.1}$$

which transform, as is easily checked, consistently with the character assigned to them above.

In analogy with the flat-space situation we impose relative to $\mathscr{H}$ the commutation relations†

$$\begin{aligned} &[B(x), B(y)] = 0 \qquad [\mathscr{E}(x), \mathscr{E}(y)] = 0 \\ &[B_{mn}(x), \mathscr{E}^k(y)] = -\mathrm{i}\, \partial_m(x) \wedge \delta_n(x, y)^k \mathbb{1} \end{aligned} \tag{3.2}$$

where $\mathbb{1}$ is the unit operator and the expression with the wedge means

$$\frac{\partial}{\partial x^m}\delta_n(x, y)^k - \frac{\partial}{\partial x^n}\delta_m(x, y)^k.$$

(To avoid confusion note that the wedge refers not to any suppressed indices but—just as in (2.15)–(2.17)—to m and n themselves.)

Together with Maxwell's equations and the reality conditions

$$F_{\mu\nu}(x) = F_{\mu\nu}(x)^*, \tag{3.3}$$

† Observe that these commutation relations possess the dualisation symmetry discussed in the footnote to (2.16) and (2.17).

the commutation relations (3.2) specify fully the quantum algebra, $\mathfrak{A}$, of our system†.

Let us rule out any obvious incompatibility‡ of (3.2) with the initial-value constraints (2.6) and (2.7) on $\mathscr{H}$ by showing that the commutators of the latter with $\mathscr{E}$ and B vanish identically.

For (2.6) the calculation of the non-trivial commutator is trivial:

$$[\partial_l \wedge B_{mn}(x), \mathscr{E}^k(y)] = -\mathrm{i}\, \partial_l \wedge \partial_m \wedge \delta_n(x, y)^k = 0$$

since $\partial \wedge \partial = 0$. For (2.7) we integrate

$$\mathrm{i}[B_{mn}(x), \partial_k \mathscr{E}^k(y)] = \partial_k(y) \partial_m(x) \wedge \delta_n(x, y)^k \tag{3.4}$$

against an arbitrary function of compact support, $f(y)$, to show that it vanishes. But by definition

$$\int f(y) \partial_k(y)\, \delta_n(x, y)^k\, \mathrm{d}^3 y = -\int \partial_k f(y)\, \delta_n(x, y)^k\, \mathrm{d}^3 y$$
$$= -\partial_n f(x)$$

so that

$$\int f(y)\, \partial_k(y)\, \partial_m(x) \wedge \delta_n(x, y)^k\, \mathrm{d}^3 y = -\partial_m \wedge \partial_n f(x) = 0$$

as required. (Notice by the way that the first half of the calculation just says $\partial_k(y)\, \delta_n(x, y)^k = -\partial_n(x)\, \delta(x, y)$, which could have been substituted directly into (3.4).)

As for compatibility with the remaining pair of Maxwell's equations, that will follow from the existence of a 'Hamiltonian' generating the deformation via ξ^μ of one hypersurface $\mathscr{H}$ to a neighbouring one. Because the 'canonical commutation relations' (3.2) take the same form on each hypersurface we can be optimistic that such a 'Hamiltonian' exists. Let us verify that, in fact, $P[\xi, \mathscr{H}]$ is the anticipated generator; i.e. that

$$[P[\xi, \mathscr{H}], X(x)] = \mathrm{i} \mathop{\pounds}_{\xi} X(x) \tag{3.5}$$

where 'X' stands for one of the 'canonical' quantities $\mathscr{E}^k$, B_{mn} and where $x \in \mathscr{H}$.

From (2.23) we have for B

$$[P[\xi, \mathscr{H}], B(x)]$$
$$= -\tfrac{1}{2} \int [\mathscr{E}^k E_k(y), B_{mn}(x)] \nu(y)\, \mathrm{d}^3 y$$
$$+ \int N^j(y) [B_{jk}(y) \mathscr{E}^k(y), B_{mn}(x)]\, \mathrm{d}^3 y$$

† I have been vague about just what sort of algebra $\mathfrak{A}$ is supposed to be. However, the concern of this paper is to discover specifically topological effects, not to worry about the mathematical problems common to *any* quantum field theory in curved space–time, which, as will appear in § 5, are neither better nor worse here than they are when topological complications are absent.

To form a more concrete idea of $\mathfrak{A}$, one could imagine it (conforming to the wording of the present section) as an algebra of operators in Hilbert space; a more 'algebraic' attitude will be adopted in § 4.

‡ From the 'algebraic' point of view of § 4 compatibility means that $\mathfrak{A}$ is not reduced to the zero algebra, i.e., that $\mathbb{1} \neq 0 \cdot \mathbb{1}$. That there is no hidden incompatibility follows from the fact that $\mathfrak{A}$ possesses non-trivial representations in Hilbert space, as will appear in § 5.

$$= \int (\nu E_k(y) - N^j B_{jk}(y))[B_{mn}(x), \mathscr{E}^k(y)]\, \mathrm{d}^3 y$$

$$= -\mathrm{i} \int (\nu E - N \lrcorner B)_k(y)\, \partial_m(x) \wedge \delta_n(x, y)^k\, \mathrm{d}^3 y$$

$$= -\mathrm{i}\, \partial \wedge (\nu E - N \lrcorner B)$$

$$= \mathrm{i} \underset{\xi}{\pounds} B(x)$$

according to (2.18). For $\mathscr{E}$ we have similarly

$$[P[\xi, \mathscr{H}], \mathscr{E}(x)]$$

$$= -\tfrac{1}{4} \int [B_{mn}\mathscr{B}^{mn}(y), \mathscr{E}^k(x)]\nu(y)\, \mathrm{d}^3 y + \int N^m(y)[B_{mn}(y), \mathscr{E}^k(x)]\mathscr{E}^n(y)\, \mathrm{d}^3 y$$

$$= \int [-\tfrac{1}{2} \nu\mathscr{B}^{mn}(y) + N^m \mathscr{E}^n(y)][B_{mn}(y), \mathscr{E}^k(x)]\, \mathrm{d}^3 y$$

$$= \mathrm{i} \int \tfrac{1}{2}(\nu\mathscr{B}^{mn} - N^m \wedge \mathscr{E}^n)(y)\, \partial_m(y) \wedge \delta_n(y, x)^k\, \mathrm{d}^3 y$$

$$= -\mathrm{i}\, \partial_m(\nu\mathscr{B} - N \wedge \mathscr{E})^{mn}$$

$$= \mathrm{i} \underset{\xi}{\pounds} \mathscr{E}$$

according to (2.19).

On the basis of (3.5) it is now easy to complete our discussion of compatibility. Let $[X, Y] = Z$ represent any one of the basic commutation relations (3.2). If $\pounds_\xi$ is as above, then since Z is in all cases an invariant bi-tensor defined in $\mathscr{H}$, $\pounds_\xi Z = 0$. As for $\pounds_\xi\, [X, Y]$ we have from (3.5) and the Jacobi identity

$$\begin{aligned}
\mathrm{i} \underset{\xi}{\pounds} [X, Y] &= \left[\mathrm{i} \underset{\xi}{\pounds} X, Y\right] + \left[X, \mathrm{i} \underset{\xi}{\pounds} Y\right] \\
&= [[P, X], Y] + [X, [P, Y]] \\
&= [P, [X, Y]] \\
&= [P, Z] \\
&= 0
\end{aligned}$$

since Z is a c number. Thus (3.2) is preserved by $\pounds_\xi$.

4. Superselection and homology

The formal quantisation just completed has furnished an algebra† $\mathfrak{A}$ purporting to be the quantum algebra of the system 'free electromagnetic field in background metric $g_{\mu\nu}$'. Ordinarily we would now try to find elements of $\mathfrak{A}$ playing the roles of creation

† Explicitly $\mathfrak{A}$ is the free $*$ algebra on generators $\{F_{\mu\nu}(x) \mid x \in M\} \cup \{\mathbb{1}\}$ modulo the relations comprised in Maxwell's equations (2.4) and (2.5), the reality conditions (3.3) and the commutation relations (3.2) (as well as the relations specifying that $\mathbb{1}$ is the unit element of $\mathfrak{A}$).

and annihilation operators and then represent $\mathfrak{A}$ in the corresponding Fock space. In other words, we would interpret $\mathfrak{A}$ as describing a system of photons. However, for curved metrics where the notion of particles becomes much more problematic, such an approach is not necessarily appropriate. Instead let us pursue our main concern—as to what becomes in quantum theory of the sort of classical solutions discussed by Sorkin (1977)—solely with reference to the quantum *algebra* $\mathfrak{A}$.

The new quantity of greatest interest which arises when space is non-orientable is undoubtedly the total charge, expressed relative to the hypersurface $\mathscr{H}$ as

$$Q = \oint_S \mathscr{E}^k \, \mathrm{d}\sigma_k \tag{4.1}$$

where S is any two-sphere enclosing the whole region K of multiple connectivity. Notice that the integrand is a true scalar only if S is interpreted as a surface with given *internal* orientation, so that a change in local orientation changes the signs of both $\mathscr{E}$ *and* $\mathrm{d}\sigma$. But to (internally) orient S is equivalent to choosing a 'right-hand rule' for the asymptotic region $\mathscr{H}-K$. Thus Q is what might be called an *axial* scalar defined at infinity. Recall also that Q depends only on the (internally oriented) two-surface S and not on the three-surface $\mathscr{H}$ in which S is embedded, as appears, e.g., in the four-dimensional expression

$$Q = \oint_S \tfrac{1}{2}(*F)_{\mu\nu} \, \mathrm{d}\Sigma^{\mu\nu},$$

both factors of whose integrand are polar four-tensors.

From (4.1) it is easy to see that Q is in the *centre* of $\mathfrak{A}$, that is that $Q \natural A$ for all $A \in \mathfrak{A}$. In fact, if x is any point of $\mathscr{H}$ then we can assume $x \notin S$: if by mischance x does meet S, just deform S slightly! This being given, (3.2) leads immediately to $Q \natural \mathscr{E}(x)$ and $Q \natural B(x)$, and thence to $Q \natural \mathfrak{A}$ since for all $x \in M$, $F_{\mu\nu}(x)$ is linearly dependent (via Maxwell's equations) on the set $\{\mathscr{E}(x), B(x) \mid x \in \mathscr{H}\}$.

Since Q commutes with every element of $\mathfrak{A}$, no measurement of such an element—in other words, no measurement of any quantity associated with the system—can induce transitions in the value of Q. Therefore for a particular representation of $\mathfrak{A}$ in a Hilbert space $\mathfrak{H}$, phase relations between sectors of $\mathfrak{H}$ characterised by different values of Q are unobservable. One says that Q is 'superselected' and one can call any (self-adjoint) central element of $\mathfrak{A}$ a 'superselection rule'. In the present theory, then, superselection of total charge, something which is often asserted on experimental grounds, appears as a theoretical consequence of Maxwell's equations and microscopic causality.

It seems, then, that the interesting features of $\mathfrak{A}$ are associated with superselection rules. In fact *any* flux integral of the form (4.1) but with S an *arbitrary* closed internally oriented two-surface (not necessarily a two-sphere!) furnishes a quantity C which, by precisely the same reasoning as before, is also superselected.

Of course, not every S yields a different C. Two surfaces S' and S'' which give the same C for every possible divergence-free axial field $\mathscr{E}^k$ are said to be *homologous*. (Conversely if some S is not homologous to zero then† there will be some (c-number) field $\mathscr{E}^k(x)$ for which $\oint_S \mathscr{E} \, \mathrm{d}\sigma \neq 0$. Using this $\mathscr{E}$ together with $B = 0$ as initial data in Maxwell's equations as given in § 2 we recover a classical solution with non-zero flux through S. This answers the question posed in the introduction about how to characterise all the classically conserved quantities associated with the topology.) Cor-

† In the notation of the appendix, $H_2(\mathscr{H}, \mathbb{R}) \simeq H(\mathscr{E}, \delta, -, 1)^*$, which is proved as equation (A.14).

responding to every homology class of two-surfaces (more precisely to every generator of $H_2(\mathscr{H}, \mathbb{R})$) there is a possible 'electric' superselection rule. We could go on now to show that no other candidates can arise, but this will emerge anyway from the more complete analysis to be made in § 5 with the aid of 'Hodge theory'.

The situation with the 'magnetic' superselection rules is similar: they also correspond to closed two-surfaces (this time externally oriented) or rather to equivalence classes of such which are homologous in an appropriate sense. The corresponding homology group turns out (A.17) to be $H_1(\mathscr{H}, \mathbb{R})$.

Finally, let us apply these insights to a particular topology discussed by Sorkin (1977)—the non-orientable handle. It is not hard to see that $H_2(\mathscr{H}, \mathbb{R})$ is one dimensional, a generator for it being given by a sphere S' surrounding either mouth of the handle. (At first sight one might conclude that the 'two-sphere at infinity', S, provides an independent generator, but as shown in effect by Sorkin (1977), this S is homologous to twice S'.) The corresponding superselected quantity is of course (half) the total charge. Similarly a generator for $H_1(\mathscr{H}, \mathbb{R})$ is any loop linking the handle whence there is exactly one magnetic superselection rule, governing the 'dipole moment' of the handle.

5. The 'radiative' part of F

The last section showed that our algebra $\mathfrak{A}$ contains superselected quantities which remain, in a sense, 'unquantised'. The present section will show how, relative to a particular Cauchy hypersurface $\mathscr{H}$, one can identify these quantities with a definite component of the field and separate them off leaving behind a 'radiation field' $\tilde{F}_{\mu\nu}$ whose understanding involves only those problems familiar from, say, the quantisation of a massless scalar field in a curved but *topologically trivial* space–time. Difficult as these problems (definition of vacuum, regularisation of $T_{\mu\nu}$, etc) are, at least the topological complexity of $\mathscr{H}$ will not have added to them. (The notation for what follows is set forth in the appendix.)

So far we have not had to worry about regularity conditions for $F_{\mu\nu}$. In the sequel, though, we will assume that both $E_k := |\gamma|^{-1/2}\gamma_{km}\mathscr{E}^m$ and B are square-integrable on $\mathscr{H}$, in other words that $E \in \overline{\mathscr{D}^{1-}}(\mathscr{H})$ and $B \in \overline{\mathscr{D}^{2-}}(\mathscr{H})$. According to (2.23) this is necessary in order that the total energy (which makes sense since $\mathscr{H}$ is asymptotically flat) be finite.

Since $B \in \bar{\mathscr{D}}$ the Kodaira theorem of the appendix decomposes B into a sum of two terms (the third being absent since $\mathrm{d}B = 0$)

$$B = \tilde{B} + \bar{B} \tag{5.1}$$

where $\tilde{B} \in \overline{\mathrm{d}\mathscr{D}^{1-}}$ and $\bar{B} \in K_2^{2-}$, so that $\mathrm{d}\bar{B} = \delta\bar{B} = 0$. Similarly

$$E = \tilde{E} + \bar{E} \tag{5.2}$$

where $\tilde{E} \in \overline{\delta\mathscr{D}}$ and $\bar{E} \in K_2$.

Now let $e \in \overline{\mathscr{D}^{1-}}$, $b \in \overline{\mathscr{D}^{2-}}$ and define

$$[b \,|\, e] = [((b, B)), ((E, e))]. \tag{5.3}$$

It is clear that this formation defined on $\overline{\mathscr{D}^{1-}} \times \overline{\mathscr{D}^{2-}}$ contains the complete commutation relations in as much as the $\mathscr{L}^2$ forms B and E can be expanded in terms of such functions

b and e. (The coefficients in the expansion will, of course, be elements of $\mathfrak{A}$ or 'q-numbers'.)

From (3.2) it is easy to evaluate (5.3):

$$\begin{aligned}[b\,|\,e] &= \iint \tfrac{1}{2}\mathscr{b}^{ab}(x)[B_{ab}(x), \mathscr{E}^k(y)]e_k(y)\,\mathrm{d}^3x\,\mathrm{d}^3y \\ &= -\mathrm{i}\iint \mathscr{b}^{ab}(x)\,\partial_a(x)\,\delta_b(x,y)^k e_k(y)\,\mathrm{d}^3x\,\mathrm{d}^3y \\ &= -\mathrm{i}\int b^{ab}\,\partial_a e_b\,\mathrm{d}V \\ &=: -\mathrm{i}\,((b, \mathrm{d}e))\end{aligned}$$

(where of course $\mathscr{b}^{ab} := |\gamma|^{1/2}b^{ab}$). Thus

$$[b\,|\,e] = -\mathrm{i}((b, \mathrm{d}e)) = \mathrm{i}((\delta b, e)). \tag{5.4}$$

In particular, if either b or e is in K_2 then $[b\,|\,e]=0$. But $\bar{B}$ (for example), being the K_2 component of B in the decomposition (A.7), can be expanded in terms of an orthonormal basis of functions in K_2:

$$\sum_n \beta_n b_n(x) \qquad (\beta_n \in \mathfrak{A}) \tag{5.5}$$

$$\beta_n = ((b_n, B)),$$

whence, if e is arbitrary,

$$\begin{aligned}[\bar{B}(x), ((E, e))] &= \sum_n b_n(x)[\beta_n, ((E, e))] \\ &= \sum_n b_n(x)[((b_n, B)), ((E, e))] \\ &= \sum_n b_n(x)[b_n\,|\,e] \\ &= 0\end{aligned}$$

since $b_n \in K_2$. Because e was arbitrary this implies $\bar{B}(x) \natural E$ so that $\bar{B}(x) \in \mathrm{centre}(\mathfrak{A})$. Similarly $\bar{E}$ is also central.

Conversely, if

$$\tilde{B}(x) = \sum_n \tilde{\beta}_n \tilde{b}_n$$

is the expansion of $\tilde{B}$ in terms of a basis for $\overline{\mathrm{d}\mathscr{D}}$ then *no* component of $\tilde{B}$ can $\natural\,\mathfrak{A}$, for if it did, then for arbitrary e and n

$$[\tilde{b}_n\,|\,e] = 0 \qquad ((\delta\tilde{b}_n, e)) = 0 \qquad \delta\tilde{b}_n = 0,$$

whence $\tilde{b}_n \in K_2$, whence $\tilde{b}_n = 0$ by (A.6).

Another way to express this result says that the 'commutator product' $[\,|\,]$ is a *non-degenerate* bilinear form on $\overline{\mathrm{d}\mathscr{D}} \times \overline{\delta\mathscr{D}}$, which means that by a proper choice of basis functions $\tilde{E}$ and $\tilde{B}$ can be resolved into linear combinations of boson creation and annihilation operators of the usual sort.

Before making this resolution explicit, let us note that we can now characterise centre($\mathfrak{A}$) as the subalgebra $\mathfrak{A}'$ of $\mathfrak{A}$ generated by $\bar{E}$ and $\bar{B}$. For on one hand we know that these commute with $\mathfrak{A}$, while on the other hand any expression involving components of $\tilde{E}$ and $\tilde{B}$ in an essential way (i.e. so that they do not drop out because of (3.2)) involves creation or annihilation operators and will not commute with all other expressions involving such operators (as follows from the usual theory of 'second quantisation'). Thus we can identify the set of superselection rules with any independent set of components of $\bar{E}$ and $\bar{B}$, that is with any basis for K_2^{1-} and K_2^{2-}. But according to the appendix (A.15), (A.16)

$$K_2^{1-} \simeq H_2(\mathcal{H}, \mathbb{R})^*$$

and

$$K_2^{2-} \simeq H_1(\mathcal{H}, \mathbb{R}),$$

which shows that every 'possible' superselection rule of the type discussed in § 4 occurs and that no other ones are possible.

Returning to the problem of diagonalising the commutation relations, let us define† for $e \in \overline{\delta\mathcal{D}^{2-}}$

$$\alpha(e) = ((e, E)) + \mathrm{i}((\mathrm{d}e, B)). \tag{5.6}$$

(Notice that e is still a *real* function.) Since $\overline{\mathrm{d}\mathcal{D}}$, $\overline{\delta\mathcal{D}}$ and K_2 are mutually orthogonal it is clear that only $\tilde{E}$ and $\tilde{B}$ will contribute to (5.6); conversely it is also clear that one can recover $\tilde{E}$ and $\tilde{B}$ from the $\alpha(e)$. We have (using (5.3) and (5.4))

$$\begin{aligned}[\alpha(e_1), \alpha(e_2)] &= \mathrm{i}[((e_1, E)), ((\mathrm{d}e_2, B))] + \mathrm{i}[((\mathrm{d}e_1, B)), ((e_2, E))] \\ &= -\mathrm{i}[\mathrm{d}e_2 | e_1] + \mathrm{i}[\mathrm{d}e_1 | e_2] \\ &= -((\mathrm{d}e_2, \mathrm{d}e_1)) + ((\mathrm{d}e_1, \mathrm{d}e_2)) \\ &= 0\end{aligned}$$

and

$$\begin{aligned}[\alpha(e_1), \alpha(e_2)^*] &= -\mathrm{i}[((e_1, E)), ((\mathrm{d}e_2, B))] + \mathrm{i}[((\mathrm{d}e_1, B)), ((e_2, E))] \\ &= ((\mathrm{d}e_2, \mathrm{d}e_1)) + ((\mathrm{d}e_1, \mathrm{d}e_2)) \\ &= 2((\mathrm{d}e_1, \mathrm{d}e_2)).\end{aligned} \tag{5.7}$$

I claim that this last expression is a *non-degenerate* quadratic form on the function space $\overline{\delta\mathcal{D}^{2-}}$. For, if $((\mathrm{d}e, \mathrm{d}e)) = 0$, then since $((\ ,\ ))$ is itself non-degenerate, $\mathrm{d}e = 0$ whence $e \in \overline{\delta\mathcal{D}} \cap K_2 = 0$. Letting (e_n) be a basis for $\overline{\delta\mathcal{D}^{2-}}$ orthonormal with respect to the form (5.7) gives at last the commutation relations in their familiar form:

$$\alpha(e_n) \natural\, \alpha(e_m) \qquad [\alpha(e_m), \alpha^*(e_n)] = \delta_{mn}.$$

Finally, consider how our decomposition of $F_{\mu\nu}$ depends on the choice of hypersurface with respect to which the decomposition is carried out. For simplicity, let us deal

† This is not the only way, but only the simplest way to diagonalise (5.3). A more useful definition in practice would be

$$a(e) = ((e, E)) - \mathrm{i}((\mathrm{d}\Delta^{-1}e, B))$$

which both destroys photons 'in state e' and accomplishes a simultaneous diagonalisation of the 'Hamiltonian'.

Notice also that, strictly speaking, it is not at all evident that (5.6) really makes sense for every $e \in \overline{\delta\mathcal{D}}$ since it can happen that $\mathrm{d}e$ is not square-integrable even though e is. In the remainder of this section I will ignore all such problems having to do with the unboundedness of the operators d, δ and Δ.

with an *irreducible* representation of $\mathfrak{A}$ in terms of operators in some Hilbert space, so that the central elements of $\mathfrak{A}$ become mere scalar multiples of the identity (c-numbers).

If we write $F = \tilde{F}' + \bar{F}'$ for the decomposition with respect to one hypersurface $\mathscr{H}'$ and propagate *separately* $\tilde{F}'$ and $\bar{F}'$ to all of M via Maxwell's equations (noting that both $\tilde{F}'$ and $\bar{F}'$ satisfy the initial-value constraints on $\mathscr{H}'$), then we get two separate solutions of the field equations, the second of which is a pure c number and the first of which therefore could itself be a complete solution to our problem (i.e. to (2.4), (2.5), (3.2) and (3.4) (and therefore to (3.5) as well)). Were we to decompose F with respect to a different hypersurface $\mathscr{H}''$, then, although $\tilde{F}$ and $\bar{F}$ would change, we would have $\tilde{F}'' - \tilde{F}' = \bar{F}' - \bar{F}'' = c$-number field. In other words the 'radiative' part $\tilde{F}$ of F is invariant up to the addition of certain c-number solutions of Maxwell's equations.

Even in flat space–time one has the freedom to add an arbitrary c-number solution to $F_{\mu\nu}(x)$, only in that case such an addition produces a *unitarily equivalent* set of operators, affecting only the identity of the vacuum state. The change from $\tilde{F}'$ to $\tilde{F}''$ is precisely of this type, as follows from (3.5) applied to $\tilde{F}$, and can therefore be interpreted in terms of creation of photons in a coherent state.

But adding to F a c-number solution with *non-zero* flux integrals—which is possible only when space–time is multiply connected—*cannot* produce a unitarily equivalent theory (the flux integrals being c numbers and therefore unitarily invariant). Instead it 'takes one to a different sector of Hilbert space'.

6. Conclusion

According to (3.2) the condition for microscopic causality, $F_{\mu\nu}(x) \natural F_{\alpha\beta}(y)$ is guaranteed whenever $x \neq y$ lie on a single Cauchy hypersurface. But because Maxwell's equations propagate information within the light-cone, it follows that in fact $F(x) \natural F(y)$ for *any* pair of points which are not joined by a causal curve.

Within its restriction to space–times, admitting Cauchy hypersurfaces (such a space–time is necessarily homeomorphic to $\mathscr{H} \times \mathbb{R}$ where $\mathscr{H}$ is any Cauchy hypersurface) our theory therefore incorporates the following features.

(1) It is defined in terms of local field variables $F_{\mu\nu}(x)$. (Thus it is trivially gauge invariant.)
(2) The $F_{\mu\nu}$ fulfill Maxwell's equations.
(3) Microscopic causality obtains.
(4) The integrated stress energy, $P[\xi, \mathscr{H}]$, generates deformations of the Cauchy hypersurface $\mathscr{H}$.

That all these features are present argues strongly (but certainly not, in the absence of a uniqueness theorem, conclusively) in its favour.

Within this theory we have been able to account fully for the influence of the topology on the quantum algebra. In particular, superselection emerges as a natural consequence of features (2) and (3) above. More generally, the algebra $\mathfrak{A}$ generated by the $F_{\mu\nu}$ splits (in a way largely independent of hypersurface) into a product of two factors, one of which is a 'Fock algebra' of the usual sort, and the other of which constitutes the centre of $\mathfrak{A}$ and reflects precisely the homology of $\mathscr{H}$.

An objection to the present theory might be that, because there is no vector potential, the field cannot interact with charged particles. However, Mandelstam (1962) has described a theory *with* interaction but *without* the vector potential. But even without this the present theory furnishes a model for charge 'coupled' *automatically* to the electromagnetic field. Full quantisation of this model (namely quantisation of the metric!) would produce a theory with the Lorentz force and therefore should incorporate somehow an effective vector potential. One could then see whether all the consequences of such a potential (e.g. the Bohm–Aharonov effect) were present. It even seems likely that the existence of an effective potential could be deduced already on the basis of the present theory, making use of the division of the field into 'radiative' and 'Coulombic' parts as described in § 5.

Finally one can ask: can the theory be freed from its $3+1$ form and especially from the need for Cauchy hypersurfaces? This question is of particular interest because it turns out, by an extension of the analysis of Sorkin (1977), that (for axial $F_{\mu\nu}$) *magnetic* monopoles can occur classically if and only if space–time lacks time orientability.

Acknowledgments

I thank John Friedman and Avner Ash for their help and suggestions.

Appendix. Asymptotically-flat Hodge theory

We consider various spaces of forms (totally skew covariant tensors) defined on a Riemannian manifold M of dimension N and with asymptotically-flat *positive-definite* metric g_{jk}. (For application to the main body of the paper M, N and g_{jk} would be respectively $\mathscr{H}$, 3, and γ_{jk}.) In more detail, we distinguish that

$$\mathscr{E}^p(M)=\{\omega\mid\omega \text{ is a } C^\infty \text{ form of rank } p\ (p\text{-form})\}$$

$$\mathscr{E}(M)=\bigoplus_{p=0}^{N}\mathscr{E}^p(M)\qquad\text{(the direct sum of the } \mathscr{E}^p)$$

$$\mathscr{E}^+(M)\ (\text{respectively } \mathscr{E}^-(M))=\{\omega\in\mathscr{E}(M)\mid\omega \text{ is polar (respectively axial)}\}$$

$$\mathscr{D}^p(M)=\{\omega\in\mathscr{E}^p(M)\mid \text{support }(\omega)\text{ is compact}\}$$

and similarly $\mathscr{D}(M)$, $\mathscr{D}^\pm(M)$, $\mathscr{E}^{p\pm}(M)$, $\mathscr{D}^{p\pm}(M)$. Clearly,

$$\mathscr{E}(M)=\mathscr{E}^+(M)\oplus\mathscr{E}^-(M).$$

There are also linear operators d, δ, Δ defined by

$$\begin{aligned}(\mathrm{d}\omega)_{jk\ldots l}&=\nabla_j\wedge\omega_{k\ldots l}\\ &=\nabla_j\omega_{k\ldots l}-\nabla_k\omega_{j\ldots l}+\ldots\pm\nabla_l\omega_{k\ldots j},\end{aligned}\tag{A.1}$$

where ∇_j is the covariant derivative,

$$(\delta\omega)_{k\ldots l}=\nabla^j\omega_{jk\ldots l}\tag{A.2}$$

$$\Delta=\mathrm{d}\delta+\delta\mathrm{d}.\tag{A.3}$$

With respect to the $\mathscr{L}^2$ scalar product

$$((\omega,\phi))=\sum_p(p!)^{-1}\int{}^p\omega_{j\ldots k}\phi^{j\ldots k}\,\mathrm{d}V\tag{A.4}$$

(where ${}^p\omega$ is the component of ω in $\mathscr{E}^p(M)$) one checks readily that the adjoint

$$\mathrm{d}^* = -\delta \tag{A.5}$$

whence also $\Delta^* = \Delta$. Also, of course, $\mathrm{d}^2 = 0 = \delta^2$ so that $\mathrm{d} \natural \Delta$, $\delta \natural \Delta$.

The operators d and δ are dual not only with respect to the scalar product (A.4) but also with respect to the so called * operator which converts between $\mathscr{E}^p$ and $\mathscr{E}^{N-p}$ by means of the alternating symbol $\epsilon_{j\ldots k}$. Note also that * preserves the $\mathscr{L}^2$ norm while reversing parity (i.e. exchanging $\mathscr{E}^+$ with $\mathscr{E}^-$).

Finally we define certain spaces of so called 'harmonic' form by

$$K = \{\omega \in \mathscr{E} \mid \mathrm{d}\omega = 0 = \delta\omega\}†$$

$$\bar{\mathscr{D}} = \text{completion of } \mathscr{D} \text{ in the } \mathscr{L}^2\text{-norm}$$

$$K_2 = K \cap \bar{\mathscr{D}}$$

$$K_2^{p\pm} = K_2 \cap \mathscr{E}^{p\pm}.$$

Note that, as discussed more fully by de Rham (1960), $\bar{\mathscr{D}}$ can also be described as the space of forms whose coefficients are $\mathscr{L}^2$-functions in the sense that $((\omega, \omega)) < \infty$. If $S \subset \bar{\mathscr{D}}$ is any subset then '$\bar{S}$' will denote the closure of S in the (real) Hilbert space $\bar{\mathscr{D}}$.

With these definitions we can now quote the theorem which authorised our decomposition of E and B into 'radiative' and 'Coulombic' parts in § 5.

Kodaira theorem.

$$\bar{\mathscr{D}} = \overline{\mathrm{d}\mathscr{D}} \oplus \overline{\delta\mathscr{D}} \oplus K_2. \tag{A.6}$$

Furthermore, if

$$\omega = \omega_1 + \omega_2 + \omega_3 \tag{A.7}$$

is the corresponding decomposition of $\omega \in \bar{\mathscr{D}}$ then $\mathrm{d}\omega = 0$ (respectively $\delta\omega = 0$) $\Leftrightarrow \omega_2$ (respectively ω_1) $= 0$.

Proof (*see theorem 24 of de Rham* (*1960*)). In the sequel we will also need the regularity results $\mathscr{E} \cap \overline{\mathrm{d}\mathscr{D}} \subset \mathrm{d}\mathscr{E}$ and dually $\mathscr{E} \cap \overline{\delta\mathscr{D}} \subset \delta\mathscr{E}$, for which see theorem 14 of de Rham (1960) together with theorem 24 and the surrounding discussion.

The rest of this appendix will discuss in detail the relation between K_2 and the homology of the topological space M. From now on we assume that $N = 3$ although everything would work for $N > 3$ and almost everything for $N = 2$.

Recall that in $\mathbb{R}^N$, d is exact in the sense that $\omega \in \mathscr{E}^p(\mathbb{R}^N)$ and $\mathrm{d}\omega = 0$ imply (unless $p = 0$) that $\omega = \mathrm{d}\Omega$ for some $\Omega \in \mathscr{E}^{p-1}(\mathbb{R}^N)$. Notice also that if, in this situation, $\Omega \in \bar{\mathscr{D}}$ then the condition that $((\Omega, \Omega))$ be minimised subject to $\mathrm{d}\Omega = \omega$ is (as is easily checked using (A.5)) that $\delta\Omega = 0$, a fact which may help to explain the proof of the following.

Lemma 1. Let $M = \mathbb{E}^3$ ($:= \mathbb{R}^3$ with Euclidean metric) and suppose that $\mathrm{d}\omega = 0$. Then
(1) $\delta\omega \in \mathscr{D} \Rightarrow \omega = \mathrm{d}\beta + \gamma$ with $\beta \in O(r^{-1})$, $\mathrm{d}\beta \in O(r^{-2})$, and $\Delta\gamma = 0$;
(2) $\delta\omega \in \mathscr{D}$ and $\omega \in \bar{\mathscr{D}} \Rightarrow \gamma = 0$ in (1);
(3) $\omega \in \mathscr{D} \Rightarrow \omega = \mathrm{d}\beta$ with $\beta \in O(r^{-2})$, $\mathrm{d}\beta \in O(r^{-3})$.

† The other common definition, $\Delta\omega = 0$, leads, in our situation, to the same notion of K_2 (theorem 26 of de Rham (1960)). Incidentally, in the present context 'harmonic' is a particularly inapt term since elements of K_2 correspond, as we have seen, precisely to those components of the field which are *not* analogous to quantum-mechanical harmonic oscillators.

Proof. (1) Suppose there are β, γ as required and suppose for a moment that $\delta\gamma = \delta\beta = 0$. Then $\Delta\beta = 0 + \delta\, \mathrm{d}\beta = \delta(\omega - \gamma) = \delta\omega$. Because in $\mathbb{E}^3$

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

We can solve this using the potential of a point charge, $G(x) = -(4\pi)^{-1}\|x\|^{-1}$:

$$\beta(x) = \int G(x-y)\delta\omega(y)\, \mathrm{d}^3 y. \tag{A.8}$$

Since the integrand vanishes for y outside the compact set $\mathrm{supp}(\delta\omega)$, (A.8) is well defined and β and $\mathrm{d}\beta$ are respectively $O(r^{-1})$ and $O(r^{-2})$ as $r \to \infty$. Moreover if $\gamma = \omega - \mathrm{d}\beta$ then (using that $\mathrm{d} \natural \Delta$ and that $\mathrm{d}\omega \in \mathrm{dd}\mathscr{E} = \{0\}$) $\Delta\gamma = \Delta\omega - \Delta\, \mathrm{d}\beta = \mathrm{d}\delta\omega - \mathrm{d}\Delta\beta = \mathrm{d}(\delta\omega - \Delta\beta) = 0$.

(2) By (1) $\mathrm{d}\beta$ has a finite $\mathscr{L}^2$-norm. If ω does as well then so will $\omega - \mathrm{d}\beta = \gamma$. But in $\mathbb{E}^3$ the only $\mathscr{L}^2$-solution of Laplace's equation is zero.

(3) Since $\mathscr{D} \subset \bar{\mathscr{D}}$ and $\delta\mathscr{D} \subset \mathscr{D}$ the conditions of (1) and (2) are fulfilled. Moreover, since ω has compact support we can switch the derivatives in (A.8) onto G, which produces the extra asymptotic power of $1/r$. □

Remark. Although we never proved (nor used in the proof) that $\delta\beta = 0$, it is in fact true since $\Delta\delta\beta = \delta\Delta\beta = \delta\delta\omega = 0$, of which the only $O(1/r)$ solution is $\delta\beta = 0$.

Corollary 1. If $\omega \in \mathscr{E}^p(\mathbb{E}^3)$, $\mathrm{d}\omega \in \mathscr{D}$ and $p > 0$ then $\exists \alpha \in \mathscr{E}^{p-1}$ and $\beta \in \overline{\mathscr{D}^p}$ such that $\omega = \beta + \mathrm{d}\alpha$.

Proof. Apply (3) of the lemma to $\Omega := \mathrm{d}\omega$ to conclude $\Omega = \mathrm{d}\beta$ with $\beta \in O(r^{-2})$. Plainly $((\beta, \beta))$ is finite so that $\beta \in \bar{\mathscr{D}}$. Moreover $\mathrm{d}(\omega - \beta) = 0 \Rightarrow \omega - \beta = \mathrm{d}\alpha$ for some $\alpha \in \mathscr{E}^{p-1}$. □

To prove the three theorems which follow we will need the last two results not only for $\mathbb{E}^3$ but also for three-manifolds which deviate slightly from perfect flatness.

Proposition. Both lemma 1 and its corollary are valid for M near enough to $\mathbb{E}^3$ in the space of asymptotically-flat three-metrics.

Proof. Our proofs used only that $G(x) \sim 1/r$ and that Laplace's equation $\Delta\omega = 0$ has no non-trivial solution which vanishes at infinity. But these seem intuitively true and in fact results† in essence the same as those we need follow from the theory elliptic operators as described by Cantor (1977).

Our aim being to relate K_2 to some purely topological invariant, let us define $H(\mathscr{E}(M), \mathrm{d})$ as the quotient space $Z/\mathrm{d}\mathscr{E}$ where $Z = \{\omega \in \mathscr{E} \mid \mathrm{d}\omega = 0\}$. (Similarly $H(\mathscr{E}, \mathrm{d}, -, p)$, for example, comprises equivalence classes of axial p forms, and would be defined as

$$\{\omega \in \mathscr{E}^{p-} \mid \mathrm{d}\omega = 0\}/\{\mathrm{d}\alpha \mid \alpha \in \mathscr{E}^{(p-1)-}\}$$

(where $\mathscr{E}^{(-1)} := \{0\}$).) Such H's are known to depend only on the topology of M in a way which we will recall later.

† For example, theorem 2.9 of Cantor (1977) implies that in any M sufficiently near to $\mathbb{E}^3$, the equation $\Delta\beta = \rho$, $\rho \in \mathscr{D}$ has a unique solution in $O(r^{-1+\epsilon})$ for each ϵ between 0 and 1.

Since $K_2 \subset Z$ the prescription '$\psi(\omega)$ = equivalence class of ω' defines a linear map $\psi: K_2 \to H(\mathscr{E}, \mathrm{d})$, which turns out to be almost an isomorphism. We will write

$$\psi(p): K_2^p \to H(\mathscr{E}, \mathrm{d}, p)$$

for the restriction of ψ to p forms and similarly $\psi(p)^{\pm}$ for its restriction to $K_2^{p\pm}$.

Theorem 1. $\psi(p)$ is surjective for $p > 0$.

Proof. Let $S \subset M$ be a sphere large enough so that, if K is the region enclosed by S then $M - K$ is nearly flat—in fact so nearly flat that one can 'fill in' the hole in $M - K$ with a nearly flat three-ball $\tilde{K}$ to produce a manifold $\tilde{M}$ to which lemma 1 and its corollary will apply. If ω is a p form on M with $\mathrm{d}\omega = 0$ (and $p > 0$) then extending arbitrarily to all of $\tilde{M}$ the restriction of ω to $M - K$ produces a p form $\tilde{\omega}$ such that $\overline{\mathrm{supp}\,(\mathrm{d}\tilde{\omega})}$ is included in the compact set $\tilde{K}$. Corollary 1 then furnishes $\tilde{\alpha} \in \mathscr{E}(\tilde{M})$ and $\tilde{\beta} \in \mathscr{D}(\tilde{M})$ such that $\tilde{\omega} = \tilde{\beta} + \mathrm{d}\tilde{\alpha}$. Let α be any form defined on all of M and agreeing with $\tilde{\alpha}$ on $M - K$ $(= \tilde{M} - \tilde{K})$ and set $\beta = \omega - \mathrm{d}\alpha$. Since $\tilde{\beta}$ is $\mathscr{L}^2$ and since $\beta = \tilde{\beta}$ on $M - K$, β is also $\mathscr{L}^2$. Appealing to the Kodaira theorem, we conclude that $\beta = \mathrm{d}\beta_1 + \gamma$ where $\gamma \in K_2$. Combining these results gives $\omega = \mathrm{d}(\alpha + \beta_1) + \gamma$, which shows that $\psi(p)(\gamma)$ is the class of ω in $H(\mathscr{E}, \mathrm{d}, p)$. □

Theorem 2. $\psi(p)$ is injective unless $p = 1$.

Proof. Suppose $p \neq 1$ and construct $\tilde{M}$ as before. Since plainly $K_2^0 = 0$ we can assume that $p \geqslant 2$. If $\psi(\omega) = 0$ then $\omega = \mathrm{d}\alpha$ for $\alpha \in \mathscr{E}^{p-1}$. Extend α to $\tilde{K}$ producing $\tilde{\alpha}$ and set $\tilde{\omega} = \mathrm{d}\tilde{\alpha}$. Since $\tilde{\omega}$ is $\mathscr{L}^2$ (because ω is) and $\delta\tilde{\omega} \in \mathscr{D}(\tilde{K})$, (2) of lemma 1 gives us $\tilde{\beta} \in O(1/r)$ for which $\mathrm{d}\tilde{\beta} = \tilde{\omega}$, assuring us in addition that $\tilde{\omega}$ (and therefore ω) is $O(1/r^2)$. Then $\mathrm{d}(\tilde{\beta} - \tilde{\alpha}) = \tilde{\omega} - \tilde{\omega} = 0$ whence (since $p \geqslant 2$) $\tilde{\beta} - \tilde{\alpha} = \mathrm{d}\tilde{\gamma}$ for some $\tilde{\gamma} \in \mathscr{E}^{p-2}(\tilde{M})$. Returning to M (extending γ any old way on K and defining $\beta = \alpha + \mathrm{d}\gamma$) we find $\beta \in (1/r)$ for which $\omega = \mathrm{d}\beta$. But then $((\omega, \omega)) = ((\omega, \mathrm{d}\beta)) = -((\delta\omega, \beta)) = ((0, \beta)) = 0$, whence $\omega = 0$. Here the integration by parts involved in the second equality is legitimate since $\omega \in O(r^{-2})$, $\beta \in O(r^{-1})$. □

Lemma 2†. Let S be any internally oriented sphere enclosing K. If M is non-orientable then there exists an axial 1-form $e \in \bar{\mathscr{D}}$ such that $\delta e = 0$ and

$$\oint_S e^k \,\mathrm{d}S_k = 1. \tag{A.9}$$

(Here, as always, K is a compact submanifold of M including the region of multiple convectivity, so that $M - K$ is diffeomorphic to $S^2 \times \mathbb{R}$.)

Proof. Notice, that according to Sorkin (1977), $\mathrm{d}S_k$ would have to be polar in order to represent an externally oriented surface element. (The question of the weight of $\mathrm{d}S$ can be ignored here since in a Riemannian manifold all weights are equivalent.) But then the integrand would not be a scalar. In order to be able to talk in familiar terms we will fix once and for all an orientation of $M - K$ chosen so that $\mathrm{d}S$ is outwardly directed relative to it.

† Neither this lemma nor most of the proof of theorem 3 are needed for the rest of the paper. They are included only for completeness and for their relevance to Sorkin (1977).

Now in the context of Sorkin (1977) or of §§ 2–4 above, we would have worked with $\mathscr{E}^k := \sqrt{\gamma}\gamma^{kj}e_j$ in place of e and interpreted (A.9) as saying that K displays unit net electric charge. From this we can see how to construct such an e. Let $\Gamma: \mathbb{R} \to M$ be a curve which comes in from and returns to infinity but only after passing through K in such a way as to reverse orientation—in other words, to reverse the handedness of any vector triad carried continuously along it. (Such a Γ exists if and only if M is non-orientable.) We can assume Γ meets S just twice, say at x_1 and x_2.

Let us interpret Γ as a 'line of electric force' of strength $\frac{1}{2}$, giving it an *external* orientation such that at x_1 it represents (relative to the chosen orientation of $M-K$) an outward flux through S. Then, by the reasoning of Sorkin (1977), it will represent an equal outward flux at x_2, giving thus a total outward flux through S of 1.

Finally we can smear Γ out into a 'flux tube' and outside K fan this tube out into a roughly spherically symmetric form looking like the $1/r^2$ Coulomb field of a unit point charge. This is our e: it is clearly $\mathscr{L}^2$ and δe (the divergence) vanishes because Γ was an *unbroken* line. $\square$

Theorem 3. Let M be an asymptotically flat Riemannian three-manifold. Then

$$\psi(p)^{\pm}: K_2^{p\pm} \to H(\mathscr{E}, \mathrm{d}, \pm, p)$$

is an isomorphism in all cases except that

(i) $K_2^0 = 0$ whereas $H(\mathscr{E}, \mathrm{d}, +, 0) \simeq \mathbb{R}$ so that $\psi(0)^+: \{0\} \to \mathbb{R}$;

(ii) when M is orientable $\psi(0)^-: \{0\} \to \mathbb{R}$;

(iii) when M is non-orientable $\psi(1)^-$ is indeed onto $H(\mathscr{E}, \mathrm{d}, -, 1)$ but has a kernel of dimension one.

Proof. According to the two previous theorems we need consider further only $\psi(p)$ for $p = 0, 1$.

Consider $p = 0$. If $\omega \in \mathscr{E}^0$ and $\mathrm{d}\omega = 0$ then ω = constant. More precisely, if ω is polar ($\omega \in \mathscr{E}^{0+}$) then $\omega = c \in \mathbb{R}$ and conversely if $\omega = c \in \mathbb{R}$ then $\mathrm{d}\omega = 0$. In other words, $Z(\mathscr{E}, \mathrm{d}, +, 0) = \mathbb{R}$. If ω is axial then given a local choice of orientation $\mathscr{O}$, ω can be considered as in $\mathscr{E}^{0+}$ and therefore $= c \in \mathbb{R}$. This locally defined and constant ω will extend consistently to all of M if ($\omega = 0$ or) $\mathscr{O}$ extends consistently to all of M, which is to say if M is orientable. Thus $Z(\mathscr{E}, \mathrm{d}, -, 0) = \mathbb{R}$ or $\{0\}$ according as M is or is not orientable. Finally if $\omega \in K_2^0$ then ω = constant whence $\omega \notin \mathscr{L}^2$ unless $\omega = 0$, confirming that $K_2^0 = \{0\}$.

Consider $p = 1$. By theorem 1 $\psi(1)$ is onto but if $\psi(\omega) = 0$ the $p = 1$ version of the proof of theorem 2 breaks down where we are to conclude that $\tilde{\beta} - \tilde{\alpha} = \mathrm{d}\tilde{\gamma}$. Nevertheless, if ω is polar then $\mathrm{d}(\tilde{\beta} - \tilde{\alpha}) = 0 \Rightarrow \tilde{\beta} = \tilde{\alpha} + c$ for $c \in \mathbb{R}$, whence $\beta := \alpha + c$ defines on M a $O(1/r)$ 0 form for which $\omega = \mathrm{d}\beta$. From here on the proof that $\omega = 0$ proceeds just as before. Even if ω is axial the proof will still work as long as M is orientable, for by orienting M we obliterate the distinction between axial and polar.

But if M is not orientable the most we can say is that by choosing an orientation for $\tilde{M}$ (or, what is the same, for $M - K$) we can regard $\tilde{\alpha} - \tilde{\beta}$ as a constant $c \in \mathbb{R}$, which, however, does not extend to an element of $Z(\mathscr{E}(M), \mathrm{d}, -, 0)$ unless $c = 0$. Call this constant $\phi(\omega)$. (In other words ϕ is defined by the relation $\phi(\mathrm{d}\alpha) = \lim_{x\to\infty} \alpha(x)$.) I claim that ϕ is a linear isomorphism of kernel $\psi(0)^-$ onto $\mathbb{R}$.

To show ϕ is well defined it suffices to show that $\phi(\omega) = 0$ if $\omega = 0$. But if $\mathrm{d}\alpha = 0$ then $\alpha \in Z(\mathscr{E}, \mathrm{d}, -, 0) = \{0\} \Rightarrow \alpha = 0 \Rightarrow \lim \alpha(x) = 0$. To show ϕ is injective suppose $\phi(\omega) =$

0. Since this means $\tilde{\alpha} = \tilde{\beta}$ it entails $\alpha \in O(1/r)$, which in turn implies $\omega = 0$ just as before. Finally, to show ϕ is onto $\mathbb{R}$ it suffices, since ϕ is into, to show that kernel ψ is not trivial. But if α is any axial 0 form which equals one in $M - K$ (relative to the chosen orientation there) and (for example) falls smoothly to zero in K then $\mathrm{d}\alpha \in \mathscr{D} \subset \bar{\mathscr{D}}$ and $\mathrm{d}(\mathrm{d}\alpha) = 0$, whence, from Kodaira, $\mathrm{d}\alpha = \gamma + \omega$ with $\omega \in K_2$, $\gamma \in \overline{\mathrm{d}\mathscr{D}}$. Now if e and S are as in lemma 2 then

$$((\mathrm{d}\alpha, e)) = \int_M (\partial_j \alpha) e^j \, \mathrm{d}V$$

$$= \oint_S \alpha e^j \, \mathrm{d}S_j - \int_M \alpha \delta e \, \mathrm{d}V$$

$$= \oint e^j \, \mathrm{d}S_j = 1.$$

On the other hand if $\beta \in \mathscr{D}$ and S includes support $(\beta) \cup K$ then $((\mathrm{d}\beta, e)) = \oint_S \beta\, e^j \, \mathrm{d}S_j = 0$, whence if $\beta_n \in \mathscr{D}$ is any sequence such that $\mathrm{d}\beta_n \to \gamma$ in the $\mathscr{L}^2$ norm then $((\gamma, e)) = ((\lim \mathrm{d}\beta_n, e)) = \lim((\mathrm{d}\beta_n, e)) = 0$. Comparing these results shows $\gamma \neq \mathrm{d}\alpha$, i.e. $\omega \neq 0$. But $\psi(\omega) = 0$ because $\gamma \in \bigcap \mathrm{d}\mathscr{D} \subset \mathrm{d}\mathscr{E} \Rightarrow \omega \in \mathrm{d}\mathscr{E}$. □

In concluding this appendix, let us recall from de Rham (1960) some basic isomorphisms subsisting among the vector spaces $H(\,,\,,\,)$ and connecting these spaces to the topology of M†. Even without reference to de Rham (1960) it is clear that the * operator induces dualities

$$H(\mathscr{E}[\text{respectively } \mathscr{D}], \mathrm{d}, \pm, p) \simeq H(\mathscr{E}[\text{respectively } \mathscr{D}], \delta, \mp, N-p) \tag{A.10}$$

$$K_2^{p\pm} \simeq K_2^{(N-p)\mp}. \tag{A.11}$$

The analogous relations corresponding to duality induced by $((\,,\,))$ are

$$H(\mathscr{E}[\mathscr{D}], \mathrm{d}, \pm, p) \simeq H(\mathscr{D}[\mathscr{E}], \delta, \pm, p)^* \tag{A.12}$$

$$K_2^{p\pm} \simeq (K_2^{p\pm})^* \tag{A.13}$$

where 'V^*' means the vector space dual of V, $\mathscr{L}(V, \mathbb{R})$.

Finally, each of the spaces $H(A, B, \pm, p)$ where $A = \mathscr{E}$ or $\mathscr{D}$, $B = \mathrm{d}$ or δ, and $0 \leqslant p \leqslant N$ can be equated to an appropriate homology or cohomology group as follows:

$\delta \leftrightarrow$ homology;

$\mathrm{d} \leftrightarrow$ cohomology;

$\mathscr{D} \leftrightarrow$ finite chains (or co-chains of compact support);

$\mathscr{E} \leftrightarrow$ infinite (but locally finite) chains (or arbitrary co-chains);

$+ \leftrightarrow$ real coefficients;

$- \leftrightarrow$ 'axial coefficients in $\mathbb{R}$' ('chaine paire' of de Rham (1960)).

In particular

$$H(\mathscr{D}, \delta, +, p) \simeq H_p(M, \mathbb{R}),$$

the ordinary p-dimensional homology group of M with real coefficients.

The above relations, along with theorem 3, allow us to equate any one of our $H(\ )$'s to one of the topological invariants $H_p(M, \mathbb{R})$). In particular, we can verify the

† But beware of de Rham's terminology. He *sometimes but not always* uses the words *pair* and *impair* oppositely to how others would use *polar* (or *true*) and *axial* (or *pseudo*).

isomorphisms referred to in §§ 4 and 5:

$$H(\mathscr{E}, \delta, -, 1)^* \simeq H(\mathscr{D}, \mathrm{d}, -, 1) \simeq H(\mathscr{D}, \delta, +, 2) \simeq H_2(M, \mathbb{R}) \tag{A.14}$$

$$K_2^{1-} \simeq K_2^{2+} \simeq H(\mathscr{E}, \mathrm{d}, +, 2) \simeq H(\mathscr{D}, \delta, +, 2)^* \simeq H_2(M, \mathbb{R})^* \tag{A.15}$$

$$K_2^{2-} \simeq (K_2^{2-})^* \simeq (K_2^{1+})^* \simeq H(\mathscr{E}, \mathrm{d}, +, 1)^* \simeq H(\mathscr{D}, \delta, +, 1) \simeq H_1(M, \mathbb{R}). \tag{A.16}$$

Finally the 'axial homology' or 'homology with externally oriented chains' which is relevant to magnetic superselection rules corresponds, by the above rules, to $H(\mathscr{D}, \delta, -, 2)$ and thus to $H(\mathscr{E}, \mathrm{d}, -, 2)^* \simeq (K_2^{2-})^*$. Comparing with (A.16)

$$H(\mathscr{D}, \delta, -, 2) \simeq H_1(M, \mathbb{R}). \tag{A.17}$$

References

Cantor M 1977 *Comm. Math. Phys.* **57** 83–96
Mandelstam S 1962 *Ann. Phys., NY* **19** 1–24
de Rham G 1960 *Varietes Differentiables* (Paris: Hermann)
Sorkin R 1977 *J. Phys. A: Math. Gen.* **10** 717–25