Nuclear Physics B333 (1990) 195–220
North-Holland

MAGNETIC WORMHOLES AND TOPOLOGICAL SYMMETRY*

Arun K. GUPTA, James HUGHES, John PRESKILL** and Mark B. WISE

California Institute of Technology, Pasadena, CA 91125, USA

Received 5 July 1989

We investigate the wormhole solutions that arise in the semiclassical analysis of euclidean gravity coupled to gauge fields. In 2 + 1 dimensions, "magnetic monopole" solutions can be constructed, for either abelian or nonabelian gauge fields. The low-energy physics induced by these wormholes qualitatively resembles, but is quantitatively distinguishable from, the low-energy physics of a gauge theory (without wormholes) that undergoes the Higgs mechanism at the wormhole mass scale. Wormholes are suppressed if matter fields are introduced that transform suitably under the gauge group. In 3 + 1 dimensions, "meron" solutions can be constructed, for nonabelian gauge fields only. We argue, however, that these wormhole solutions do not contribute to the semiclassical evaluation of the path integral.

1. Introduction

The proposal [1–3] that fluctuations in the topology of space-time may occur in quantum gravity has attracted great attention recently, because the implications of this proposal appear to be quite profound. In the euclidean path integral approach to quantum gravity [4], such fluctuations in topology can be systematically studied. In particular, at least for the case of gravity coupled to matter fields with suitable properties, interesting solutions to the euclidean field equations can be constructed [3, 5]. These solutions have the topology of two asymptotically flat euclidean space-times connected together by a narrow throat, which is known as a "wormhole". If the asymptotically flat regions are identified at infinity, the topology is that of a four-sphere with a handle attached to it. Although the regions connected together by the wormhole are distantly separated on the four-sphere, the proper distance between them as measured along the wormhole is very short, comparable to the thickness of the wormhole throat.

Since wormhole solutions are stationary points of the euclidean action, one may argue that they make an important contribution to the euclidean path integral in the semiclassical (small $\hbar$) limit. Furthermore, the semiclassical interpretation that can

* This work supported in part by the U.S. Department of Energy under Contract No. DE-AC0381-ER40050.

** NSF Presidential Young Investigator.

be assigned to the wormhole contribution to the path integral qualitatively distinguishes it from the contribution due to geometries of trivial topology; it may be interpreted as the amplitude, in the WKB approximation, for a quantum tunneling event in which a tiny closed "baby universe" spontaneously nucleates [3]. The fluctuating microscopic baby universes can have a remarkable impact on macroscopic physics – they might subject the fundamental constants of Nature to an intrinsic quantum mechanical uncertainty [6, 7], or might force the cosmological constant to vanish [8].

Although the foundations that underlie the euclidean path integral approach to quantum gravity appear shaky at present, and the semiclassical approximation to quantum gravity may be unjustified in any case, it nonetheless seems worthwhile to investigate the properties of the wormhole solutions. Even if semiclassical reasoning cannot be justified in detail, one might gain useful insights into the qualitative features of wormhole physics by applying semiclassical methods. An instructive analogy can be invoked with quantum chromodynamics. The semiclassical approximation to QCD, though not strictly valid, provides a deep understanding of the emergence of the vacuum angle θ and the resolution of the U(1) problem. These qualitative insights continue to apply even after the semiclassical approximation is discarded. Furthermore, the *semiclassical* predictions that can be inferred from the euclidean path integral formalism may well survive even if the formalism itself requires serious modification. It is interesting to ask, for example, whether wormhole solutions exist in the standard model coupled to gravity, or whether it is instead necessary to introduce nonstandard matter with exotic properties in order to ensure the existence of solutions.

Yet another motivation for further study of wormhole solutions is provided by the large wormhole problem, which afflicts most recent attempts to extract information about low-energy physics from wormhole considerations. The most promising proposed resolution of this difficulty [9, 10] appears to rely very heavily on semiclassical reasoning. To begin to assess the viability of the proposal made in refs. [9, 10], we should scrutinize it carefully within the context of the semiclassical approximation.

It appears that the requirement that wormhole solutions exist in (euclidean) general relativity coupled to matter may indeed place significant restrictions on the matter content of the theory. One can appreciate that wormhole solutions to the field equations are not trivial to construct by noting that the throat of the wormhole acts as a *defocussing* gravitational lens; hence, negative stress-energy is required to support the throat. Fortunately, negative stress-energy is not nearly so difficult to achieve in euclidean space as in Minkowski space. For example, the euclidean Lagrange density of an electromagnetic field is

$$\mathscr{L}_{\mathrm{E}} = \tfrac{1}{2}(\boldsymbol{E}^2 + \boldsymbol{B}^2), \tag{1.1}$$

and the corresponding euclidean "energy density" is

$$(T_{00})_{\mathrm{E}} = \tfrac{1}{2}(\boldsymbol{E}^2 - \boldsymbol{B}^2). \tag{1.2}$$

Therefore, a magnetic field acts as a source of negative "energy" in the euclidean Einstein equation, and is capable of supporting the throat of a wormhole.

This observation was recently exploited by Hosoya and Ogura [11], who described wormhole solutions for gravity coupled to gauge fields in both $2+1$ and $3+1$ space-time dimensions, and provided the impetus for the work reported here. Our main objective in this paper will be to elucidate the physical implications of the wormholes described in ref. [11], and of some related solutions. Most of the analysis will concern the wormholes in $2+1$ dimensions, but we will also make some comments about $(3+1)$-dimensional wormholes at the end of the paper.

The remainder of this paper is organized as follows: In sect. 2, we review the properties of the magnetic wormhole solutions of $(2+1)$-dimensional electrodynamics coupled to gravity, and note the existence of corresponding solutions in $(2+1)$-dimensional Yang–Mills theory. The throat thickness of the electrodynamic wormholes may be arbitrarily large, but the stable Yang–Mills wormholes have a maximum size.

In sect. 3, we observe that magnetic wormholes violate the topological conservation laws of $(2+1)$-dimensional gauge theories, and we explore some consequences. We show that wormholes convert ordinary noncompact electrodynamics to compact electrodynamics, and drive electric confinement. We find that wormhole solutions do not occur in $(2+1)$-dimensional Yang–Mills theory if matter fields are introduced that transform faithfully under the center of the gauge group. (Similarly, no wormholes occur in $(2+1)$-dimensional electrodynamics if two matter fields are introduced that have incommensurate charges.) We compare the low-energy physics induced by magnetic wormholes with the low-energy physics of an ordinary non-abelian gauge theory (without wormholes) that undergoes the Higgs mechanism at a large mass scale. Topological symmetry is intrinsically broken in both cases, but the symmetry breaking generated by wormholes need not be small, while the symmetry breaking in the corresponding Higgs theory is necessarily very weak.

In sect. 4, we describe a class of wormhole solutions of $(3+1)$-dimensional Yang–Mills theory that includes the solution constructed in ref. [11]. All of these solutions are unstable with respect to small deformations of the gauge field, however, and we argue that they do not correspond to any semiclassical tunneling process.

2. Magnetic wormholes in 2 + 1 dimensions

For a U(1) gauge field (photon) coupled to gravity in $2+1$ dimensions, wormhole solutions can be constructed corresponding to each nonzero value of an integer n, where n is the number of units of quantized magnetic flux that disappear down the throat of the wormhole. That is, a slice through the throat is a two-sphere, and the magnetic flux integrated over the two-sphere is $\Phi = n\Phi_0$, where Φ_0 is the flux quantum. Of course, flux is actually unquantized in pure electrodynamics without charged particles, but we anticipate that particles of unit charge will eventually be introduced on the wormhole background. The quantum mechanics of these particles can be consistently defined only if the magnetic flux that disappears down the throat is restricted to an integer multiple of $\Phi_0 = 2\pi$ ($\hbar = c = 1$). Semiclassically, we interpret the wormhole solution as describing the quantum mechanical nucleation of a baby universe that carries magnetic flux $\Phi = n\Phi_0$. (Note that a closed universe is permitted to carry a nonvanishing magnetic flux, because the magnetic field is a scalar quantity in two spatial dimensions.)

This "magnetic wormhole" in $2+1$ dimensions is actually the obvious analog of a $(3+1)$-dimensional wormhole that was described by Giddings and Strominger [3]. The magnetic field $F = \mathrm{d}A$ is a two-form that can be expressed as the curl of a one-form potential A. In their construction, Giddings and Strominger considered a field strength $H = \mathrm{d}B$ that can be expressed as the curl of a two-form potential B. Just like F, H enters the $(3+1)$-dimensional euclidean Einstein equation as a source of negative euclidean energy density, and can support the throat of a wormhole. Semiclassically, we interpret this wormhole solution as describing the quantum mechanical nucleation of a three-dimensional baby universe that carries a nonvanishing H flux. In general, the H flux on a three-sphere is unquantized, but if we anticipate that *strings* will eventually be introduced on the wormhole background that couple to the potential B, then the H flux disappearing down the throat of the wormhole is required to be an integer multiple $n\Phi_0$ of an elementary flux quantum Φ_0 [12]. Wormhole solutions can be constructed for each nonvanishing value of n. This construction has an obvious generalization to D-dimensional euclidean space [13], for $D \geqslant 2$. (Incidentally, the Giddings–Strominger wormhole may also be described in terms of a massless scalar field, the "axion," that is dual to H. This duality applies for arbitrary dimensionality; we will exploit it in the $(2+1)$-dimensional case shortly.)

Similar wormhole solutions can be constructed for Yang–Mills gauge fields coupled to gravity in $2+1$ dimensions. These solutions are nearly identical to the corresponding solutions in electrodynamics, for only one abelian component of the gauge field is excited. However, the Yang–Mills solutions have some new features that distinguish them from the solutions in electrodynamics. First, since Yang–Mills fields themselves carry nonzero electric charges, the magnetic flux that disappears down the throat is automatically quantized; there is no need to contemplate introducing additional charged matter fields. And second, the magnetic flux does not take arbitrary nonzero integer values. Although solutions can indeed be constructed with $\Phi = n\Phi_0$, where n is any integer, most of these solutions are unstable [14,15]. The instability afflicts the long-range Yang–Mills field far from the wormhole throat. If slightly perturbed, the field strength decays, with the excess energy being radiated to infinity. (Here we have described the instability of a field configuration in three-dimensional euclidean space in a language appropriate to three space and one time dimension.) The stable solutions can be characterized by the value of a topologically conserved magnetic flux on a two-sphere. This flux takes values in the *center* of the gauge group G. For example, if $G = SU(N)$ it takes values in Z_N, and there are $N-1$ distinct wormhole solutions, one for each nontrivial element of Z_N. It is sensible to include only these stable wormhole solutions in the semiclassical evaluation of the path integral*.

This topologically conserved magnetic quantum number arises from the requirement that the "Dirac string" of the Yang–Mills potential A on the two-sphere not be visible to the Yang–Mills fields. Thus, the "Aharanov–Bohm phase" associated with the string must take values in the center of the group. No smooth deformation of the gauge field can change the value of this element of the center, since the center is a discrete group. Time evolution is a particular type of smooth deformation, and so the topological magnetic flux is a constant of the motion. In the case $G = SU(N)$, if the flux Φ on a two-sphere is N times the flux quantum, there is no need for a string at all, and therefore a configuration with $\Phi = N\Phi_0$ can be smoothly deformed to a trivial configuration with vanishing field strength. This topological property is crucially important for understanding the difference between an abelian and a nonabelian gauge theory, as we will discuss further in sect. 2.

In both the U(1) and Yang–Mills cases, the magnetic field in the asymptotically flat region far from the wormhole throat is the field of a magnetic monopole that carries magnetic charge $g = ng_0$, where g_0 is the Dirac magnetic charge. Thus, these wormholes provide a magnetic realization in three-dimensional euclidean space of an old conception due to Misner and Wheeler [16] concerning the origin of electric charge. Euclidean space-time with many wormholes attached to it appears to be sprinkled with many magnetically charged particles to an observer who is unable to resolve the tiny wormhole throat. Upon closer inspection, though, one finds that the magnetic field is actually divergence-free everywhere; what appears at first to be a monopole–antimonopole pair is really a wormhole threaded with magnetic flux. Furthermore, the infinite self-energy of point monopoles is avoided. Each monopole has a finite core radius provided by the wormhole throat.

Before examining further the impact of these magnetic wormholes on low-energy physics, let us first consider more explicitly the properties of the solutions themselves. For the case of U(1) gauge field coupled to gravity in 2 + 1 dimensions, the euclidean action is**

$$S = \int d^3x \sqrt{g} \left[-\frac{1}{16\pi G} R + \frac{1}{4e^2} g^{ij} g^{kl} F_{ik} F_{jl} \right]. \tag{2.1}$$

* See sect. 4 for a further discussion of the stability of wormhole solutions.

** We are considering electrodynamics without a "topological" mass term for the photon.

Here G is Newton's constant, R is the scalar curvature, e is the electromagnetic coupling and $F_{ij} = \partial_i A_j - \partial_j A_i$ is the magnetic field strength. (Although F_{ij} is actually the euclidean electromagnetic field strength, with two electric components and one magnetic component, we find it convenient to describe it in a language appropriate for a three-component magnetic field in three-dimensional space.) In 2 + 1 dimensions, both G^{-1} and e^2 have the dimensions of mass (with $\hbar = c = 1$).

The field equations of this theory have a spherically symmetric solution in which the metric takes the form

$$ds^2 = dt^2 + a(t)^2(d\theta^2 + \sin^2\theta\, d\phi^2); \tag{2.2}$$

the slice at constant "euclidean time" t is a two-sphere of radius $a(t)$. The magnetic field on this slice points in the $\hat{t}$ direction and carries magnetic flux $\Phi = 2\pi n$; thus the field strength is

$$F = n/2a^2, \tag{2.3}$$

and the electromagnetic Lagrange density on the two-sphere of radius a is

$$\frac{1}{2e^2}F^2 = \frac{n^2}{8e^2a^4}. \tag{2.4}$$

The scale factor a satisfies the Einstein equation

$$\dot{a}^2 = 1 - \frac{\pi G n^2}{e^2 a^2}, \tag{2.5}$$

which is solved by

$$a(t)^2 = b_n^2 + t^2, \qquad b_n^2 = \pi G n^2/e^2. \tag{2.6}$$

Here b_n is the minimal value of the scale factor, or the thickness of the throat of the wormhole; b_1 is roughly the geometric mean of the two length scales G and e^{-2} in the problem. Semiclassically, we may interpret b as the radius of a baby universe that spontaneously nucleates. It is, in fact, the maximum radius of a (2 + 1)-dimensional Friedmann–Robertson–Walker cosmology that is supported by the magnetic flux $\Phi = 2\pi n$. This radius is large in Planck units as long as the gauge coupling e^2 is small in those units. The euclidean action of a "semiwormhole" (comprised of one of the asymptotically flat regions and half of the throat) is

$$\begin{aligned} S_{E,n} &= 2\int d^3x\sqrt{g}\,\frac{1}{2e^2}F^2 = \int_0^\infty dt\, 4\pi a^2 \frac{n^2}{4e^2a^4} \\ &= 2\pi^2\left(\frac{n}{2e}\right)^2\frac{1}{b_n} = \frac{\pi^{3/2}}{2}\frac{n}{e\sqrt{G}}. \end{aligned} \tag{2.7}$$

(Half of $S_{E,n}$ comes from the scalar curvature term in S, and half from the magnetic field strength term.) Roughly, this is the "self-energy" of a monopole with magnetic charge $g = ng_0$, and core size b_n.

For the case of a Yang–Mills gauge field, the metric has the same form, but the magnetic field strength is

$$F = \frac{1}{2a^2} Q, \tag{2.8}$$

where Q is an element of the Lie algebra of G such that $\exp(2\pi i Q)$ is a nontrivial element of the center of G. For example, if G = SU(N), the values of Q that satisfy the Brandt–Neri–Coleman [14, 15] stability criterion are, in a particular gauge,

$$Q_n = \mathrm{diag}\Bigg(\underbrace{\frac{n}{N}, \dots \frac{n}{N}}_{N-n \text{ times}}, \underbrace{\frac{n-N}{N}, \dots \frac{n-N}{N}}_{n \text{ times}} \Bigg), \qquad n = 1, 2, \dots, N-1. \tag{2.9}$$

The Yang–Mills Lagrange density on the two-sphere of radius a is

$$\frac{1}{e^2} \mathrm{tr}\, F^2 = \frac{1}{4e^2 a^4} \frac{n(N-n)}{N}, \tag{2.10}$$

and an analysis identical to that described above shows that the thickness of the wormhole throat is

$$b_n^2 = \frac{2\pi G}{e^2} \frac{n(N-n)}{N}, \tag{2.11}$$

and the action of a semiwormhole is

$$S_{E,n} = \sqrt{\frac{\pi^3}{2}} \frac{1}{e\sqrt{G}} \sqrt{\frac{n(N-n)}{N}}. \tag{2.12}$$

In both the U(1) and Yang–Mills cases, there are also wormhole solutions if the gauge fields are coupled to Higgs fields, and the gauge symmetry is realized in a Higgs phase. Suppose, for example, that all gauge fields acquire masses of order μ via the Higgs mechanism. Then, if μ^{-1} is large compared to the throat thickness b computed above, the wormhole solution closely resembles that described above in the vicinity of the throat (for $a \ll \mu^{-1}$). But for $a \gg \mu^{-1}$, it becomes favorable for the magnetic flux to collapse to a vortex with thickness of order μ^{-1}, rather than be distributed uniformly over the two-sphere. As a result, the action of the semiwormhole diverges linearly like

$$nm_{\mathrm{v}}\int \mathrm{d}t\,, \tag{2.13}$$

where m_{v} is precisely the mass of the stable vortex particle that appears in the spectrum of the theory in the Higgs phase. Now the integer n may be interpreted as the number of vortices that disappear down the wormhole throat.

To conclude this section, we would like to stress one feature of the wormhole solutions that occur in a (2 + 1)-dimensional Yang–Mills theory that distinguishes them from the wormholes of the abelian gauge theory. Because the topological magnetic flux takes only a finite number of nontrivial values in the nonabelian case, the wormholes have a maximal throat thickness. The same feature should apply to any theory with a global discrete Z_N symmetry, for we expect to be able to construct solutions in which any nontrivial value of the Z_N charge flows down the wormhole. Indeed, the (2 + 1)-dimensional Yang–Mills theory admits a dual description in terms of a scalar field that carries such a Z_N charge. Wormhole solutions with a maximum thickness could likewise be constructed in a (3 + 1)-dimensional scalar field theory with a Z_N symmetry. This observation may be relevant to the large wormhole problem, since in a model of this kind no large wormholes exist, at least at the level of the semiclassical approximation.

3. Intrinsic breaking of topological symmetry, and its consequences

We now wish to consider the influence of the wormholes on physics at low energy, or at distances large compared to the thickness of the throat. For this purpose, we "integrate out" the wormholes, incorporating their effects into an effective field theory that is quasilocal on a cutoff distance scale comparable to the throat thickness b. Since the wormhole is a source of magnetic flux, the local operators in the effective action that reproduce the effects of the wormhole must likewise act as a source of magnetic flux.

Indeed, we can easily define a local operator $\phi(x)$ that creates (or destroys) a unit of magnetic flux [17]. We will find it convenient to describe how the operator $\phi(x)$ is defined in both the canonical formalism and in the euclidean path integral formalism. In the canonical formalism, we impose the gauge condition $A_0 = 0$. Then a large Hilbert space H is constructed that is spanned by the eigenstates of the operators $A_i(x)$, and the physical states are those states that are invariant under (time-independent) local gauge transformations; they define a subspace H_P of H. To define the action of the (Schrödinger picture) operator $\phi(x)$ on H_P we first specify that $\phi(x)$ acts on a basis for H according to

$$\phi(x)|A_i(y)\rangle = |A_i^{\Omega(x)}(y)\rangle\,. \tag{3.1}$$

Here, $\Omega(\boldsymbol{x})$ is a gauge transformation that is smooth everywhere except at $\boldsymbol{x}$, but is singular at $\boldsymbol{x}$ and has a minimal winding number around the point $\boldsymbol{x}$. For example, if the gauge group is $\mathrm{G}=\mathrm{SU}(N)$, then $\Omega \in \mathrm{SU}(N)$ has a discontinuity

$$\Omega_{\text{above cut}} = \mathrm{e}^{2\pi i/N}\, \Omega_{\text{below cut}} \tag{3.2}$$

across a cut in the plane that terminates at the point $\boldsymbol{x}$. Since the gauge field actually transforms as a representation of $\mathrm{SU}(N)/\mathrm{Z}_N$, the discontinuity cannot be detected by the gauge field.

Now, the action of $\phi(\boldsymbol{x})$ on the large Hilbert space H depends on the detailed form of the gauge transformation Ω, but its action on the space $\mathrm{H_P}$ of physical gauge-invariant states is specified entirely by the location $\boldsymbol{x}$ of the singularity of Ω, and the winding number about the singularity. Indeed, acting on $\mathrm{H_P}$, $\phi(\boldsymbol{x})$ is a gauge-invariant operator that commutes at spacelike separation with all gauge-invariant smeared polynomials in the gauge fields; it is qualified for admission to the algebra of gauge-invariant local operators of this field theory. Furthermore, it is not redundant to include $\phi(\boldsymbol{x})$ in the local field algebra, because its action on $\mathrm{H_P}$ cannot be duplicated by any smeared polynomial in the gauge fields. To appreciate this point, note that one can introduce an infrared cutoff by, for example, compactifying two-dimensional space on a two-sphere. Then the states of $\mathrm{H_P}$ divide into sectors that are characterized by the value of the topologically conserved magnetic flux; this quantity takes values in Z for $\mathrm{G}=\mathrm{U}(1)$ and in Z_N for $\mathrm{G}=\mathrm{SU}(N)$. While the smeared gauge-invariant polynomials in the gauge fields preserve these sectors, the operator $\phi(\boldsymbol{x})$ interpolates between sectors by creating or destroying a unit of magnetic flux.

Corresponding to this canonical construction is a (Heisenberg picture) operator $\phi(t, \boldsymbol{x})$ that is conveniently described in euclidean path integral language [17]. In this language, a correlation function with an insertion of the operator $\phi(t, \boldsymbol{x})$ is computed by summing over all those gauge field configurations that have a Dirac charge magnetic monopole singularity located at the point $(t, \boldsymbol{x})$. Of course, in the definition of ϕ in either the canonical language or the path integral language, a short-distance regulator is implicitly assumed. For example, in the path integral formalism, the singular core of the monopole must be smeared out over a finite cutoff length scale; this cutoff prevents the monopole from having an infinite self-energy. If we wish to remove the cutoff, then a field renormalization of ϕ must be performed.

The field ϕ is a complex scalar field, and the *phase* of ϕ is the analog of the "axion" field a that is dual to the three-form field strength H in the $(3+1)$-dimensional theory that was discussed by Giddings and Strominger [3]. Just as ϕ creates or destroys a quantum of F magnetic flux, so e^{ia} creates or destroys a quantum of H flux, in the case considered by Giddings and Strominger.

Having seen how a gauge-invariant local operator ϕ can be constructed that acts as a source of magnetic flux, we now return to the task of incorporating the effects of magnetic wormholes into an effective field theory description of low-energy physics. By integrating out wormholes that carry n units of quantized magnetic flux, we evidently generate a term in the effective action that has the form, in the leading semiclassical approximation,

$$\mathscr{L}_{E,n} = C_n e^{-S_{E,n}}\left(a_n^\dagger + a_{-n}\right)\phi^n(x) + \text{h.c.}. \tag{3.3}$$

Here ϕ is the local operator defined above, $a_n^\dagger$ is an operator that creates a baby universe that carries n units of magnetic flux, and a_{-n} annihilates a baby universe that carries $-n$ units of magnetic flux; $S_{E,n}$ is the euclidean action of the corresponding semiwormhole solution, eq. (2.7). (More precisely, $S_{E,n}$ is the part of action coming from a region near the wormhole throat; this distinction is important in the case where the gauge symmetry is spontaneously broken and there are stable vortices, since then the total action of the semiwormhole diverges in the infrared.) The constants C_n can in principle be determined by matching the Green functions of our effective field theory to Green functions computed on the wormhole background, in the leading semiclassical approximation. (The required formalism has been outlined in refs. [18, 19].) To interpret eq. (3.3), one observes that the Hilbert space of baby universes is spanned by eigenstates of the operators $a_n^\dagger + a_{-n}$, with eigenvalues α_n. Thus, the low-energy physics described by our effective field theory divides into superselection sectors [6, 7]. The distinct sectors are labeled by $\{\alpha_n\}$, and each sector has distinct physics, for the coefficient of the local operator $\mathscr{L}_{E,n}$ in the effective action depends on the value of α_n. For the purpose of discussing the qualitative effects of wormholes on low-energy physics, it suffices to consider the physical effects of the interactions $\mathscr{L}_{E,n}$.

As was previously noted, both abelian and nonabelian gauge theories, in the absence of wormholes, respect topological conservation laws; the magnetic flux, which takes values in Z for G = U(1) and in Z_N for G = SU(N), is a constant of the motion. Since the action of ϕ on a state changes the value of the topologically conserved quantity, we may define a unitary symmetry operator U that commutes with the hamiltonian according to

$$U(\alpha)\phi(x)U^{-1}(\alpha) = e^{i\alpha}\phi(x), \qquad \alpha \in [0, 2\pi] \tag{3.4}$$

in the case G = U(1), or

$$U_k\phi(x)U_k^{-1} = e^{2\pi i k/N}\phi(x), \qquad k \in \{0, 1, 2, \dots, N-1\} \tag{3.5}$$

in the case G = SU(N). Thus we may say that the gauge theory respects a "topological symmetry". This is a continuous global U(1) symmetry if the gauge group is U(1), and a discrete Z_N symmetry if the gauge group is SU(N). Evidently, the terms in eq. (3.3) that arise in the effective action when wormholes are integrated out break this topological symmetry explicitly. We have thus identified an interesting qualitative feature of the effects of wormholes on a gauge theory – wormholes that occur semiclassically in a gauge theory coupled to gravity in 2 + 1 dimensions violate intrinsically the topological conservation laws.

In fact, the topological symmetry plays a central role in the theory of "quark" confinement in 2 + 1 dimensions, as was stressed by Polyakov [20] and 't Hooft [17]. One is thus prompted to consider the impact of wormholes on this theory, which we will now discuss in detail. We will first analyze the realization of the topological symmetry in an *abelian* gauge theory in 2 + 1 dimensions, and then investigate the consequences of intrinsic breaking of the symmetry. Next, we will repeat this analysis for the case of a nonabelian gauge theory.

In 2 + 1 dimensions, the Hilbert space of a U(1) gauge theory with an infrared cutoff divides into sectors labeled by the magnetic flux $n \in \mathbb{Z}$. In ordinary electrodynamics with a massless photon, the energy of the ground state $|n\rangle$ of sector n tends to zero as the infrared cutoff is removed. Thus, the vacuum state of the theory is infinitely degenerate in the infinite-volume limit. The basis $\{|n\rangle\}$ for the vacuum states is not a convenient basis for the evaluation of the correlation functions of the local operator ϕ, because ϕ is not diagonal in this basis; instead,

$$\langle m|\phi(x)|n\rangle = v\delta_{m,n-1}, \tag{3.6}$$

and the correlation functions evaluated in the n-basis do not satisfy the cluster decomposition property. It is preferable to use the basis

$$|\theta\rangle = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} e^{in\theta}|n\rangle \tag{3.7}$$

such that

$$\langle\theta'|\theta\rangle = \delta(\theta' - \theta) \quad \text{and} \quad \langle\theta'|\phi(x)|\theta\rangle = v\, e^{i\theta}\, \delta(\theta - \theta'). \tag{3.8}$$

The action of the topological U(1) symmetry on the vacuum states is

$$U(\alpha)|n\rangle = e^{-in\alpha}|n\rangle \quad \text{or} \quad U(\alpha)|\theta\rangle = |\theta - \alpha\rangle. \qquad (3.9),(3.10)$$

Thus we see that in massless electrodynamics the topological U(1) symmetry is spontaneously broken. According to Goldstone's theorem, the spectrum of the theory must therefore contain a massless spin-zero particle, the Goldstone boson. We may envision a finite-wavelength Goldstone excitation as a local fluctuation in θ. But in fact the Goldstone boson is simply the photon; in two spatial dimensions massless particles have no helicity, and the representation of the Poincaré group according to which the photon transforms does not distinguish it from a scalar particle.

We may also contemplate coupling a Higgs scalar to electrodynamics, and arranging for the U(1) gauge symmetry to be realized in a Higgs mode. In that event, there is a massive stable vortex particle in the spectrum of the theory that carries a unit of magnetic flux. Thus, the states $|n\rangle$ no longer become degenerate in the infinite-volume limit, and the true ground state is in the $n=0$ sector. In the Higgs phase, therefore, the topological symmetry is manifest. The operator ϕ acting on the vacuum creates a vortex particle.

Let us now imagine (following Polyakov [20]) that finite-action magnetic monopole configurations are permitted to contribute to the euclidean path integral of a U(1) gauge theory in 2 + 1 dimensions. Indeed, such monopoles will occur if a compact U(1) gauge theory is defined on a space-time lattice, or if U(1) is embedded in a nonabelian gauge group that breaks to U(1) via the Higgs mechanism. The monopoles provide a mechanism for the magnetic flux to change, and they therefore break the topological symmetry *intrinsically*. Hence, in the phase in which the topological symmetry is spontaneously broken, the would-be Goldstone boson actually acquires a small mass suppressed by the factor e^{-S_M}, where S_M is the euclidean action of the magnetic monopole. In the language of electrodynamics, the plasma of monopoles and antimonopoles induces Debye screening; correlation functions therefore decay exponentially and the theory has a mass gap. From either viewpoint, we see that magnetic monopoles prevent the photon from being exactly massless, in a U(1) gauge theory in 2 + 1 dimensions.

A further consequence of the plasma of magnetic monopoles is electric confinement. Electric fields cannot penetrate the plasma; they are confined to flux tubes. Electric confinement is easy to demonstrate by invoking Wilson's area-law criterion. We associate with an oriented closed loop C in euclidean space-time the Wilson loop operator

$$W(\mathrm{C}) = \exp\left(i\oint_{\mathrm{C}} A_\mu \mathrm{d}x^\mu \right). \tag{3.11}$$

If the vacuum expectation value of $W(\mathrm{C})$ exhibits the behavior

$$\langle 0|W(\mathrm{C})|0\rangle \sim \exp[-\kappa(\mathrm{Area})] \tag{3.12}$$

as the minimal area of a surface bounded by C grows asymptotically large, then we may identify κ as the string tension, or mass per unit length, of an electric flux tube. This area-law behavior of the Wilson loop vacuum expectation value is inevitable if magnetic fluctuations occur in the vacuum that become weakly correlated at large separation, for in that event, $\langle 0|W(\mathrm{C})|0\rangle$ behaves like a product of uncorrelated factors, with the number of factors growing like the area of the loop. And a dilute plasma of magnetic monopoles in euclidean space-time provides an ideal source of such magnetic fluctuations [20]. (In spite of their long-range magnetic fields, the monopoles are weakly correlated at large separation because of Debye screening.)

It is also amusing to note that we may interpret the area-law behavior of the Wilson loop in terms of the scalar field ϕ that is dual to the gauge field. Consider the euclidean Green function

$$\langle 0 | W(\mathrm{C}) \phi(x) | 0 \rangle , \tag{3.13}$$

where C is a very large loop, and imagine that the point x traverses a path in euclidean space-time that winds once around C and then returns to its starting point. Because a Dirac-charge magnetic monopole has in effect passed through C, and because the argument of the exponential in eq. (3.11) is the magnetic flux through a surface bounded by C, the phase of the Green function (3.13) advances by 2π as x traverses the closed path. In massless electrodynamics, the topological symmetry is spontaneously broken, and $\phi(x)$ has a nonvanishing vacuum expectation value. We may therefore interpret $W(\mathrm{C})$ as an operator that creates a *global string* along C, for the phase of the expectation value of ϕ has a nontrivial winding number on a loop that encircles C. If there are finite-action magnetic monopoles, however, the topological symmetry is intrinsically broken, and the phase of ϕ has a preferred value in the vacuum. It therefore becomes favorable for the advance of the phase of ϕ by 2π to be confined to a *wall* of finite thickness that terminates on the loop C. The action of the loop created by $W(\mathrm{C})$ thus grows like the area of the loop, and the area-law behavior of $\langle 0 | W(\mathrm{C}) | 0 \rangle$ follows. The "axion domain wall" that is bounded by C is just the world-sheet traced out in euclidean space-time by the electric flux tube.

While magnetic monopoles prevent $(2+1)$-dimensional electrodynamics from having a Coulomb phase in which the photon is exactly massless, they do not prevent a Higgs phase, in which the photon is massive but the Wilson loop exhibits perimeter-law behavior. In the Higgs phase, monopoles and antimonopoles are typically joined in euclidean space-time by *magnetic* flux tubes with nonvanishing action per unit length. Hence monopoles and antimonopoles are very strongly correlated; the vacuum is not magnetically disordered, and the electric string tension vanishes. We may think of the monopoles as sources that create or destroy the stable vortex particles. Because of the monopoles, the topological symmetry is intrinsically broken, and the vortex number is not exactly conserved. In the dual description of the Higgs phase, the operator $W(\mathrm{C})$ no longer creates a global string. Instead, the advance by 2π of the phase of the Green function (3.13) occurs when a magnetic flux tube that terminates at x crosses the loop C.

All of the above discussion applies directly to the $(2+1)$-dimensional abelian gauge theory with magnetic wormholes. The wormholes provide a plasma of magnetic monopoles and antimonopoles; in effect, the monopoles have a core size of order b_1, given by eq. (2.6), and a finite euclidean action $S_{\mathrm{E},1}$ of order $1/e^2 b_1$. In other words, when we couple ordinary noncompact electrodynamics to gravity in $2+1$ dimensions, magnetic wormholes convert it to compact electrodynamics with an effective "lattice spacing" of order the wormhole throat thickness.

As we saw above, the wormholes generate a photon mass and an electric string tension, both of order $\exp(-S_{E,1})$. In the Higgs phase, however, a wormhole swallows or emits the massive stable vortex particle, and therefore wormholes and antiwormholes are very strongly correlated [18, 19]. The wormholes thus fail to generate magnetic fluctuations that are uncorrelated at asymptotically large separation, and the electric string tension vanishes. The wormholes *do* break intrinsically the topological U(1) symmetry, so that vortex number is not precisely conserved.

This analysis also generalizes straightforwardly to the $(3+1)$-dimensional magnetic wormholes considered by Giddings and Strominger [3]. As we have already noted, the Giddings–Strominger wormhole acts like a monopole source for the three-form magnetic field strength $H = dB$, and the monopole strength must be quantized if we wish to couple the potential B to string sources. Associated with an oriented closed two-dimensional surface Σ in euclidean space-time we may define a Wilson surface operator

$$W(\Sigma) = \exp\left(i \int_{\Sigma} B_{\mu\nu}\, dx^{\mu}\, dx^{\nu} \right). \tag{3.14}$$

The vacuum expectation value of $W(\Sigma)$ probes the response of the vacuum to an external string source, just as the expectation value of $W(C)$ probes the response to an external charged particle source. The plasma of monopoles generated by magnetic wormholes causes the vacuum expectation value of $W(\Sigma)$ to exhibit the volume-law behavior

$$\langle 0 | W(\Sigma) | 0 \rangle \sim \exp[-\sigma(\text{volume})] \tag{3.15}$$

as the minimal volume bounded by Σ grows asymptotically large. The volume law signifies that the action of the object created by $W(\Sigma)$ grows like the volume enclosed by Σ. Again, this behavior is easy to interpret in the dual language. There is a Goldstone field $a(x)$ that is dual to H, and a spontaneously broken topological symmetry such that

$$a(x) \to a(x) + \alpha . \tag{3.16}$$

We may interpret $W(\Sigma)$ as an operator that creates a "global string" on Σ, such that the expectation value of the Goldstone field $a(x)$ advances by a nonzero constant as x traverses a closed path that encircles Σ. (Here we use the term "string" to denote an object that is of codimension two in euclidean space-time.) This means that $a(x)$ must be a periodic variable, and that is how, in the dual description, a string source enforces quantization of charge. The magnetic wormholes intrinsically break the topological symmetry, so that the periodic variable a has a preferred expectation value in the vacuum, and it becomes favorable for the advance of a by a constant to be confined to a wall of codimension one that is bounded by Σ. The action of the object created by $W(\Sigma)$ therefore grows like the volume enclosed by Σ, and the volume-law behavior of $\langle 0|W(\Sigma)|0\rangle$ follows.

Having completed our discussion of magnetic wormholes in an abelian gauge theory, we now turn to the magnetic wormholes of (2 + 1)-dimensional Yang–Mills theory coupled to gravity. To set the stage for this analysis, let us first briefly recall some qualitative features of ordinary (2 + 1)-dimensional Yang–Mills theory, without wormholes. As we have already emphasized, pure Yang–Mills theory with gauge group $G = SU(N)$ has a crucial topological property. Consider a gauge field configuration on a two-dimensional plane such that the field strength vanishes outside a compact region, and the compact region contains N units of magnetic flux. It is possible, then, to smoothly deform the field strength in the compact region to zero, even while the field strength remains vanishing at all times outside the compact region. Among the gauge field histories that contribute to the euclidean path integral, therefore, are histories in which magnetic fluctuations occur that are weakly correlated at large distances [21]. These fluctuations have no analog in the noncompact abelian gauge theory, but play a role closely analogous to that of the magnetic monopoles of the compact abelian theory – they magnetically disorder the vacuum and generate electric confinement. Thus, a (2 + 1)-dimensional Yang–Mills theory that does not undergo the Higgs mechanism is a confining theory with a mass gap. The mass scale of the theory is controlled by the gauge coupling e^2, which has the dimensions of mass. If Higgs fields are introduced and the Higgs mechanism occurs, however, there is a stable massive vortex particle, and magnetic flux in euclidean space-time is confined to tubes with nonvanishing action per unit length. Thus, the magnetic fluctuations are strongly correlated, and the electric string tension vanishes. There is still a mass gap, but electric fields are screened by Higgs fields, rather than confined to flux tubes.

It was emphasized by 't Hooft [17] that the choice between a confining or Higgs realization of the SU(N) gauge symmetry can be just as well described in terms of the realization of the topological Z_N symmetry. To appreciate this point, consider the euclidean Green function

$$\langle 0|W(C)\phi(x)|0\rangle, \tag{3.17}$$

where

$$W(C) = \mathrm{tr}\, P\exp\left(i\oint_C A_\mu \mathrm{d}x^\mu\right) \tag{3.18}$$

is the nonabelian Wilson loop operator associated with the oriented closed loop C, and $\phi(x)$ is the local operator that destroys a quantum of Z_N magnetic flux. The

Green function (3.17) has the important property of being *multivalued* [17]. If the point x traverses a path in euclidean space-time that winds once around the loop C (in the appropriate sense) and returns to its starting point, then the Green function is modified by the nontrivial multiplicative phase factor $\exp(2\pi i/N)$. This multivaluedness arises because, by winding x around C, we have in effect passed a Z_N magnetic monopole through the loop C, and $W(C)$ is an operator that detects the Z_N magnetic flux through a surface bounded by the loop C.

Multivalued functions are awkward to deal with, so it is for many purposes more convenient to force the Green function (3.17) to be single-valued by arbitrarily restricting it to one of its N branches. The price we pay for single-valuedness is that the Green function has a cut; for example, we may select an arbitrary surface bounded by the loop C and demand that the Green function jump discontinuously by the multiplicative factor $\exp(2\pi i/N)$ when x crosses the surface. Of course, although the Green function is either multivalued or has a kinematic discontinuity, if we make use of the Green function to address a well-posed physical question, the answer will be unambiguous and independent of the arbitrarily selected position of the cut.

Now, if the topological Z_N symmetry is spontaneously broken, and $\phi(x)$ has a nonvanishing vacuum expectation value, then the phase of the expectation value takes N possible values, corresponding to the N degenerate vacuum states. If we restrict the multivalued Green function to a single branch as described above, then we expect cluster decomposition to apply,

$$\langle 0|W(C)\phi(x)|0\rangle \to \langle 0|W(C)|0\rangle\langle 0|\phi(x)|0\rangle, \tag{3.19}$$

when x is far from the loop C. But if the Green function has a *kinematic* phase discontinuity across some surface bounded by C, then there must be a compensating physical discontinuity on a surface bounded by C, across which the phase of the Green function rotates by $-2\pi i/N$. This physical discontinuity is a domain wall with finite action per unit area, and area-law behavior of $\langle 0|W(C)|0\rangle$ follows [17]. Indeed, the domain wall is just the world-sheet of an electric flux tube. Just as one expects, then, a vacuum expectation value for $\phi(x)$, which signifies that magnetic fluctuations are copious in the vacuum, implies that electric flux is confined. (In fact, 't Hooft [17] also discovered a corresponding relation between magnetic disorder and electric confinement that applies in $3+1$ dimensions.) It is rather amusing to note that, in order that the Wilson loop operator create a string along C that is the boundary of a domain wall, it is necessary for the topological U(1) symmetry to be *intrinsically* broken in the abelian gauge theory, while it is necessary for the topological Z_N symmetry to be *spontaneously* broken in the nonabelian gauge theory.

If, on the other hand, the Higgs mechanism occurs, then $\phi(x)$ creates or destroys a stable massive vortex particle, and the topological Z_N symmetry is manifest. In this case, the advance by $2\pi/N$ of the phase of the Green function (3.17) actually occurs when a magnetic flux tube that terminates at x crosses the loop C [17].

We must now consider how the above picture becomes modified in the presence of magnetic wormholes. Because a quantum of Z_N magnetic flux disappears down the wormhole, the magnetic wormholes of an SU(N) gauge theory *intrinsically* break the topological Z_N symmetry, just as the magnetic wormholes of an abelian gauge theory intrinsically break the topological U(1) symmetry. The wormholes, then, provide magnetic fluctuations that tend to magnetically disorder the vacuum. Indeed, one may wonder whether intrinsic breaking of the Z_N symmetry, like spontaneous breaking, results in electric confinement. Could it be that a Higgs realization of the nonabelian gauge symmetry is not possible, in the presence of magnetic wormholes?

In contemplating this question, we confront a subtlety, namely that the Wilson area-law criterion for electric confinement no longer makes sense. The problem is that, as we noted above, $W(\mathrm{C})$ is a multivalued function on a gauge field background that contains Z_N magnetic charges. Hence, there is no satisfactory way of defining $\langle 0|W(\mathrm{C})|0\rangle$ if the path integral includes a fluctuating plasma of monopoles and antimonopoles. (If we restrict $W(\mathrm{C})$ to a single branch, as described before, then its value depends on the arbitrarily selected position of the branch cut.)

We have met with this difficulty because, while a wormhole that swallows a quantum of Z_N magnetic flux is perfectly acceptable in the pure gauge theory, such a wormhole ought to be disallowed when the gauge field couples to matter fields that transform according to the fundamental representation of SU(N). The fundamental-representation matter fields do not propagate consistently on the wormhole background, because they can detect the Dirac string of a Z_N magnetic monopole. (In fact, the same difficulty would have arisen in the abelian gauge theory, if we had attempted to evaluate in the presence of wormholes the expectation value of a Wilson loop for an electric source that carries, say, half the minimal electric charge allowed by the Dirac quantization condition.)

We see that if we wish to include in a nonabelian gauge theory matter that transforms faithfully under the center of the gauge group, then we must *not* include in the euclidean path integral any wormhole configurations in which a nontrivial amount of topologically conserved magnetic flux flows through the wormhole throat. We are left, then, with no magnetic wormhole solutions at all, for we have already rejected as unstable the solutions in which the magnetic flux is topologically trivial. (Incidentally, the situation encountered here is reminiscent of restrictions on quantum gravity that arise when gravity is coupled to spinors [22]; if spinors are included, then the sum over topologies must be restricted to *spin manifolds* on which the spinors can be consistently defined.)

This phenomenon, that the representation content of the matter fields places restrictions on the allowed wormhole configurations, can also be explained in a somewhat different language. If the gauge symmetry is realized in the Higgs mode, there is a stable vortex particle that carries a quantum of Z_N magnetic flux; in euclidean space-time, magnetic flux is confined to tubes with nonzero action per unit length. A magnetic wormhole can swallow a Z_N vortex, and so in the presence of such wormholes, the Z_N vortex number is not precisely conserved. In euclidean space-time, the Z_N flux tubes can terminate on magnetic monopoles. Now suppose that matter fields are introduced that transform nontrivially under the center of $SU(N)$. These matter fields exhibit a nontrivial Aharanov–Bohm effect when a Z_N vortex is circumnavigated. A paradoxical situation would arise if the vortex were permitted to spontaneously disappear down a wormhole; thus, such a wormhole must not be allowed. When phased this way, the restriction on wormholes described here is seen to closely resemble an observation made by Krauss and Wilczek [23]. They emphasized that wormholes cannot intrinsically break a discrete symmetry if the discrete symmetry is embedded in a gauge symmetry that has undergone the Higgs mechanism. The Z_N symmetry that we are considering here, however, is not a gauge symmetry; rather, it is a global symmetry with a topological origin.

Let us return now to our discussion of the pure nonabelian gauge theory, or to a theory in which the matter fields transform trivially under at least some subgroup of the center of the gauge group, so that magnetic wormhole solutions are allowed. We have seen that such wormholes induce an intrinsic breakdown of the topological symmetry of the theory. While this effect sounds novel, one should recognize that a similar phenomenon can be achieved without resorting to wormholes, by embedding the gauge group in a larger gauge group that undergoes the Higgs mechanism at a large mass scale. Indeed, the same remark applies to the U(1) gauge theory that we considered earlier. For example, we may imagine embedding a U(1) gauge theory in an SU(2) gauge theory with a hierarchy of symmetry breakdown

$$SU(2) \xrightarrow{v_1} U(1) \xrightarrow{v_2} \mathbf{1}; \tag{3.20}$$

here $v_{1,2}$ are the mass scales that characterize the two stages of symmetry breaking, and satisfy $v_1 \gg v_2$. Nonsingular magnetic monopoles of finite action are generated in the first stage of symmetry breakdown, and vortices, or magnetic flux tubes, are generated in the second stage. If we were aware of only the second stage of symmetry breakdown that occurs at low energy, we would conclude that the theory possesses a manifest topological U(1) global symmetry, and that vortex number is exactly conserved. But in fact, the magnetic flux tubes can terminate on the heavy magnetic monopoles that are generated at the first stage; thus, the topological symmetry is intrinsically broken. This symmetry breaking is a small effect for $v_1 \gg v_2$, because the magnetic monopoles are heavily suppressed by a factor of order e^{-S_M} where $S_M \sim \sqrt{v_1}/e$.

Actually, the above scenario applies only if the Higgs field that drives the second stage of symmetry breakdown transforms nontrivially under the center Z_2 of SU(2). If the Higgs fields transform trivially under the center, then there is an absolutely stable Z_2 vortex that survives in spite of the magnetic monopoles; the monopoles do not break the topological U(1) symmetry of the abelian gauge theory completely, but instead break it to a Z_2 subgroup. In this case, a pair of magnetic flux tubes, rather than a single tube, terminates on a monopole. A monopole is not the end of a flux tube; it slides along the tube like a bead on a string [24]. The same phenomenon would occur in a U(1) gauge theory with magnetic wormholes, if we introduced a matter field with electric charge equal to half the charge of the Higgs field, so as to disallow the wormhole with minimal magnetic flux.

By a similar strategy, we can construct a model (*without* wormholes) that duplicates the low-energy physics of a nonabelian gauge theory *with* magnetic wormholes. For example, an SU(2) gauge theory can be embedded in an SU(3) gauge theory with the symmetry breaking hierarchy

$$\mathrm{SU}(3) \xrightarrow{v_1} \mathrm{SU}(2)/\mathrm{Z}_2 \xrightarrow{v_2} \mathbf{1}. \tag{3.21}$$

Here SU(2) is embedded so that the fundamental 3 of SU(3) transforms as a 3 of SU(2); Higgs fields must transform nontrivially under the center Z_3 of SU(3)*, but of course all representations of SU(3) transform trivially under the center of SU(2). In this model, a Z_2 vortex arises in the second stage of symmetry breakdown. But once again the topological Z_2 symmetry is intrinsically broken, because the Z_2 magnetic flux tube can terminate on the heavy Z_2 magnetic monopole that arises at the first stage of symmetry breakdown. Notice that our difficulty in defining a Wilson loop operator in the fundamental representation of SU(2) has been cleverly evaded; any matter that we add to the theory must be in a representation of the underlying SU(3) gauge group, and hence transforms trivially under the center of SU(2).

These examples are nicely representative of the general case, which we will now briefly describe [15, 25]. Consider the symmetry breaking hierarchy

$$\mathrm{G} \xrightarrow{v_1} \mathrm{H} \xrightarrow{v_2} \mathrm{K}. \tag{3.22}$$

Magnetic monopoles of finite action that arise in the first stage of symmetry breakdown are associated with closed loops in the gauge group H that *cannot* be contracted to a point in H but *can* be contracted to a point in the larger group G. The homotopy class of the noncontractible loop in H may be identified as the magnetic charge of the monopole; it characterizes the behavior of the gauge field in the vicinity of the Dirac string. For the magnetic monopole to have a nonsingular core, and hence finite action, it is necessary that the nontrivial topological H magnetic charge carried by the monopole correspond to a trivial topological G

* For the first stage of symmetry breaking, a suitable choice is a Higgs field ϕ_{ij} in the two-index symmetric tensor representation of SU(3), with vacuum expectation value $\langle\phi_{ij}\rangle = \sqrt{v_1}\,\delta_{ij}$.

magnetic charge when H is embedded in G. Only then is it possible for a singularity to be avoided by exciting the G gauge fields in the core of the monopole. This is why the noncontractible loop in H that characterizes the Dirac string must be contractible in G. And because this loop is contractible in G, we are assured that no matter fields in any representation of G experience a nontrivial Aharanov–Bohm effect when they circumnavigate the Dirac string.

Now, when the second stage of symmetry breakdown occurs, the monopoles generated at the first stage may or may not survive. If the noncontractible loop in H that defines the monopole charge can be deformed to a loop that is entirely contained inside the new unbroken gauge group K, then the monopole does survive, with its long-range field duly converted to the field of a K monopole. But if this noncontractible loop in H cannot be deformed to a loop in K, then the monopole does not survive. The heavy H gauge fields must be excited in the vicinity of the Dirac string, and so the string becomes a physical object, a magnetic flux tube with nonzero energy per unit length. Thus, a flux tube (or perhaps several flux tubes) that is generated at the second stage of symmetry breakdown terminates on a monopole that was generated at the first stage. We emphasize again that no matter field in any representation of G experiences a nontrivial Aharanov–Bohm effect upon circumnavigating a flux tube that terminates on a monopole, in accord with the restriction that we placed on magnetic wormholes.

This analysis shows that the low-energy physics induced by magnetic wormholes in $2+1$ dimensions is not really so novel as it at first appears to be; the same physics can arise in an ordinary nonabelian gauge theory, without wormholes, that undergoes the Higgs mechanism at a large mass scale. Wormholes enable the gauge theory to contain nonsingular magnetic monopoles, without any enlargement of the gauge symmetry, and the corresponding violation of topological symmetry is in principle detectable at low energy. But a low-energy observer who is unable to resolve the core of the monopole cannot directly distinguish a monopole that is nonsingular due to a wormhole throat of finite thickness from a monopole that is nonsingular due to heavy gauge fields that are excited in its core.

It may seem, then, that the physical effects of magnetic wormholes are completely indistinguishable at low energy from the effects of conventional nonsingular magnetic monopoles that arise as a consequence of the Higgs mechanism. We wish to emphasize, though, an important distinction between these two possible sources of intrinsic breaking of topological symmetry. If the Higgs mechanism occurs at the mass scale v, and the gauge coupling e^2 is small compared to v, then the topological symmetry breaking induced by magnetic monopoles is necessarily extremely weak, since all effects of magnetic monopoles are suppressed by the very small factor e^{-S_M} where $S_M \sim \sqrt{v}/e$ is the euclidean action of the monopole. In contrast, the topological symmetry breaking due to magnetic wormholes need not be so weak, even if the gauge coupling is much smaller than b^{-1}, where b is the width of the wormhole throat.

Recall that the truly novel implication of wormholes is that low-energy physics is actually described by an *ensemble* of effective field theories labeled by the set of parameters $\{\alpha_n\}$. The expansion parameters of the semiclassical approximation are $|\alpha_n|e^{-S_n}$, so that, strictly speaking, semiclassical reasoning can be justified only for $|\alpha_n|e^{-S_n} \ll 1$. We may nonetheless imagine that αe^{-S} is comparable to one, even if e^{-S} is tiny. In this event, the coefficients of the terms in the effective action that break the topological symmetry are *not* small.

Indeed, it is reasonable to anticipate that αe^{-S} will be of order one in a scenario in which wormhole effects account for the vanishing of the cosmological constant [8]. While the terms in the effective action that are linear in α_n are restricted by the topological symmetry to be of the form (3.3), the terms of quadratic and higher order in α are not so severely restricted. Terms of quadratic order can be induced by correlations of wormhole pairs, or by conventional renormalization as the floating ultraviolet cutoff descends [26]. We expect, for example, that a contribution to the cosmological constant occurs in quadratic order in α of the form [10]

$$\delta\Lambda = \sum_n \lambda_n b_n^{-4} |\alpha_n|^2 e^{-2S_n} + \cdots . \tag{3.23}$$

Here b_n is the thickness of the wormhole of type n, and S_n is the corresponding semiwormhole action; the λ's are numerical constants of order one. If $\delta\Lambda$ in eq. (3.23) is to cancel a bare cosmological constant that is of order one in units of the wormhole length scale, then αe^{-S} must assume a value of order one.

4. Magnetic wormholes in 3 + 1 dimensions

In the concluding portion of this paper, we will discuss some wormhole solutions that occur in (3 + 1)-dimensional Yang–Mills theory coupled to gravity. These solutions, unlike the (2 + 1)-dimensional solutions that we have considered previously, rely for their existence on the nonabelian character of the gauge group; they have no abelian analog. The enormous virtue of these solutions is that they occur in the standard model of elementary particle interactions (including gravity); furthermore, they share with our solutions in (2 + 1)-dimensional Yang–Mills theory the desirable property of having a maximum throat thickness. Unfortunately, these solutions are not stable, and we do not believe that they relate to the semiclassical tunneling amplitude for any physically realizable quantum mechanical process.

We have observed that the throat of a wormhole can be supported by a magnetic field. Some years ago, in fact, de Alfaro et al. [27] described a spherically symmetric singular solution to the classical Yang–Mills equations on flat four-dimensional euclidean space with a (nonabelian) magnetic field that is well suited to support a wormhole. Their solution, which became known as a "meron", has much in common with the singular monopole solutions of gauge theories in three-dimensional euclidean space. Like the three-dimensional monopole, the four-dimensional meron has a field strength that varies as a function of the distance r from its center like r^{-2}. Furthermore, the field strength is purely "magnetic"; that is,

$$F^a_{ri} = 0, \tag{4.1}$$

where $\hat{r}$ denotes the radial direction and $\hat{i}$ is a direction orthogonal to $\hat{r}$. Thus, the "electric" field strength vanishes if we identify the radius r as a euclidean "time."

When (3 + 1)-dimensional Yang–Mills theory is coupled to gravity, then, we may construct a spherically symmetric nonsingular wormhole solution that is very similar to our (2 + 1)-dimensional magnetic wormholes; the singularity at the center of the meron is avoided due to the finite thickness of the wormhole throat. This solution was described by Hosoya and Ogura [11].

Let us disregard gravity for the time being, and consider in more detail the properties of the meron solution on flat euclidean space. For the gauge group G = SU(2), the meron gauge potential may be written as

$$A(r, \Omega) = \tfrac{1}{2} g^{-1} \mathrm{d} g, \tag{4.2}$$

where $g(r, \Omega)$ is the standard identity map from the three-sphere (parametrized by three angles that we have schematically denoted by Ω) to the group SU(2). Because of the crucial factor of $\frac{1}{2}$ in eq. (4.2), this potential is not merely a gauge transformation of $A = 0$. Indeed, the components of the field strength $F = \mathrm{d}A + A^2$ take the values

$$F^a_{ri} = 0, \qquad F^a_{ij} = \epsilon_{aij}/r^2. \tag{4.3}$$

Note that all three noncommuting components of the SU(2) magnetic field strength are excited. For this reason, the commutator term in the Yang–Mills equation is nonvanishing, and the meron is an intrinsically nonabelian object; it solves the nonlinear Yang–Mills equation but not the linear Maxwell equation. Obviously, we can find a meron solution for any nonabelian gauge group G, by choosing an embedding of SU(2) in G.

Since the meron field strength behaves like r^{-2}, its euclidean action diverges logarithmically at both $r = 0$ and $r = \infty$. The two singularities at $r = 0$ and $r = \infty$ are really identical in form, and we may by a conformal transformation convert the spherically symmetric solution to a "meron pair" configuration that is nonsingular at $r = \infty$ and has two singular points that are separated by a finite distance R. (Since the classical Yang–Mills equations are conformally invariant, a conformal transformation always takes a classical solution to another classical solution.) If the two singular meron cores are smeared slightly, then the new configuration (which is not an exact solution) has a finite action that increases logarithmically with the separation R.

We may gain a deeper appreciation of the interpretation of this object if we return to the spherically symmetric form of the meron and evaluate on each concentric three-sphere the "charge"

$$Q = \int_{S^3} \omega_{cs} . \tag{4.4}$$

Here ω_{cs} is the Chern–Simons form

$$\omega_{cs} = \frac{1}{8\pi^2} \operatorname{tr}\left(A\, dA + \tfrac{2}{3}A^3\right), \tag{4.5}$$

which satisfies

$$d\omega_{cs} = \frac{1}{8\pi^2} \operatorname{tr}(F^2) . \tag{4.6}$$

The quantity on the right-hand side of eq. (4.6) is the Pontriagin topological charge density, or instanton density. Therefore, the total instanton number in the four-volume bounded by two nested closed three-surfaces is the difference between the values of the charge Q on the two surfaces.

For the spherically symmetric meron, a simple calculation shows that $Q = \frac{1}{2}$ on each three-sphere of radius r, for $0 < r < \infty$; hence, the instanton density vanishes almost everywhere. (This is quite clear, in fact, since the electric field vanishes.) However, half a unit of instanton number is localized at each of the two singular points, at $r = 0$ and $r = \infty$.

Therefore, we may think of the conformally-equivalent meron pair configuration, in which the singular points are separated by a finite distance R, as a distorted instanton. To construct the meron pair, we slice an instanton in half, separate the two fragments by the distance R, and then shrink the fragments to zero size. The long-range field of an instanton, at a distance from the instanton center that is large compared to the size of the instanton core, approaches a mere gauge transformation of $A = 0$. Thus, instantons (and anti-instantons) of zero size are noninteracting. But the long-range field of *half* an instanton is not a pure gauge, so that merons, unlike instantons, have logarithmic long-range interactions.

When Yang–Mills theory is coupled to gravity, the magnetic field of the meron supports the throat of a wormhole, and the singular core of the meron expands to a nonsingular throat of finite width [11]. The width b of the throat is given by

$$b^2 = 4\pi G/e^2 , \tag{4.7}$$

where G is Newton's constant and e is the Yang–Mills coupling; the width is large compared to the Planck length if the gauge coupling is weak. Since the long-range field strength of the meron decays like r^{-2}, the euclidean action of this wormhole diverges logarithmically in the infrared. The action behaves like

$$S \sim \frac{3\pi^2}{e^2} \ln R \,, \tag{4.8}$$

if R is the separation, as measured on the background space-time to which the wormhole attaches, between the two ends of the wormhole.

We saw that the $(2+1)$-dimensional magnetic wormholes could be characterized by the value of the topologically conserved magnetic flux that disappears down the wormhole. Similarly, we might characterize the $(3+1)$-dimensional meron wormhole by saying that half a unit of the charge Q defined by eq. (4.4) disappears down the wormhole. This characterization is not entirely satisfactory, however, for two reasons.

First, the charge Q is not gauge invariant. Though Q is *not* changed by "small" topologically trivial gauge transformations, it *is* changed by "large" topologically nontrivial gauge transformations. Fortunately, the large gauge transformations change Q by just an integer. Thus, Q modulo an integer is a well-defined gauge-invariant quantity that takes values in the unit interval $[0, 1)$. One might expect, then, that our meron wormhole can be generalized – that it is possible to construct a wormhole solution in which the charge Q that disappears down the wormhole takes any nontrivial value in $(0, 1)$. This generalization does indeed exist.

In fact, Callan et al. [28] described a construction of spherically symmetric "generalized merons" in which the instanton number localized at the center takes any desired value. All of these solutions, when coupled to gravity, are capable of supporting the throat of a wormhole. For each value of $Q \in (0, 1)$, then, a spherically symmetric wormhole solution may be constructed such that that value of Q disappears down the throat. These solutions may be chosen so that the electric field vanishes at the middle of the throat; we require that the wormhole meet this condition so that a semiwormhole can be continuously matched up with real-time evolution, which is necessary in order that the wormhole admit a semiclassical interpretation in terms of the quantum mechanical nucleation of a baby universe. Except for the meron ($Q = \frac{1}{2}$) solution, though, the electric field is nonvanishing elsewhere, away from the middle of the throat. All of the wormholes for $Q \neq \frac{1}{2}$ have throat thickness less than b given by eq. (4.7); in fact, the thickness approaches zero as Q approaches zero or one. All also have logarithmically divergent euclidean action.

The more serious problem with characterizing the wormholes in terms of the charge Q is that Q is not conserved. Indeed, we have already noted that the change in Q between two nested three-surfaces is the instanton number enclosed by the two surfaces, and this instanton number need not be an integer. Since the charge that disappears down the throat is not a conserved charge, our $(3+1)$-dimensional wormhole solutions are actually closely analogous to the $(2+1)$-dimensional wormholes that are afflicted by the Brandt–Neri–Coleman [14,15] instability. One suspects, then, that all of the $(3+1)$-dimensional magnetic wormholes are unstable, and in fact they are. A widely separated meron pair, for example, has a euclidean action that increases logarithmically with the separation R. It is clear that the action can be reduced by dressing each meron with half a unit of instanton number [28]. The dressed merons then have the long-range fields of instantons, and hence are noninteracting for large separation; the action of the dressed meron pair is independent of R. Furthermore, the electric and magnetic fields are of equal strength in the core of the dressed meron, so that the core cannot support the throat of the wormhole. Instead, the throat prefers to pinch off. A similar instability afflicts all of the generalized meron wormhole solutions.

Because of this instability, it is difficult to make sense of the sum over the small fluctuations of the gauge field about the wormhole solution. The instability signifies that the meron wormhole is not the stationary configuration of lowest euclidean action for any reasonable choice of boundary conditions. We are therefore reluctant to interpret the meron wormhole solution in terms of the semiclassical evaluation of a quantum mechanical tunneling process. Note that the gauge field instability has a quite different status than the well-known tendency [4] of fluctuations in the conformal degree of freedom of the metric to decrease the gravitational action; the latter is not typically indicative of the existence of another classical solution of lower action.

We have some confidence that a wormhole solution is relevant to semiclassical physics if, by imposing suitable initial and final boundary conditions on a single connected component of the three-dimensional geometry, we can ensure that the solution contributes appreciably to the euclidean path integral. Our unstable $(3+1)$-dimensional magnetic wormholes do not have this property. But our stable $(2+1)$-dimensional wormholes do, because they are the solutions of lowest action that interpolate between connected three-geometries that carry different values of the topologically conserved magnetic flux.

An unstable wormhole will be the solution of lowest action only if boundary conditions are imposed on disconnected three-geometries, and in particular only if we require that the initial and final three-geometries have a different number of connected components. It seems to us perverse to claim that these solutions relate to a semiclassically allowed tunneling process, for the throat of the wormhole is locally unstable, and would prefer to pinch off if it were permitted to.

This viewpoint may lead one to the hypothesis that the only wormhole solutions that correspond to the semiclassical nucleation of a baby universe are those such that a nonzero quantity of a global charge flows down the throat of the wormhole, where the charge would be exactly conserved in the absence of wormhole effects. We expect, actually, that it is also possible to attach a semiclassical interpretation to wormholes that are supported by a charge that is not exactly conserved, if charge conservation holds to a sufficient approximation. Indeed, a systematic analysis of wormholes supported by approximately conserved global charges may shed light on the viability of the solution to the large wormhole problem that was proposed in refs. [9, 10].

The generalized meron wormholes that we have described here were also discussed recently by Rey [29]. We thank him and Joe Polchinski for useful discussions.

Note added in proof

We have learned that meron wormholes have also been constructed by Verbin and Davidson [30], and by Tomimatsu [31].

References

[1] S.W. Hawking, Phys. Lett. B195 (1987) 337; Phys. Rev. D37 (1988) 904
[2] G.V. Lavrelashvili, V. Rubakov and P.G. Tinyakov, JETP Lett. 46 (1987) 167
[3] S.B. Giddings and A. Strominger, Nucl. Phys. B306 (1988) 890
[4] S.W. Hawking, *in* Relativity, groups, and topology II, ed. B. DeWitt and R. Stora (North-Holland, Amsterdam, 1983)
[5] K. Lee, Phys. Rev. Lett. 61 (1988) 263
[6] S. Coleman, Nucl. Phys. B307 (1988) 867
[7] S.B. Giddings and A. Strominger, Nucl. Phys. B307 (1988) 854
[8] S. Coleman, Nucl. Phys. B310 (1988) 643
[9] B. Grinstein, Nucl. Phys. B321 (1989) 439
[10] S. Coleman and K. Lee, Phys. Lett. B221 (1989) 242
[11] A. Hosoya and W. Ogura, Phys. Lett. B225 (1989) 117
[12] C. Teitelboim, Phys. Lett. B167 (1986) 63,69;
R. Rohm and E. Witten, Ann. Phys. (N.Y.) 170 (1986) 454
[13] R.C. Myers, Phys. Rev. D38 (1988) 1327
[14] R. Brandt and F. Neri, Nucl. Phys. B161 (1979) 253
[15] S. Coleman, *in* The unity of the fundamental interactions, ed. A. Zichichi (Plenum, London, 1983)
[16] C.W. Misner and J.A. Wheeler, Ann. Phys. (N.Y.) 2 (1957) 525
[17] G. 't Hooft, Nucl. Phys. B138 (1978) 1
[18] S. Coleman and K. Lee, Wormholes made without massless matter fields, Harvard preprint HUTP-89/A022 (1989)
[19] L.F. Abbott and M.B. Wise, Nucl. Phys. B325 (1989) 687
[20] A.M. Polyakov, Phys. Lett. B59 (1975) 82; Nucl. Phys. B120 (1978) 477
[21] R.P. Feynman, Nucl. Phys. B188 (1981) 479
[22] S.W. Hawking and C.N. Pope, Phys. Lett. B73 (1978) 42
[23] L.M. Krauss and F. Wilczek, Phys. Rev. Lett. 62 (1989) 1221
[24] M. Hindmarsh and T.W.B. Kibble, Phys. Rev. Lett. 55 (1985) 2398
[25] J. Preskill, Ann. Rev. Nucl. Part. Sci. 34 (1984) 461; *in* Architecture of fundamental interactions at short distances, ed. P. Ramond and R. Stora (North-Holland, Amsterdam, 1987)
[26] J. Preskill, Nucl. Phys. B323 (1989) 141;
A.K. Gupta and M.B. Wise, Phys. Lett. B218 (1989) 21
[27] V. de Alfaro, S. Fubini and G. Furlan, Phys. Lett. B65 (1976) 163
[28] C.G. Callan, R. Dashen and D.J. Gross, Phys. Rev. D17 (1978) 2717
[29] S.-J. Rey, Spacetime wormholes with Yang–Mills fields, UCSB-TH-89/18 (1989)
[30] Y. Verbin and A. Davidson, Phys. Lett. B229 (1989) 364
[31] A. Tomimatsu, Prog. Theor. Phys. 59 (1978) 2096