Charge quantization, the topology and 3-dimensionality of the universe, and abolishing monopoles

Warren D. Smith *
WDSmith@fastmail.fm

14 Jan., 2000

Abstract —

Why do all electrons have the same charge and why do all protons have exactly the opposite charge? Dirac provided a possible answer by proposing the existence of magnetic monopoles.

We propose a simple possible topological explanation of why charge is quantized, involving a tiny permanent magnetic field trapped in the topology of the universe. The idea apparently works for every possible compact 3-manifold topology for the universe except for the 3-sphere S^3 and elliptic 3-geometry. This picture does not need to assume magnetic monopoles exist, and indeed looks incompatible with their existence.

Which topologies for the universe are consistent (or not) with laws of physics? We present an argument that all orientable 3-manifolds should be consistent with a very wide class of possible laws of physics; but almost all n-manifolds for each $n \neq 3$ won't be. This is perhaps a "reason the world is 3 dimensional."

Next we argue that if point monopoles exist, and Dirac's wave equation of quantum mechanics in static (scalar and vector) potentials holds, then a contradiction arises. Hence monopoles cannot exist – or: they are not points; or: Dirac's equation for quantum mechanics is not correct. Meanwhile non-point monopoles lead to other unpalatabilities or contradictions.

Hence there are reasons to prefer our topological explanation of charge quantization to Dirac's monopole hypothesis.

(In the event that non-point monopoles do exist, we show how, by very elementary reasoning, to deduce certain of their properties, e.g. their moment of inertia.)

Keywords — Monopoles, Dirac, topology of the universe, dimensionality of the universe, combing hair on manifolds, charge quantization, Hopf fibration, moment of inertia.

*Temple University Math. Dept. Wachman Hall, 6th floor, 1805 North Broad Street Philadelphia, PA 19122.

1 Review of previous arguments about magnetic monopoles and charge quantization

No magnetic monopoles have ever been found, despite extensive experimental searches [5][11][25][26][1]. There are nevertheless three reasons why many people believe (or hope) that monopoles exist:

1. The quantization of charge would be explained (§1.2)[2].

[1] Cabrera [11] found a single monopole candidate event with his superconducting magnetic loop detector in 1982. It was consistent with a Dirac monopole (EQ 4) with $Z = 1$. However, this observation has never been reproduced despite running much bigger and better detectors for longer, and Schwinger's theoretical arguments [73] suggest that Z must be a multiple of 4. Hence this candidate event is regarded as dubious. A possible future monopole search might be conducted with huge superconducting loops placed somewhere cold in the outer solar system.

[2] "Kaluza-Klein theories" [43][45][70][7][86][21] postulate extra "compact" dimensions of space. The simplest K-K theory involves one extra 5th "cylindrical" dimension, which has the same circumference everywhere (of order the Planck length). The electromagnetic potential 4-vector A^μ is actually the 5th row of the metric tensor $g_{\alpha\beta}$ (except that we impose the demand that g_{55} be a constant).

Supposedly, in K-K theory the Einstein equations obeyed by $g_{\alpha\beta}$, under certain assumptions about "how things look" in the 4D "projection" when we "average over microscopic details in the 5th dimension x^5" and assume $g_{\alpha\beta}$ does not depend on x^5, yield the usual 3+1D Einstein gravity equations as well as the Maxwell electromagnetism equations, at least in vacuum. Electric charge in K-K theory is actually momentum in the 5th dimension. Clockwise motion around the extra dimension corresponds to positive charge; counterclockwise to negative. Charge reversal and parity reversal are the same thing (and it seems to be difficult or impossible to incorporate parity violation into a K-K theory). Charge conservation arises from conservation of momentum. Angular momentum conditions lead to charge quantization automatically [45]. My brief examination of K-K theory suggests that it is quite complicated and the fundamental claims above have not been justified rigorously, nor has a careful examination of the fundamental assumptions behind K-K theory been made (e.g.: which ones are really necessary?).

Certain GUTs also yield charge quantization automatically; thus on p.430 of [13]: "whenever the unification gauge group is simple, charge quantization will follow... because the eigenvalues of a simple non-Abelian group are discrete while those corresponding to the Abelian $U(1)$ group are continuous. For example in $SO(3)$ the eigenvalues of the third component of angular momentum can take on only integer or half integer values while... [for] $U(1)$ symmetry of translation invariance in time... no restriction..." Although I do not understand

2. Maxwell's equations would get a new symmetry (§1.1).

3. The simplest Grand Unified Theory [29] predicted the existence [68][80][13] of monopoles of mass $\gtrsim 10^{16}\text{GeV}/c^2 \approx 2\times 10^{-11}$kg. (Actually, 't Hooft [80] had claimed the mass would only be at most about $7\text{TeV}/c^2$, but his calculation was silently ignored by all later authors, who said 10^{16}-$10^{17}\text{GeV}/c^2$.) After SU(5) GUT was refuted by proton lifetime measurements [9][3], other "supersymmetric GUT" theories and "superstring" theories were devised which continue to predict monopoles, although now of larger masses $\gtrsim 10^{17}\text{GeV}/c^2$. These GUT monopoles are predicted to have at least *twice* Dirac's minimal magnetic charge value (EQ 4).

In the present paper, we are going to show how to explain charge quantization without any need for monopoles. We are going to argue for this new explanation of charge quantization, and against monopoles:

Our new explanation removes rationale #1. We'll show that monopole existence would lead to contradictions in present physical laws – undercutting rationale #3 to the extent that present day physical theories are regarded as correct. Of course if they are regarded as overruled by GUTs (at least in appropriate regimes) my objections too are presumably overruled – but still interesting[4].

1.1 Maxwell's equations

Maxwell's equations feature some remarkable symmetries (e.g. they are invariant under Lorentz transformations, as well as a much larger, albeit lesser known, continuous group of symmetries discovered by Cunningham [16] in 1910). But they are not symmetric under the interchange of electricity and magnetism. By introducing magnetic monopoles we gain such a symmetry[5].

this, note that it is demonstrated by a logical pathway that does not ever involve a monopole. That gives one the impression that it might be possible to explain charge quantization without either our picture, or monopoles. *But*, Georgi and Glashow [30] clarify that magnetic monopoles exist "in gauge theories based on simple groups." Therefore, that impression was misleading. If we ignore GUTs and restrict ourselves to presently well accepted physical theories, then Dirac's and mine are the only explanations I know of.

[3] According to the 1999 summary of particle data by the Particle Data Group, the mean proton lifetime for decay into a neutral pion and positron is $> 5.5\times 10^{32}$ years with 90% confidence. Ongoing measurements by the super-Kamiokande detector should increase this bound by an order of magnitude; indeed an unsubstantiated claim is made on the super-K web page that the proton lifetime is now known to exceed 10^{33} years.

[4] In that case my objections at least serve to focus attention on how major consequences can follow from as yet unknown and speculative physical laws in regimes of the very small. Also, the possibility and perhaps desirability of abolishing monopoles will now give more freedom to the creators of GUTs.

[5] This gain comes at the cost of invalidating the "PT" symmetry arising from space inversion and time reversal [71] – but the "CPT" symmetry, also incorporating charge reversal for *both* magnetic and electric charges, would remain valid. The gain is greater than the loss, because actually we also would gain a *continuous* symmetry allowing not only the *interchange* (with appropriate unit conversion and sign changes: $\vec{E}\to c\vec{B}$, $c\vec{B}\to -\vec{E}$, $c\vec{J}_e\to -\vec{J}_m$, $\vec{J}_m\to c\vec{J}_e$) of magnetic and electric fields and charges, but in fact any 2D rotation $\vec{E}\to \vec{E}\cos\theta + c\vec{B}\sin\theta$, $c\vec{B}\to c\vec{B}\cos\theta - \vec{E}\sin\theta$, $cq\to cq\cos\theta - g\sin\theta$, $g\to cq\sin\theta + g\cos\theta$ among them (the interchange arising as the special case when the rotation angle θ is $\theta=\pi/2$) [40].

Define the "SIpole," a new SI unit (pending approval!) of magnetic charge, as follows: "One SIpole is the amount of magnetic charge that would experience a force of 1 Newton if placed in a 1 Tesla magnetic field." (Note, 1 Coulomb is the amount of electrical charge that would experience a force of 1 Newton in an electrical field of 1 volt/meter. 1 SIpole= 1meter×Coulomb/second.) The Maxwell equations with monopoles then are

$$\vec{\nabla}\cdot\vec{E} = \rho_e/\epsilon_0,\ \ \vec{\nabla}\cdot\vec{B} = \rho_m\mu_0, \qquad (1)$$
$$\vec{\nabla}\times\vec{B} - c^{-2}\dot{\vec{E}} = \mu_0\vec{J}_e,\ \ \vec{\nabla}\times\vec{E} + \dot{\vec{B}} = \mu_0\vec{J}_m.$$

Here $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields, $\vec{J}_e$ and $\vec{J}_m$ are the current densities for electric and magnetic charges respectively, ρ_e and ρ_m are the charge densities, $\mu_0 = 4\pi\times 10^{-7}$Henry/meter is the permittivity of free space, and $\epsilon_0\mu_0 = c^{-2}$ where $c = 299792458$meter/second is the speed of light. The Lorentz force experienced by an electric charge q moving with velocity $\vec{v}$ is then $\vec{E}q + \vec{v}\times\vec{B}q$. The neo-Lorentz force on a *magnetic* charge g is $\vec{B}g - c^{-2}\vec{v}\times\vec{E}g$.

1.2 Charge quantization

The argument that the existence of even one magnetic monopole anywhere in the universe would imply (to keep physics self consistent) the quantization of both electric and magnetic charges (thus explaining the facts that all electrons have the same charge, and that protons and electrons have exactly opposite charges) is due to P.A.M.Dirac [19]. There are three ways to look at this argument.

1. An integration first performed by J.J.Thomson [82] shows that the angular momentum

$$\vec{A} = \frac{1}{c\mu_0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \vec{r}\times\vec{E}\times\vec{B}\,\mathrm{d}^3\vec{r} \qquad (2)$$

arising from the magnetic field $\vec{B} = \mu_0 g\vec{r}/(4\pi r^3)$ of a point monopole g and the electric field $\vec{E} = q\vec{r}/(4\pi\epsilon_0 r^3)$ of a point charge q, is $|\vec{A}| = qg\mu_0/(4\pi)$, where $\vec{A}$ points from the charge toward the monopole, *regardless* of the distance between them. If one assumes that this angular momentum is $Z\hbar$ where Z is an integer, we conclude that

$$qg = \frac{h}{\mu_0}Z. \qquad (3)$$

Assuming the electric charge quantum is $e\approx 1.602\times 10^{-19}$ coulomb (i.e. the electron charge)[6], the quantum of magnetic charge is

$$g = \frac{h}{\mu_0 e}Z \approx 3.29106\times 10^{-9}Z \quad \text{SIpole}. \qquad (4)$$

[6] Actually, quarks have charges that are integer multiples of $e/3$. This would cause the quantum of magnetic charge to be 3 times larger. However, it is generally conjectured to be impossible to isolate a free quark (or anything of nonzero "color" or subintegral charge) more than at most $.86\times 10^{-15}$meter away from other quarks which neutralize these anomalies.

2. During a flyby of a charge q past a fixed monopole g the charge will be pulled by Lorentz forces in the manner of a "twist," in such a way that the total angular momentum it acquires is $2qg\mu_0$, *regardless* of the charge's velocity, its mass, and the minimum distance between the two [13]. Assuming this increment in angular momentum is quantized in units of $\hbar$ again leads to (EQ 3).
3. Consider [19][13] the Dirac wavefunction of an electron in the presence of a point monopole. Dirac's quantum relativistic equation is actually inapplicable in this situation because it[7] involves a "vector potential."

However, Dirac argued that a smooth vector potential *can* be created that will work everywhere *except* on one point on each sphere that encloses the monopole. His approach may be thought of as regarding the monopole as one end of a very long and very thin solenoid, and taking the limit as this solenoid becomes infinitely long and thin (Dirac calls the resulting mathematical object a "string"). The vector potential is singular on the curve describing this solenoid. Dirac then wanted the charge not to "see" the solenoid and wanted the wavefunction to be independent (except for symmetries which do not matter) of the solenoid's location. He argued this and an assumption of continuity would force the "Aharonov Bohm phase shift" arising for a trajectory enclosing the string, to be an integer multiple of 2π; this led to (EQ 3). The precise argument is given in [13]; for the analogous argument in the present paper's alternative topology-based charge quantization explanation, see §**??**.

Later Schwinger [73] attempted to formulate a theory of (or at least, some constraints on a theory of) relativistic quantum mechanics which would work in the presence of monopoles. In order to remove "arbitrariness in physical predictions" in relativistically invariant "photon exchange between different source types," he deduced that Z in (EQ 4) must be a multiple of 4.

2 Our alternative: Charge quantization from topologically trapped magnetism

It is possible to make a similar argument, but without any need to assume a monopole exists, by assuming a "topologically trapped" magnetic field in the universe. I will prepare the ground by describing easy-to-understand special cases of my idea in §2.1-2.4, then give a fully-rigorous and general description in §2.5.

Throughout this paper, when we speak of "topology of the universe" we shall mean 3D spatial topology, regarded as a 3-manifold. **Definition:** *In this paper "n-manifold" means "smooth compact*[8] *connected boundaryless Riemannian n-manifold."* Until §5, We will mostly ignore the fact that, in general relativity, space and time are in fact inseparably entangled to form a $3+1$D Lorentzian manifold. As a partial compromise, we will allow thinking of the Riemannian 3-manifold as slowly varying with time.

2.1 The argument in a 3-Torus universe

A simple reification of our idea is a 3D "flat torus" universe $T^3 = S^1 \times S^1 \times S^1$, i.e. a Euclidean box with sides L_1, L_2, L_3 with "wraparound" periodic boundary conditions. Assume there is a constant magnetic field $\vec{B}$ in the L_3 direction. (Of course, any *constant* $\vec{B}$-field is an exact solution of the vacuum Maxwell equations in this 3-torus.)

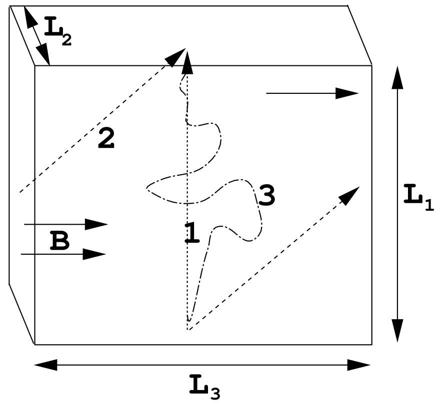

Figure 1: Rectangular 3-Torus universe with constant trapped magnetic field $\vec{B}$. Consider a charge q moving on trajectory 1 around the universe (and returning to its starting point). Other possible trajectories are 2 and 3.

Assume a charge q moves along some trajectory up through the upper box face (coming back again through the bottom face) eventually returning to the same location it started. The total momentum transferred to the charge by Lorentz forces is qBL_1 in the L_2 (transverse) direction.

[7] If the $\vec{B}$ field is allowed to have nonzero divergence, no such potential exists. The experimental confirmation of the "Aharonov-Bohm effect" tells us that a vector potential really is needed in quantum mechanics, so we have no clue how to proceed without it.

[8] Throughout this paper we'll assume a compact (i.e. spatially closed & finite) universe. It is unknown if the universe actually *is* spatially closed & finite. By "closed" here, we do not mean the same thing as cosmologists mean. we mean what mathematicians mean. Cosmologists mean: rebounding to a "big crunch." We do not care whether the universe crunches or expands forever. What we care about is whether the universe is finite in spatial extent at any particular time (e.g.: the 3D surface S^3 of a 4D sphere, or a 3-torus $S^1 \times S^1 \times S^1$, or a finite-volume hyperbolic 3-manifold) versus whether it is infinite spatially at any particular time (e.g.: euclidean 3-space R^3, or 3D hyperbolic space H^3, or an "infinite cylinder" $S^1 \times R^2$). It is presently totally unknown what the topology of the universe is – the best that can be said [49][76] is there are unreliable statistical indications disfavoring universes with positive overall curvature (e.g. S^3). The question of whether the universe is spatially compact is totally unknown.

Other, bizarre, possibilities are also thinkable: boundaryless 3-manifolds of finite volume but nevertheless noncompact [1], i.e. having unboundedly large distances! I'll now demonstrate how to construct such universes. Consider a 4D convex polyhedron all of whose vertices lie on the 3D surface of a ball in R^4. (The simplest example: a 4-simplex, the convex hull of 5 generic points on a sphere in R^4.) The 3D surface of such a polyhedron is – if the interior of the ball is now regarded (via the "projective," also called "gnomonic," model of hyperbolic space) as hyperbolic 4-space – a hyperbolic (i.e. negative constant curvature) 3-manifold of finite 3-volume. But as you approach the vertices of the polyhedron, you get infinitely far from everything else in the hyperbolic metric. (Note, the "edges" of the polyhedron are invisible to the intrinsic surface metric.) The whole theory of the present paper probably would allow many noncompact manifolds (including these) too, but I have not examined that question.

The important thing to notice is that this transverse momentum formula holds *regardless* of the charge's mass, its velocity, and its trajectory $\vec{x}(t)$ (provided it has "unit winding number;" otherwise we have to multiply by an appropriate integer).

Hence with respect to a fixed observer in the center of the box, the charge has acquired *angular* momentum increment $qBL_1^2/2$. If this is quantized in steps of $\hbar$, then q must be quantized in steps of

$$2\hbar/(BL_1^2). \qquad \text{(WRONG)} \tag{5}$$

We have deduced charge quantization.

Actually, though, it is wrong to rely on angular momentum. Angular momentum is not a well defined concept in a compact, not rotationally symmetric universe. For example its sign depends on "which way the observer is looking." Also, if the observer looks "several times round the universe" he can artificially lengthen the moment arm and get an infinite number of different angular momentum values.

It is better to rely on the fact that momentum itself has a quantal property in a compact universe, namely there have to be an integer number of De Broglie wavelengths $\lambda = h/p$ on a suitable path round the universe in the direction of that momentum $\vec{p}$. With $p = qBL_1$, we find that $L_2 p$ must be h times an integer, therefore *charge must be quantized in steps of*

$$\frac{h}{BL_1L_2} = \frac{h}{\Phi} \tag{6}$$

where $\Phi = BL_1L_2$ is the total magnetic flux "round the universe." Note, if the quantum of charge is assumed to be the electron's charge e, and $L_1 = L_2 = 10^{10}$ lightyears, the the flux round the universe must be an integer multiple of the "flux quantum" $\Phi_0 = h/e \approx 4.136 \times 10^{-15}$Weber and the trapped magnetic field of the universe must be (an integer multiple of) $B = h/(eL_1L_2) \approx 4.6 \times 10^{-67}$ Tesla. Thus, an undetectably small B-field suffices to do the job. (Cf. [88] for the best known observational upper bound 3×10^{-15} Tesla.)

Note that there is no need for the universe to be a rectangular box – any parallelipiped would do.

Now suppose this universe expands, i.e. L_1, L_2, and/or L_3 increase. Are we then going to get a quantum of charge that depends on the time-varying size of the universe? (Bad!) No.

As the L_k increase, B must decrease in such a way that $\Phi = BL_1L_2$ remains constant. This is because the number of "lines" of magnetic flux must be fixed – lines cannot be created or destroyed since they must be closed loops and can have no endpoints in the absence of magnetic monopoles. (Each "line" represents a fixed infinitesimal amount of magnetic flux.) More precisely: electric currents *can* (and will) create new closed B-field loops, but such loops must be homotopic to the identity, i.e. contractible to nothing. Uncontractible loops such as ours are "trapped" and can't be created or destroyed. Hence the charge quantum (EQ 6) would not be affected.

One could similarly argue that were our torus universe "wider" in some places than in others, BL_1L_2 would still remain invariant no matter which cross section one were on. (Anyhow, it would suffice logically for just *one* closed trajectory to work, since we could consider moving a charge along a path P until reaching the approved trajectory, using it, then moving back along P^{-1}.)

Indeed we are going to show in §2.5 that the same argument will work in *any* 3-torus, not just these "standard flat" ones – and also in other topologies besides tori.

2.2 Some physical remarks

Physically, one would certainly expect some topologically trapped magnetism in any universe in which it is topologically possible – i.e. it would be surprising to have none.

My trapped $\vec{B}$-field hypothesis seems *incompatible* with Dirac's hypothesis monopoles exist, since a monopole could keep accelerating along the $\vec{B}$-field loop forever, sucking up energy forever. The "potential energy" associated with the $\vec{B}$-field would be nonconservative. It seems that a gas of monopoles in such a universe could keep getting hotter and hotter forever, with conservation of energy being disobeyed. To avoid this, we must abolish monopoles. (Presumably the universe is free of trapped $\vec{E}$-field for the same reason.)

The total energy of a single flux quantum B-field trapped in our $L_1 \times L_2 \times L_3$ model universe is

$$E_{\rm mag} = \frac{B^2 L_1 L_2 L_3}{2\mu_0} = \frac{h^2 L_3}{2\mu_0 e^2 L_1 L_2}. \tag{7}$$

For comparison, the total energy of a single photon with wavelength L is hc/L. Observe that if $L_1 = L_2 = L_3 = L$, then the former energy is $h/(2\mu_0 c) = 4/\alpha \approx 34.25$ times the latter, i.e., of the same order. This holds regardless of the size of the universe. This perhaps provides a somewhat better explanation of "why the trapped field is there" than merely "because zero is an infinitely unlikely real number." Namely: suppose the universe was created with approximately 1 photon worth of energy in each low-frequency EM mode. Then if we regard constant B-field as the lowest frequency "mode" we get the right order of magnitude answer.

2.3 Analogue of Dirac's "string" argument

We have seen that trying to argue about angular momentum runs into difficulties. Trying to argue about *ordinary* momentum works for the flat 3-torus universe, but gets tricky when one tries to consider general curved 3-manifolds. So then I was pulled toward Dirac's own notion (cf. #3 of §1.2) that the best kind of argument depends on (1) the continuity of physics – under infinitesimal perturbations, observable effects should change infinitesimally – and (2) the Aharonov-Bohm phase-angle shift.

This Dirac-like idea does work and seems the clearest and best way to proceeed, *but* even here there are some traps to snare the unwary. E.g., in the 3-torus example we just discussed, one can explain the $\vec{B}$-field as being generated by a "vector potential" $\vec{A}$ with $\vec{\nabla} \times \vec{A} = \vec{B}$. Thus the constant horizontal field $\vec{B} = (1, 0, 0)$ could be thought of as due to a vector potential $\vec{A} = (0, 0, y)$. Note that this $\vec{A}$ field necessarily has some discontinuities if we consider it "wrapped" into the flat 3-torus, e.g. we could make it discontinuous on the plane $y = L_1$ (which by vertical "wraparound" is the same as the plane $y = 0$) but this $\vec{A}$ works everywhere else. One might then imagine making some kind of charge-quantization argument by considering electron trajectories which cross or go near this discontinuity-plane. But this approach *does not work*! There are several reasons. First, $\vec{A} = (0, -z, 0)$ *also* works to explain $\vec{B} = (1, 0, 0) = \vec{\nabla} \times \vec{A}$. This has an entirely different natural discontinuity plane ($z = L_1$ or $z = 0$). It seems as though one can always use gauge freedom in $\vec{A}$ to make it stay continuous in any neighborhood of whatever electron trajectory one is considering. Second, $\vec{A}$ *is* continuous and well behaved everywhere in the *covering space* (here $\mathbf{R}^3$) of our manifold. These facts seem to prevent making any argument about discontinuities or singularities in $\vec{A}$.

The *right* and *maximally simple* approach, which avoids all that mess, is to base everything on the following:

$\vec{A}$-free formulation of the Aharonov-Bohm phase shift principle: Suppose an electron travels along a closed path ∂D that may be regarded as the boundary of a topological disk D. I.e., D is homeomorphic to the set $\{(x, y)$ such that $x^2 + y^2 \leq 1\}$, and ∂D is homeomorphic to $\{(x, y)$ such that $x^2 + y^2 = 1\}$. Then the electron's wave function will be multiplied by a complex phase $e^{i\theta}$ where $\theta = e\Phi/\hbar$, where e is the electron charge and Φ is the magnetic flux "through the loop," i.e. crossing D.

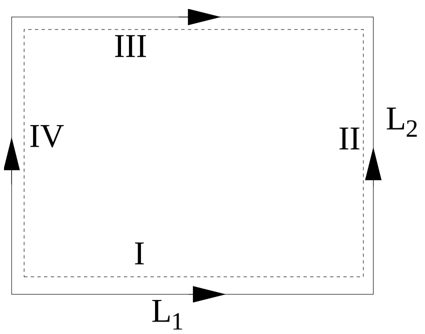

Figure 2: Rectangular 2-Torus; magnetic field $\vec{B}$ is perpendicular to the plane of the figure. Consider the electron trajectory shown dashed. Regard it as infinitesimally close to the boundary of the rectangle.

The point is that this avoids employing any vector potential $\vec{A}$ as in the usual expression $\theta = (e/\hbar) \int \vec{A} \cdot d\vec{\ell}$.

Now in our flat 3-torus universe example, simply consider the electron trajectory in figure 2. Obviously, this trajectory encloses a topological disk passing flux $\Phi = L_1 L_2 B$. Therefore, the Aharonov-Bohm phase shift angle is infinitesimally close to $\theta = e\Phi/\hbar$. *But*, also obviously, this trajectory should have Aharonov-Bohm phase shift angle infinitesimally close to *zero*, because of our assumption of the continuity of physics, and the fact that any phase shift on part I of the trajectory should be exactly canceled by the phase shift experienced on the (infinitesimally nearby, and opposite) part III; and similarly II and IV should cancel. The only way to resolve this contradiction is to demand that θ must be an integer multiple of 2π. This yields our usual quantization condition EQ 6.

2.4 *But the argument does not work in a 3-sphere universe*

It is possible to make a nontrivial topologically trapped constant $\vec{B}$-field on a 3-sphere S^3 (I.e. a nonzero exact time-independent solution of Maxwell's vacuum equations). There is a topologically unique way to do this, called the "Hopf fibration" (described below). However, it turns out that the total Lorentz momentum increment experienced by an electron on a closed trajectory in this universe, is always *zero* – and no interesting deductions can be made by considering the Aharonov-Bohm phase shift and continuity – so neither approach to deducing charge quantization works.

The Hopf fibration ([85] p.103-108; [37][32]) is best described using "quaternions." Quaternions are like complex numbers except they are 4D instead of 2D because there are 3 square roots of -1 (called i,j,k) not just one root i. Arithmetic: $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$, $ijk = i^2 = j^2 = k^2 = -1$. Quaternion addition is an abelian group, and multiplication of nonzero quaternions is a noncommutative group (indeed $\{i, j, k, 1\}$ either commute or anticommute), and the distributive law holds. Quaternions $Q = a + bi + cj + dk$ (a,b,c,d, real) have "norm" $|Q|^2 \equiv a^2 + b^2 + c^2 + d^2$ and norms are multiplicative. Multiplication by a fixed unit-norm quaternion performs a rigid rotation on 4-space.

We now are ready to describe the Hopf fibration on the 3-sphere $|Q| = 1$.

If you rotate a 4-vector by rotating in the $(1, i)$ plane by i (90 degrees) *and* rotate in the (j, k) plane by i (also 90 degrees), then the resulting vector is orthogonal to the original vector, no matter what that original vector was. (In contrast it is impossible to devise a fixed 3D rotation with this property, since any 3D rotation leaves its "pole" unaffected. Indeed, this is impossible in all odd dimensions by an eigenspace argument, but plainly possible in even dimensions.) Thus if Q is a unit-norm quaternion, then iQ is also, and $(iQ) \cdot Q = 0$ where $\cdot$ denotes the ordinary real 4-vector inner product.

Geometrically this means iQ is a continuous unit 4-vector field tangent to the 3-sphere $|Q| = 1$. I.e. you can "comb the hairs on a 3-sphere." This combing is called the "Hopf fibration[9]." It has some remarkable properties, e.g. every flowline of the vector field is a great circle geodesic; any two such circles are linked; any two of the circles are "parallel" in the sense that they are at a constant distance from one another. It is the *unique*[10]fibering of S^3 into S^1's in the sense that any other one on any other S^3 is diffeomorphic to the Hopf fibration of the standard round S^3.

[9] Of course, we could have used vQ instead of iQ, where v is any unit-norm quaternion with zero real part.

[10] This uniqueness may be shown with the aid of the fact that any fibration of S^3 into S^1's must have base space S^2 (for S^3 to be simply connected, the base must be also, hence by Poincare's proof of the 2D version of the generalized Poincare conjecture, it must be S^2) and the "classification theorem for fiber bundles" mentioned in section 24.4 of [20]. But there are smooth nowhere zero flows on the 3-sphere with *no* periodic orbits [48].

In contrast, you cannot comb the hairs on a basketball (2-sphere) [66][23][59][11]; but you can of course on a 1-sphere.

Hence if the universe were a 3-sphere and had a magnetic field along the Hopf fibration, this magnetic field would be "topologically trapped" and unable to go away. This $\vec{B}$-field also is an exact solution of the time independent source free Maxwell equations, indeed essentially the unique one [36][32], so it will stay there.

Unfortunately, if a charge q moves around a great circle geodesic of the universe, say the equator in the $(1, i)$ plane, then the $\vec{B}$-field will be everywhere tangent to the trajectory, and hence the charge will acquire momentum increment 0. In other cases such as the great circle in the $(1, j)$ plane the $\vec{B}$-field will be everywhere orthogonal to the trajectory, *but* rotating around it in such a way that the net effect is still zero. The fundamental reason for all the zeros is that every closed trajectory is contractible.

Yet another remarkable property of the Hopf fibration (viewed as a magnetic field) is that it is (up to a constant factor) its *own* vector potential, i.e. its curl is *itself.*

All of the above argument concerning S^3 and the Hopf fibration also applies to 3D elliptic geometry (i.e., S^3 with antipodal points identified).

2.5 Mathematically rigorous generalization to all topologies

Theorem 1 (Trapped static $\vec{B}$-fields solving Maxwell equations exist) *Every 3-manifold with uncontractible closed loops has some topologically trapped magnetic field $\vec{B}$ solving the time-independent vacuum Maxwell equations.*

Proof. Clearly, in any 3-manifold with an uncontractible closed loop one may construct *some* topologically trapped $\vec{B}$-field: simply use a "bundle" of fluxlines going around some smooth uncontractible loop, and zero elsewhere. (We can even make this field smooth, without difficulty, e.g. consider $|\vec{B}| = \exp(1/[d^2 - \epsilon^2])$ where d is the distance to the uncontractible loop, $0 < d < \epsilon$, and ϵ is sufficiently small.) Assume from now on that the total amount of flux going round said loop, is normalized to 1.

Now among all such fields $\vec{B}$, consider[12] the one with *minimum energy* $\int |\vec{B}|^2 d^3$volume.

Any $\vec{B}$ obeying the time independent source free Maxwell equations $\vec{\nabla} \times \vec{B} = \vec{0}$ and $\vec{\nabla} \cdot \vec{B} = 0$, is the gradient (at least locally) of some "potential" field Φ obeying Laplace's equation $\nabla^2 \Phi = 0$. But minimizing energy $\int |\vec{\nabla}\Phi|^2 d^3$volume yields solutions of Laplace's equation. QED.

Theorem 2 (Charge quantization) *In any 3-manifold containing uncontractible closed loops, there are connected, but non-simply-connected, 2D submanifolds S. Under the assumption that Aharonov-Bohm phase shifts vary infinitesimally if the electron's trajectory is varied infinitesimally, we conclude that electron charge must be an integer multiple of h/Φ_S (or equivalently if the electron charge is regarded as fixed, that Φ_S must be an integer multiple of the flux quantum Φ_0).*

Proof. Consider an electron trajectory corresponding to the boundary ∂D of a topological disk D consisting of all S except for an infinitesimal amount. We find that the amount Φ_S of magnetic flux passing through S must be an integer number of flux quanta Φ_0, implying that charge must be quantized in steps of h/Φ_S. (Should all be obvious after following the flat torus argument in §**??**[13]). QED.

2.5.1 Some remarks

Remark. *Poincare's conjecture* (one of the biggest open problems in topology) states that a compact simply connected 3-manifold is homeomorphic to the 3-sphere.

Generalizations of the Poincare conjecture to n-manifolds have been proven for $n = 2$ (H.Poincare), $n \geq 5$ (S.Smale, E.C.Zeeman 1961), and $n = 4$ (M.Freedman 1981), but the original $n = 3$ problem remains open[14].

However, "nonsimply connected" is *not* the critical criterion for our argument to work (contrary to my original impression). Instead, what is critical is the condition that the 3-manifold contain nonseparating hypersurfaces. The two conditions are the same for connected 2-manifolds, but are different on connected 3-manifolds.

[11] It often is claimed incorrectly that any smooth vector field on a 2-sphere must have at least *two* zeros. Wrong: one suffices. Consider the lines parallel to the x-axis on the plane; place unit vectors parallel to those lines; scale them by $\exp(-x^2 - y^2)$; and stereographically project them up onto the sphere.

[12] Assuming such a minimum exists should be adequate for physicists. Mathematicians will be quick to observe that this argument, used by Riemann in the mid-1800s and called by him the "Dirichlet principle," is not rigorous because this existence assumption needs to be justified [61]. However, fortunately, for smooth maps from a compact n-manifold to a manifold with nonpositive curvature (in particular flat space $\mathbf{R}^3$) the fact that we may always deform the map into a harmonic, i.e. energy minimizing, map, has been justified [22] by the "heat flow method."

[13] Incidentally, Stokes's theorem does *not* necessarily hold (at least, without modification) if the region enclosed by the loop is *not* topologically equivalent to a disk $x^2 + y^2 < 1$ in $\mathbf{R}^2$. (Mentioned, e.g., in [60] box 4.1 page 96, 4c, and is the reason for the use of "star shaped domains" in [55].) As a simple counterexample, consider, instead of the disk $\{x, y, z | x^2 + y^2 < 1,\ z = 0\}$, the toroid

$$z^2 + \left(x - \frac{(1-\epsilon)x}{\sqrt{x^2+y^2}}\right)^2 + \left(y - \frac{(1-\epsilon)y}{\sqrt{x^2+y^2}}\right)^2 = \epsilon^2$$

where the case $\epsilon \to 0^+$ makes the invalidity of Stokes's theorem completely obvious. (A referee had used Stokes's theorem in this wrong way to try to come to the [ridiculous] conclusion that the magnetic flux around a 3-torus "must" be zero, thus "invalidating" the present paper!)

[14] Prof. Everett Pitcher, former secretary of the Amer. Math. Soc., claims to have a proof of Poincare's conjecture. He gave a lecture on it at Lehigh University, Bethlehem PA, on 16 October 2002, and he submitted a 40-page paper on it to Trans. of the AMS. According to the *New York Times* 15 April 2003, page F3, Grigori Perelman of the Steklov Math. Institute, St.Petersburg, claims to have a partly-unpublished proof of Thurston's geometrization conjecture (which is more general than Poincare) and lectured on it at M.I.T. [65]. I was not at either lecture and I do not know if their proofs are correct. A $\$10^6$ prize is being offered by the Clay Math. Institute for a proof or disproof.

Then every thinkable compact 3D universe containing topologically trapped magnetism must have quantized charge – except for the so-called "rational homology spheres" (which are precisely the connected 3-manifolds which do not contain nonseparating hypersurfaces. I had originally thought (assuming Poincare's conjecture) that only S^3 would be excluded, but the rational homology spheres are a superset of S^3.

Remark. "Conformal[15] transformations" preserve solutions of Maxwell's equations. Thus a conformal map applied to the electron's trajectory will yield in the new transformed universe, a new trajectory, still with the same Aharonov-Bohm phase shift angle, and a new magnetic field, still a solution of the time-independent Maxwell equations. Thus, at least as far as conformal transformations of the metric are concerned, the charge quantum in a universe is a *topological invariant* of that universe. Schoen [72] showed that *every* Riemannian n-manifold may be conformally transformed to one with constant curvature. There is good reason to believe [75] that our universe is, essentially, a 3-manifold with constant scalar curvature.

Remark. Note, it would be possible to have *two* homotopically inequivalent uncontractible loops (in some topologies, such as a two-holed torus) and this would yield 2 different charge quantization conditions, both of which would have to be satisfied simultaneously by any charge in that universe. (Actually, to be more precise, the n-torus has n homotopically inequivalent uncontractible loops, but $n-1$ of them do not matter for us because we are only interested in the one which points in the direction of the magnetic field. More-holed tori have more homotopically inequivalent uncontractible loops.) If the two quantum step values had an irrational ratio (which, generically, they would), this should make nonzero charges impossible in that universe. This strongly suggests that our universe does *not* have two homotopically inequivalent uncontractible loops, which is a very strong constraint on the topology of the universe[16].

2.6 Which topologies are compatible with laws of physics?

2.6.1 Weak interaction

Both the unidirectionality of time, and the fact that the weak interaction violates parity symmetry, seem to rule out universes which are *non-orientable* manifolds, such as the Cartesian product of a circle with 2D "elliptic geometry." (Any particles which had in the past traveled "round the universe" would be chirally reversed. Such particles have not been observed. It perhaps would be worthwhile to study weak interactions of extragalactic cosmic rays to see if any are chirally reversed.)

2.6.2 The combing of hairs

A necessary condition for our charge quantization argument was that it be possible to "comb the hair" in the 3-manifold representing our universe in some manner homotopically inequivalent to doing nothing, i.e. so that the $\vec{B}$-field can't go away.

A survey of foliations is [28]; for vector fields on manifolds see [46][81]. Also possibly of interest is [64].

It is possible to comb the hairs on a closed 2-manifold (more precisely, foliate it smoothly into 1-manifolds)[17] exactly when it has Euler characteristic $\chi = 0$, i.e., for the torus and the Klein bottle only. This was shown by Hopf [38] in 1927. See figure 3. Thurston [84] found an extension to n dimensions.

Every smooth open [35] or closed 3-manifold, orientable [51] or not [92], has a smooth[18] 2-foliation. (By "smooth" we mean we may demand C^∞.) Furthermore [92][83], any smooth plane field is homotopic to a 2-foliation, so that any smooth vector field (perpendicular to the plane field) is homotopic to one transverse to a 2-foliation. **Fact:** Consequently, you can always "comb the hairs" on any smooth closed 3-manifold.

Hence, the hair combing condition is hardly restrictive. But now let us consider a more interesting condition.

Definition 3 *An n-manifold will be called "m-combable" if on it, there can exist m orthonormal smooth vector fields.*

Remark. In order for an n-manifold to be m-combable for any particular m, $1 \le m \le n$, it suffices for it to have m everywhere *linearly independent* nowhere zero smooth vector fields, because by taking appropriate linear combinations (with coefficients which are appropriate smooth functions of position) we may orthonormalize them.

Observation. Certainly, the standard 3-torus and the standard 3-sphere are 3-combable. **Proof:** let the 3-torus be any parallelipiped with periodic boundary conditions. Now simply use vectors parallel to the axes of an (arbitrarily rotated) 3D Cartesian coordinate system. Let the 3-sphere be the unit norm quaternions $|Q| = 1$, with some arbitrary pre-rotation of the 4D coordinate system. Now use as the 3 mutually orthogonal unit-length vector fields iQ, jQ, and kQ. QED.

Remark. The n-sphere is n-combable[19] exactly

[15] Two smooth manifolds are said to be "conformally equivalent" if, under some coordinatizations, their metric tensors g_{ab} are proportional with a smooth positive multiplicative pure-scalar function as the proportionality factor.

[16] See [74][85] for a survey of Thurston's geometrization conjecture which (if true) would classify the topologies of all compact orientable 3-manifolds. (Thurston's conjecture now has allegedly been proven by Perelman [65], see footnote 14.) I have not worked through this classification trying to determine which 3-manifolds are compatible with this constraint, but certainly very few are, and perhaps the 3-torus is the only one.

[17] An n-manifold Ω will be said to have a "k-foliation" for some $k < n$ if it is a disjoint union of immersed k-dimensional submanifolds. If each point of Ω lies in a ball diffeomorphic to a ball in $\mathbf{R}^n$ with the leaves of the foliation intersecting that ball being diffeomorphic to the standard k-dimensional sets with x_j constant for each $j > k$, then the foliation is "smooth."

[18] However, for the 3-sphere it is not possible to find an *analytic* 2-foliation [34].

[19] n-combable n-manifolds are often called "parallelizable" in the literature. Some at least as ill-advised words, which have sometimes also been employed, to describe the maximum m such that a manifold in m-combable, are its "span" and "rank." The present paper turns out not to be the first physics paper using notions of n-combability; see [31].

[3][44][58][81] when $n = 1, 3, 7$, which is related to the existence and uniqueness of the complexes, quaternions, and octonions. [Indeed by a construction analogous to Hopf's one may construct 3 orthonormal vector fields on S^n if $4|(n+1)$ by using quaternions, and 7 orthonormal vector fields on S^n if $8|(n+1)$ by using octonions.] Let $f(n)$ be the maximum possible number of linearly independent smooth nowhere zero vector fields on a sphere S^n. Then $f(n) = 2^c + 8d - 1$ where $n + 1 = (2a + 1)2^b$ and $b = c + 4d$ and $0 \le c \le 3$. Note $f(n) = 0$ if n is even. For odd n, we have:

```
  n =1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33
f(n)=1 3 1 7 1  3  1  8  1  3  1  7  1  3  1  9  1
```

The lower bounds are due to constructions by A.Hurwitz [39] and J.Radon [69] who actually showed that there are $f(n)$ orthogonal $(n+1)\times(n+1)$ matrices A_p with $A_p^2 = -I$ and $A_pA_q + A_pA_q = 0$ for $p \neq q$. Then the matching upper bound was proven by J.F.Adams [2] using K-theory, a branch of cohomology theory.

Do any other 3-tori and 3-spheres (besides these particularly geometrically nice ones) work? Yes. Any 3-torus or 3-sphere conformally equivalent to the standard ones is 3-combable. Indeed certain diffeomorphisms more general than conformal ones will work[20]: it will suffice for n-combability if we can recoordinatize the n-manifold so that its metric tensor $g_{\alpha\beta}$ becomes diagonal. The "Cotton-Darboux theorem" [12][15] shows that *every* 3-manifold is so recoordinatizable, at least in finite patches locally.

Is there going to be any obstacle preventing extending this patch to cover the whole manifold (provided we are transforming to a compatible topology)? There is no obstacle: every 3-manifold has an everywhere orthogonal curvilinear coordinate system[21]:

Theorem 4 (n-combability) *All orientable 3-manifolds are 3-combable. The only 2-combable 2-manifolds are 2-tori. For $n \ge 4$, generically, n-manifolds are not n-combable (even locally) although certain specific special n-manifolds can be (e.g. flat n-tori $S^1 \times S^1 \times \cdots \times S^1$).*

Proof. As we already mentioned Hopf's [38] and Thurston's [85] theorems show that the 2-manifolds are combable precisely if they have $\chi = 0$, i.e. are a torus or Klein bottle. (Smoothly foliatable into 1D manifolds is the same as having a smooth nowhere zero vector field, because you can use the flowlines of the vector field, which by uniqueness and existence theorems for ordinary differential equations, works.) Now, on *orientable* 2-manifolds, combable implies 2-combable; simply take the second vector field to be orthogonal to the first (rotated 90°). But in contrast, even the usual flat Klein bottle (in figure 3) is *not* 2-combable. Proof: The "2-combing" shown in figure 3 is invalid because the vertical vector field would have to flip direction 180° discontinuously, at the right edge of the figure. If the two vector fields, near the right edge of the figure, were oriented at any other angles besides normal and parallel to that edge, then they would both discontinuously change direction by some nonzero angle as we crossed the edge.

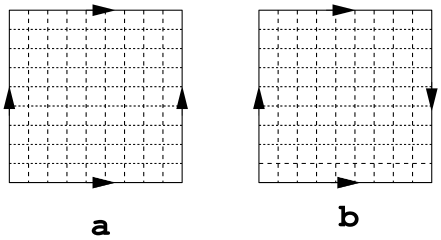

Figure 3: 2-combing of 2-Torus (a) and failed 2-combing for Klein bottle (b). In both figures a valid 1-combing arises by considering only the horizontal dashed lines.

n-Combability is generically false, even locally[22] when $n \ge 4$ due to insufficient "gauge freedom." Specifically, it is necessary that the metric tensor $g_{\alpha\beta}$ be diagonal (in some coordinatization) and there are n degrees of gauge freedom you have in selecting your favorite coordinatization, which is enough to make it diagonal if $n \le 3$, but not if $n \ge 4$.

The result that every orientable 3-manifold is 3-combable was shown by Stiefel [77]. QED.[23]

2.7 Why is the world 3-dimensional?

2.7.1 3-manifolds are compatible with photons

The connection of theorem 4 to physics is as follows.

Suppose we regard it as desirable for "momentum-polarization eigenstates" of photons to exist in our universe. In flat space these are just plane waves – with equal wavevector $\vec{k}$ everywhere, and polarization vectors $\vec{E}$ and $\vec{B}$ existing everywhere and everywhere perpendicular both to each other and to $\vec{k}$.

These properties in curved space are possible only if the universe's topology is 3-combable! Note, incidentally, that 3-combability is necessary if a momentum-polarization eigenstate "photon" (with any desired momentum direction and 2 orthogonal polarization directions at any particular one point) is to exist. (If one multiplies the 3 orthonormal vector fields by an appropriate fixed orthogonal 3×3 matrix, one can make these 3 directions be whatever one wants at any particular point.)

[20] It has often been claimed incorrectly that a Riemannian metric is conformally flat if and only if its Weyl tensor vanishes. This is true if the dimension $n \ge 4$, but when $n \le 3$, the Weyl tensor automatically vanishes but not all 3-manifolds are conformally flat, not even locally [47] (see also [60] p.550 ex.21.22.). All 2-manifolds are locally conformally flat, and so are the standard round n-spheres (by considering the stereographic projection map).

[21] The reader might want to look at the discussions of orthogonal curvilinear coordinate systems in $\mathbf{R}^3$ found in [57][62][63].

[22] A different way to try to prove this would be simply to fall back on the Kervaire-Milnor-Adams results for the n-sphere to prove impossibility in every dimension $n \ge 4$ except for $n = 7$. However, this would only yield a *global* obstacle – the standard round n-sphere is n-combable locally (for each $n \ge 1$) everywhere except at a single point, as may be seen by stereographic projection mapping the surface of that sphere to flat n-space. Presumably, however, the surface of a generic *ellipsoid* in $\mathbf{R}^{n+1}$ is not n-combable even locally.

[23] Could it be that such topological constraints could also be employed to restrict the possible ways in which "extra dimensions are curled up" in "string theories?"

Remark. Indeed, arguably n-combable n-manifolds are globally compatible with *every* physical law we can formulate in flat n-space, since we can make an everywhere orthogonal curvilinear coordinate system for such a manifold, and the physical laws will be able to operate locally everywhere as though they were in a flat space coordinate system. However, n-manifolds not meeting such a condition, although they would be *locally* compatible with flat space physical laws, might have some global incompatibility – as in the photon eigenstate example above.

We now see that theorem 4 represents a remarkable property of the dimension "3." Perhaps we can even brand it as a "reason the world is 3-dimensional."

Summary. We have identified a topological property: n-combability. We've shown: "Random" n-manifolds are n-combable, and hence compatible (with "probability 1") with a wide class of thinkable physical laws, if $n = 3$. However, if $n \geq 2$ but $n \neq 3$, then random n-manifolds will *not* be n-combable (with "probability 1") and hence would yield a self-contradiction in a wide class of thinkable physical laws. Thus, "with probability 1," physics can only happen[24] in 3D.

Further, I point out that it is known that the notion that a Lorentzian $(3 + 1)$-manifold is "globally hyperbolic" is known to be equivalent to the claim that it can be foliated into spacelike 3-manifolds. Thus, globally hyperbolic $(n + 1)$D spacetimes can be $(n + 1)$-combed with n of the combings being spacelike and 1 timelike everywhere. That however is not the case for $(n + 1)$-manifolds for $n \neq 3$ and $n \geq 2$.

2.7.2 Other arguments the world "must" be 3D

I'll quickly summarize previous arguments [79][10][78] for why the universe "had" to be 3D.

Assuming Maxwell's and Schrödinger's (or Dirac's) equations hold, hydrogen has no bound states if $n \geq 4$ (atoms implode, indeed with infinite energy release for pointlike electrons and nuclei), but has no unbound states if $n \leq 2$. Assuming Newtonian mechanics with r^{-n} force laws hold, planetary orbits are unstable if $n \geq 4$ (no orbit which oscillates between a maximum r_2 and a minimum radius value r_1 can exist with $0 < r_1 < r_2$; the weaker result that circular orbits are unstable if $n > 3$ is in the undergraduate textbook [33]), but with $n \leq 2$ there are no unbound orbits since the escape velocity is infinite. Attempts to generalize the general relativistic Schwarzschild solution to $n \geq 4$ space dimensions exhibit the same behavior. Lovelock [53] pointed out that if $n+1 \leq 3+1$ there is a unique Einsteinian gravity field equation of order ≤ 2 arising from a scalar Lagrange density depending only on the metric tensor and its first two derivatives. If $n \geq 4$, this uniqueness statement no longer holds. Assuming general relativity holds, gravity cannot exist in vacuum if $n \leq 2$ since the 4-indexed Riemann tensor would be everywhere zero in vacuum. Mariwalla [56] showed that the TCP theorem for Dirac's equation would only hold if n were odd; and if we demand the TCP operation be continuously connected to the Lorentz group identity, that forces $n \geq 3$; he also says, "An arbitrary geodesic in a curved [Riemannian] space... can be looked upon as the motion of a particle under a force in flat spacetime, and vice versa" only if $n + 1 = 4$.

Lovelock [54] pointed out that there are various "dimension dependent identities" in differential geometry that hold only in certain dimensions. In particular, Rainich's identity that underlies "geometrodynamics" (unification of electromagnetism and gravity into a single geometric theory) works in 4D only. The Weyl conformal curvature tensor vanishes in dimensions ≤ 3 only. Misner, Thorne, and Wheeler [60] point out remarkable properties of the Einstein tensor that hold only in 4D. All of classical electromagnetism is based on the properties of the curl and div operators, many of which (since they arise from 4D quaternion algebra) hold only in 3 dimensions.

2.7.3 Some dubious arguments

Philosophers have claimed that since nerves cannot cross in 2D, intelligent observers would be impossible in a 2D universe. (The point being that, while perhaps there are universes of all dimensions, if intelligent observers can only exist in 3D universes, it is no surprise that our universe is 3D.) But that would seem to be refuted by Conway's construction [8] of "life," a Turing universal computer in a simple 2-state 2D cellular automaton. An earlier and more complicated Turing universal cellular automaton (29 states at each vertex of a 2D square grid) had been constructed by von Neumann [89]. It was, however, simpler than "life" in the sense that communication only occurred between a site and its 4 nearest neighbors on a square grid, rather than its 8 nearest. Also: optical, audible, or time multiplexed signals *can* cross through each other.

Another similar biological-philosophical idea: organisms cannot have both a mouth and an anus without being disconnected. But this too seems refuted: there are sea organisms which use the same opening as both a mouth and an anus.

Fluid turbulence could not happen in 2D, which perhaps somehow would mitigate against the evolution of life, but this is hardly convincing.

2.7.4 Comparison with my, more abstract, argument

All of these previous arguments that the world "has to be" 3-dimensional suffer from the flaw that they assume the physical laws in an n-dimensional universe would be the same as here. (Perhaps the inhabitants of a 7D universe are busily arguing that 7 was the only possibility for the same reason.) They also all are aimed at demonstrating the nonexistence of intelligent observers, rather than the logical inconsistency of physics itself – although the arguments for

[24] Recently "string theories" and "M-brane theory" have been proposed [67], which claim the universe is $9 + 1$ (or $10 + 1$) dimensional. Our argument does *not* rule these out, because they postulate a *fixed, rigid* metric, namely flat $(3+1)$-space cartesian producted with some particular fixed compact 6D (or 7D) manifold M (for which, there are presently at least about 10^5 viable candidates). By choosing M to be an atypical, n-combable metric such as the flat n-torus $(S^1)^n$, our argument, which only applies to *generic* metrics, is defeated.

atomic implosion with infinite energy release for the collapse of two oppositely charged point particles, may qualify as an exception.

All the known arguments that 2D universes featuring intelligent life are impossible leave me unconvinced.

It perhaps would be more credible and satisfying if one could argue at a highly abstract level, allowing any of a large class of kinds of physical laws but without specifying what they must be, and deriving a logical contradiction, rather than merely arguing for the unlikelihood of intelligent life. Aside from the atomic implosion arguments, my argument is the only such argument that I know of.

Nevertheless, my argument also is not completely satisfying in 2D since it would still permit a 2D torus universe, although all other compact 2-manifolds would be forbidden.

3 Skeptical Reexamination of Monopoles

In this section I'll consider a sequence of attacks on the monopole idea. The initial attacks (§3.1-3.2) can be defended against, and we present the defenses. The later attacks (§3.3-3.7) which, e.g., use Dirac's own equation against him, seem harder to refute unless the defenders rely on new speculative theories of physics in ways that have never been analyzed.

3.1 Curved space

Dirac's 3 arguments in §1.2 all pertained assuming a Euclidean geometry, i.e. "flat space." It is not immediately clear any of them will work in curved spaces. However, one could argue that Maxwell's equations force charge (and magnetic charge) conservation[25] and hence we could consider transporting an electric and magnetic charge to a flat region of space far from everything else, applying Dirac's arguments there to see the charges must be quantized, and then transporting them back. This works (at least, to high accuracy) assuming there is anyplace big and flat to go to.

3.2 Point charges versus distributed charges

At first glance, Dirac's arguments seem to work only if the elementary charges are *points*. And indeed, all experimental evidence so far is fully consistent with electrons and quarks being point particles, and QED and QCD would have to undergo massive revision, wholly changing and greatly complicating their mathematical structure, if non-point electrons and quarks were allowed[26].

But ignore that. Suppose the monopole's magnetic charge is distributed over a region of nonzero volume. Then Thomson's integral (EQ 2) would no longer give a result independent of the separation of the two charges. The charge-monopole flyby argument would yield a change in the electron's angular momentum now dependent on the electron's "impact parameter" (distance of closest approach) if the electron's trajectory were allowed to penetrate the monopole. And Dirac's argument would seem to be totally inapplicable because the divergence of $\vec{B}$ would occur throughout a region of nonzero volume, causing Dirac's wave equation to be inapplicable everywhere in that region, and with no way to get rid of the problem by having a vector potential singular only on a string of measure zero. So in all three cases, the charge quantization argument seems to break down in the presence of non-point charges.

But at second glance, it is again possible to rescue all three arguments. Thomson's integral *would* be independent of the separation asymptotically for *large* separations. Hence charges and monopoles far from one another would be quantized according to (EQ 3). We then again use charge conservation and consider moving them closer together. As we move them closer together, conservation of angular momentum would force the monopole to start spinning. Similarly the flyby argument is rescued by the fact that the monopole will acquire spin if the electron trajectory penetrates it.

Finally, Dirac's "string" argument still applies for spheres completely enclosing the monopole (the continuity of the wavefunction forces charge quantization). Assuming Dirac's wave equation holds in this exterior region and some as yet unknown equation holds in the interior region, suffices to force (EQ 3).

3.3 There cannot be both a North pole and South pole (both points) in a universe in which Dirac's equation holds

Suppose that both a North monopole, and a South monopole, both point particles, exist[27]. Assume they both are minimally charged (have $Z = 1$ in EQ 4) since the below argument will only work better otherwise. Transport them to a region of flat space far from everything else.

Now assume Dirac's wave equation governs their wavefunctions. Due to the symmetry of Maxwell's revised equations, we may treat these two poles as electric charges of the same masses, *but* with the value of the fine structure constant $\alpha = e^2/(4\pi\epsilon_0\hbar c) \approx 1/137.036$ changed [19] to $\alpha_{\text{new}} \approx 137.036/4 \approx 34.259$.

Now it is well known that Dirac's wave equation for positronium is *unstable* if $\alpha > 4$. (The full set of exact solutions of the Dirac equation is available in this case.) That is, there will be a nonzero probability amplitude that, during any time interval, the two particles will "fall down the hole" into each other's singularities. The Hamiltonian operator is not self-adjoint in this case.

Note, if $\alpha < 1/N$, then the hydrogenic atom (with infinitely massive point nucleus of charge Ne) is stable and such "falling in" is impossible.

Also, Lieb and collaborators [52], showed that under various kinds of "relativistic Schrödinger equations," if $N\alpha < 2/\pi$ and $\alpha < 1/94$, then "matter" (*any* configuration of electrons amidst massive point nuclei of maximum charge Ne)

[25] And so do the Einstein-Maxwell equations in curved space.

[26] Feynman diagrams would no longer have points as vertices and path integration over many extra degrees of freedom, perhaps an infinite number, would be required.

[27] Indeed, if the universe has a compact topology, it is topologically *necessary* for the universe to have exact charge (and magnetic charge) neutrality. This may be shown using the generalization of Gauss's divergence theorem to manifolds.

is "thermodynamically stable" i.e. obeys a lower bound on its energy, proportional to the number of charges in the system. On the other hand if $\alpha > 2.72$, matter (even if subintegral nuclear charges are allowed, provided there are enough nuclei) is always unstable and will implode. In later papers of these authors, externally imposed magnetic fields and spin-field interactions were allowed, although the bounds became weaker.

The conclusion we draw from all these results is that the effective value for α for monopole "positronium" is far larger than the bounds of Lieb et al., i.e. the two monopoles *would* fall into one another.

Two point charges falling into each other's singularity under Dirac's wave equation would produce an *infinite* energy release. This is not the same thing as positron-electron annihilation, which releases only $2m_ec^2$ energy. There will always be a finite nonzero probability current "down the hole", leading, no matter how incredibly small this current is, to an infinite average rate of energy production.

If we regard it as forbidden for there to be a positive probability that the universe will self-destruct at any time, or forbidden for there to be non-self-adjoint Hamiltonians, then we conclude that either

1. Monopoles are not point particles, or
2. There are not both North and South poles in the universe[28], or
3. Dirac's wave equation in flat space is not applicable (and the revised equation exhibits different behavior)[29].

3.4 Another reason point monopoles cannot exist

The potential between a point monopole and a point magnetic dipole (such as a spinning electron) would behave proportionally to $-r^{-2}$ for separation r. As is well known (as we've just discussed, even $-Kr^{-1}$ potentials are forbidden, if K is a large enough constant), this is too severe a singularity for quantum mechanics to be defined [50]. Again the electron would "fall in" with infinite energy release in finite time[30].

3.5 Another monopole problem pointed out by Weinberg

Steven Weinberg (section VIII of [90]) after an examination, on very fundamental grounds, of quantum field theories involving photons, charges, and monopoles, concluded that

1. Monopoles must, under either P, or T, or C, change into their antiparticles.
2. It is impossible for a monopole to carry electric charge, i.e. a so-called "dyon" is impossible[31].
3. Any quantum field theory of photons, charges, and monopoles must be acausal and violate local Lorentz invariance.

The former two constraints, while rather peculiar, perhaps are acceptable. But the latter seems an extremely serious criticism rendering monopoles unacceptable. Although Weinberg never explicitly mentioned the issue, I presume he had in mind *point* monopoles and that his analysis does not apply to non-points; this is the escape hatch.

3.6 Another monopole problem pointed out by Fronsdal

Christian Fronsdal [27] examined quantum field theories not only in flat spacetime, but also allowed the possibility of a small cosmical constant Λ (perturbing the metric to become a "de Sitter space") He claimed

> If physics is stable with respect to a class of perturbations of the spacetime metric, including that of "small" constant four-dimensional curvature, then it may be shown that (1) left-handed and right-handed neutrinos are distinguished by a superselection rule; (2) magnetic monopoles cannot exist; (3) the conformal symmetry associated with the field equations for massless particles with spin 0, 1/2, and 1 is spontaneously broken except in the case of neutrinos with fixed chirality.

More precisely, Fronsdal's claim (2) [his "theorem II"] was that magnetic monopoles could not "coexist with electric charges." This was because the "field associated with a magnetic monopole source describes a state that is not in the domain of the Hamiltonian." Fronsdal's program of investigating the consequences of assuming "a principle of continuity with respect to the [constant] curvature" on the laws of physics were investigated in a series of 5 papers [27]. As is noted in paper V especially, *flat* spacetime is actually

[28]Note: Gauss's divergence theorem shows that, if the universe is a compact manifold, then it must contain an equal number of + and − charges and an equal number of North and South monopoles; i.e. it is exactly neutral. So in this case we could conclude that a compact universe cannot contain a monopole.

[29]Note: when I speak here of "Dirac's equation" I mean his equation for the wavefunction of an electron in a static 4-potential field. This equation plainly is self-consistent in certain situations, e.g. the "hydrogen atom" (electron in a Coulomb field) where a full set of well-behaved exact solutions are known [17]. However, Dirac's equation plainly is *not* self-consistent in certain other situations, such as with point monopoles, where singularity and infinite energy release occur in finite time. It could be argued that perhaps this is because Dirac's 1-electron equation alone is not the full physics (although even if so, the fact that the very equation Dirac used to argue for monopoles, now is being used against him, still carries considerable clout). Indeed, gravitational and "higher order Feynman diagram" effects might well become important at very small length scales (although it is unclear what effect they have on stability). It is very difficult to make rigorous statements about quantum field theory, so it is presently impossible to say whether it it mathematically self-consistent either with or without monopoles. (And at present, attempts to combine gravity with QFT are completely inadequate.)

[30]The "exact solution" [87] of the Dirac equation in a magnetic monopole and Coulomb potential is not applicable here since that solution pertained to a spinless (i.e. unrealistic) "monopole" having no dipole moment, i.e. having only r^{-1}, and not r^{-2}, potential singularities.

[31]Weinberg's impossibility theorem in no way prevented prominent later authors [42] from showing the existence of dyons in non-Abelian gauge theories, and indeed even in Weinberg's own electroweak theory [14]; in no case did the later authors mention this contradiction.

a degenerate case from the point of view of axiomatic formulations of quantum field theories. Fronsdal's arguments do not seem to be using any assumption that monopoles are points, in which case they must be taken extremely seriously. However, the monopole community has ignored them.

3.7 What if monopoles are not points?

It would be easy just to dismiss the idea of non-point monopoles since they would overthrow QED and QCD.

However, there are physicists who feel that monopoles do exist, and they are not points. (As we saw in §3.2, this is logically conceivable.) They are motivated by GUT and superstring theories. If we now analyze non-point monopoles – but using comparatively pedestrian physics and without using any speculative GUT and superstring theories; we will see that some rather disagreeable conclusions follow.

A monopole with distributed charge is necessarily held together by some other (non-electromagnetic) force. This force is strong enough to prevent the monopole from ever breaking apart, no matter how hard you whack it (since otherwise we would violate magnetic charge quantization and/or conservation). The weak and strong forces are inadequate to this task[32].

Of course, the monopole could be held together by some yet unknown "fifth force." That is essentially what is going on in the GUTs. So suppose monopoles really are both indivisible, and distributed. Then the distribution cannot have compact support because that would violate the special relativistic prohibition of rigid objects[33]. Now severe logical/physical problems arise when, e.g., a monopole flies by a black hole. A small part P of the monopole then must go inside the black hole. It therefore follows from indivisibility that the entire monopole must fall into the hole – no matter how small P was, i.e. no matter how far away the monopole's "planned" central trajectory was from ever penetrating the event horizon of the hole, and no matter how fast the monopole was flying! Are we to conclude, then, that monopoles can only exist inside black holes (in which case, in some sense, they do not exist, or at least, are inseparable)? Are we then further to conclude that (since a monopole presumably would extend slightly inside both of *two* black holes) that there can only be one black hole in the universe (contrary to experimental fact)?

Now ignore the "fifth force" and only allow the 4 known forces. The gravitational force is the only one which seems adequate – but only if the monopole is a black hole: A well known theorem [12] in classical general relativity states that it is impossible to pull a black hole apart, i.e. divide its event horizon into two. But this would require that the monopole actually *be* a black hole, i.e. have an event horizon. The well known Kerr-Newman exact solution [12] of charged black holes, nowadays thought to be unique, does not have an event horizon unless the mass, in Planck units, exceed the charge (and actually, in the presence of rotation, there is an even more restrictive condition). So the mass m_m of a Dirac-minimal magnetic pole must be at least

$$m_m \geq \frac{2m_P}{\sqrt{\alpha}} \approx 23.41 m_P \approx .51\text{milligram} \tag{8}$$

(i.e.[34] $m_m c^2 \geq 4.6\times 10^{10}$Joules) where $m_P = \sqrt{\hbar c/G}$ is the "Planck mass." .5 milligrams is a macroscopic-scale mass.

With a mass this large, gravity at Earth's surface would exert 5 millinewtons. To counteract this would require a huge magnetic field of 1550 Tesla. (Larger charged monopoles, such as Schwinger's, would have correspondingly larger masses, so the critical $\vec{B}$ field would remain unchanged.) We conclude that any thermalized monopoles this massive on our planet must be resting near its center and attempts to find them in accessible iron ore deposits, moon rocks, etc. [25][26][5] are doomed to failure. (Asteroids, which contain magnetic material and have low gravity, might be a better place to look for accessible monopoles.)

The Schwarzschild radius of a mass m is $2Gm/c^2$, which for this minimal mass would be 7.6×10^{-34} meters. *This is 19000 times larger* than the "Bohr radius"

$$r_{\text{mono Bohr}} = \frac{2\hbar}{\alpha_{\text{new}} m_m c} = \frac{8\hbar\alpha}{m_m c} \tag{9}$$

for monopole positronium, which with the mass value in (EQ 8) would be 4.0×10^{-38} meters. (The discrepancy would be even larger for Schwinger's 4× or 12× larger charged monopoles.)[35] This suggests that *any two opposite monopoles could combine into a magnetically neutral state impossible to pull apart.* This suggests in turn that it is impossible to create a separated monopole pair out of pure energy (even granting the feasibility of creating energies of this ridiculously large magnitude), which again suggests that monopoles cannot exist, since they could never have been created. (Dirac's charge quantization arguments in §1.2 don't work for non-point monopoles only available as inseparable neutral pairs.)[36]

[32] Actually, the strong force is supposed to allow unbreakable attractions between quarks in the sense that, if you pull two quarks apart further than some critical distance, a new quark/anti-quark pair will be created by the "elastic strain energy" and one member of the pair will attach itself to each of your two original quarks to neutralize them. However, the quark pair formation mechanism seems irrelevant to preventing single monopole breakup.

[33] Pull on one side of a "maximally stretched" monopole. The support of its distribution will stretch further.

[34] Actually, if the monopole is assumed to have angular momentum $\hbar/2$, its mass would have to be slightly larger even than this bound, which was computed under the assumption of a nonrotating monopole. Also, Schwinger's minimal monopoles with 4 times (or 12 times, if we view $e/3$ as the minimal electric charge, not e) larger magnetic charges, would have 4 or 12 times larger minimal masses.

[35] Note, we have used a factor of 2 times the usual Bohr radius formula because we have a "positronium-like" system in mind here, rather than an "infinitely massive nucleus." The alert reader should justifiably object that Bohr's radius formula may not be applicable here since it was derived nonrelativistically, while here the monopoles have so large charges that relativistic effects should be important. However, special relativistic effects would in fact *decrease* the Bohr radius by multiplying it by $\sqrt{1-v^2/c^2}$, where v is the orbital speed, i.e. my argument only becomes more true.

[36] Datta [18] invented related, but different (and also valid) arguments to argue for the impossibility of creating a monopole North-South pair.

In conclusion, both point monopoles and non-point monopoles lead to unpalatable contradictions and self-defeating conclusions. To accept them, one must have considerable faith in the ability of highly speculative and/or future theories of physics to resolve those contradictions.

So our alternative hypothesis (§2.1), which can explain charge quantization without need for monopoles, therefore might be preferable.

4 The moment of inertia of a distributed monopole – a relation between its charge and mass distributions

A distributed monopole must be a nonrigid object, since rigid objects are impossible in special relativity (the speed of sound would have to exceed c). Hence it must have a spectrum of excited modes of vibration. Since it is impossible to pull the monopole apart without violating magnetic charge quantization, this spectrum must be infinite.

There will also be excitations arising just from spinning the monopole with nonzero angular momentum quantum numbers.

It might be thought impossible to reason about the *unknown* forces holding the monopole together (or even about gravity). Remarkably, we can, by using very elementary physical arguments.

Consider shooting a (pointlike) electron, initially far away from a monopole, through its center, out the other side, and eventually ending up far away in the other direction. By consideration of the electric and magnetic fields before and after (and considering Thomson's integral EQ 2 and conservation of angular momentum), we see that this process will increase the angular momentum of the monopole by $Z\hbar$.

An alternative process which will accomplish the same thing would be to shoot in a positron from one side and an electron from the other and let them annihilate inside the monopole (see figure 4).

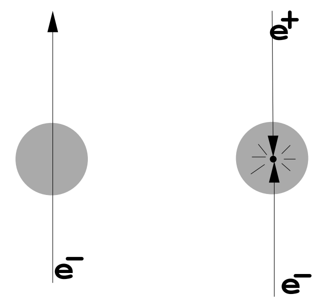

Figure 4: Two alternative processes for spinning up non-point monopoles.

Suppose this process (or the alternative one) is applied n times. We will then have an (excited) monopole with angular momentum $n\hbar$. What will its energy be?

Well, we can deduce the energy of the monopole by deducing the energy required to push the electron along its trajectory. The electron will always move parallel to the $\vec{B}$ field and hence will not experience any magnetic Lorentz forces. However, if the monopole is rotating, it will have an electric dipole moment. This will cause an electric field directly opposing the motion of the electron.

The electric dipole moment D_n of a mass-m_m charge-gZ (where g here is Dirac's minimum magnetic charge) monopole will be $D_n = gZn\hbar/(2m_m c^2)$ (not counting any intrinsic moment it would have with $n = 0$ due to spin; this will be negligible in comparison when n is large). This holds both in general relativity and in Dirac's wave equation. This corresponds[37] to an energy of $D_n e\langle R_g^{-2}\rangle/(\pi\epsilon_0)$ we need to supply to the electron, where $\langle R_g^{-2}\rangle$ is the mean squared-reciprocal radius (about the axis of rotation) of the monopole's magnetic charge distribution.

Actually, instead of D_n it would presumably be better to use $(D_n + D_{n+1})/2$ since the dipole moment changes from D_n to D_{n+1} during the process. We conclude that the energy E_n of excitation of the monopole after n shoot-throughs (when it has acquired angular momentum $L = n\hbar$) is

$$E_n = \frac{gZe\hbar\mu_0\langle R_g^{-2}\rangle}{2\pi m_m}\sum_{k=0}^{n-1}(k+\frac{1}{2}) = \frac{gZe\hbar\mu_0\langle R_g^{-2}\rangle}{2\pi m_m}\frac{n^2}{2} = \frac{Zn^2\hbar^2\langle R_g^{-2}\rangle}{2m_m}. \quad (10)$$

In comparison, the quantum energy levels of a nonrelativistic "rigid rotator" with moment of inertia I are

$$E_n = \frac{\hbar^2}{2I}(n+1)n. \quad (11)$$

The two equations are the same (except for the details of the behavior for small n, which anyway depended on whether the monopole was assumed to have intrinsic spin and how much that affects its dipole moment, which are issues we have not tried to address). We conclude that

The moment of inertia of a mass-m_m monopole of Z times Dirac's minimal magnetic charge is

$$I_{\text{monopole}} = \frac{m_m}{Z\langle R_g^{-2}\rangle}. \quad (12)$$

where $\langle R_g^{-2}\rangle$ is the mean squared reciprocal radius of the monopole's magnetic charge distribution about its axis of rotation. But of course

$$I_{\text{monopole}} = m_m\langle R_m^2\rangle \quad (13)$$

where $\langle R_m^2\rangle$ is the mean squared radius of the monopole's mass distribution about its axis of rotation, if the monopole behaves enough like a rigid rotator.

Hence $\langle R_m^2\rangle\,\langle R_g^{-2}\rangle = 1$. We conclude that the mass and charge distributions of the monopole must differ, unless

[37] It may help to consider the following analogous problem. Consider a current loop of radius R carrying current i. The magnetic dipole moment of the loop is $\mu = i\pi R^2$. The B field along the axis of the loop is $B = \frac{1}{2}\mu_0 iR^2/(R^2+x^2)^{3/2}$ where x is the distance out of the plane of the loop. Thus a magnetic charge g would gain energy $g\int_{-\infty}^{+\infty} B\mathrm{d}x = \mu_0 ig = \mu_0\mu g/(\pi R^2)$ by flying along this axial line. We have the exact same problem but with electricity and magnetism interchanged.

they both are a "cylindrical delta function" at one particular radius value. Due to nonrigidity both these distributions must vary with n, which seems to make miraculous the fact that this equality must continue to hold, and hence a priori unlikely – and also the prospect of an extremely excited monopole (n very large), which would be a particle of very large macroscopic mass and size, seems appalling [38].

5 Lorentzian manifolds

The present paper has modeled the universe as a Riemannian 3-manifold with separate "absolute time." But in fact, in Einstein's theory of general relativity, the universe is a complicated Lorentzian $3+1$-manifold. This oversimplification is a serious gap in our rigorous reasoning – I believe the most serious one. What effect does this have on our picture? There are several remarks to be made.

0. The purely topological nature of the arguments in §?? and 2.5 hopefully should force them to remain valid – although some of the other arguments may not be as easy to repair. (Thus, slow changes in the 3-manifold with absolute time, should still leave everything valid.)

1. At present I think there is no hope that the present picture of charge quantization can be made into a rigorous theorem in Einstein's $(3+1)$D Lorentzian model. That is because our argument depends on both differential geometry (of the manifold) *and* quantum mechanics (to get quantization), and no satisfactory theory of "quantum gravity" has yet been advanced combining these two things. Our whole notion of an electron "coming back to the same point it started from" no longer has any meaning. It seems hard to assign a meaning to the concept of the "spectrum of the Hamiltonian" of a particle in a Lorentzian universe. The rigorous theorems in this paper have only been obtained by using a 3-manifold with separate absolute time, and that restriction, in the present theoretical climate, appears necessary.

2. Nevertheless, some parts of our argument *do* work in Einstein's $(3+1)$D manifolds. In the "Einstein-Maxwell equations" governing electromagnetism and gravity there still is an analogue of the 4-potential A (see [60] EQ 22.19) in charge-free regions of spacetime.

I do not think the presence of, or creation of, "black holes" need cause any damage to our argument: charges which avoid the holes obey our argument, and those which fall into a hole either are irrelevant or become relevant if they continue their journey around the universe by dragging the hole along for the ride – either way, our arguments remain valid.

[38] In riposte, it has been argued to me that a rotationally excited monopole would quickly decay by emitting e^+e^- pairs. One may verify that if $m_m \gg m_e$ and if the characteristic length scales (such as $\langle R_g^{-2}\rangle^{-1/2}$) of the monopole are smaller than or comparable to its Compton wavelength $h/(m_m c)$, then this decay mechanism is energetically highly favored even when $n=1$. This might prevent any attempt to spin a monopole up to high rotational quantum numbers.

6 Summary

Dirac had argued that if monopoles exist, then if charge were unquantized, contradictions would result. The easiest such contradiction to understand is the nonquantization of angular momentum, which is experimentally quantized in units of $\hbar$. This was taken as evidence for the existence of monopoles. I am similarly arguing that, even without monopoles, one may get a similar contradiction provided the universe has a suitable topology and contains a "topologically trapped" magnetic field. Amazingly enough, the momentum increment, times L_2, experienced by electrons on distance-L_1 trips round a flat 3-torus $L_1 \times L_2 \times L_3$ universe (with magnetic field in the L_3 direction) does not depend on the precise shape either of the universe, or of its magnetic field, or of the electron's trajectory, and it is invariant if L_1, L_2, or L_3 are slowly altered. Consideration of De Broglie wavelengths then forces charge quantization. This may all be generalized by demanding that the Aharonov-Bohm phase shift angle for certain electron trajectories must, modulo 2π, change only infinitesimally when the electron trajectory changes infinitesimally. That generalization works for *almost any* compact topology for the universe and almost any magnetic field.

More precisely, the argument works for any compact topology with uncontractible closed loops passing through a nonseparating hypersurface, i.e. everything but the rational homology spheres. Of these, the latter apparently may be logically ruled out by orientability considerations [75], and present experimental evidence seems to mitigate against universes with positive curvature, i.e. against both. Further, our argument works for any magnetic field which contains nonzero amounts of flux "trapped" in uncontractible loops.

Hence, this too would "explain" charge quantization. It is important to realize that no actual electron needs to take any such trip; merely the logical possibility of it is enough to force the necessary inconsistency in the laws of physics.

I'll now argue that my hypothesis for charge quantization via topology is actually less outrageous[39] than monopoles.

On what assumptions does our proposal rest?

1. The universe has a topology. (Indisputable.)

2. There is a "topologically trapped magnetic field." (If the topology is such that this is permitted, it would seem surprising if there were *no* trapped $\vec{B}$-field. The onus certainly would seem to be on opponents of the present idea, to think of any reason why there should be none.)

3. Aharonov-Bohm phase shift varies continuously as trajectory varies. (Very hard to dispute. Otherwise physics would admit peculiar discontinuities at arbitrary locations.)

4. Certain theorems about topology work. (Indisputable.)

[39] Albeit Dirac's and my proposals can be regarded as spiritually related, except that now "the universe *is* the monopole."

These are not wild and crazy assumptions. The only questionable one is whether the universe has a *suitable* topology, but it has been made clear that "most" topologies will work, and there is presently little or no experimental evidence favoring any particular topology for the universe, so nothing experimental mitigates against it.

Now, let us compare with the monopole camp. Monopoles have never been found despite extensive searches. We've presented arguments sugggestign they cannot exist. We've shown that if they do exist they necessarily would have some extremely peculiar properties. The sole basis for supposing they exist (and the only way they could exist) is highly speculative theories of physics (grand unified theories, superstring theories) which so far have essentially no experimental support.

So in light of present experimental evidence, the present paper's proposal is the more conservative choice, since it assumes the least new physics.

7 Acknowledgements

I'd like to thank Alfred S. Goldhaber, William P. Thurston, and Wolfgang Ziller for some helpful correspondence and William Bialek and A.Peter Blicher for useful conversations. (They may not agree with some of the conclusions of this paper.) Dennis Sullivan explained rational homoology spheres to me, thus correcting a misimpression I had about Poincare's conjecture (cf. §2.5.1).

References

[1] Colin Adams: The Noncompact Hyperbolic 3-Manifold of Minimal Volume, Proc. Amer. Math. Soc. 100,4 (Aug. 1987) 601-606.

[2] J.Frank Adams: Vector fields on spheres, Bull.Amer.Math.Soc. 68 (1962) 39-41; Annals of Math. 75 (1962) 603-632.

[3] J.F.Adams: On the nonexistence of elements of Hopf invariant one, Ann.Math. 72,1 (1960) 20-104; J.F.Adams & M.F.Atiyah: K-theory and the Hopf invariant, Quart.J.Math.Oxford (2 ser) 17 (1966) 31-38.

[4] J.W.Alexander: An Example of a Simply-Connected Surface Bounding a Region which is not Simply-Connected, Proc. Nat. Acad. Sci. USA 10 (1924) 8-10.

[5] L.W.Alvarez et al: Search for magnetic monopoles in the lunar sample, Science 167 (1970) 701-703.

[6] W.Ballmann, G.Thorbergsson, W.Ziller: Existence of closed geodesics on positively curved manifolds, J.Differential Geometry 18,2 (1983) 221-252.

[7] P.G.Bergmann: Introduction to the theory of relativity, Prentice Hall NY 1942.

[8] E.R.Berlekamp, J.H.Conway, R.K.Guy: Winning ways, for your mathematical plays (2 vols), Academic Press 1982.

[9] G.Blewitt, J.M.LoSecco, & 29 others: Experimental limits on the free proton lifetime for two and three body decay modes, Phys.Rev.Lett. 55 (1985) 2114-2117.

[10] W.Büchel: Why is space 3 dimensional?, English version (I.M.Freedman) Amer.J.Physics 37,12 (1969) 1222-1224; original German: Physikalische Blätter 19,12 (1963) 547-549.

[11] B.Cabrera: First positive result from a superconductive detector for moving magnetic monopoles, Phys.Rev.Lett. 48,20 (1982) 1378-1381.

[12] S.Chandrasekhar: Mathematical theory of black holes, Oxford Univ. Press 1983.

[13] T-P.Cheng & L-F.Li: Gauge theory of elementary particle physics, Oxford Univ. Press 1984, reprinted with corrections 1994.

[14] Y.M.Cho: Analytic electroweak dyon, `hep-th 0210298`.

[15] É.Cotton: Sur les variétés à trois dimensions, Ann. Fac. Sci. Toulouse (ser. 2) 1 (Paris 1899) 385-.

[16] E.Cunningham: The principle of relativity in electrodynamics and an extension thereof, Proc. London Math'l. Soc. 8 (1910) 77-98.

[17] Charles G. Darwin: The Wave Equations of the Electron, Proc. Royal Soc. London A118 (1928) 654-680 (see also Walter Gordon: Zeitschrift Phys. 48 (1928) 11-).

[18] T.Datta: The fine-structure constant, magnetic monopoles, and the Dirac charge-quantization conditions, Lettere Nuovo Cimento(2) 37,2 (19833) 51-54.

[19] P.A.M.Dirac: Quantized singularities in the electromagnetic field, Proc.Royal Soc. London A 133 (1931) 60-72; The theory of magnetic poles, Phys.Rev. 74,7 (1948) 817-830.

[20] B.Dubrovin, A.T.Fomenko, S.P.Novikov: Modern geometry methods and applications, Springer (GTM #104) 1985.

[21] M.J. Duff: Kaluza-Klein Theory in Perspective, in Proceedings Oskar Klein Centenary Symposium (19-21 September 1994) Sponsored by Royal Swedish Academy of Sciences through its Nobel Committee for Physics and by Stockholm University. World Scientific (ISBN 981-02-2332-3). Also available electronically as `hep-th/9410046`.

[22] J.Eells Jr. & J.Sampson: Harmonic mapping of Riemannian manifolds, Amer.J.Math. 86 (1964) 109-160.

[23] M.Eisenberg & R.Guy: A proof of the hairy ball theorem, Amer.Math.Monthly 86,7 (1979) 572-574.

[24] H.Flanders: Differential forms, with applications to the physical sciences, Academic Press, NY 1963. (Mathematics in science and engineering #11.)

[25] R.L.Fleischer,H.R.Hart,I.S.Jacobs,P.B.Price: Search for magnetic monopoles in deep ocean deposits, Phys.Rev. 184 (1969) 1393-1397

[26] R.L.Fleischer, P.B.Price, R.T.Woods: Search for tracks of massive multiply charged magnetic poles, Phys. Rev. 184 (1969) 1398-1401.

[27] Christian Fronsdal: Elementary particles in a curved space, IV: massless particles: Phys. Rev. D 12 (1975) 3819-3830. The previous papers in this series were I: Rev.Mod.Phys. 37 (1965) 221-224; II: Phys.Rev. D 10 (1974) 589-598; III (with R.B.Haugen): Phys.Rev.D 12 (1975) 3810-3818. Also there was a paper which some might consider to be "V," (with J.Fang): Phys.Rev.D 22 (1980) 1361-1367.

[28] D.Fuks: Foliations, J.Soviet Math. 18,2 (1982) 255-291 [Itogi Nauki i Techniki Algebra Topology Geometry 18 (Moscow 1981) 151-213 in Russian].

[29] H.Georgi & S.L.Glashow: Unity of all elementary particle forces, Phys.Rev.Lett. 32 (1974) 438-441.

[30] H.Georgi & S.L.Glashow: Unified theory of elementary particle forces, Physics Today (Sept. 1980) 30-. (Reprinted pp.214-238 in S.L.Glashow: The charm of physics, AIP 1991.)

[31] R.P.Geroch: Spinor structures of space-time in general relativity I, J.Math'l. Phys. 9,11 (1968) 1739-1744.

[32] H.Gluck, F.Warner, W.Ziller: The geometry of the Hopf fibrations, L'Enseignement Math. 32 (1986) 173-198; Fibrations of spheres by parallel great spheres, and Bergers' rigidity theorem, Annals Global Analytic Geometry 5,1 (1987) 53-82.

[33] Herbert Goldstein: Classical Mechanics, Addison-Wesley 1980. [Contains extensive discussion of "Bertrand's theorem," for which he cites J.Bertrand: Comptes Rendus 77 (1873) 849-853.]

[34] A.Haefliger: Sur les feuilletages analytiques, Comptes Rendus Acad. Sci. Paris 242,25 (1956) 2908-2910; Comment. Math. Helvetica 32 (1958) 249-329; Varietes feuilletees, Ann.Scuola Norm. Sup. Pisa 16 (1962) 367-397;

[35] A.Haefliger: Feuilletages sur les varietes ouvertes, Topology 9,2 (1970) 183-194.

[36] D-S. Han & J-W.Kim: Unit vector fields on spheres, which are harmonic maps, Mathematische Zeitschrift 227,1 (1998) 83-92.

[37] Heinz Hopf: Über die Abbildungen der dreidimendionalen Sphäre auf die Kugelfläche, Mathematische Annalen 104,5 (1931) 637-665.

[38] Heinz Hopf: Vektorfelder in n-dimensionalen Mannigfaltigkeiten, Math. Annalen 96 (1927) 225-250.

[39] A.Hurwitz: Über die Komposition der quadratischen formen, Math.Annalen 88 (1923) 1-25. (Posthumous.) This result had previously been sketched by Hurwitz in: Über die komposition der quadratischen Formen von beliebig vielen Variablen, Nachrichten der Königl. Ges. der Wissen. Göttingen (1898) 309-316.

[40] J.D. Jackson: Classical electrodynamics (3rd ed.), Wiley 1999.

[41] Proofs of the Jordan curve theorem are in:
L.V. Ahlfors: Complex Analysis, McGraw-Hill 1979.
I.Fary & E.M.Isenberg: On a converse of the Jordan curve theorem, Amer. Math. Monthly 81 (1974) 636-639.
R.Maehara: The Jordan curve theorem via the Brouwer fixed point theorem, Amer. Math. Monthly 91,10 (1984) 641-643.
C.Thomassen: A link between the Jordan curve theorem and the Kuratowski planarity condition, Amer. Math. Monthly 97,3 (1990) 216-218.
Helge Tverberg: A proof of the Jordan curve theorem, Bull.London Math. Soc. 12,1 (1980) 34-38.
K. Venkatachaliengar: A simple proof of the Jordan curve theorem, Quart. J. Math. Oxford 8 (1937) 241-244.
G.T.Whyburn: Topological analysis, Princeton Univ. Press 1958.

[42] B.Julia & A.Zee: Poles with both electric charge and magnetic charge in non-Abelian gauge theory, Physical Review D 11 (1975) 2227-. (Reprinted in C.Rebbi & G.Soliani: Solitons and Particles, World Scientific 1984.)

[43] T.F.E.Kaluza: Zum Unitätsproblem der Physik (On the Unification Problem of Physics), Sitzungsberichte Preussische Akademie der Wissenschaften 54 (1921) 966-976.

[44] M.Kervaire: Non-parallelizability of the sphere for $n > 7$, Proc. Nat'l Acad. Sci. USA 44 (1958) 280-283.

[45] Oskar Klein: Quantum Theory and Five-Dimensional Relativity (in German), Zeit. Physik 37 (1926) 895-; Nature 118 (1926) 516-.

[46] J.Korbas & P.Zvengrowski: The vector field problem; a survey with emphasis on specific manifolds, Expositiones Mathematicae 12,1 (1994) 3-20.

[47] K.S.Kulkarni & U.Pinkall (eds.): Conformal geometry, Vieweg Braunschweig 1988.

[48] K.M.Kuperberg: A smooth counterexample to the Seifert conjecture, Ann. of Math. 140 (1994) 723-732. G.Kuperberg: A volume-preserving counterexample to the Seifert conjecture, Comment. Math. Helv. 71

(1996) 70-97; G.Kuperberg & K.M.Kuperberg: Generalized counterexamples to the Seifert conjecture, Ann. of Math.144 (1996), 239-268.

[49] M.Lachiez-Rey, J.M. Luminet: Cosmic topology, Physics Reports 254 (1995) 135-214.

[50] L.D.Landau & E.M.Lifshitz: Quantum mechanics, nonrelativistic theory, Pergamon 1976. (Section 35, "Fall of a particle to the centre.")

[51] W.B.R.Lickorish: A foliation for 3-manifolds, Annals of Math. 82 (1965) 414-420.

[52] Elliott H. Lieb: Selecta (ed. by W.Thirring) "The stability of matter from atoms to stars," Springer (second ed.) 1997.

[53] D.Lovelock: The uniqueness of the Einstein field equations in a 4 dimensional space, Arch. Rat. Mech. & Anal. 33 (1969) 54-70.

[54] David Lovelock: Dimensionally dependent identities, Proc. Cambridge Philos. Soc. 68 (1970) 345-350.

[55] D.Lovelock & H.Rund: Tensors, differential forms, and variational principles, J.Wiley 1978; Dover corrected reprint 1989.

[56] K.H.Mariwalla: Dimensionality of space-time, J.Mathematical Physics 12 (1971) 96-99.

[57] W.Miller Jr.: Symmetry and Separation of Variables, Addison-Wesley, London 1977

[58] J.Milnor: Some consequences of a theorem of Bott, Annals Math. 68 (1958) 444-449.

[59] J.Milnor: Analytic proofs of the 'hairy ball theorem' and the Brouwer fixed point theorem, Amer.Math.Monthly 85,7 (1978) 521-524.

[60] C.W.Misner, K.S.Thorne, J.A.Wheeeler: Gravitation, Freeman 1973.

[61] A.F. Monna: Dirichlet's principle: a mathematical comedy of errors and its influence on the development of analysis, Oosthoek, Scheltema & Holkema, Utrecht 1975.

[62] P. Moon & D.E. Spencer: Field Theory Handbook: Including Coordinate Systems, Differential Equations and Their Solutions, Springer-Verlag, Berlin 1961.

[63] P.M.Morse & H.Feshbach: Methods of Theoretical Physics, vols I & II, McGraw-Hill 1953.

[64] K.R.Nagarajan & T.Soundararajan: On the module of 1-forms on a differentiable manifold and parallelizable manifolds, Indian J. Math. 41,1 (1999) 115-129.

[65] Grisha Perelman: The entropy formula for the Ricci flow and its geometric applications, 39 pages, `arXiv:math.DG/0211159`.

[66] H.Poincaré: Sur les courbes définies par les équations différentielles, J.Math.Pures Appl. (4) 1 (1885) 167-244 (chapter 13).

[67] J. Polchinski: String theory (2 vols), Cambridge University Press 1998.

[68] A.M.Polyakov: Particle spectrum in quantum field theories, JETP Letters 20 (1974) 194-195.

[69] J.Radon: Lineare Scharen Orthogonaler Matrizen, Abh.Math.Sem.Univ.Hamburg 1 (1922) 1-14

[70] A. Salam & J. Strathdee: On Kaluza Klein theory, Annals of Physics 141 (1982) 316-352.

[71] L.I.Schiff: Classical examples of time reversal, American J. Physics 32 (1964) 812.

[72] Richard M. Schoen: Conformal deformation of a Riemannian metric to constant scalar curvature, J. Differential Geometry 20,2 (1984) 479–495.

[73] J.Schwinger: Magnetic charge and quantum field theory, Phys.Rev. 144 (1966) 1087-1093; Sources and magnetic charge, Phys.Rev. 173 (1968) 1536-1543.

[74] Peter Scott: The geometries of 3-manifolds, Bull. London Math. Soc. 15 (1983), 401-487.

[75] Warren D. Smith: On the shape of the universe, `http://math.temple.edu/~wds/homepage/unishape.{dvi,ps`

[76] Glenn D. Starkman: Topology and cosmology, Classical Quantum Gravity 15,9 (1998) 2529-2538. This is the introduction to a special issue devoted to questions about the geometry and topology of the universe, which is recommended in its entirety.

[77] E.Stiefel: Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten, Comment. Math. Helv. 8,4 (1936) 305-353.

[78] F.R.Tangherlini: Schwarzschild field in n dimensions and the dimensionality of space problem, Il Nuovo Cimento 27,3 (1963) 636-651.

[79] Max Tegmark: On the dimensionality of spacetime, Classical & Quantum Gravity 14,4 (1997) L69-L75.

[80] G. 't Hooft: Magnetic monopoles in unified gauge theories, Nucl.Phys. B79 (1974) 276-284.

[81] Emery Thomas: Vector fields on manifolds, Bull.Amer.Math.Soc. 75 (1969) 643-683.

[82] J.J. Thomson: Elements of electricity and magnetism, Cambridge Univ. Press (3rd ed) 1904. (Section 284).

[83] William P. Thurston: A local construction of foliations for 3-manifolds, Proc. Sympos. Pure Math. 27 part 1 (1975) 315-319.

[84] William P. Thurston: Existence of codimension-1 foliations, Ann.Math. (2) 104 (1976) 249-268.

[85] William P. Thurston: Three-dimensional geometry and topology (edited by Silvio Levy), Princeton University Press (mathematical series #35) 1997.

[86] Marie-Antoinette Tonnelat: Les théoriés unitaires de l'électromagnétisme et de gravitation, Gauthier-Villars, Paris 1965.

[87] G.F.Torres del Castillo & L.C.Cortés-Cuautli: Solution of the Dirac equation in the field of a magnetic monopole, J.Math.Phys. 38,6 (1997) 2996-3006.

[88] J.P.Vallee: Intergalactic and galactic magnetic fields – an updated test, Astrophys.Lett. 23 (1983) 85-94.

[89] J. von Neumann & A.W.Burks: Theory of self reproducing cellular automata, Univ. of Illinois Press, Urbana 1966.

[90] Steven Weinberg: Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations, Phys. Rev. 138, 4B (1965) 988-1002.

[91] S.Weinberg: Gravitation and cosmology, Wiley 1972.

[92] J.W.Wood: Foliations on 3-manifolds, Annals of Math. 89 (1969) 336-358; Foliations of codimension 1, Bull.Amer.Math.Soc. 76,5 (1970) 1107-1111.