

## Closed timelike curves

Kip S. Thorne

California Institute of Technology, Pasadena, California 91125, and  
Institute for Theoretical Physics, UCSB, Santa Barbara, California 93106

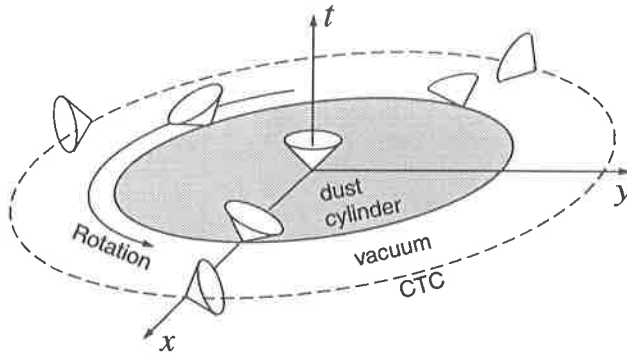
**Abstract.** This lecture reviews recent research on closed timelike curves (CTCs), including these questions: Do the laws of physics prevent CTCs from ever forming in classical spacetime? If so, by what physical mechanism are CTCs prevented? Can the laws of physics be adapted in any reasonable way to a spacetime that contains CTCs, or do they necessarily give nonsense? What insights into quantum gravity can one gain by asking questions such as these?

### 1. Introduction

Much of the forefront of theoretical physics deals with situations so extreme that there is no hope to probe them experimentally. Such, largely, was the case nearly a century ago for Einstein's formulation of general relativity, and such is the case today for the attempt to quantize gravity. In these situations, thought experiments can be helpful. Of all thought experiments, perhaps the most helpful are those that push the laws of physics in the most extreme ways. A class of such thought experiments, which I and others have found useful in the last few years, asks [1] *What constraints do the laws of physics place on the activities of an arbitrarily advanced civilization?* In asking this question, we have in mind *all* the laws of physics that govern our universe, taken together—including those, such as quantum gravity, that are not yet well understood, and others, such as classical general relativity, that are, and with each set of laws holding sway only in its own domain of validity.

An especially fruitful question of this type is [1] *Do the laws of physics prevent arbitrarily advanced civilizations from constructing "time machines" (machines for backward time travel), and if so, by what physical mechanism are they prevented?* Hawking [2] has given the name *chronology protection* to the conjecture that there is such a mechanism, and that therefore *closed timelike curves* (CTCs) can never be created in the real Universe, no matter how hard advanced civilizations might try. In this lecture I shall review recent research on the chronology protection conjecture and related issues.

The laws of general relativity by themselves do not enforce chronology protection: it is easy to find solutions of the Einstein field equation that have closed timelike curves



**Figure 1.** Van Stockum's spacetime.

(Section 2). However, the combination of general relativity's laws and the laws of quantum fields in curved spacetime may well provide a chronology protection mechanism—though we might not be sure of this until we understand the laws of quantum gravity much more deeply than today (Section 3).

Independently of whether chronology protection is correct, much insight into the laws of physics might be gained by studying how they behave in the presence of closed timelike curves (Section 4).

## 2. Spacetimes with closed timelike curves

A number of spacetimes with closed timelike curves have been exhibited in the literature, and much is now understood about the generic chronological structure of such spacetimes.

### 2.1. Spacetimes with eternal CTCs

The earliest example of a spacetime with CTCs is Van Stockum's 1937 solution of the Einstein field equation [4, 5], which represents an infinitely long cylinder made of rigidly and rapidly rotating dust. The dust particles are held out against their own gravity by centrifugal forces, and their rotation drags inertial frames so strongly that the light cones tilt over in the circumferential direction in the manner shown in Figure 1, causing the dashed circle in the figure to be a CTC. CTCs pass through every event in the spacetime, even an event on the rotation axis where the light cone is not tilted at all: one can begin there, travel out to the vicinity of the dashed circle (necessarily moving forward in  $t$  as one travels), then go around the cylinder a number of times traveling backward in  $t$  as one goes, and then return to the rotation axis, arriving at the same moment one departed. For the mathematical details of Van Stockum's solution see, e.g. Bonnor [5].

Physicists (but not science fiction writers) have generally dismissed Van Stockum's solution as "unphysical" because its source is infinitely long. Whether a finite-length, rotating body can also produce CTCs is not known; I shall return to this in Sec. 2.2.

A second old, famous example of a spacetime with CTCs is Gödel's solution of the Einstein equation [6], which describes a stationary, homogeneous cosmological model

with nonzero cosmological constant, filled with rotating dust. Again, the rotation tilts the light cones, creating CTCs. Because the spacetime is homogeneous and stationary, CTCs pass through every event. For the mathematical details of Gödel's spacetime, see, e.g. Hawking and Ellis [3], especially Figure 31.

Physicists have generally dismissed Gödel's solution as unphysical because it requires a nonzero cosmological constant and/or it doesn't resemble our own universe (whose rotation is small or zero).

## 2.2. Spacetimes with compactly generated chronology horizons

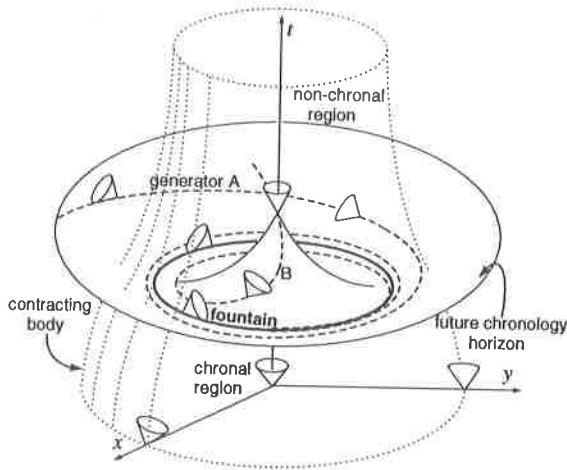
A spacetime whose CTCs are not eternal can be divided into *chronal regions* that are free of CTCs, and *nonchronal regions* that contain CTCs everywhere. The boundaries between the choral and nonchronal regions are called *chronology horizons*; choral regions end and CTCs are created at *future* chronology horizons; CTCs are destroyed and choral regions begin at *past* chronology horizons.

A future chronology horizon is a special type of future Cauchy horizon, and as such it is subject to all the laws that govern any such horizon [3]; most importantly, it is generated by null geodesics that have no past endpoints but can leave the horizon when followed into the future. If the generators, when followed into the past, enter one or more compact regions  $\mathcal{K}$  of spacetime and never thereafter leave them, the future chronology horizon is said to be *compactly generated* [2]; otherwise it is *non-compactly generated*. (If an arbitrarily advanced civilization were to create a time machine in a compact region of spacetime, then its chronology horizon obviously would be compactly generated.) A past chronology horizon is, in essence, the time reversal of a future one; and it therefore is generated by null geodesics that have no future endpoints but can leave the horizon when followed into the past.

When a future chronology horizon, at which CTCs arise, is compactly generated, its generators, followed into the past, can become confined into their compact region  $\mathcal{K}$  in either of two ways: They can wander ergodically around  $\mathcal{K}$  or some portion of it; or they can asymptote to one or more smoothly closed null geodesics in  $\mathcal{K}$ . Such smoothly closed null geodesics are called *fountains* because, when the generators are followed forward in time, they are seen to originate in the fountains and spew out of them, like streams of water, into the surrounding spacetime.

Hawking [2] has proved that in the generic case,  $\mathcal{K}$  will contain such fountains, and it seems likely to me that generically all or almost all the horizon generators will emerge from them. Thus, fountains are generic, while ergodic wandering in  $\mathcal{K}$  probably is not.

It is tempting to conjecture that, if a finite-sized, rapidly rotating body were to contract in some carefully designed, axially symmetric manner, it might create CTCs around itself in the manner of the Van Stockum solution without forming an event horizon. Figure 2 shows the chronological structure in the body's equatorial plane for such an evolution. At early times, when the body is large, there are no CTCs, so spacetime is choral. At late times, when the body has settled down into its final, smaller state, the light-cone and chronology structures are similar to Van Stockum's solution, so spacetime is nonchronal. A future chronology horizon separates the choral and nonchronal regions; it is shaped like a bowl with a small mountain in the center, i.e. like a "Mexican hat". The horizon's generators all originate on a single fountain. Two generators are shown. One, labeled *A*, spirals outward from the fountain and remains



**Figure 2.** The chronological structure of a spacetime that might result from an axially symmetric contraction of finite sized, rotating body. This diagram is confined to the body's equatorial plane.

always on the horizon, eventually reaching future null infinity. The other,  $B$ , spirals inward, and then leaves the horizon at the tip of the Mexican hat. Hawking [2] has given mathematical details of a spacetime with this chronological structure.

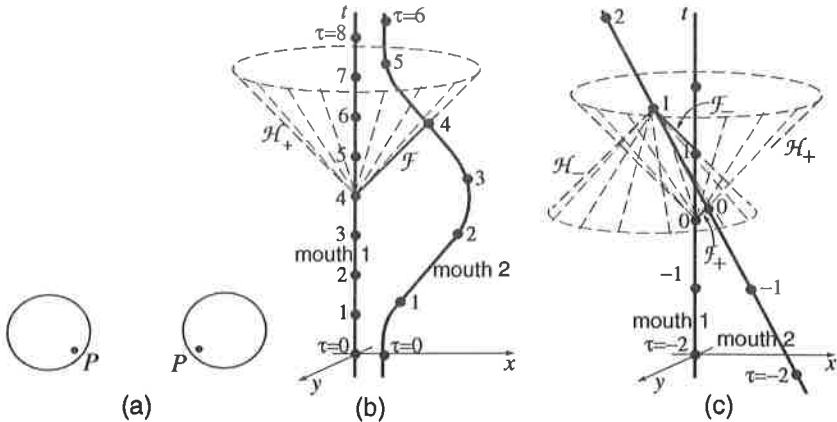
In order for horizon generators to emerge from the fountain, there must be a net *defocusing* of any bundle of null geodesics that travels around the fountain. By the equations of geometric optics together with the Einstein field equation, this requires that

$$\oint_{\mathcal{F}} T_{\alpha\beta} l^{\alpha} l^{\beta} d\zeta < 0, \quad (1)$$

where the integral is around the fountain  $\mathcal{F}$ ,  $T_{\alpha\beta}$  is the total stress-energy tensor for all matter and fields on  $\mathcal{F}$ ,  $\zeta$  is an affine parameter along  $\mathcal{F}$ , and  $l^{\alpha} = dx^{\alpha}/d\zeta$  is the tangent to  $\mathcal{F}$ . Equation (1) says, in words, that *the averaged null energy condition (ANEC) must be violated around the fountain*; i.e., the integral in (1) must be negative. All ordinary, familiar forms of matter satisfy ANEC; therefore, no imploding body made of such matter can create CTCs in the manner of Figure 2. In Section 3 we shall return to the issue of whether ANEC can *ever* be violated, and shall learn that the answer is yes.

Lorentzian wormholes constitute a class of simple, explicit spacetimes that have generic-type, compactly generated chronology horizons [1, 7, 8]; as such, they have become a useful testbed for studies of chronology issues.

The simplest such wormholes are obtained by removing two balls from Euclidean space and identifying their surfaces in the manner of Figure 3a; the surfaces then become the wormhole's mouths. Such a wormhole necessarily violates ANEC: Any bundle of radially traveling null geodesics that passes through the wormhole is converging as it enters and diverging as it leaves, and therefore gets defocused by the wormhole, which means that  $\oint T_{\alpha\beta} l^{\alpha} l^{\beta} d\zeta < 0$  along the bundle, with the negative contribution coming from a delta function  $T_{\alpha\beta}$  at the junction between the two mouths. One can show more generally that every traversable wormhole, regardless of its shape or motion, violates ANEC [1].



**Figure 3.** (a) A wormhole formed by removing two balls from Euclidean space and identifying their surfaces (the wormhole “mouths”); the identified points are reflections of each other in the midplane between the balls. (b) Chronology structure for a wormhole, one of whose mouths makes a “twins paradox trip.” (c) Chronology structure for a wormhole whose mouths move past each other at uniform speed.

One can construct wormhole spacetimes, whose wormhole mouths travel along arbitrarily chosen world lines in Minkowski spacetime, by removing world tubes along those lines in a manner analogous to Figure 3a, and identifying their surfaces with each other [7]. Of course, the identification must be done in such a way that the intrinsic geometries of the two mouths’ world tubes are the same. This may require a distortion of the spacetime geometry near the mouths if they are accelerated, but the distortion becomes vanishingly small in the limit that  $(\text{acceleration}) \times (\text{mouth radius}) \rightarrow 0$  [1, 7]. Since the mouths’ intrinsic geometries are the same, the proper time interval  $\Delta\tau$  between two identified neighboring events on the mouths must be same as seen through either mouth; it is this that dictates the form of the time markings in Figures 3b,c.

Figures 3b,c show the chronological structures of two wormhole spacetimes constructed in this way. In Figure 3b, one mouth remains at rest in a chosen Lorentz frame, while the other makes a “twins-paradox-type trip” into the external universe and returns [1]. As seen in the external universe there is a dilation of proper time on the moving mouth relative to the static one, but as seen through the wormhole there is no such time dilation. As a result, the relative motion of the mouths changes the manner in which time hooks up to itself through the wormhole. Initially the hookup is such that there are no CTCs; spacetime is choral. After the trip, the hookup entails CTCs; spacetime is nonchronal. The future chronology horizon (denoted  $\mathcal{H}_+$  in the figure) is the future light cone of the event  $\tau = 4$  at the center of the right face of the left (static) mouth. The generators of this horizon (long-dashed lines) all originate in a single fountain (labeled  $\mathcal{F}$ ): the smoothly closed null geodesic that travels from  $\tau = 4$  on the left mouth to  $\tau = 4$  on the right mouth, then through the infinitesimally short wormhole and back to where it started; cf. Figure 10 of Ref. [7].

In Figure 3c, one mouth moves past the other, creating CTCs that are confined to a

bounded nonchronal region of spacetime: the region that begins at the future chronology horizon  $\mathcal{H}_+$  and ends at the past horizon  $\mathcal{H}_-$ . The generators of  $\mathcal{H}_+$  (the future light cone of  $\tau = 0$  on the static mouth) emerge from the fountain  $\mathcal{F}_+$  and leave the horizon when they pass through  $\mathcal{H}_-$ . The generators of  $\mathcal{H}_-$  (the past light cone of  $\tau = 1$  on the moving mouth) enter the horizon at its intersection with  $\mathcal{H}_+$  and ultimately asymptote to the fountain  $\mathcal{F}_-$ .

Figures 3b,c are prototypes for the consequences of generic relative motions of a wormhole's mouth: such motions will always produce CTCs [1], as will the gravitational redshifts that result from placing a wormhole in a generic external gravitational field [8]. Most physicists react to this by asserting that the laws of physics must prevent the existence of classical, traversable wormholes—perhaps by forbidding the existence of material that violates ANEC (“exotic material”).

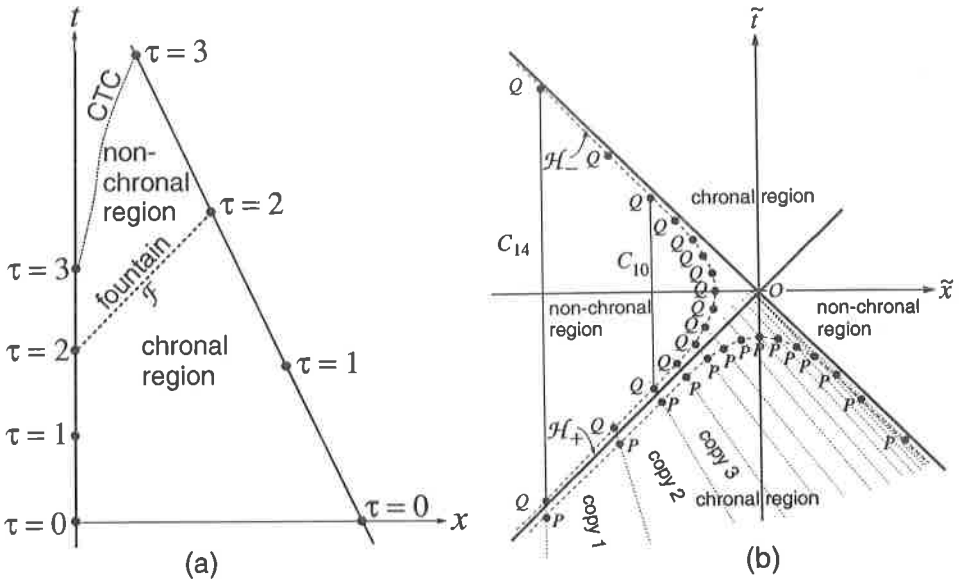
Not all compactly generated chronology horizons have the generic “generators-emerge-from-fountains” structure of Figures 2 and 3. An example that is different is Taub-NUT space [9]—a vacuum solution of the Einstein equation with a spatially compact chronal region, followed by a compact future chronology horizon, followed in turn by a non-compact nonchronal region with CTCs. Being a vacuum solution, Taub-NUT space satisfies ANEC and also satisfies the local null energy condition (NEC),  $T_{\alpha\beta}l^\alpha l^\beta \geq 0$  everywhere. This means that the generators of the chronology horizon cannot peel off of fountains in the manner of Figures 2 and 3. Instead, every generator is itself a fountain (a smoothly closed null geodesic).

A simpler spacetime with this special type of chronology horizon is Misner space [10]. The relevant variant of Misner space can be obtained as follows: go into your bedroom in Minkowski spacetime, identify the back wall with the front wall (so when you walk into the back you find yourself emerging from the front), similarly identify the floor with the ceiling and the left wall with the right, and set the right wall moving toward the left. In other words, Misner space is Minkowski spacetime with identification under translations along  $y$  and  $z$ , and under a boost along  $x$ . Figure 4a shows the  $x$ - $t$  portion of this spacetime. It initially is chronal, and then becomes nonchronal at a future chronology horizon whose generators are the closed null geodesics (fountains)  $y = \text{const}$ ,  $z = \text{const}$ ,  $x = t - 2$ .

Hawking [2] shows that *every* fountain  $\mathcal{F}$  on a compactly generated chronology horizon must have a non-positive ANEC integral,  $\int_{\mathcal{F}} T_{\alpha\beta}l^\alpha l^\beta \leq 0$ . The generic case where generators peel off the fountain (Figs. 2 and 3) corresponds to “ $< 0$ ” for this integral and thus to a violation of ANEC; the special Taub-NUT and Misner cases correspond to “ $= 0$ ”. Hawking points out that as soon as one allows energy of any sort to flow through the Taub-NUT or Misner horizon, or through any other horizon whose fountains have  $T_{\alpha\beta}l^\alpha l^\beta = 0$  everywhere, that energy flow will carry a nonzero local value of  $T_{\alpha\beta}l^\alpha l^\beta$ , and therefore in order to keep the ANEC integral nonpositive, the local null energy condition (NEC) must be violated somewhere along each perturbed fountain. This means that on any physically realistic, compactly generated chronology horizon, NEC must be violated, even if ANEC is not. (This conclusion strengthens an earlier result due to Tipler [11].)

### 2.3. Spacetimes with non-compactly generated chronology horizons

The simplest example of a spacetime with a non-compactly generated chronology horizon is Grant space [12], which is a slight generalization of Misner space. Go into your bedroom

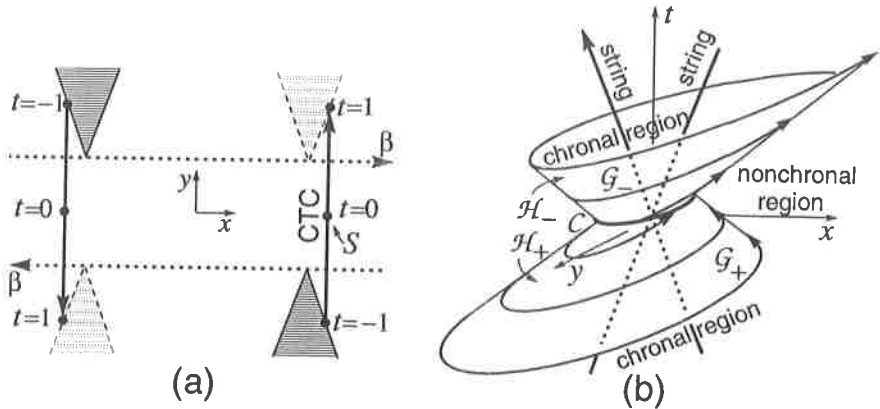


**Figure 4.** (a) The  $t$ - $x$  portion of Misner space. (b) The covering space for Misner space (with the points  $P$  identified and  $Q$  identified) and for Grant space (with each successive  $P$  or  $Q$  identified only after a displacement by distance  $a$  into the paper).

in Minkowski spacetime, translate the left wall relative to the right by a distance  $a$ , and then identify them and set the right wall moving toward the left with speed  $\beta$ ; the result is Grant space. In other words Grant space is Minkowski spacetime identified along the  $x$  direction then boosted in  $x$  and translated in  $y$ . Misner space (without identification in  $y$  and  $z$ ) is the same as Grant, but with vanishing  $y$  translation ( $a = 0$ ).

Figure 4b is the  $t$ - $x$  portion of the covering space for Grant space. In this covering space, a sequence of copies of Grant space (labeled “copy 1”, “copy 2”, etc.) are lined up side by side, each one boosted by  $\beta$  with respect to the previous one. The (fictitious) wall at which the boosts occur is shown dotted. The events  $P$  and  $Q$  lie on the wall, with each successive copy of  $P$  or  $Q$  displaced into the paper by a distance  $a$  relative to the preceding one.

With the aid of Figure 4b, one can show that the translation along  $y$  does not change the location of the chronology horizon; it is the dark line labeled  $\mathcal{H}_+$  for  $a = 0$  (Misner space) and also for  $a$  finite (Grant space). To show this for finite  $a$ , we need only demonstrate that through any event  $Q$  which is in the claimed nonchronal region but arbitrarily close to the claimed  $\mathcal{H}_+$ , there passes a CTC. One can connect  $Q$  to itself by many geodesics  $C_n$ , with each one circling around Grant space a different number of times,  $n$ . The figure depicts projections of the geodesics  $C_{10}$  and  $C_{14}$  on the  $t$ - $x$  plane; they also extend distances  $10a$  and  $14a$  down the  $y$  axis (into the paper). Since  $C_{14}$  has twice as long a temporal duration  $\Delta\tilde{t}$  in the covering space as  $C_{10}$ , but only makes 40% more trips around Grant space and thus has only a 40% longer extent in the  $y$  direction,  $dy/d\tilde{t}$  is  $1.4/2.0 = 0.7$  as large on  $C_{14}$  as on  $C_{10}$ . As one goes to an ever larger number  $n$



**Figure 5.** (a) Spatial diagram of Gott space. (b) Spacetime diagram showing the chronological structure of Gott space. For ease of visualization, the wedges are not removed around the strings, and to compensate for this, the chronology horizons and generators are shown as curved rather than flat.

of traversals,  $dy/d\tilde{t}$  gets ever smaller (as one can readily show by a detailed calculation), so that eventually, for a sufficiently large number  $n$  of traversals,  $dy/d\tilde{t}$  is small enough for the geodesic  $C_n$  to be timelike.

Although the translation  $a \neq 0$  leaves the location of the horizon unchanged, it alters radically the character of the horizon generators. For  $a = 0$  (Misner space), the generators are all smoothly closed null geodesics (fountains) and the horizon is compactly generated. For  $a \neq 0$ , the generators originate at past null infinity in the covering space and, never intersecting themselves or each other, they travel to the covering space's spacetime origin  $\mathcal{O}$ . Correspondingly, after the identification that produces Grant space, the generators travel around and around Grant space an infinite number of times without ever intersecting themselves or each other. When followed to the past, through an infinite affine parameter, they never leave the future chronology horizon  $\mathcal{H}_+$ . When followed to the future, after a finite affine parameter and an infinite number of circuits, they reach the end of  $\mathcal{H}_+$  and leave it.

A famous example of a spacetime with a noncompactly generated chronology horizon is Gott space [13], which is a solution of the Einstein field equation representing two infinitely long, parallel, straight cosmic strings that move past each other at high speed. Figure 5a depicts the strings in a spatial diagram; they are at the vertices of the wedges and extend into and out of the paper ( $z$  direction) infinitely far. The figure is drawn in the strings' mean rest frame; the upper string moves rightward at speed  $\beta$  and the lower, leftward at speed  $\beta$ . Each string is surrounded by a flat but conical spatial geometry, which can be obtained by removing the indicated wedge from Euclidean space and identifying its edges. The identification is synchronous in the rest frame of the string, which means that for the upper, rightward moving string the event labeled  $t = 1$ , at Lorentz time  $t = 1$ , is identified with that labeled  $t = -1$ , at Lorentz time  $t = -1$ , and similarly for the lower string.

For a suitable choice of parameters, the dark vertical line in the diagram is a CTC.



It begins at the “starting point” labeled  $S$ , at time  $t = 0$  in the mean rest frame, when the two strings are just passing each other. It moves upward to meet the right edge of the upper string’s wedge at  $t = 1$ . It passes through the wedge, emerging at  $t = -1$  when the right string was near the left edge of the diagram. It then travels downward to meet the lower, left-moving string’s wedge at  $t = 1$ , passes through the wedge emerging at  $t = -1$  when the wedge was near the right edge of the diagram, and then travels upward to its starting event  $S$ .

Cutler [14] has deduced the chronological structure of Gott space. It is shown in Figure 5b, topologically correctly but not geometrically correctly. (The geometrically correct depiction, shown in Fig. 3 of Cutler’s paper, takes some work to decipher because of the string wedges that are removed; their removal permits the chronology horizons and their generators to be flat planes and straight null lines instead of curved surfaces and curved lines as here.) The future chronology horizon  $\mathcal{H}_+$  has null geodesic generators  $\mathcal{H}_+$  that originate at spatial infinity and, spiraling around and around the moving strings, work their way in to the closed spacelike geodesic  $\mathcal{C} = \mathcal{H}_+ \cap \mathcal{H}_-$ , where they leave  $\mathcal{H}_+$ . The past chronology horizon  $\mathcal{H}_-$  is generated by null geodesics  $\mathcal{G}_-$  that enter  $\mathcal{H}_-$  at  $\mathcal{C}$ , and then spiral their way around and around the moving strings until they reach spatial infinity. The CTCs are confined to the nonchronal region outside  $\mathcal{H}_+ \cup \mathcal{C} \cup \mathcal{H}_-$ , which means that they are bounded away from the strings. In the mean rest frame—indeed, in any Lorentz frame—the horizons extend to temporal infinity; thus, at all times there are CTCs, but at arbitrarily early or late times they are confined to arbitrarily large radii.

If one parallel transports a set of vectors around the strings and back to their starting event, the local Lorentz transformation that relates the returning vectors to the starting vectors is called the *holonomy* of Gott space. One can similarly transport vectors around the closed  $x$ -dimension of Grant space and compute the resulting holonomy. Grant [12] has shown that for suitable choices of parameters, the two holonomies, that of Gott space and that of Grant space, are identical; and he has argued that this, plus the fact that both spaces are flat (except at the string locations) implies that Grant space must actually be the same as a portion of Gott space.

Because of its translation and boost invariance in the  $z$  direction, Gott’s two-string space can be regarded as a solution of the  $2 + 1$  dimensional vacuum Einstein equation for two point masses moving past each other at high speed. This has enabled Deser et. al. [15] and Carroll et. al. [16] to infer, using ideas from the  $2+1$  dimensional theory, that, despite the fact that each of the strings moves at less than the speed of light, taken together they have a *tachyonic* total momentum. Stated more precisely, the strings’ holonomy (in a suitable Lorentz frame) is a pure boost and not a rotation, and this implies in the  $2+1$  theory that their total momentum is spacelike. Deser et. al. and Carroll et. al. argue that this means Gott space is *unphysical* within the framework of  $2+1$  dimensional theory, and it suggests to them that Gott space might also be unphysical (not creatable by realizable initial conditions) in our real  $3+1$  dimensional universe. (For further interesting results on CTCs in  $2+1$  dimensional, point-particle spacetimes, both spatially closed and spatially open, see the references in Carroll et. al. [16].)

Two other arguments have been used to cast doubt on cosmic strings as generators of CTCs: Gott [13], by order-of-magnitude estimates, suggests that, if one tries to make CTCs by the relative motion of two *curved* strings, the strings’ energies in their center of mass frame will become so great that they *might* form a black hole around themselves before the CTCs can arise. More firmly and convincingly, Hawking [2] points

out that finite loops of cosmic string, by themselves, cannot create CTCs because their stress-energy tensor satisfies NEC, and any physically realizable, compactly generated chronology horizon must violate NEC (see the end of Sec. 2.2 above).

Another famous vacuum solution of the Einstein equation that has a non-compactly generated chronology horizon is Kerr spacetime. The exterior of Kerr's outer horizon ( $r > r_+$  in the usual notation) and the region between the outer and inner horizons ( $r_+ > r > r_-$ ) are choral; the inner horizon ( $r = r_-$ ) is a chronology horizon; and the region inside there ( $r < r_-$ ) is nonchronal. It is conventionally argued that, although the choral region is likely to occur in our real universe as the exterior and interior of an old rotating black hole, the spacetime near and inside the chronology horizon will be altered by an instability due to infalling, blueshifted perturbations; and this alteration (hopefully) will prevent CTCs from arising [17].

### 3. Is there a Chronology Protection Mechanism?

The examples in Section 2 show that, according to classical general relativity, a wide variety of circumstances can give rise to CTCs. What attitude should a physicist take to this? The most common attitude is to assert that all such circumstances are *unphysical*: Infinitely long, rotating cylinders are unphysical and (presumably) finite ones will not produce CTCs; our universe does not rotate as fast as the Gödel universe, so Gödel is unphysical; traversable wormholes are unphysical; infinitely long, straight cosmic strings are unphysical; . . .

I do not find such assertions at all satisfying. Physicists' past records in labeling various things as unphysical are not good. For example, Oppenheimer, Wheeler, and others in the 1930s through the 1950s claimed on physical grounds that the trace of the stress-energy tensor cannot be negative, and therefore superdense matter cannot have a pressure that exceeds  $1/3$  its energy density. They were wrong, as Zel'dovich showed in the early 1960s by a simple quantum-field-theory model, and nowadays several plausible equations of state for nuclear matter entail  $T_\alpha^\alpha < 0$ . As another example, it was widely asserted several decades ago that negative energy densities are unphysical, but we now know they are not: quantum field theory predicts negative renormalized energy densities under a variety of circumstances—e.g. in the Casimir vacuum between two electrically conducting plates and in squeezed vacuum states of light, both of which are realized in the laboratory.

This poor record cautions us to keep an open mind about CTCs until we have found a concrete chronology protection mechanism (or mechanisms): a mechanism that will prevent CTCs from arising under *all* conceivable circumstances—e.g. when a hypothetical arbitrarily advanced civilization is using all means at its disposal to produce CTCs. It seems likely to me that the search for such a firm chronology protection mechanism may teach us much about the laws of physics.

It would be rather surprising to me if Nature uses one protection mechanism in one situation (e.g., collapsing, spinning bodies), a different one in another situation (e.g., moving cosmic strings), and a third mechanism in a third situation (e.g., the interior of a spinning black hole). More likely there is one universal mechanism, that always does the job if other mechanisms fail. (Visser [18] has argued for a number of universal

mechanisms, i.e. a “defense in depth” against CTCs. On this I am agnostic; I would be happy to find just one firm, universal mechanism).

In the following subsections I shall discuss the three mechanisms that have seemed most promising in recent years: An enforcement of NEC or ANEC by quantum field theory, a classical instability of future chronology horizons, and a quantum-field instability.

### 3.1. Enforcement of NEC or ANEC

Because compactly generated chronology horizons across which energy flows must always violate NEC, if we knew that NEC is always enforced by the laws of physics, then we could rule out CTCs ever being generated in compact regions of spacetime. This might not be a fully universal chronology protection mechanism, but it would come close.

Unfortunately, quantum field theory—the ultimate arbiter of obedience to energy conditions—insists that NEC *can* be violated; for example, it is violated in the Casimir vacuum and in squeezed states of light.

It may well be that in the real universe, for reasons that we do not yet understand firmly, compactly generated chronology horizons must be of the generic sort illustrated in Figure 2 (contracting, rotating cylinder) and Figure 3 (wormholes): the horizon generators emerge from fountains that necessarily violate ANEC. This makes enforcement of ANEC an attractive possible chronology protection mechanism.

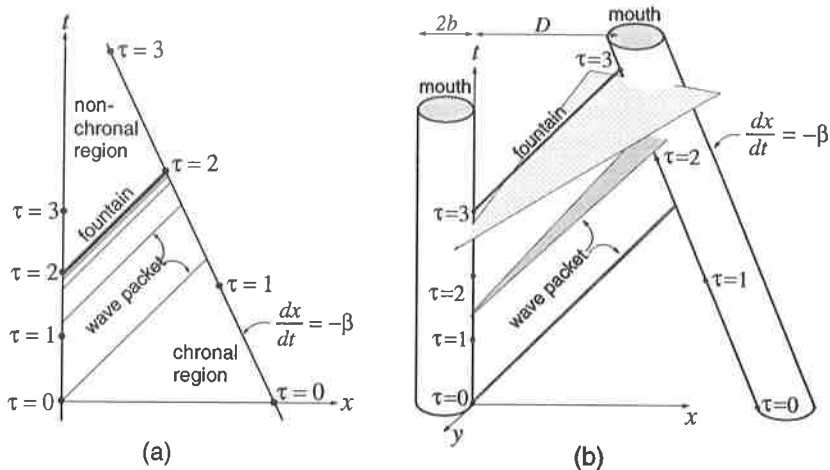
With this motivation, there has been considerable effort in the last several years to determine quantum field theory’s attitude toward ANEC. It has been shown that ANEC is enforced for noninteracting quantized scalar and electromagnetic fields in Minkowski spacetime [19, 20], and in generic, curved 1+1 dimensional spacetimes [21]. On the other hand, in 3+1 dimensions (the real universe), both nontrivial topology [19] and spacetime curvature [21] can induce ANEC violations. Indeed, as Wald and Yurtsever have shown, there are generic classes of curved spacetimes in which quantum fields violate ANEC.

It could still turn out that ANEC is enforced under all circumstances where CTCs try to form, thereby protecting chronology; for example, it might be impossible for quantum fields ever to produce the specific ANEC-violating stress-energy tensors that are required to hold a wormhole open, and therefore wormhole-based CTCs might be forbidden. However, the fact that ANEC *can* be violated under a wide class of generic situations suggests to me that ANEC enforcement is *not* a very promising, universal chronology protection mechanism.

### 3.2. Classical instability of future chronology horizons

The future chronology horizons in the Kerr, Taub-NUT, and Misner spaces are infamously classically unstable. Particles or fields falling into a Kerr black hole, or traveling around the spatially closed Taub or Misner space, become infinitely blue shifted as they near the horizon; and it seems reasonable to hope that the resulting divergent energy density will always act back on the spacetime, via the Einstein equation, to prevent the CTCs from forming.

The example of Misner space is depicted in Figure 6a: A high-frequency electromagnetic wave packet moving along the solid world line gets blue shifted by a factor  $\xi \equiv \sqrt{(1 + \beta)/(1 - \beta)}$  with each passage around the “universe”, and it traverses the universe an infinite number of times as it nears the chronology horizon’s fountain, thereby



**Figure 6.** (a) The motion of a high-frequency electromagnetic wave packet in Misner space. (b) The same wave packet in a wormhole spacetime where CTCs are forming.

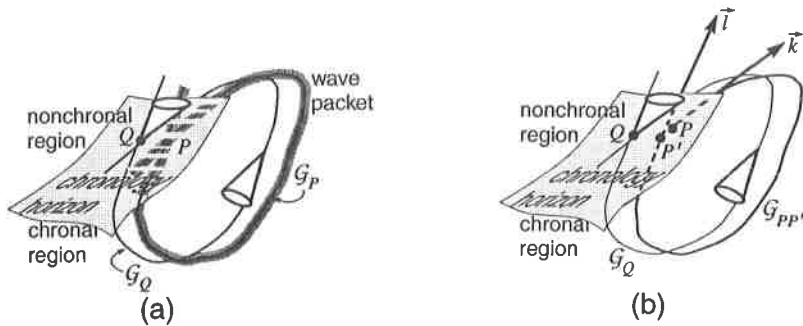
piling up on itself in spacetime and producing a divergent energy density just before the chronology horizon forms.

Until a few years ago, it was widely thought that such instabilities *always* occur at future chronology horizons, thereby protecting chronology. However, wormhole spacetimes provide a counterexample [1], and generalizing this result, Hawking [2] has shown that a generic subset of all compactly generated chronology horizons are counterexamples: they are all classically stable.

A wormhole counterexample is depicted in Figure 6b. This wormhole spacetime is identical to Misner space (Fig. 6a), except that Misner's identified flat walls are converted into spherical "walls" (the wormhole's mouths). The high-frequency wave packet still gets blue shifted by a factor  $\xi \equiv \sqrt{(1+\beta)/(1-\beta)}$  with each circuit through the wormhole, and still tries to pile up on itself at the fountain. However, the wormhole's ANEC-induced diverging-lens action causes the wave packet to spread laterally, driving its amplitude down by a factor  $b/2D$  with each circuit (where  $b$  is the wormhole radius and  $D$  the distance between the mouths, as seen in the left mouth's reference frame, when the horizon forms). If  $(b/2D)\xi < 1$ , i.e. if the distance between the mouths is large enough, then the packet's energy density decreases with each circuit, and the total energy density at the horizon remains finite, despite the pileup. Chronology is not protected—at least not by this mechanism.

### 3.3. Quantum-field instability of future chronology horizons

Our greatest hope—indeed, it seems, a very realistic hope—for universal chronology protection lies in a quantum field instability of all future chronology horizons. This instability was first discovered in 1982, in the context of Misner space, by Hiscock and Konkowski [22]. After Morris, Yurtsever, and I discovered that Misner space's *classical* instability is removed by curving its walls (i.e. by going to a wormhole spacetime) [1], we presumed the same would be true of the quantum instability. We were wrong, as



**Figure 7.** (a) Heuristic explanation of the quantum-field instability of the chronology horizon. (b) Geometric construction for point-splitting computation of the renormalized stress-energy tensor, which induces the instability.

Kim and I [23], Frolov [24], and Gnedin and Kompaneets [26] all independently realized in 1989. The quantum instability is universal; it must arise at every location on every future chronology horizon in any spacetime.

This instability can be described heuristically as due to a piling up of the vacuum fluctuations of any quantum field in the vicinity of any chronology horizon. This pile up causes the fluctuations to have a nonzero renormalized energy density (nonzero *vacuum polarization*) that diverges as one approaches the horizon. The diverging energy density in turn, via the semiclassical Einstein equation, *might* distort the spacetime geometry in such a way as to protect chronology. I shall return to the “might” in the next subsection.

To understand this heuristic explanation in greater detail, consider an arbitrary location on any future chronology horizon  $\mathcal{H}_+$ . Since the horizon is the dividing line between a region with CTCs and one with none, arbitrarily close to  $\mathcal{H}_+$ , on its nonchronal side, there is an event  $Q$  through which passes a CTC. As one pushes  $Q$  closer to  $\mathcal{H}_+$ , the CTC through  $Q$  becomes more nearly null, then null, and then a null geodesic that I shall call  $\mathcal{G}_Q$ ; see Figure 7a. This  $\mathcal{G}_Q$  travels from  $Q$  around a closed loop and back to  $Q$ , but in general does *not* return pointing in the same direction as it started; for this reason it is sometimes called a *self-intersecting null geodesic*.

Now, let  $P$  be an event very close to  $Q$ , but on the chronal side of  $\mathcal{H}_+$ . There will be an almost closed null geodesic  $\mathcal{G}_P$  that starts out at  $P$ , travels along nearly the same route as  $\mathcal{G}_Q$ , and returns very near  $P$ , but cannot quite close itself up at  $P$  because  $P$  is in the chronal region. High-frequency wave-packet modes of any massless quantum field can travel along this  $\mathcal{G}_P$ ; the world tube of such a mode is shown as a dark strip in Figure 7a. The closer  $P$  is to the chronology horizon  $\mathcal{H}_+$ , the closer will  $\mathcal{G}_P$  come to closing up on itself, and if it comes close enough, then the wave packet, with its finite size, will pile up on itself in spacetime near  $P$ , and its piled-up vacuum fluctuations will interfere with themselves in such a way as to produce a nonzero energy density after renormalization.

As  $P$  is pushed closer and closer to  $\mathcal{H}_+$  (and thence to  $Q$ ),  $\mathcal{G}_P$  comes closer and closer to closing up on itself, and correspondingly modes of higher and higher frequency manage to pile up on themselves, with each contributing a nonzero amount to the vacuum polarization. With more and more modes of higher and higher frequency contributing, the renormalized energy density grows larger and larger in magnitude, as  $P$  is pushed

up to  $\mathcal{H}_+$ .

This heuristic picture has been justified by a point-splitting calculation of the renormalized stress-energy tensor  $T^{\alpha\beta}$  for a quantized, noninteracting, massless scalar field  $\hat{\phi}$  [23, 24]. The single point  $P$  is split into two points  $P$  and  $P'$  (Fig. 7b), and the field's regularized Hadamard function  $G_{\text{reg}}^{(1)}(P, P') = \langle \hat{\phi}(P)\hat{\phi}(P') + \hat{\phi}(P')\hat{\phi}(P) \rangle$  is evaluated. The dominant contribution—one due to vacuum fluctuations and therefore independent of the state of the field—comes from scalar-wave propagation of  $\hat{\phi}(P)$  and thence  $G^{(1)}$  around routes close to the null geodesic  $\mathcal{G}_P$  (see Refs. [23] and [24] for careful justifications); it has the usual Hadamard normal form

$$G_{\text{reg}}^{(1)} = \frac{\Delta^{1/2}}{4\pi^2} \left( \frac{1}{\sigma_{PP'}} + \frac{1}{\sigma_{P'P}} \right). \quad (2)$$

Here  $\Delta$  is the Van Vleck-Morette determinant, which measures the amount of focusing ( $\Delta > 1$ ) or defocusing ( $\Delta < 1$ ) that occurs around  $\mathcal{G}_P$  or equally well around  $\mathcal{G}_Q$ ;  $\sigma_{PP'}$  is the geodetic interval along the spacelike (but nearly null) geodesic  $\mathcal{G}_{PP'}$  (Fig. 7b) that leads from  $P$  to  $P'$  by a route that is very close to the null geodesic  $\mathcal{G}_P$ ; and  $\sigma_{P'P}$  is the interval along the similar route that begins at  $P'$  and ends at  $P$ . More specifically,  $\sigma_{PP'} = \int_{\mathcal{G}_{PP'}} g_{\alpha\beta} (dx^\alpha/d\zeta)(dx^\beta/d\zeta) d\zeta$ , with  $\zeta$  an affine parameter that goes from 0 at  $P$  to 1 at  $P'$ , and similarly for  $\sigma_{P'P}$ . The closer is  $P$  (and thus also  $P'$ ) to  $Q$  (which was arbitrarily close to the horizon), the closer will the  $\sigma$ 's be to zero, and thus the larger will be the Hadamard function.

The renormalized stress-energy tensor is computed from the Hadamard function by the standard point-splitting relation

$$T_{\mu\nu} = \ell_{\text{Pl}}^2 \lim_{P' \rightarrow P} \left( \frac{2}{3} \nabla_\mu \nabla_{\nu'} - \frac{1}{3} \nabla_\mu \nabla_\nu - \frac{1}{6} g_{\mu\nu} \nabla_\alpha \nabla^{\alpha'} \right) G_{\text{reg}}^{(1)}, \quad (3)$$

where  $\ell_{\text{Pl}}$  is the Planck length and  $G = c = 1$  so  $\ell_{\text{Pl}} = \sqrt{\hbar}$ . The dominant contribution to  $T_{\mu\nu}$  comes from differentiating twice the nearly-zero and sharply varying  $\sigma$ 's; the result is

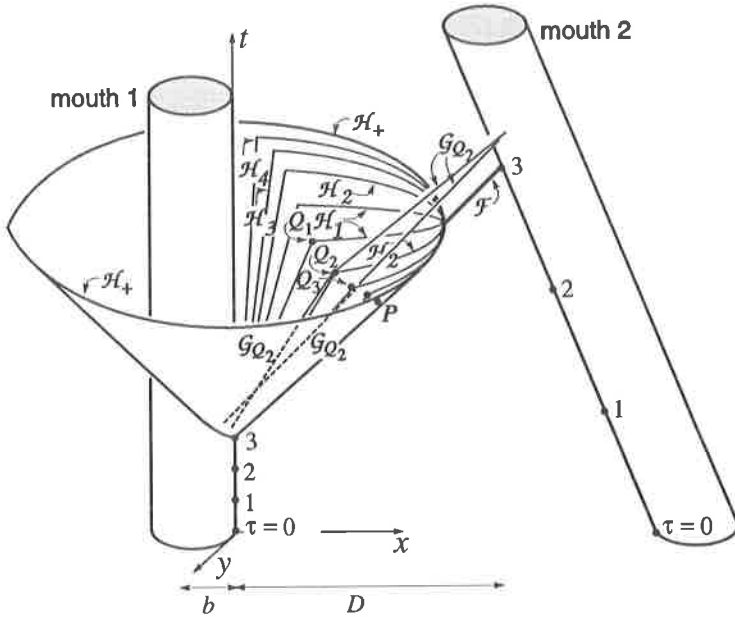
$$T_{\mu\nu} = -\frac{\Delta^{1/2} \ell_{\text{Pl}}^2}{6\pi^2 \sigma^3} (2k_\mu l_\nu + 2l_\mu k_\nu + k_\mu k_\nu + l_\mu l_\nu + g_{\mu\nu} l_\alpha k^\alpha). \quad (4)$$

Here  $k_\alpha = -\lim_{P' \rightarrow P} \nabla_\alpha \sigma_{PP'}$  is the outgoing tangent to  $\mathcal{G}_{PP}$  or equally well to the self-intersecting null geodesic  $\mathcal{G}_Q$ , and  $l_\alpha = +\lim_{P' \rightarrow P} \nabla_\alpha \sigma_{PP'}$  is the returning tangent; see Figure 7b. For  $Q$  arbitrarily close to the horizon (as we have assumed),  $\sigma$  becomes arbitrarily small as  $P$  approaches the horizon, while  $\Delta$ ,  $l_\mu$ , and  $k_\nu$  remain finite; and therefore the renormalized stress-energy tensor becomes arbitrarily large.

The divergence is actually a little more complicated than this, because there is an infinite sequence of events  $Q_1, Q_2, Q_3, \dots$  in the nonchronal region that asymptote to any chosen event on  $\mathcal{H}_+$ , and each of which is connected to itself by a self-intersecting null geodesic  $\mathcal{G}_{Q_n}$ . Each of these events  $Q_n$  with its own  $\mathcal{G}_{Q_n}$  gives rise to a term of the form (4) in the renormalized stress-energy tensor, and the total stress-energy tensor is a sum over all these contributions

$$T^{\mu\nu} = -\sum_{n=1}^{\infty} \frac{\Delta_n^{1/2} \ell_{\text{Pl}}^2}{6\pi^2 \sigma_n^3} (2k_n^\mu l_n^\nu + 2l_n^\mu k_n^\nu + k_n^\mu k_n^\nu + l_n^\mu l_n^\nu + g^{\mu\nu} l_{n\alpha} k_n^\alpha). \quad (5)$$

One can see that there is such an infinite sequence of  $Q$ 's by the following argument [23]: Take the original  $Q$ , and construct a new causal curve that connects this  $Q$  to itself



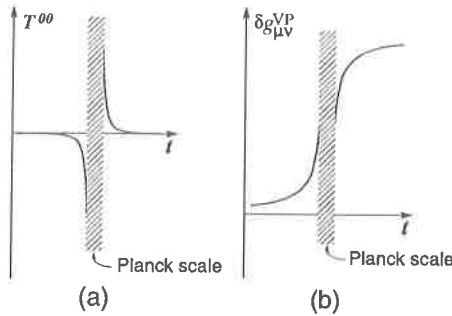
**Figure 8.** The future chronology horizon  $\mathcal{H}_+$  and polarized hypersurfaces  $\mathcal{H}_n$  for the simple wormhole spacetime of Figure 6b. The kink in the left edge of  $\mathcal{H}_+$  is due to a caustic there.

by traveling around the original  $\mathcal{G}_Q$  twice. That curve (call it  $\mathcal{G}_{Q_2}$ ) has a kink in its middle (a discontinuous jump from the null direction  $l_\alpha$  to  $k_\alpha$ ). By moving  $Q$  nearer the horizon (and giving it the new name  $Q_2$ ), while keeping  $\mathcal{G}_{Q_2}$  null, one forces the kink to smooth out and converts  $\mathcal{G}_{Q_2}$  into a null geodesic. Repeating the process indefinitely, one obtains a subsequence of the infinite sequence of  $Q$ 's alluded to above.

Each  $Q_n$  lies on a distinct hypersurface  $\mathcal{H}_n$  made of events that are connected to themselves by self-intersecting null geodesics; and these  $\mathcal{H}_n$ , which are called *polarized hypersurfaces*, asymptote to the future chronology horizon  $\mathcal{H}_+$  in the limit  $n \rightarrow \infty$ . We can regard the order- $n$  term in the vacuum-polarization stress-energy tensor (5) as produced by the presence nearby of the  $n$ 'th polarized hypersurface  $\mathcal{H}_n$ , with its specific event  $Q_n$  and associated self-intersecting null geodesic  $\mathcal{G}_{Q_n}$ .

The wormhole spacetime of Figure 8 provides an example [23]. The polarized hypersurfaces  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4, \dots$  are nested, one inside the next, within the chronology horizon. The event  $Q_n$  on  $\mathcal{H}_n$  is connected to itself by a null geodesic  $\mathcal{G}_{Q_n}$  that traverses the wormhole  $n$  times, and that gives rise to the order- $n$  term in the stress-energy tensor (5) at the event  $P$ .

Explicit evaluations of the vacuum fluctuational stress-energy tensor (5) have been carried out near the fountain of the wormhole spacetime of Figure 8 by Kim and Thorne [23] (and for arbitrarily slow wormhole motions,  $\beta \rightarrow 0$ , by Visser [18]), near the fountain of other wormhole spacetimes by Frolov [24], near the fountain of a generic, compactly generated horizon by Klinkhammer [25], and near the non-compactly-generated horizon



**Figure 9.** The vacuum-polarization-induced stress-energy tensor, and the metric perturbations it produces as one passes through  $\mathcal{H}_+$  or  $\mathcal{H}_n$ .

of the Grant/Gott spacetime by Grant [12]. A central issue in these and all other cases is this:

### 3.4. Does the back-action of the vacuum-polarization energy protect chronology?

By inserting the  $T_{\mu\nu}$  of Eq. (3) into the semiclassical Einstein equation and performing a rough order-of-magnitude integration, one obtains an estimate for the metric perturbations created by the vacuum-polarization energy of the quantized scalar field (and also of any other nongravitational field):

$$\delta g_{\mu\nu}^{VP} \sim G_{\text{reg}}^{(1)} \sim \sum_n \Delta_n^{1/2} \frac{\ell_{\text{Pl}}^2}{\sigma_n}. \quad (6)$$

Although these metric perturbations diverge at the chronology horizon and at each subsequent polarized hypersurface, the divergences can be remarkably slow—so slow that it is conceivable, under some circumstances, that quantum gravity will invalidate the above analysis before the spacetime has been altered substantially [23]. Note that, as one passes through the  $n$ 'th polarized hypersurface,  $\sigma_n$  passes through zero and reverses sign. Correspondingly, if quantum gravity were simply to smooth out the divergences in Eqs. (5) and (6), one would see the vacuum polarization produce, on the nonchronal side of  $\mathcal{H}_+$  and  $\mathcal{H}_n$ , a  $T_{\mu\nu}$  equal in magnitude but opposite in sign to that on the chrontal side. Therefore, as observers approach and then pass through  $\mathcal{H}_+$  or  $\mathcal{H}_n$ , they might see this  $T_{\mu\nu}$  first distort the spacetime geometry, and then undo the distortions it had produced, as illustrated in Figure 9.

Such a scenario is highly speculative, but seems to me plausible *if* the divergence is sufficiently weak.

Just how strong is the divergence? The most important place to ask this question is at the chronology horizon  $\mathcal{H}_+$  rather than at a polarized hypersurface  $\mathcal{H}_n$ , because that is where CTCs first arise. For a compactly generated  $\mathcal{H}_+$ , the divergence is much stronger at the horizon's fountains than away from the fountains. This is because the polarized hypersurfaces  $\mathcal{H}_n$  (at which the  $\sigma_n = 0$ ) are all tangent to  $\mathcal{H}_+$  at any fountain  $\mathcal{F}$ , but each  $\mathcal{H}_n$  is finitely separated from  $\mathcal{H}_+$  away from the fountains [23]; see Figure 8 for an example. If, as I have conjectured in Section 2.2, almost all the horizon generators originate on fountains, then a divergence that is strong enough at the fountains to distort



the spacetime geometry significantly there will have its influence propagate over the entire horizon and perhaps thereby protect chronology everywhere. Conversely, if the divergence is too weak to protect chronology at the fountains, it probably is too weak to protect chronology elsewhere.

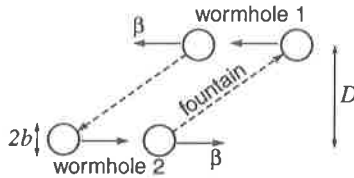
As an example, consider the wormhole spacetime of Figure 8, as examined by an observer who sits on the left mouth and on the  $x$ -axis, where the fountain  $\mathcal{F}$  arises. Assuming the mouth speed  $\beta$  is not too close to zero or one, at a time  $\Delta t$  before crossing the horizon this observer sees a value  $\sigma_n \sim D\Delta t$  for the geodetic interval along the closed, spacelike geodesic that traverses the wormhole  $n$  times, and he sees  $\Delta_n^{1/2} \sim (b/2D)^{n-1}$  for the amount of defocusing around that geodesic [23]. (There is no net defocusing on the geodesic's first trip because it first passes through the wormhole—the “diverging lens”—only at the end of that trip; however, the first wormhole passage produces a net defocusing of  $b/2D$  during the second trip, the second wormhole passage produces another defocusing  $b/2D$  during the third trip, etc.) Correspondingly, the spacetime distortion measured on the wormhole throat is

$$\delta g_{\alpha\beta}^{\text{VP}} \sim \frac{\ell_{\text{P1}}^2}{D\Delta t} \sum_n \left( \frac{b}{2D} \right)^{n-1}. \quad (7)$$

When Kim and I first computed this back-action of the vacuum polarization, it seemed to me to be extremely weak. “Surely,” I said to myself, “the analysis will break down during a time interval  $\Delta t \sim \ell_{\text{P1}}$  around the passage through the horizon (the ‘Planck region’ of Fig. 9), since ‘time’ does not make classical sense on such short scales.” If this were true, and quantum gravity were to smooth out the divergence, then the metric distortion just before smooth-out would be much too small,  $\delta g_{\alpha\beta}^{\text{VP}} \sim \ell_{\text{P1}}/D \sim 10^{-35}$  for  $D \sim 1$  meter, to protect chronology.

Hawking [2] has convinced me that this assessment is wrong [23, 27]. The distance  $D$  between the mouths and the time  $\Delta t$  until the Chronology horizon depend on the observer's reference frame, he points out, but the product  $D\Delta t$  does not (as one can see from the fact that  $\sigma_n \sim D\Delta t$  is an invariant). Therefore, he conjectures, it may well be that the spacetime remains classical, near the chronology horizon, and the computed  $\delta g_{\mu\nu}^{\text{VP}}$  of Eq. (7) remains correct, until the product  $D\Delta t$  gets as small as  $\ell_{\text{P1}}^2$ , and correspondingly  $\delta g_{\mu\nu}^{\text{VP}}$  reaches unity. The resulting distortion of the classical spacetime geometry might then be sufficient to protect chronology, Hawking speculates.

If Hawking were right, and the relevant Planck region were  $D\Delta t \sim \ell_{\text{P1}}^2$ , then there is a strategy that an arbitrarily advanced civilization could use to circumvent chronology protection. The civilization need only make sure that the fountain encounters two or more widely separated regions of defocusing, instead of only one as in Figure 8. This could be done, for example, by using two wormholes to make CTCs, with time through each wormhole synchronously identified in the wormhole's rest frame, and the two wormholes moving in the manner of Figure 10. (Such a spacetime was suggested to Mike Morris and me several years ago by Tom Roman.) The fountain would have the indicated form, the defocusing as measured on a wormhole mouth would be  $\Delta_n^{1/2} \sim (b/2D)^{2n-1}$  rather than the  $(b/2D)^{n-1}$  of Figure 8, and thus for a large wormhole separation,  $D \gg b$ , the resulting metric perturbation on the mouth (which is dominated by  $n = 1$ ) would be  $\delta g_{\mu\nu}^{\text{VP}} \sim (b/2D)(\ell_{\text{P1}}^2/D\Delta t)$  [28]. By making  $b/2D$  arbitrarily small, the advanced civilization could force  $\delta g_{\mu\nu}^{\text{VP}}$  to be arbitrarily small at the beginning of Hawking's conjectured Planck



**Figure 10.** Spatial diagram of Roman's wormhole spacetime.

region,  $D\Delta t \sim \ell_{\text{Pl}}^2$ . If quantum gravity were then to provide a cutoff and smooth-out, chronology would not be protected.

Even weaker is the vacuum-polarization divergence in Gott/Grant spacetime, as computed by Grant [12]. Since the chronology horizon is non-compactly generated, it has no fountains, and every polarized hypersurface is everywhere separated from the horizon by a finite distance. (More specifically, in the covering space of Figure 4b, the  $n$ 'th polarized hypersurface  $\mathcal{H}_n$  consists of events connected to themselves by self-intersecting null geodesics such as  $\mathcal{C}_{10}$  of the figure, that circle around the  $\tilde{x}$ -direction of the universe  $n$  times; this  $\mathcal{H}_n$  is the hyperbola  $\tilde{x}^2 - \tilde{t}^2 = a^2 n^2 \xi^{-n} (1 - \xi^{-n})^{-2}$ , where  $\xi = \sqrt{(1 + \beta)/(1 - \beta)}$  is the blue shift produced by each propagation around the universe and  $a$  is the  $y$ -translation that makes the horizon non-compactly generated.) When Grant sums over the vacuum polarization contributions from all the polarized hypersurfaces, each one finitely displaced from the horizon but approaching the horizon in the limit  $n \rightarrow \infty$ , and when he then computes the resulting back action on the spacetime metric, he obtains

$$\delta g_{\alpha\beta}^{\text{VP}} \sim \frac{\ell_{\text{Pl}}^2}{a^2} \ln \left( \frac{\tilde{t}^2 - \tilde{x}^2}{a^2} \right). \quad (8)$$

Although this metric perturbation diverges as one approaches the horizon,  $\tilde{t} \rightarrow \tilde{x}$ , the divergence, being logarithmic in time with a coefficient  $\ell_{\text{Pl}}^2/a^2$  that can be made arbitrarily small, is extremely weak. It is even harder here than in the Roman spacetime to see how such a divergence can protect chronology.

Nevertheless, I suspect that it may do so. It may well be that quantum gravity invalidates the semiclassical analysis only when metric fluctuations, treated as a spin-two field on the classical background, develop mean-square fluctuations of order unity *as a result of the same pileup process as induces the vacuum polarization of nongravitational fields*. If so, then the semiclassical analysis, just before *every* chronology horizon, might remain valid up to the location where  $\delta g_{\mu\nu}^{\text{VP}} \sim 1$ , and only then fail.

To determine whether this is so, and to determine the nature of the subsequent evolution of spacetime, will require an understanding of quantum gravity. Indeed, it may be that efforts to decipher these issues will teach us useful things about quantum gravity.

#### 4. Physics in the presence of closed timelike curves

It may turn out that on macroscopic lengthscales chronology is *not* always protected, and even if chronology *is* protected macroscopically, quantum gravity may well give

finite probability amplitudes for microscopic spacetime histories with CTCs [29]. For these reasons, some effort has been devoted recently to exploring whether and how the laws of physics might adapt themselves to CTCs [7]. In this concluding section, I shall summarize very sketchily what has been learned.

The cleanest of such explorations are carried out in spacetimes, such as Figure 3c, that have a chronal “IN” region, followed by a compact nonchronal region, followed by a chronal “OUT” region. Initial data are posed in the IN region for some physical system, and the system is then evolved from the IN region through the nonchronal region and into the OUT region. The evolutionary laws are generally chosen to be the most conservative possible—the same laws, at least locally, in the nonchronal region as one is accustomed to in everyday, chronal physics—and one asks whether the evolution problem is well posed, i.e. whether standard initial data in the IN region produce a unique evolution through the chronal region and into the OUT region.

For *noninteracting, classical* systems (particles [7] and fields [30]) the answer appears to be yes; there does exist a unique evolution. However, just as interactions produce evolutionary problems in science fiction (e.g., one can go back in time and kill one’s younger self), so also interactions produce trouble for classical particles [31] and presumably also for classical fields: One finds that a large number of classical evolutions can follow from a single, standard set of initial data. It was thought, at first, that for some initial data there might be *no* self-consistent evolutions; but thus far no clean examples of such a thing have been exhibited in classical, continuum physics [31, 32]. (On the other hand, there *are* examples in simple, highly idealized, discrete models [33].)

Of course, physics is quantum mechanical at heart, not classical, and it is in the quantum domain that these studies become especially fruitful. Just as in quantum cosmology, where there is *no a priori* notion of “time”, so also in nonchronal spacetimes, where CTCs alter the nature of time, the only viable approach to quantum mechanics seems to be Feynman’s sum over histories. Indeed, spacetimes with CTCs have become a useful testbed for the sum-over-histories formulations of quantum theory that are being developed for use in quantum cosmology [34].

It turns out that for nonrelativistic particles [35] and also for relativistic fields [37], the sum-over-histories formalism enables one to compute unique probabilities for the outcomes of all measurements that one might reasonably try to make, even in the nonchronal region of spacetime. However, when the particles or fields are self-interacting, their interactions produce peculiar phenomena: (i) the propagators from the IN region to the OUT region are not unitary—but nevertheless, there is no loss of probability [35, 36, 37]; and (ii) although one recovers standard Hamiltonian quantum mechanics in the chronal OUT region, one does not recover it in the chronal IN region, and the fact that CTCs exist to the future of the IN region influences probabilities in the IN region itself [37]. The strength of this influence, and how it grows as one approaches the future chronology horizon, are not as yet understood.

In summary, these studies are giving us glimpses of how CTCs influence physics; but whether those glimpses are teaching us something deep and important, or we are just playing fun mental games, is far from clear.

## Acknowledgments

The author's research on this topic was supported by the National Science Foundation Grant AST-8817792 until February 1992. Since then, both the Astronomy and the Physics Divisions of NSF have declined to support this research, so it has been supported solely by Caltech's Feynman Research Fund. This manuscript was written at the Institute for Theoretical Physics in Santa Barbara, California, with support from National Science Foundation Grant PHY-8904035.

## References

- [1] Morris M S, Thorne K S, and Yurtsever, U 1988 *Phys. Rev. Lett.* **61** 1446–1449
- [2] Hawking S W 1992 *Phys. Rev. D* **46** 603–611
- [3] Hawking S W and Ellis G F R 1973 *The large scale structure of space-time* (Cambridge: Cambridge University Press)
- [4] Van Stockum W J 1937 *Proc. Roy. Soc. Edin.* **57** 135–154
- [5] Bonnor W B 1980 *J. Phys. A: Math. Gen.* **13** 2121–2132
- [6] Gödel K 1949 *Rev. Mod. Phys.* **21** 447
- [7] Friedman J, Morris M S, Novikov I D, Echeverria F, Klinkhammer G, Thorne K S, and Yurtsever U *Phys. Rev. D* **42** 1915–1930
- [8] Frolov V P and Novikov I D 1990 *Phys. Rev. D* **42** 1057–1065
- [9] Taub A H 1951 *Ann. Math.* **53** 472  
Newman E T, Tamburino L, and Unti T J 1963; *J. Math. Phys.* **4** 915–923.
- [10] Misner C W 1967 in *Relativity Theory and Astrophysics I. Relativity and Cosmology* ed J Ehlers (Providence RI: American Mathematical Society) pp. 160–169  
Thorne K S 1993 in *Directions in General Relativity* eds BL Hu et al (Cambridge: Cambridge University Press)
- [11] Tipler F J 1977 *Ann. Phys.* **108**, 1–36.
- [12] Grant J D E 1993 *Phys. Rev. D* in press
- [13] Gott J R 1991 *Phys. Rev. D* **66** 1126–1129
- [14] Cutler C 1992 *Phys. Rev. D* **45** 487–494
- [15] Deser S, Jackiw R., and 't Hooft, G 1992 *Phys. Rev. Lett.* **68** 267–269
- [16] Carroll S M, Farhi E, and Guth, A H 1993 *Phys. Rev. D* in press
- [17] Ori A 1992 *Phys. Rev. Lett.* **68** 2117 and references therein
- [18] Visser M 1993 *Phys. Rev. D* **47** 554–565
- [19] Klinkhammer G 1992 *Phys. Rev. D* **43** 2542–2548
- [20] Folacci A 1993 *Phys. Rev. D* **47** 2726–2729
- [21] Wald R M and Yurtsever U 1991 *Phys. Rev. D* **44** 403–416
- [22] Hiscock W A and Konkowski D A 1982 *Phys. Rev. D* **26** 1225–1230
- [23] Kim S-W and Thorne K S 1991 *Phys. Rev. D* **43** 3929–3947
- [24] Frolov V P 1991 *Phys. Rev. D* **43** 3878–3894

- [25] Klinkhammer G 1992 *Phys. Rev. D* **46** 3388–3394
- [26] Gnedin N N 1991 private communication  
Kompaneets D A 1991 private communication.
- [27] Thorne K S 1991 *Ann. New York Acad. Sci.* **631** 182–193.
- [28] Lyutikov M 1993 research in progress.
- [29] Friedman J 1992 in *Proceedings of the 4th Canadian Conf. on General Relativity and Relativistic Astrophysics* eds G Kunstatter et al (Singapore: World Scientific) pp. 183–199
- [30] Friedman J L and Morris M S 1991 *Phys. Rev. Lett.* **66** 401–404
- [31] Echeverria F, Klinkhammer G, and Thorne, K S 1991 *Phys. Rev. D* **44** 1077–1099
- [32] Novikov I D 1992 *Phys. Rev. D* **45** 1989–1994; Lossev A and Novikov I D 1992 *Class. Quant. Grav.* **9** 2309–2321
- [33] Deutsch D 1991 *Phys. Rev. D* **44** 3197–3217
- [34] Hartle J B 1993 “Unitarity and causality in generalized quantum mechanics for acausal spacetimes” in preparation
- [35] Politzer H D 1992 *Phys. Rev. D* **46** 4470–4476
- [36] Boulware D G 1992 *Phys. Rev. D* **46** 4421–4441
- [37] Friedman J L, Papastamatiou N J, and Simon J Z 1992 *Phys. Rev. D* **46** 4442–4455, and **46** 4456–4469

