Wormholes, baby universes, and causality

Matt Visser*
Theoretical Division, T-8, Mail Stop B-285, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
(Received 27 December 1988; revised manuscript received 2 November 1989)

In this paper wormholes defined on a Minkowski signature manifold are considered, both at the classical and quantum levels. It is argued that causality in quantum gravity may best be imposed by restricting the functional integral to include only causal Lorentzian spacetimes. Subject to this assumption, one can put very tight constraints on the quantum behavior of wormholes, their cousins the baby universes, and topology-changing processes in general. Even though topology-changing processes are tightly constrained, this still allows very interesting geometrical (rather than topological) effects. In particular, the laboratory construction of baby universes is *not* prohibited provided that the "umbilical cord" is never cut. Methods for relaxing these causality constraints are also discussed.

I. INTRODUCTION

Recently the contemplation of wormholes and baby universes has become a growth industry. Most significantly, baby universes have been invoked to explain the vanishing of the cosmological constant.[1–6] These efforts have largely avoided discussing the constraints imposed by causality. Indeed, in most discussions the calculations are performed in "Wick rotated" Euclidean quantum gravity, where causality is not manifest. From the viewpoint of Euclidean quantum gravity the process of baby-universe emission "followed" by reabsorption can at best be interpreted as a schematic description of a quantum tunneling process.

In this paper we shall pursue a different objective. We shall be interested in processes taking place in ordinary Minkowski signature spacetime. This discussion extends a program carried out by Morris and Thorne,[7] by Morris, Thorne, and Yurtsever,[8] and by the present author.[9–12] We shall not demand that the Einstein field equations be satisfied, so our conclusions will also apply to quantum gravity, insofar as quantum gravity may be viewed as a functional integral over Minkowski signature spacetimes. We shall be very conservative, and take causality constraints extremely seriously. As the principal causality constraint we shall use the notion of "stable causality" as defined, for instance, by Hawking and Ellis[13] and by Wald.[14] Adopting the constraint of stable causality severly inhibits topology-changing processes. In particular we shall see that "transient" wormholes are forbidden by causality. Further, in the process of baby-inverse parturition we shall see that the "umbilical cord" connecting a baby universe with its parent universe can never be severed. The nonsevering of the umbilical cord is a necessary, but not sufficient, condition for the prevention of causality-violating processes in baby-universe production and absorption.

We shall also discuss the constraints that causality places on "permanent" wormholes.[15] Causality constraints imply that "permanent wormholes" cannot be created or destroyed, but must be built into the multiverse *ab initio.*[16] If a "permanent" wormhole connects two distant regions of the same universe then severe causality violations may occur. We shall discuss ways of mollifying such potentially serious problems.

II. QUANTUM GRAVITY

Quantum gravity is currently a patchwork of rather ill-understood techniques. Bearing this in mind, all quantum-mechanical comments made in this paper should be taken with a grain of salt. It is fair to say that the functional integral approach to quantum gravity is presently ascendant. In this approach, quantization is viewed as the process of performing a functional integration over the set of "all metrics." Unfortunately, there is very little agreement as to exactly what constitutes the set of "all metrics."

In the popular approach pioneered by Hawking,[17] one "Wick rotates" the Einstein-Hilbert action to Euclidean signature manifolds and performs the functional integration over all Euclidean metrics. Unfortunately, this approach has serious drawbacks. Firstly, because one is dealing with Euclidean signature metrics, causality is not manifest, and one does not know how to formulate, discuss, or understand the causal properties of Euclidean quantum gravity. Secondly, the unboundedness of the Euclidean action leads to the so-called "conformal factor problem," for which a number of *ad hoc* and unsatisfactory resolutions have been suggested. The situation has recently been clarified by Mazur and Mottola,[18] who have argued that the conformal factor problem can be avoided by correctly identifying the (physical) Lorentzian degrees of freedom. Then if one really insists on a Euclidean partition function, one may analytically continue the Lorentzian degrees of freedom. Attacking the problem from a different direction, Suen and Young[19] and Cline[20] have shown that minisuperspace quantum cosmology is much better behaved when the functional integral is taken over Lorentzian signature Robertson-Walker metrics. Similar comments may be found in work due to Farhi.[21]

Formulating quantum gravity as a functional integral over Lorentzian spacetimes unfortunately does seem to entail the abandonment of (at least the current understanding of) the rather beautiful Hartle-Hawking prescription[22] for the boundary condition on the wave function of the Universe. This would seem to lend support to the alternative boundary conditions proposed by Suen and Young,[19] Cline,[20] and Vilenkin.[23]

The suggestion made above lends itself to further refinement: it seems unreasonable to integrate over *all* Lorentzian signature metrics. It would appear to be much safer to restrict the integration only to that set of metrics that does not violate causality. Thus one views quantum gravity as the functional integral:

$$Z = \int_L \mathcal{D}g \exp\left[-i \int \sqrt{g}\, R\, dx\right], \tag{2.1}$$

the set L consisting of the set of causality-preserving Lorentzian spacetimes. A large part of this paper will be devoted to elaborating on this simple proposal.

III. LOCAL CAUSALITY: LORENTZIAN MANIFOLDS

The existence of an everywhere Lorentzian metric on a manifold severely constrains the topology of the manifold (even before any questions of global causality are addressed). The appropriate mathematical machinery is that of the theory of "Lorentzian cobordisms," as discussed by Sorkin[24] and Borde.[25] The main results can be summarized in the following lemmas.

Lemma. Any compact (with or without boundary) Lorentzian even-dimensional spacetime has Euler characteristic zero.

Proof. If the spacetime is time orientable, this follows by noting that any time-orientable spacetime admits (by definition) a globally defined timelike vector field. But the existence on an even-dimensional manifold of any globally defined vector field (be it timelike or otherwise) implies that the Euler characteristic is zero. If the spacetime is not time orientable one proceeds by considering the double cover (which is time orientable). By the previous argument it follows that the double cover has Euler characteristic zero, and thus that the original spacetime also has Euler characteristic zero.

Lemma. Any compact Lorentzian odd-dimensional spacetime has a boundary that is the disjoint union of two sets whose Euler characteristics are equal. That is,

$$\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2,\ \partial\Omega_1 \cap \partial\Omega_2 = \varnothing,\ \chi(\partial\Omega_1) = \chi(\partial\Omega_2)\ .$$

Proof. See Sorkin.[24] A discussion of the consequences of this lemma may also be found in Borde.[25]

For the remainder of the paper we shall assume that all spacetimes are both time orientable and space orientable (and so are "completely" orientable). The assumption of time orientability is justified by the simple observation that we can (macroscopically) tell the difference between past and future. At the quantum level, the observed breaking of time-reversal invariance argues in favor of a time-orientable spacetime. The assumption of space orientability is justified by the observation of parity violation in elementary-particle physics. A more detailed discussion of this topic is given in Hawking and Ellis,[13] pp. 181 and 182. This restriction serves to exclude some of the more interesting examples discussed by Sorkin[24] and Borde.[25] While the use of time (and space) nonorientable examples serves to clarify some of the concepts they discuss, we shall view such models as not physically relevant. Application of these comments to (transient) wormholes is immediate, though we shall have to address various dimensionalities separately.

0+1 dimensions. There are only four distinct manifolds in $0+1$ dimensions: $\mathcal{R}$, $S^1 \equiv \mathcal{R}/Z$, $[0,\infty)$, and $[0,1]$. These are all trivial and clearly any topology change is excluded.

1+1 dimensions. If an everywhere Lorentzian metric is placed on a two-dimensional orientable manifold without boundary then (1) if the manifold is not compact it is diffeomorphic either to the Minkowski plane $\mathcal{R}^2 = \mathcal{R}\times\mathcal{R}$ or to a cylinder $\mathcal{R}\times S^1$. (2) if the manifold is compact it is diffeomorphic to a torus $T^2 = S^1 \times S^1$. In none of these cases is there any possibility of topology change in $1+1$ dimensions.

This discussion has an amusing consequence when applied to string theories. Let us work in a Lorentzian signature spacetime ($25+1$ for bosonic strings, $9+1$ for superstrings, or $3+1$ for "physical" strings). Let us further demand that the world-sheet metric be everywhere Lorentzian. Then there cannot be any string interactions, and the string must be a free string. This follows by observing that the previous discussion implies that the world sheet swept out by the string must have the topology of a cylinder. It follows that in an interacting string theory in a Lorentzian spacetime the world-sheet metric *cannot* be everywhere Lorentzian. There must be at least a finite number of points where the Lorentzian signature of the world-sheet metric breaks down. This may be viewed as a breakdown of the world-sheet equivalence principle (this point of view has been emphasized by Sorkin,[24] see also the discussion of Hartle[26]). Whether a breakdown of the world-sheet equivalence principle leads to a breakdown of the spacetime equivalence principle is presently unknown, but seems somewhat unlikely.

2+1 dimensions. If a Lorentzian $2+1$ cobordism interpolates between an initial spacelike hypersurface $\partial\Omega_1$ and a final spacelike hypersurface $\partial\Omega_2$ then their Euler characteristics must be equal. Since the Euler characteristic completely characterizes the topology of compact orientable manifolds in two dimensions it follows that the initial and final spacelike hypersurfaces are homeomorphic. Thus topology change is completely prohibited in $2+1$ classical/quantum gravity.

3+1 dimensions. The situation now becomes considerably more complicated. By assumption, far away from the wormhole the metric settles down to the Minkowski metric. Apply periodic boundary conditions to compactify the spacetime. The resulting spacetime is compact without boundary, and by application of the previous lemma has an Euler characteristic of zero. On the other hand, existence of the (transient) wormhole would seem to guarantee that the resulting compactified spacetime has nonzero Euler characteristic. This, unfortunately, is not the case. The counting for Euler characteristics in $3+1$ dimensions is such that any nonzero Euler number arising from the wormhole itself can be compensated by surgically grafting other fixtures onto the spacetime. This procedure is discussed in detail by Borde,[25] to whom much of the discussion of this section is due. We cannot conclude that the existence of a Lorentzian metric on a four-dimensional manifold by itself prohibits topology change. To see this, consider the connected sum $\mathcal{M}\#\mathcal{N}$ of two four-dimensional manifolds defined by removing a four-dimensional disk from each of $\mathcal{M},\mathcal{N}$ and identifying the boundaries. Then $\chi(\mathcal{M}\#\mathcal{N})=\chi(\mathcal{M})+\chi(\mathcal{N})-2$. Now $\chi(T^4)=0$, $\chi(S^2\times S^2)=4$, and $\chi(\mathrm{CP}^2)=3$. (Recall that CP^2 is orientable.) Thus,

$$\begin{aligned}
&\chi(\mathcal{M}\#\mathrm{CP}^2)=\chi(\mathcal{M})+1\ ,\\
&\chi(\mathcal{M}\#T^4)=\chi(\mathcal{M})-2\ ,\\
&\chi(\mathcal{M}\#(S^2\times S^2))=\chi(\mathcal{M})+2\ ,
\end{aligned} \tag{3.1}$$

and we see that repeated acts of surgery of this type serve to bring the Euler characteristic to zero.

$(4k+1)+1$ dimensions. The situation in $(4k+1)+1$ dimensions is interesting for a number of reasons. Firstly this is the relevant dimensionality for $(9+1)$-dimensional superstrings, and for $(25+1)$-dimensional bosonic strings. Secondly, within the more general context of Kaluza-Klein theories, Witten has shown[27] that only in $(4k+1)+1$ dimensions does one obtain chiral fermions after compactifying to $3+1$ dimensions. The novelty in this case is that there are no orientable closed $(4n+2)$-dimensional manifolds of odd Euler characteristic.[25] One could try to use RP^{4k+2} which has $\chi(RP^{4k+2})=1$, but this manifold is not orientable, and so is excluded by the previous discussion. Equation (3.1) is replaced by

$$\begin{aligned}
&\chi(\mathcal{M}\#T^{4k+2})=\chi(\mathcal{M})-2\ ,\\
&\chi(\mathcal{M}\#(S^2\times S^{4k}))=\chi(\mathcal{M})+2\ .
\end{aligned} \tag{3.2}$$

We see that we can bring any even Euler characteristic down to zero by repeated acts of surgery. This places a rather mild constraint on the existence of a Lorentzian metric in $(4k+1)+1$ dimensions.

Summary. In $0+1$, $1+1$, and $2+1$ dimensions, classical and quantum topology changing processes are forbidden merely by the existence of an everywhere Lorentzian metric. In $3+1$ dimensions the existence of an everywhere Lorentzian metric is not in and of itself a sufficient condition to prevent topology change. In $(4k+1)+1$ dimensions there is a very mild restriction on topology change. In the next section we shall investigate the much stronger effects that global causality constraints have on this situation.

IV. GLOBAL CAUSALITY

It is fair to say that most conservative physicists have very serious reservations about the admissibility and reality of causality-violating processes. Causality violation (i.e., the existence of a "time machine") is such an extreme violation of our understanding of the cosmos that it behooves us to be as conservative as possible about introducing such unpleasant effects into our models. In this section we shall collect a number of well-known technical results relevant to our discussion.

At a minimum we should exclude from our models the existence of closed timelike loops. (This is known as the chronology condition.[13,14]) The existence of closed timelike loops leads to such unpleasant situations as meeting oneself five minutes ago. It is argued by many authors (e.g., Hawking and Ellis[13] and Wald[14]) that it is prudent to impose the stronger condition known as stable causality. The stable causality conditions states, roughly, that the Universe is not "on the verge of" violating the chronology condition. More precisely, this may be formulated as following.

Definition. A spacetime $(\mathcal{M},g)$ is stably causal if and only if there exists a continuous timelike vector field t^α such that the metric $\tilde{g}_{\alpha\beta}=g_{\alpha\beta}-t_\alpha\cdot t_\beta$ satisfies the chronology condition.

This definition implies that the set of all stably causal metrics constitutes an open subset (in the C^0 open topology) of the space of all Lorentzian metrics.[13] From the functional integral point of view, the imposition of stable causality on the multiverse is nothing more complicated than a simple restriction on the domain of the functional integral. A little work now suffices to prove the standard theorem.

Theorem. A spacetime $(\mathcal{M},g)$ is stably causal if and only if there exists a function τ (not unique), such that $\nabla\tau$ is a timelike vector field.

The hypersurfaces $\Sigma_t=\{p\in\mathcal{M}\,|\,\tau(p)=t\}$ are spacelike hypersurfaces. If *all* the spatial sections Σ_t are compact then they are diffeomorphic to each other and the manifold $\mathcal{M}$ has the topology $\mathcal{M}\sim\mathcal{R}\times\Sigma$. This already severely constrains topology-changing processes in a stably causal spacetime in that it indicates that topology change can occur only by invoking noncompact spatial sections. (More crudely, by processes at spatial infinity.[28]) These comments may be refined by a simple lemma.

Lemma. Let $(\mathcal{M},g)$ be a stably causal spacetime, and let Ω be a compact subset of $\mathcal{M}$ whose boundary $\partial\Omega$ is diffeomorphic to S^3, then Ω is diffeomorphic to B^4.

Proof. Since $\partial\Omega$ is diffeomorphic to S^3 we can perform a smooth cut and paste operation. Consider spatially compactified Minkowski space $\mathcal{R}\times[0,1]^3=\mathcal{R}\times T^3$, i.e., Minkowski space with spatially periodic boundary conditions. By inspection, spatially compactified Minkowski space is stably causal. Cut a ball out of this space and smoothly insert Ω into the hole. This yields a stably causal manifold all whose sections are compact. By the previous comment all of these sections are diffeomorphic, hence $\Omega\approx B^4$ as claimed.

This may (loosely) be interpreted as saying that topology changes cannot occur in bounded regions of a stably causal spacetime. Similar theorems date back (at least) to the pioneering work of Geroch[30] and Tipler.[31] Essentially similar results may be obtained from the somewhat stronger requirement of *global hyperbolicity*.[13,14] Global hyperbolicity may actually be a slightly too strong condition in that it completely forbids topology-changing processes (i.e., $\mathcal{M}\sim\mathcal{R}\times\Sigma$).

The preceding discussion implies that the usual picture of baby-universe creation engenders severe causality violations. It has recently become an article of folklore that "causality violations are suppressed by the Planck scale." This assertion should be viewed with extreme distrust. Crudely put, the occurrence of Euclidean wormholes of radius ρ are expected to be suppressed by a Boltzman factor: $\exp[-(\rho/L_{\rm Planck})^2]$. Fischler and Susskind[4] have shown that, contrary to this naive expectation, large Euclidean wormholes are not suppressed and in fact occur at all scales up to macroscopic scales. The controversy is continued in Refs. 5 and 6. This particular argument does not seem to have any simple analogue in Minkowski signature. The objection raised in this paper is a more fundamental matter of principle. It seems unreasonable to assert that a quantum average over an inconsistent (causality-violating) microphysics leads to a consistent causal low-energy effective theory. Any theory that is "just a little bit causality violating" is "just a little bit inconsistent."

With this machinery under our belts we shall now turn to the implications of these results for wormhole physics. The canonical picture of a wormhole (transient variety) is of the parent universe emitting a (perhaps virtual) baby universe "followed" by "subsequent" (of course for Euclidean wormholes the words "followed" and "subsequently" are meaningless) reabsorption of the baby universe (Fig. 1). Such a process, however, *cannot* take place in a stably causal spacetime, by application of the preceding lemma. This simple point is of sufficient import to warrant extensive additional commentary.

To summarize this section, the canonical picture of transient wormholes leads to serious causality violations. The processes described by Fig. 1 is qualitatively unphysical, and cannot occur in physically acceptable classical spacetimes. If one grants the previously argued viewpoint, that quantum gravity should be thought of as a functional integration over causal Lorentzian spacetimes, then this conclusion will also hold at the quantum level. To avoid causality violations, and retain some semblance of the usual wormhole physics, one must then make qualitative and severe changes to our picture of the phenomenon.

V. UMBILICAL CORDS

Having decided that the canonical picture of transient wormholes (Fig. 1) is unphysical, we shall now be interested in deducing the minimal violence that must be done to this picture in order to avoid causality-violating processes.

We shall impose stable causality as the basic causality constraint on the multiverse. One of the earlier lemmas implies that localized topology-changing processes are forbidden. Thus, to quote Hawking and Ellis, "topology (is) very dull." Two points need to be made: (1) even though the topology is boring, the geometry can be very interesting indeed and (2) the boring topology is the mathematicians' topology. We shall argue that physicists should really consider a much more interesting energy-dependent "physicists' topology." The closest causality-preserving analogue of the canonical transient wormhole picture is that exhibited in Figs. 2 and 3. These diagrams are to be interpreted as follows. One starts by "blowing a bubble in spacetime." This bubble then grows while its connection with the parent universe shrinks. However, one does not permit the connection between the parent universe and its baby to ever be severed. An "umbilical cord" at all times connects the parent and daughter

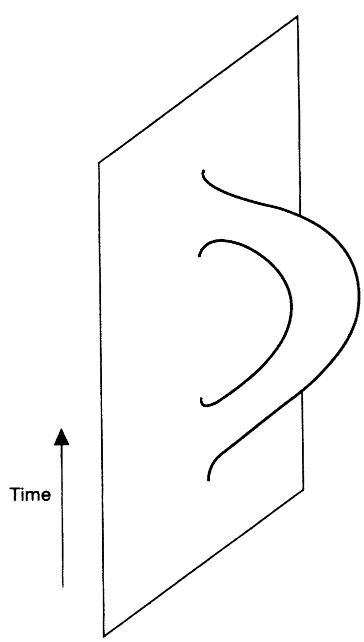

FIG. 1. A transient wormhole. The parent universe emits and "subsequently" reabsorbs a baby universe.

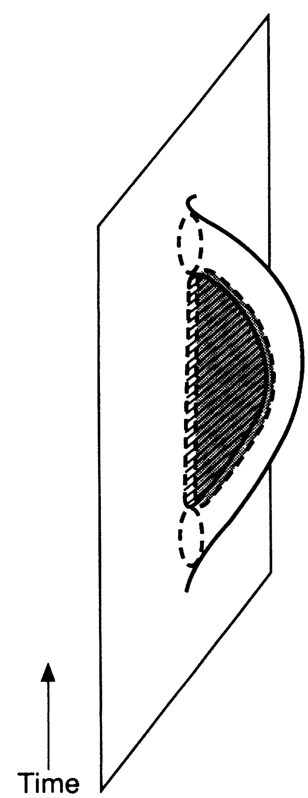

FIG. 2. Baby universe with umbilical cord. Note that the topology is trivial.

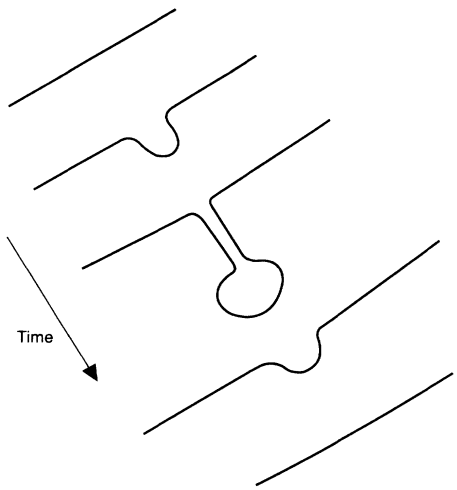

FIG. 3. Spatial sections of a baby universe with umbilical cord.

universes. The existence of this umbilical cord is a *necessary* consequence of stable causality. The existence of the umbilical cord is, by itself, not a sufficient condition for the prevention of causality violations.

The basic picture described in Figs. 2 and 3 has some interesting variants. For instance one may conceive of a process of baby-universe emission that is *not* followed by reabsorption. The umbilical cord connecting the parent and daughter universes might then be assumed to be "small." The only relevant scale would appear to be the Planck scale, so that one might expect the radius of the umbilical cord to be quantum-mechanically stabilized at around $R \simeq L_{\text{Planck}} \simeq 10^{-35}$ m. This suggestion is in fact supported by a recent minisuperspace calculation.[12] If this picture is even approximately correct, it becomes useful to distinguish between the mathematicians' topology and a suitable "coarse grained" physicists' topology. Even though the topology is mathematically trivial, physical probes will not be able to pass through the umbilical cord to see the daughter universe unless their energy is greater than $\hbar c/R$. Thus we should view the physical topology as an energy-dependent concept. For $E > \hbar c/R$ the spatial topology is $\mathcal{R}^3$, while for $E < \hbar c/R$ the effective spatial topology is $\mathcal{R}^3 \otimes S^3$. This concept of an energy-dependent physicists' topology is very important. At low energy the umbilical cord will appear to behave as an elementary particle. Presumably, such an umbilical cord will have a mass of order the Planck mass, and might be distinguished from a Planck mass primordial black hole by the absence of Hawking radiation. There is after all no need for the umbilical cord to possess an event horizon. Indeed, by adapting the (permanent) wormholes discussed by Morris and Thorne,[7] Morris, Thorne, and Yurtsever,[8] and the present author,[9–12] it is easy to model umbilical cords which possess no event horizons. It is of course conceivable that although the radius of the umbilical cord be of order the Planck length, its mass might be quite small compared to the Planck mass.[12] In such a situation, the intriguing possibility arises that these umbilical cords might be related to Wheeler's viewpoint that elementary particles might be thought of as (permanent) Minkowski wormholes with trapped electric flux (or some other flux?).[32]

We take this opportunity to mention that the analysis of Morris and Thorne[7] implies that the region near the throat of a classical umbilical cord contains "exotic matter" which violates the weak, strong, and dominant energy hypotheses. This should not perturb one. It cannot be emphasized strongly enough that the weak, strong, and dominant energy hypotheses have been experimentally tested in the laboratory, and have all been experimentally shown to be *false*. It is not commonly appreciated, but it is in fact true, that the observation of the Casimir effect between parallel plates[33] experimentally disproves the weak, strong, and dominant energy hypotheses. For analyses of the form of the stress-energy tensor between parallel plates see Gibbons[34] and DeWitt.[35] Further comments along these lines may be found in Roman,[36] Morris and Thorne,[7] and in Morris, Thorne, and Yurtsever.[8] In this regard it is perhaps somewhat embarrassing to realize that the experimental observations disproving the energy hypotheses predate the formulation of the energy hypotheses by some 25 years.

Even though quantum effects can violate the weak energy hypothesis, it is still very much an open question as to whether or not quantum effects can *in general* violate the *averaged* weak energy hypothesis. Roman has pointed out[37] that, at least for some geodesics, quantum effects do violate the averaged weak energy hypothesis. Merely consider a geodesic that is parallel to (and lies between) the conducting plates that generate the Casimir effect. Although this geometry is of course very special, it does serve to show that any potential nonexistence proof for exotic matter (by implication this would be a nonexistence proof for semiclassical wormholes) will require a very delicate statement of hypotheses in terms of "generic" geodesics.

In summary, the condition of stable causality enforces the continued existence of a remnant umbilical cord connecting parent and daughter universes. In such a situation topology should be viewed as an energy dependent concept.

VI. PERMANENT WORMHOLES

We shall now discuss the causality properties of "permanent wormholes" (Fig. 4). These objects have been extensively discussed in the excellent paper by Morris and Thorne.[7] These "permanent" wormholes are entirely unconnected with the transient wormholes of Refs. 1–3. We shall be interested in extending the discussion of their causality properties. Note that the "permanent" wormholes of this section would separate a $(1+1)$-dimensional spacetime into disconnected components. Thus "permanent" wormholes are interesting only in $2+1$ dimensions and higher.

The first observation is this: since stable causality forbids localized topology changing processes, it follows that permanent wormholes (if they exist at all) must date back to the "first cause." (One should say "first cause" and not "big bang," since the big bang, in the cosmogony out-

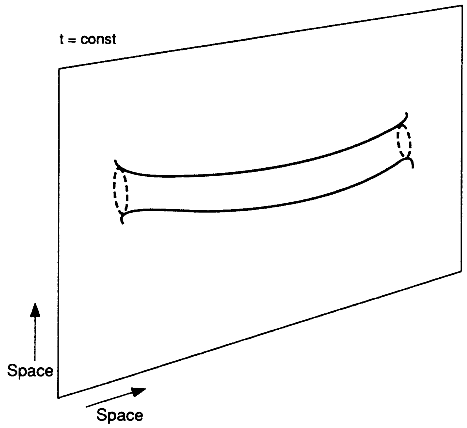

FIG. 4. Spatial section (t=const) of a "permanent" wormhole.

lined in this paper, is just the moment at which *our* Universe branched off from the multiverse. From this viewpoint, the big bang has nothing to do with questions of the ultimate origin of the multiverse.) Under these circumstances, an arbitrarily advanced civilization would still be unable to *manufacture* wormholes. Such a civilization would have to be satisfied with finding an already existing wormhole and "adapting" it to their purposes. Since wormholes, like their siblings the umbilical cords, presumably will shrink to a natural diameter of order the Planck length when left to their own devices, this suggests a new industry. Advanced civilizations may find it advantageous to mine the centers of stars and large planets looking for Planck mass wormholes and umbilical cords to be used as the raw material for their transportation networks.

This naturally brings us to consider a related topic: what is the difference between a umbilical cord and a wormhole? Umbilical cords connect parent universes to their offspring, while wormholes are permanent intrauniverse or interuniverse connections, but from a low-energy perspective they should be very similar. A signal that may serve to distinguish the two is this: if the process of parturition of a baby universe traps some electric charge in the baby universe, then the associated umbilical cord will carry electric charge. Wormholes, on the other hand, allow the possibility of permanently trapping some magnetic flux lines by having them thread the wormhole. In this case we would expect the low-energy physics to mimic a (Planck mass?) magnetic monopole. This would certainly be an exotic and interesting signal.

Passing to the issue of causality constraints, we first note that the end points of a (permanent) wormhole should be "spacelike separated." Otherwise closed timelike loops trivially occur. This situation may easily be modeled by taking 2+1 Minkowski space and identifying the two lines $(0,0,t)$ and (x_0,y_0,t_0+t). Causality enforces the constraint that $|t_0|<\sqrt{x_0^2+y_0^2}$. More serious is the conflict that arises when two permanent wormholes are placed close to each other. In this case, even though the permanent wormholes individually are well behaved (because their end points are spacelike separated), the combination still admits closed timelike loops. This situation is modeled by taking 2+1 Minkowski space with two sets of lines identified. For the first wormhole, identify the line $(0,0,t)$ with the line $(x_0,0,t_1+t)$. For the second wormhole identify the line $(0,y_0,t)$ with (x_0,y_0,t_2+t). Consider the lightlike trajectory $(0,0,0)\rightarrow(0,y_0,y_0)\equiv(x_0,y_0,(t_2+y_0))\rightarrow(x_0,0,(t_2+2y_0))\equiv(0,0,(t_2-t_1+2y_0))$. From the previous argument we know that $|t_1|\le x_0$ and $t_2\le x_0$, but this is not a strong enough condition to prevent $(t_2-t_1+2y_0)$ from becoming negative. One could of course assert that permanant wormholes do not exist, but that would be a pity. It would also be quite difficult to reconcile with the quite benign explicit solutions obtained by Morris and Thorne.[7]

Another serious causality constraint arises when one has a singe wormhole whose end points are free to move with respect to each other.[8] As a model, consider (2+1)-dimensional Minkowski space with two generic (nongeodesic) world lines identified. Let the world lines be parametrized in terms of their respective proper times by $x_1^\mu(\tau_1)$ and $x_2^\mu(\tau_2)$. We choose τ_1 and τ_2 in such a manner that at $\tau_1=\tau_2=0$ the points $x_1^\mu(0)\equiv x_2^\mu(0)$ are identified, i.e., connected via the wormhole. Firstly, notice that the points $x_1^\mu(0)$ and $x_2^\mu(0)$ must be spacelike separated; otherwise closed causal curves trivially exist. Secondly, we let imaginary clocks that follow the world lines x_1 and x_2 evolve in time. By remembering that these two (imaginary) clocks are in fact one clock (because we are identifying the world lines), it is easy to see that

$$x_1^\mu(\tau)\equiv x_2^\mu(\tau)\ .$$

What this is telling us is this: if we construct a wormhole by identifying two world lines in Minkowski space, then special relativity tightly constrains the identification process. World lines may only be identified in a manner that is compatible with their respective proper times. This argument, of course, trivially generalizes to the curved geometries of general relativity. Unfortunately this simple argument also swiftly leads to causality violations. For suppose that one end of the wormhole moves with respect to the other. One may then very easily obtain the wormhole version of the so-called "twin paradox." In the wormhole case this is a true paradox involving causality violations, whereas in (special or general) relativity without wormholes no paradox exists. The paradox in this case arises in this fashion: by moving one end of the wormhole with respect to the other, the moving end may be made to age more slowly than the other, until eventually there comes a proper time τ_0 such that the points $x_1^\mu(\tau_0)\equiv x_2^\mu(\tau_0)$ are lightlike separated (i.e., a closed lightlike curve exists). For $\tau>\tau_0$, we can in fact arrange for $x_1^\mu(\tau>\tau_0)\equiv x_2^\mu(\tau>\tau_0)$ to be timelike separated, so that closed timelike loops exist.

There is an indication, though not a proof, that problems of this particular type may be mollified by dynamical effects. The wormholes of the previous paragraph were pointlike and structureless. A more realistic model of a wormhole may be constructed by use of the "thin-wall approximation."[9–12] To make the discussion concrete, consider flat (3+1)-dimensional Minkowski space, and excise from it two *identical* regions Ω_1 and Ω_2. Each of these regions is taken to have the topology $\mathcal{R}\times B^3$. The boundaries $\partial\Omega_1$ and $\partial\Omega_2$ are then timelike hypersurfaces with topology $\mathcal{R}\times S^2$. Now identify the boundaries $\partial\Omega_1\equiv\partial\Omega_2$. The resulting spacetime is by construction multiply connected and geodesically complete. The throat of the wormhole occurs at $\partial\Omega$. The Reimann tensor is almost everywhere zero, except on the throat of the wormhole, where it is proportional to a delta function. The important point is that in order for the identification process $\partial\Omega_1\equiv\partial\Omega_2$ to work, it is necessary that the first fundamental forms (i.e., the induced metrics) of the hypersurfaces $\partial\Omega_1$ and $\partial\Omega_2$ be equal. This has a very important consequence: if one end of the wormhole is accelerated, then this acceleration influences the induced metric on the throat. But the induced metric on $\partial\Omega_1$ is forced to be identical to the induced metric on $\partial\Omega_2$. Thus the other end of the wormhole must also accelerate. In fact, the accelerations of both ends of the wormhole must be identical, so that there is no way to "time dilate" one end of the wormhole with respect to the other, and the argument of the preceding paragraph is obviated. Now this particular argument depends critically on the thin-wall approximation used to construct the wormhole, and so cannot be said to hold in all generality. The argument does serve to suggest that dynamical effects may serve to mollify some of the causality problems considered by Morris, Thorne, and Yurtsever.[8]

The causality violation problems discussed above are extremely serious. Recall that to avoid the causality problems associated with baby universes we had only to enforce the condition of stable causality, thereby deducing the existence of umbilical cords. Similar considerations now apply in multiply connected spacetimes (i.e., spacetimes containing "permanent" wormholes). The condition of stable causality implies the existence of many (highly nonunique) global time functions on the spacetime (i.e., τ a globally defined scalar function, $\nabla\tau$ everywhere timelike). To make wormholes compatible with (stable) causality one enforces the condition that the end points of all (permanent) wormholes connect regions of the multiverse that are simultaneous in terms of some restricted subclass of these cosmic time functions. This condition now excludes causality violations, but at a rather high cost. We have had to reintroduce the notion of "distant simultaneity," which runs counter to all of our intuition learned from both special and general relativities. Note, however, that this concept of "distant simultaneity" does not directly conflict with experiment. *Local* experiments will always in this formalism see exactly the behavior expected on the basis of general realitivity. It is only when a civilization is sufficiently advanced so as to be able to set up a network of traversable wormholes that the discovery of "distant simultaneity" becomes possible.

There is an alternative, which is safe but relatively boring. One could restrict the functional integral to only run over spacetimes of the topology $\mathcal{R}\times S^3$. This constraint is weak enough to allow the development of baby universes (with umbilical cords), and to allow (variants of) almost all interesting classical solutions of the Einstein field equations to be considered, but forbids the existence of permanent wormholes. In such a multiverse one could in principle travel to other universes, but beating the light-speed barrier in our own Universe would be impossible.

Summarizing: the existence of (permanent) wormholes without causality violations is possible, but only at the cost of adding rather strong assumptions to general relativity. In this regard the hypothesis of the existence of permanent wormholes should be regarded as a much more radical idea than baby universes. Note that these causality constraints affect only wormholes which connect a universe to itself. Wormholes that connect two distinct universes are unaffected by these particular problems.

VII. EVASIONS

In this section we shall look at ways around the problems and constraints discussed in this paper. At least two alternatives present themselves. (1) Admit metrics that are "almost everywhere" Lorentzian; (2) try to formulate a consistent viewpoint that permits causality violations.

One alternative is to permit the spacetime metric to be "almost everywhere" Lorentzian, i.e., the signature is Lorentzian except at a number of discrete isolated points. An almost everywhere Lorentzian metric of this type can be constructed by combining a Euclidean metric (g_E) with a Morse function. [A Morse function is simply a scalar function $\tau(x)$ whose gradient is almost everywhere nonzero.] Given Euclidean metric and a Morse function, a suitable almost Lorentzian metric is

$$g_L=g_E-2\frac{\nabla\tau\otimes\nabla\tau}{|\nabla\tau|^2} . \qquad (7.1)$$

By inspection, g_L is a Lorentzian signature metric except at points where $\nabla\tau$ is zero. The Morse function plays a role analogous to the "cosmic time" function encountered in discussing stable causality. In fact, it is easy to see that a vector V is timelike with respect to g_L if and only if $|V\cdot\nabla\tau|>\sqrt{\tfrac{1}{2}}|V|\,|\nabla\tau|$. Here $|X|$ denotes the length of the vector X in the Euclidean metric g_E. This is sufficient to indicate that along any timelike curve (not necessarily a geodesic) the Morse function is strictly monotone increasing (or decreasing). Thus closed timelike curves do not exist in this almost Lorentzian metric, and the causal behavior is suitably pleasing. The problem, of course, lies in the interpretation of the nature of the points where $\nabla\tau=0$. At these points, it is clear that no possible change of coordinates can locally bring the metric into the Minkowski form. That is, critical points of the Morse function correspond to points where the geometry is not locally that of Minkowski space. This can best be interpreted as a breakdown of the spacetime equivalence principle.[24,26] (That is, if transient wormholes exist, they violate the equivalence principle.) It would be a rather interesting research problem to attempt a quantitative statement of possible experimental consequences of such a breakdown, but we shall not attempt such an analysis here. We should emphasize, however, that the equivalence principle is so well established in gravity theory that the possibility of its abandonment should be viewed with extreme caution.

Secondly, we should consider the possibility that the multiverse *does* violate causality. This possibility is exceedingly unpleasant, and rather worse for the state of physics than any of the (relatively conservative) ideas discussed above. The most well thought out version of a causality-violating cosmos is, in my opinion, the non-Hausdorff cosmos discussed by Penrose.[38]

VIII. DISCUSSION AND CONCLUSIONS

We have argued that the path-integral method of quantizing gravity makes most sense when one restricts the integration to those Minkowski signature metrics obeying the stable causality condition. Subject to this assumption, we have discussed at some length the consequences of causality (non)violation for wormhole physics. We have identified three basic objects as being of interest: transient wormholes, umbilical cords (which connect parent universes to their offspring), and permanent wormholes (which provide permanent intrauniverse and interuniverse connections). We have shown that transient wormholes are forbidden by causality, and that baby-universe production is compatible with causality only if an umbilical cord at all times connects the parent and baby universes. Although causality very tightly constrains the possibility of (mathematical) topology change, we have emphasized the notion of an energy-dependent physicists' topology. Permanent wormholes are a more complicated topic, these wormholes may also be made compatible with causality by imposition of the stable causality constraint. At the classical level it should be emphasized that the results of this paper are very general in nature, depending in a crucial way only on the locally Lorentzian nature of spacetime. At the quantum level, the results of this paper depend on the choice of acceptable spacetimes to be used in the functional integral.

ACKNOWLEDGMENTS

I wish to thank the Aspen Center for Physics for its hospitality during early and late phases of this work. Useful and encouraging discussions with Henry Kendrup and Tom Roman are also acknowledged. I particularly wish to thank Arvind Borde for making available some of his unpublished work.

---

*Present address: Physics Department, Washington University, St. Louis, Missouri 63130-4899.

[1]S. Coleman, Nucl. Phys. **B310**, 643 (1988).

[2]S. Giddings and A. Strominger, Nucl. Phys. **B321**, 481 (1989).

[3]I. Klebanov, L. Susskind, and T. Banks, Nucl. Phys. **B317**, 665 (1989).

[4]W. Fischler and L. Susskind, Phys. Lett. B **217**, 48 (1989).

[5]S. Coleman and K. Lee, Phys. Lett. B **221**, 242 (1989).

[6]J. Polchinski, Nucl. Phys. **B325**, 619 (1989).

[7]M. S. Morris and K. S. Thorne, Am. J. Phys. **56**, 395 (1988).

[8]M. S. Morris, K. S. Thorne, and U. Yurtsever, Phys. Rev. Lett. **61**, 1446 (1988).

[9]M. Visser, Phys. Rev. D **39**, 3182 (1989).

[10]M. Visser, Los Alamos National Laboratory Report No. LA-UR-89-1008, 1989 (unpublished).

[11]M. Visser, Nucl. Phys. **B328**, 203 (1989).

[12]M. Visser, Los Alamos National Laboratory Report No. LA-UR-89-1745, 1989 (unpublished).

[13]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, England, 1973).

[14]R. M. Wald, *General Relativity* (University of Chicago, Chicago, 1984).

[15]It is unfortunate that the word "wormhole" means different things to different authors. Some authors (typically high-energy theorists) use the word "wormhole" to mean a transient phenomenon wherein a baby universe is emitted from a parent universe and "subsequently" reabsorbed (Refs. 1–3). Other authors (typically relativity theorists) use the word "wormhole" to mean an essentially permanent connection between different universes or different parts of the same universe (Refs. 7 and 8). The phrase "transient wormhole" shall be used to describe the first possibility, while the phrase "wormhole" used without a qualifer will indicate a "permanent" wormhole.

[16]The multiverse is defined to be the collection of all universes. The universes comprising the multiverse are connected by umbilical cords and "permanent" wormholes. The multiverse is by definition a connected manifold. A universe is defined to be any reasonably large, reasonably flat region of spacetime.

[17]S. W. Hawking, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979), p. 746.

[18]P. Mazur and E. Mottola, Nucl. Phys. B (to be published).

[19]W. M. Suen and K. Young, Phys. Rev. D **39**, 2201 (1989).

[20]M. Cline, Phys. Lett. B **224**, 53 (1989).

[21]E. Farhi, Phys. Lett. B **219**, 403 (1989).

[22]J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).

[23]A. Vilenkin, Phys. Rev. D **33**, 3560 (1986).

[24]R. D. Sorkin, Phys. Rev. D **33**, 978 (1986).

[25]A. Borde (in preparation).

[26]J. B. Hartle, Class. Quantum Grav. **2**, 707 (1985).

[27]E. Witten, in *Shelter Island II*, edited by R. Jackiw, N. N. Khuri, S. Weinberg, and E. Witten (MIT Press, Cambridge, MA, 1983), p. 227.

[28]We should emphasize that compactness of the spatial sections is almost universally assumed by authors working in the field of quantum gravity, though in many cases this is not explicitly mentioned. For example, the usual formulation of the Wheeler-DeWitt equation (Ref. 29) implicitly assumes both compact spatial sections and stable causality.

[29]B. S. DeWitt, Phys. Rev. **160**, 1113 (1967).

[30]R. P. Geroch, J. Math. Phys. **8**, 782 (1967); Ph.D. thesis, Princeton University, 1967.

[31]F. J. Tipler, Ann. Phys. (N.Y.) **108**, 1 (1977); Phys. Rev. Lett. **37**, 879 (1976).

[32]C. W. Misner and J. A. Wheeler, Ann. Phys. (N.Y.) **2**, 525 (1957); reprinted in J. A. Wheeler, *Geometrodynamics*

(Academic, New York, 1962); J. A. Wheeler, in *Battelle Rencontres: 1967 Lectures in Mathematics and Physics,* edited by B. S. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[33]H. B. G. Casimir, Proc. Kon. Nederl. Akad. Wet. **51**, 793 (1948).

[34]G. W. Gibbons, in *General Relativity: An Einstein Centenary Survey* (Ref. 17), p. 639.

[35]B. S. DeWitt in *General Relativity: An Einstein Centenary Survey* (Ref. 17), p. 680.

[36]T. A. Roman, Phys. Rev. D **33**, 3526 (1986).

[37]T. A. Roman (private communication).

[38]R. Penrose, in *General Relativity: An Einstein Centenary Survey* (Ref. 17), p. 581.

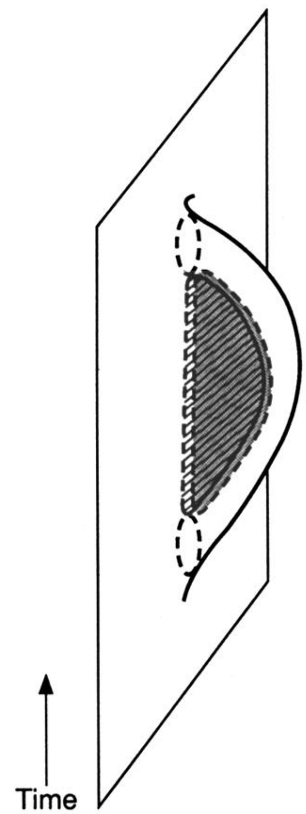

FIG. 2. Baby universe with umbilical cord. Note that the topology is trivial.