

Can the Universe create itself?

J. Richard Gott III and Li-Xin Li

Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544

(Received 30 December 1997; published 29 May 1998)

The question of first-cause has troubled philosophers and cosmologists alike. Now that it is apparent that our universe began in a big bang explosion, the question of what happened before the big bang arises. Inflation seems like a very promising answer, but as Borde and Vilenkin have shown, the inflationary state preceding the big bang could not have been infinite in duration—it must have had a beginning also. Where did it come from? Ultimately, the difficult question seems to be how to make something out of nothing. This paper explores the idea that this is the wrong question—that *that* is not how the Universe got here. Instead, we explore the idea of whether there is anything in the laws of physics that would prevent the Universe from creating itself. Because spacetimes can be curved and multiply connected, general relativity allows for the possibility of closed timelike curves (CTCs). Thus, tracing backwards in time through the original inflationary state we may eventually encounter a region of CTCs—giving *no* first-cause. This region of CTCs may well be over by now (being bounded toward the future by a Cauchy horizon). We illustrate that such models—with CTCs—are *not* necessarily inconsistent by demonstrating self-consistent vacuums for Misner space and a multiply connected de Sitter space in which the renormalized energy-momentum tensor does not diverge as one approaches the Cauchy horizon and solves Einstein's equations. Some specific scenarios (out of many possible ones) for this type of model are described. For example, a metastable vacuum inflates producing an infinite number of (big-bang-type) bubble universes. In many of these, either by natural causes or by action of advanced civilizations, a number of bubbles of metastable vacuum are created at late times by high energy events. These bubbles will usually collapse and form black holes, but occasionally one will tunnel to create an expanding metastable vacuum (a baby universe) on the other side of the black hole's Einstein-Rosen bridge as proposed by Farhi, Guth, and Guven. One of the expanding metastable-vacuum baby universes produced in this way simply turns out to be the original inflating metastable vacuum we began with. We show that a Universe with CTCs can be stable against vacuum polarization. And it can be classically stable and self-consistent if and only if the potentials in this Universe are retarded—which gives a natural explanation of the arrow of time in our universe. Interestingly, the laws of physics may allow the Universe to be its own mother.

[S0556-2821(98)00614-6]

PACS number(s): 98.80.Cq, 04.20.Gz, 04.62.+v

I. INTRODUCTION

The question of first-cause has been troubling to philosophers and scientists alike for over two thousand years. Aristotle found this sufficiently troubling that he proposed avoiding it by having the Universe exist eternally in both the past and future. That way, it was always present and one would not have to ask what caused it to come into being. This type of model has been attractive to modern scientists as well. When Einstein developed general relativity and applied it to cosmology, his first cosmological model was the Einstein static universe, which had a static S^3 spatial geometry which lasted forever, having no beginning and no end [1].

As we shall discuss, since the big bang model's success, models with a finite beginning have taken precedence, even when inflation and quantum tunneling are included. So the problem of first-cause reasserts itself. The big question appears to be how to create the universe out of nothing. In this paper we shall explore the idea that this is the wrong question. A remarkable property of general relativity is that it allows solutions that have closed timelike curves (CTCs) [2–8] (for review see [9,10]). Often, the beginning of the universe, as in Vilenkin's tunneling model [11] and Hartle and Hawking's no-boundary model [12], is pictured as being like the south pole of the Earth and it is usually said that asking

what happened before that is like asking what is south of the south pole [13]. But, suppose the early universe contains a region of CTCs. Then, asking what was the earliest point might be like asking what is the easternmost point on the Earth. You can keep going east around and around the Earth—there is no eastern-most point. In such a model every event in the early universe would have events that preceded it. This period of CTCs could well have ended by now, being bounded by a Cauchy horizon. Some initial calculations of vacuum polarization in spacetimes with CTCs indicated that the renormalized energy-momentum tensor diverged at the Cauchy horizon separating the region with CTCs from the region without closed causal curves, or at the polarized hypersurfaces nested inside the Cauchy horizon [14–18]. Some of these results motivated Hawking [19,20] to propose the chronology protection conjecture which states that the laws of physics do not allow the appearance of CTCs. But, a number of people have challenged the chronology protection conjecture by giving counter-examples [16,9,21–30]. In particular, Li and Gott [30] have recently found that there is a self-consistent vacuum in Misner space for which the renormalized energy-momentum tensor of vacuum polarization is zero everywhere. (Cassidy [31] has independently given an existence proof that there should be a quantum state for a conformally coupled scalar field in Misner space, for which

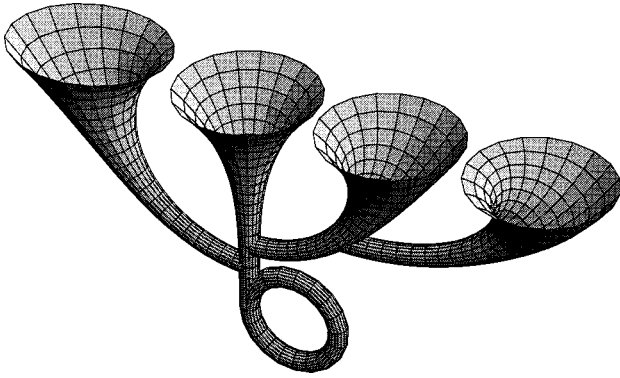


FIG. 1. A self-creating Universe scenario. Four inflating baby universes are shown—A, B, C, and D—from left to right. Universes A and D have not created any baby universes so far. Universe C has created universe D. Universe B has created three universes: universe A, universe C and itself. The toroidal—shaped region at the bottom is a region of CTCs (closed timelike curves). The region is bounded to the future by a Cauchy horizon, after which, there are no CTCs. Universes A, C, and D, for example, are formed after the Cauchy horizon when the epoch of CTCs is already over.

the renormalized energy-momentum tensor is zero everywhere, but he has not shown what state it should be. Li and Gott [30] have found that it is the “adapted” Rindler vacuum.) In this paper we give some examples to show how it is possible in principle to find self-consistent vacuum states where the renormalized energy-momentum tensor does not blow up as one approaches the Cauchy horizon. To produce such a region of CTCs, the universe must, at some later time, be able to reproduce conditions as they were earlier, so that a multiply connected solution is possible. Interestingly, inflation is well suited to this. A little piece of inflationary state expands to produce a large volume of inflationary state, little pieces of which resemble the starting piece. Also there is the possibility of forming baby universes at late times where new pieces of inflating states are formed. Farhi, Guth, and Guven [32], Harrison [33], Smolin [34,35], and Garriga and Vilenkin [36] have considered such models. If one of those later inflating pieces simply turns out to be the inflating piece that one started out with, then the Universe can be its own mother. Since an infinite number of baby universes are created, as long as the probability of a particular multiple connection forming is not exactly zero, then such a connection might be expected, eventually. Then the Universe neither tunneled from nothing, nor arose from a singularity; it created itself (Fig. 1).

Before discussing this approach to the first-cause problem, let us review just how troublesome this problem has been. As we have noted, Einstein [1] initially tried to avoid it by siding with Aristotle in proposing a model which had an infinite past and future. The Einstein static universe appears to be the geometry Einstein found *a priori* most aesthetically appealing, thus presumably he started with this preferred geometry and substituted it into the field equations to determine the energy-momentum tensor required to produce it. He found a source term that looks like dust (stars) plus a term that was proportional to the metric which he called the cosmological constant. The cosmological constant, because of

its homogeneous large negative pressure, exerts a repulsive gravitational effect offsetting the attraction of the stars for each other; allowing a static model which could exist (ignoring instabilities, which he failed to consider) to the infinite past and future. If one did not require a static model, there would be no need for the cosmological constant. Friedmann [37] calculated models without it, of positive, negative or zero curvature, all of which were dynamical. When Hubble [38] discovered the expansion of the universe, Einstein pronounced the cosmological constant the biggest blunder of his life.

But now there was a problem: all three Friedmann models ($k=0$, $k=1$, and $k=-1$) that were expanding at the present epoch had a beginning in the finite past (see e.g. [39,40]). In the Friedmann models the universe began in a singularly dense state at a finite time in the past. The equations could not be pushed beyond that finite beginning singularity. Furthermore, if today’s Hubble constant is H_0 , then all of the Friedmann models had ages less than $t_H = H_0^{-1}$. The universe thus began in a big bang explosion only a short time ago, a time which could be measured in billions of years. The universe was not infinitely old. Gamow [41,42] and his colleagues Alpher and Herman [43] calculated the evolution of such a big bang cosmology, concluding correctly that in its early phases it should have been very dense and very hot, and that the thermal radiation present in the early universe should still be visible today as microwave radiation with a temperature of approximately 5 K. Penzias and Wilson’s discovery of the radiation with a temperature of 2.7 K [44] cinched the case for the big bang model. The Cosmic Background Explorer (COBE) results which have shown a beautifully thermal spectrum [45,46] and small fluctuations in the temperature $\delta T/T \approx 10^{-5}$ [47], fluctuations that are of approximately the right magnitude to grow into the galaxies and clusters of galaxies that we see at the present epoch, have served to make the big bang model even more certain. With the big bang model in ascendancy, attention focused on the initial singularity. Hawking and Penrose proved a number of singularity theorems [48,49,40] showing that, with some reasonable constraints on the energy-momentum tensor, if Einstein’s equations are correct and the expansion of the universe is as observed today, there is no way to avoid an initial singularity in the model; that is, initial singularities would form even in models that were not exactly uniform. So the initial singularity was taken to be the first-cause of the Universe. This of course prompted questions of what caused the singularity and what happened before the singularity. The standard answer to what happened before the big bang singularity has been that time was created at the singularity, along with space, and that there was no time before the big bang. Asking what happened before the big bang was considered to be like asking what is south of the south pole. But particularly troublesome was the question of what caused the initial singularity to have its almost perfect uniformity—for otherwise the microwave background radiation would be of vastly different temperatures in different directions on the sky. Yet the initial singularity could not be exactly uniform, for then we would have a perfect Friedmann model with no fluctuations which would form no galaxies. It needed to be

almost, but not quite perfectly uniform—a remarkable situation—how did it get that way? These seemed to be special initial conditions with no explanation for how they got that way.

Another problem was that singularities in physics are usually smeared by quantum effects. As we extrapolated back toward the initial singularity (of infinite density), we would first reach a surface where the density was equal to the Planck density and at this epoch classical general relativity would break down. We could not extrapolate confidently back to infinite density, we could only say that we would eventually reach a place where quantum effects should become important and where classical general relativity no longer applied. Since we do not have a theory of quantum gravity or a theory-of-everything we could honestly say that the singularity theorems only told us that we would find regions in the early universe where the density exceeded the grand unified theory (GUT) or Planck densities beyond which we did not know what happened—rather much like the Terra Incognita of old maps. We could not then say how our universe formed.

So, questions about how the initial big bang singularity was formed and what preceded it remained. The closed Friedmann model, popular because it is compact and therefore needs no boundary conditions, re-collapses in a finite time in the future to form a big crunch singularity at the end. Singularity theorems tell us that in a collapsing universe the final big crunch singularity cannot be avoided. Classical general relativity tells us that a closed universe begins with a singularity and ends with a singularity, with nothing before and nothing after. Nevertheless, many people speculated that there could be more than one connected cycle—after all, the singularities only indicated a breakdown of classical general relativity and the quantum Terra Incognita at the Planck density might allow a cosmology collapsing toward a big crunch to bounce and make another big bang [50–52]. In support of this is the fact that de Sitter space (representing the geometry of a false vacuum—an inflationary state as proposed by Guth [54]—with a large cosmological constant) looks like a spatially closed S^3 universe whose radius as a function of proper time is $a(t) = r_0 \cosh(t/r_0)$, where $r_0 = (3/\Lambda)^{1/2}$ is the radius of the de Sitter space and Λ is the cosmological constant (throughout the paper we use units $G = c = \hbar = k_B = 1$), which is a collapsing cosmology which bounces and turns into an expanding one. Thus if quantum gravitational effects make the geometry look like de Sitter space once the density reaches the Planck density as some have suggested [55–57], then a big crunch singularity might be avoided as the closed universe bounced and began a big bang all over again. This bouncing model avoids the first-cause problem. The answer to what caused our universe in this model is “the collapse of the previous universe,” and so on. An infinite number of expansion and contraction cycles make up the Universe (note the capital U—in this paper this denotes the ensemble of causally connected universes) which consists of an infinite number of closed big bang models laid out in time like pearls on a string. The Universe (the infinite string of pearls) has always been in existence and will always be in existence, even though our cycle, our standard closed big bang cosmology

(our pearl) has a finite duration. So we are back to Aristotle, with an eternal Universe, and close to Einstein with just an oscillating (rather than static) closed Universe that has infinite duration to the past and future. Thus in this picture there is no first-cause because the Universe has existed infinitely far back in the past.

The oscillating universe was thought to have some problems with entropy [53]. Entropy is steadily increasing with time, and so each cycle would seem to be more disordered than the one that preceded it. Since our universe has a finite entropy per baryon it was argued, there could not be an infinite number of cycles preceding us. Likewise it was argued that each cycle of the universe should be larger than the preceding one, so if there were an infinite number preceding us, our universe would have to look indistinguishable from flat (i.e., closed but having an infinite radius of curvature). The real challenge in this model is to produce initial conditions for our universe (our pearl) that were as uniform and low entropy as observed. COBE tells us that our universe at early times was uniform to one part in a hundred thousand [47]. At late times we expect the universe at the big crunch to be very non-uniform as black hole singularities combine to form the big crunch. In the early universe the Weyl tensor is zero, whereas at the big crunch it would be large [58,59]. How does the chaotic high-entropy state at the big crunch get recycled into the low-entropy, nearly uniform, state of the next big bang? If it does not, then after an infinite number of cycles, why are we not in a universe with chaotic initial conditions?

Entropy and the direction of time may be intimately tied up with this difference between the big bang and the big crunch. Maxwell’s equations (and the field equations of general relativity) are time-symmetric, so why do we see only retarded potentials? Wheeler and Feynman addressed this with their absorber theory [60]. They supposed that an electron shaken today produces half-advanced-half-retarded fields. The half-advanced fields propagate back in time toward the early universe where they are absorbed (towards the past the universe is a perfect absorber) by shaking charged particles in the early universe. These charged particles in turn emit half-advanced-half-retarded fields; their half-retarded fields propagate toward the future where they: (a) perfectly cancel the half-advanced fields of the original electron, (b) add to its retarded fields to produce the electron’s full retarded field, and (c) produce a force on the electron which is equal to the classical radiative reaction force. Thus, the electron only experiences forces due to fields from other charged particles. This is a particularly ingenious solution. It requires only that the early universe is opaque—which it is—and that the initial conditions are low-entropy; that is, there is a cancellation of half-advanced fields from the future by half-retarded fields from the past, leaving no “signals” in the early universe from later events—a state of low-entropy. (Note that this argument works equally well in an open universe where the universe may not be optically thick toward the future—all that is required is that the universe be a perfect absorber in the past, i.e., toward the state of low-entropy.) Wheeler and Feynman noted that entropy is time-symmetric like Maxwell’s equations. If you find an ice cube

on the stove, and then come back and re-observe it a minute later, you will likely find it half-melted. Usually an ice cube gets on a stove by someone just putting it there (initial conditions), but suppose we had a truly isolated system so that the ice cube we found was just a statistical fluctuation. Then if we asked what we would see if we had observed one minute *before* our first observation, we will also be likely to see a half-melted ice cube, for finding a still larger ice cube one minute before would be unlikely because it would represent an even more unlikely statistical fluctuation than the original ice cube. In an *isolated* system, an (improbable) state of low-entropy is likely to be both followed and preceded by states of higher-entropy in a time-symmetric fashion. Given that the early universe represents a state of high order, it is thus not surprising to find entropy increasing after that. Thus, according to Wheeler and Feynman [60], the fact that the retarded potential arrow of time and the entropy arrow of time point in the same direction is simply a reflection of the low-entropy nature of the big bang. The big crunch is high-entropy, so time follows from past to future between the big bang and the big crunch.

Thus, in an oscillating universe scenario, we might expect entropy to go in the opposite direction with respect to time, in the previous cycle of oscillation. In that previous universe there would be only advanced potentials and observers there would sense a direction of time opposite to ours (and would have a reversed definition of matter and anti-matter because of *CPT* invariance). Thus the cycle previous to us would, according to *our* definition of time, have advanced potentials and would end with a uniform low-entropy big crunch and begin with a chaotic high-entropy big bang (see Gott [61] for further discussion). Thus, an infinite string of oscillating universes could have alternating high and low-entropy singularities, with the direction of the entropy (and causality—via electromagnetic potentials) time-reversing on each succeeding cycle. Every observer using the entropy direction of time would see in his “past” a low-entropy singularity (which he would call a big bang) and in his “future” a high-entropy singularity (which he could call a big crunch). Then the mystery is why the low-entropy big bangs exist—they now look improbable. An oscillating universe with chaotic bangs and crunches and half-advanced-half-retarded potentials throughout would seem more likely. At this point anthropic arguments [62] could be brought in to say that only low-entropy big bangs might produce intelligent observers and that, with an infinite number of universes in the string, eventually there would be—by chance—a sufficiently low-entropy big bang to produce intelligent observers. Still, the uniformity of the early universe that we observe seems to be more than that required to produce intelligent observers, so we might wonder whether a random intelligent observer in such a Universe would be expected to see initial conditions in his/her big bang as uniform as ours. (Among intelligent observers, the Copernican principle tells us that you should not expect to be special. Out of all the places for intelligent observers to be there are by definition only a few special places and many nonspecial places, so you should expect to be in one of the many nonspecial places [63].)

II. INFLATION AS A SOLUTION

Guth’s proposal of inflation [54] offered an explanation of why the initial conditions in the big bang should be approximately, but not exactly uniform. (For review of inflation see [64,65].) In the standard big bang cosmology this was always a puzzle because antipodal points on the sky on the last scattering surface at $1+z \approx 1000$ had not had time to be in communication with each other. When we see two regions which are at the same temperature, the usual explanation is that they have at some time in the past been in causal communication and have reached thermal equilibrium with each other. But there is not enough time to do this in the standard big bang model where the expansion of the scale factor at early times is $a(t) \propto t^{1/2}$. Grand unified theories (GUT) of particle physics suggest that at early times there might have been a non-zero cosmological constant Λ , which then decayed to the zero cosmological constant we see today. This means that the early universe approximates de Sitter space with a radius $r_0 = (3/\Lambda)^{1/2}$ whose expansion rate at late times approaches $a(t) = r_0 \exp(t/r_0)$. Regions that start off very close together, and have time to thermally equilibrate, end up very far apart. When they become separated by a distance r_0 , they effectively pass out of causal contact—if inflation were to continue forever, they would be beyond each other’s event horizons. But eventually the epoch of inflation ends, the energy density of the cosmological constant is dumped into thermal radiation, and the expansion then continues as $a(t) \propto t^{1/2}$ as in a radiation-dominated big bang cosmology. As the regions slow their expansion from each other, enough time elapses so that they are able to interchange photons once again and they come back into effective causal contact. As Bill Press once said, they say “hello,” “goodbye,” and “hello again.” When they say “hello again” they appear just like regions in a standard big bang cosmology that are saying “hello” for the first time (i.e., are just coming within the particle horizon) except that with inflation these regions are already in thermal equilibrium with each other, because they have seen each other in the past. Inflation also gives a natural explanation for why the observed radius of curvature of the universe is so large ($a \geq cH_0^{-1} \approx 3000h^{-1}$ Mpc; here $H_0 = 100h$ km s⁻¹ Mpc⁻¹ is the Hubble constant). During the big bang phase, as the universe expands, the radius of the universe a expands by the same factor as the characteristic wavelength λ of the microwave background photons, so $a/\lambda = \text{const} \geq e^{67}$. How should we explain this large observed dimensionless number? Inflation makes this easy. The energy density during the inflationary epoch is $\Lambda/8\pi$. Let λ be the characteristic wavelength of thermal radiation which would have that density. Even if a started out of the same order as λ , by the end of the inflationary epoch $a \geq \lambda e^{67}$, providing that the inflationary epoch lasts at least as long as $67r_0$, or 67 e -folding times. At the end of the inflationary epoch when the inflationary vacuum of density $\Lambda/8\pi$ decays and is converted into an equivalent amount of thermal radiation, the wavelength of that radiation will be λ and the ratio of a/λ is fixed at a constant value which is a dimensionless constant $\geq e^{67}$, retained as the universe continues to expand in the radiation and matter-dominated epochs. Thus, even a

short run of inflation, of 67 e -folding times or more, is sufficient to explain why the universe is as large as it is observed to be.

Another success of inflation is that the observed Zeldovich-Peebles-Yu-Harrison fluctuation spectrum with index $n=1$ [66–68] has been naturally predicted as the result of random quantum fluctuations [69–72]. The inflationary power spectrum with cold dark matter (CDM) has been amazingly successful in explaining the qualitative features of observed galaxy clustering (cf. [73–82]). The amount of large scale power seen in the observations suggests an inflationary CDM power spectrum with $0.2 < \Omega h < 0.3$ [83–88].

III. OPEN BUBBLE UNIVERSES

Gott [89] has shown how an open inflationary model might be produced. The initial inflationary state approximates de Sitter space, which can be pictured by embedding it as the surface $W^2 + X^2 + Y^2 + Z^2 - V^2 = r_0^2$ in a five-dimensional Minkowski space with metric $ds^2 = -dV^2 + dW^2 + dX^2 + dY^2 + dZ^2$ [40,90]. Slice de Sitter space along surfaces of $V = \text{const}$, then the slices are three-spheres of positive curvature $W^2 + X^2 + Y^2 + Z^2 = a^2$ where $a^2 = r_0^2 + V^2$. If t measures the proper time, then $V = r_0 \sinh(t/r_0)$ and $a(t) = r_0 \cosh(t/r_0)$. This is a closed universe that contracts then re-expands—at late times expanding exponentially as a function of proper time. If slices of $V + X = \text{constant}$ are chosen, the slices have a flat geometry and the expansion is exponential with $a(t) = r_0 \exp(t/r_0)$. If the slices are vertical ($W = \text{const} > r_0$), then the intersection with the surface is H^3 , a hyperboloid $X^2 + Y^2 + Z^2 - V^2 = -a^2$ living in a Minkowski space, where $a^2 = W^2 - r_0^2$. This is a negatively curved surface with a radius of curvature a . Let t be the proper time from the event E ($W = r_0$, $X = 0$, $Y = 0$, $Z = 0$, $V = 0$) in the de Sitter space. Then the entire future of E can be described as an open $k = -1$ cosmology where $a(t) = r_0 \sinh(t/r_0)$. At early times, $t \ll r_0$, near E, $a(t) \propto t$, and the model resembles a Milne cosmology [91], but at late times the model expands exponentially with time as expected for inflation. This is a negatively curved (open) Friedmann model with a cosmological constant and nothing else. Note that the entire negatively curved hyperboloid (H^3), which extends to infinity, is nevertheless causally connected because all points on it have the event E in their past light cone. Thus, the universe should have a microwave background that is isotropic, except for small quantum fluctuations. At a proper time τ_1 after the event E, the cosmological constant would decay leaving us with a hot big bang open ($k = -1$) cosmology with a radius of curvature of $a = r_0 \sinh(\tau_1/r_0)$ at the end of the inflationary epoch. If $\tau_1 = 67r_0$, then Ω is a few tenths today; if $\tau_1 \gg 67r_0$, then $\Omega \approx 1$ today [89].

Gott [89] noted that this solution looks just like the interior of a Coleman bubble [92]. Coleman and de Luccia [93] showed that if a metastable symmetric vacuum (with the Higgs field $\phi = 0$), with positive cosmological constant Λ were to decay by tunneling directly through a barrier to reach the current vacuum with a zero cosmological constant

(where the Higgs field $\phi = \phi_0$), then it would do this by forming a bubble of low-density vacuum of radius σ around an event E. The pressure inside the bubble is zero while the pressure outside is negative (equal to $-\Lambda/8\pi$), so the bubble wall accelerates outward, forming in spacetime a hyperboloid of one sheet (a slice of de Sitter space with $W = \text{const} < r_0$). This bubble wall surrounds and is asymptotic to the future light cone of E. If the tunneling is direct, the space inside the bubble is Minkowski space (like a slice $W = \text{const} < r_0$ in the embedding space, which is flat). The inside of the future light cone of E thus looks like a Milne cosmology with $\Omega = 0$ and $a(t) = t$. Gott [89] noted that what was needed to produce a realistic open model with Ω of a few tenths today was to have the inflation continue inside the bubble for about 67 e -folding times. Thus, our universe was one of the bubbles and this solved the problem of Guth's inflation that in general one expected the bubbles not to percolate [94,95]. But, from inside one of the bubbles, our view could be isotropic [89].

It was not long before a concrete mechanism to produce such continued inflation inside the bubble was proposed. A couple of weeks after Gott's paper appeared Linde's [96] proposal of new inflation appeared, followed shortly by Albrecht and Steinhardt [97]. They proposed that the Higgs vacuum potential $V(\phi)$ had a local minimum at $\phi = 0$ where $V(0) = \Lambda/8\pi$. Then there was a barrier at $\phi = \phi_1$, followed by a long flat plateau from ϕ_1 to ϕ_0 where it drops precipitately to zero at ϕ_0 . The relation of this to the open bubble universe's geometry is outlined by Gott [98] (see Fig. 1 and Fig. 2 in [98]). The de Sitter space outside the bubble wall has $\phi = 0$. Between the bubble wall, at a spacelike separation σ from the event E, and the end of the inflation at the hyperboloid H^3 which is the set of points at a future timelike separation of τ_1 from E, the Higgs field is between ϕ_1 and ϕ_0 , and τ_1 is the time it takes the field (after tunneling) to roll along the long plateau [where $V(\phi)$ is approximately equal to $\Lambda/8\pi$ and the geometry is approximately de Sitter]. After that epoch, $\phi = \phi_0$ where the energy density has been dumped into thermal radiation and the vacuum density is zero (i.e., a standard open big bang model). In order that inflation proceeds and the bubbles do not percolate, it is required that the probability of forming a bubble in de Sitter space per four volume r_0^4 is $\epsilon < \epsilon_{\text{cr}}$ where $5.8 \times 10^{-9} < \epsilon_{\text{cr}} < 0.24$ [95]. In order that there be a greater than 5% chance that no bubble should have collided with our bubble by now, so as to be visible in our past light cone, $\epsilon < 0.01$ for $\Omega = 0.4$, $\Lambda = 0$, $h = 0.63$ today [88], but this is no problem since we expect tunneling probabilities through a barrier to be exponentially small. This model has an event horizon, which is the future light cone of an event E' ($W = -r_0$, $X = 0$, $Y = 0$, $Z = 0$, $V = 0$) which is antipodal to E. Light from events within the future light cone of E' never reaches events inside the future light cone of E. So we are surrounded by an event horizon. This produces Hawking radiation; and, if r_0 is of order the Planck length, then the Gibbons-Hawking thermal state [100] (which looks like a cosmological constant due to the trace anomaly [101]) should be dynamically important [89].

If we observe $\Omega < 1$ and $\Omega_\Lambda = 0$, then $k = -1$ and we need inflation more than ever—we still need it to explain the isotropy of the microwave background radiation and we would now have a large but *finite* radius of curvature to explain, which 67 e -folds of inflation could naturally produce. When Gott told this to Linde in 1982, Linde said, yes, if we found that $\Omega < 1$, he would still have to believe in inflation but he would have a headache in the morning. Why? Because one has to produce a particular amount of inflation, approximately 67 e -folds. If there were 670 e -folds or 670 million e -folds, then Ω currently would be only slightly less than 1. So there would be what is called a “fine tuning of parameters” needed to produce the observed results.

The single-bubble open inflationary model [89] discussed above has recently come back into fashion because of a number of important developments. On the theoretical side, Ratra and Peebles [102,103] have shown how to calculate quantum fluctuations in the H^3 hyperbolic geometry with $a(t) = r_0 \sinh(t/r_0)$ during the inflationary epoch inside the bubble in the single bubble model. This allows predictions of fluctuations in the microwave background. Bucher, Goldhaber, and Turok [104,105] have extended these calculations, as well as Yamamoto, Sasaki and Tanaka [106]. Importantly, they have explained [104,105] that the fine tuning in these models is only “logarithmic” and, therefore, not so serious. Linde and Mezhlumian [107,108] have shown how there are reasonable potentials which could produce such bubble universes with different values of Ω . In a standard chaotic inflationary potential $V(\phi)$ [109], one could simply build in a bump, so that one would randomly walk to the top of the curve via quantum fluctuations and then roll down till one lodged behind the bump in a metastable local minimum. One would then tunnel through the bump, forming bubbles that would roll down to the bottom in a time τ_1 . One could have a two-dimensional potential $V(\phi, \sigma) = \frac{1}{2}g^2\phi^2\sigma^2 + V(\sigma)$, where g is a constant and there is a metastable trough at $\sigma = 0$ with altitude $V(\phi, 0) = \Lambda/8\pi$ with a barrier on both sides, but one could tunnel through the barrier to reach $\sigma > 0$ where $V(\phi, \sigma)$ has a true minimum, and at fixed σ , is proportional to ϕ^2 [107,108]. Then individual regions could tunnel across the barrier at different values of ϕ , and hence have different roll-down times τ_1 and thus different values of Ω . With a myriad of open universes being created, anthropic arguments [62] come into play and if shorter roll-down times were more probable than large ones, we might not be surprised to find ourselves in a model which had Ω of a few tenths, since if Ω is too small, no galaxies will form [110].

A second reason for the renaissance of these open inflationary models is the observational data. A number of recent estimates of h (the present Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$) have been made (i.e., $h = 0.65 \pm 0.06$ [111], $0.68 \leq h \leq 0.77$ [112], $0.55 \leq h \leq 0.61$ [113], and $h = 0.64 \pm 0.06$ [114]). Ages of globular cluster stars have a 2σ lower limit of about 11.6 billion years [115], we require $h < 0.56$ if $\Omega = 1$, but a more acceptable $h < 0.65$ if $\Omega = 0.4$, $\Omega_\Lambda = 0$. Models with low Ω but $\Omega + \Omega_\Lambda = 1$ are also acceptable. Also, studies of large scale structure have shown that

with the inflationary CDM power spectrum, the standard $\Omega = 1$, $h = 0.5$ model simply does not have enough power at large scales. A variety of observational samples and methods have suggested this: counts in cells, angular covariance function on the sky, power spectrum analysis of 3D samples, and finally topological analysis, all showing that $0.2 < \Omega h < 0.3$ [82–88]. If $h > 0.55$ this implies $\Omega < 0.55$, which also agrees with what one would deduce from the age argument as well as the measured masses in groups and clusters of galaxies [116]. With the COBE normalization there is also the problem that with $\Omega = 1$, $(\delta M/M)_{8h^{-1} \text{Mpc}} = 1.1 - 1.5$ and this would require galaxies to be anti-biased [since for galaxies $(\delta M/M)_{8h^{-1} \text{Mpc}} = 1$] and would also lead to an excess of large-separation gravitational lenses over those observed [117]. These things have forced even enthusiasts of $k = 0$ models to move to models with $\Omega < 1$ and a cosmological constant so that $\Omega + \Omega_\Lambda = 1$ and $k = 0$ [118]. They then have to explain the small ratio of the cosmological constant to the Planck density (10^{-120}). Currently we do not have such a natural explanation for a small yet finite Λ as inflation naturally provides for explaining why the radius of curvature should be a big number in the $k = -1$ case.

Turner [119] and Fukugita, Futamase, and Kasai [120] showed that a flat $\Omega_\Lambda = 1$ model produces about 10 times as many gravitational lenses as a flat model with $\Omega = 1$, and Kochanek [121] was able to set a 95% confidence lower limit of $0.34 < \Omega$ in flat models where $\Omega + \Omega_\Lambda = 1$, and a 90% confidence lower limit $0.15 < \Omega$ in open models with $\Omega_\Lambda = 0$. Thus, extreme- Λ dominated models are ruled out by producing too many gravitational lenses.

Data on cosmic microwave background fluctuations for spherical harmonic modes from $l = 2$ to $l = 500$ will provide a strong test of these models. With $\Omega_B h^2 = 0.0125$, the $\Omega = 1$, $\Omega_\Lambda = 0$ model power spectrum reaches its peak value at $l = 200$; an $\Omega = 0.3$, $\Omega_\Lambda = 0.7$ model reaches its peak value also at $l = 200$ [122]; while an $\Omega = 0.4$, $\Omega_\Lambda = 0$ model reaches its peak value at $l = 350$ [123]. This should be decided by the Microwave Anisotropy Probe (MAP) and PLANCK satellites which will measure this range with high accuracy [124].

For the rest of this paper we shall usually assume single-bubble open inflationary models for our big bang universe (while recognizing that chaotic inflationary models [64] and models with multiple epochs of inflation are also possible; it is interesting to note that Penrose also prefers an open universe from the point of view of the complex-holomorphic ideology of his twister theory [125]). If the inflation within the bubble is of order 67 e -folds, then we can have Ω of a few tenths; but if it is longer than that, we will usually see Ω near 1 today. In any case, we will be assuming an initial metastable vacuum which decays by forming bubbles through barrier penetration. The bubble formation rate per unit four volume r_0^4 is thus expected to be exponentially small so the bubbles do not percolate. Inflation is thus eternal to the future [128–131]. Borde and Vilenkin have proved that if the Universe were infinitely old (i.e., if the de Sitter space were complete) then the bubbles would percolate immediately and inflation would never get started (see

[126,127] and references cited therein). Recall that a complete de Sitter space may be covered with an S^3 coordinate system (a $k=1$ cosmology) whose radius varies as $a(t) = r_0 \cosh(t/r_0)$ so that for early times ($t < 0$) the universe would be contracting and bubbles would quickly collide preventing the inflation from ever reaching $t=0$. Thus Borde and Vilenkin have proved that in the inflationary scenario the universe must have a beginning. If it starts with a three-sphere of radius r_0 at time $t=0$, and after that expands like $a(t) = r_0 \cosh(t/r_0)$, the bubbles do not percolate (given that the bubble formation rate per four volume r_0^4 is $\epsilon \ll 1$) and the inflation continues eternally to $t = \infty$ producing an infinite number of open bubble universes. Since the number of bubbles forming increases exponentially with time without limit, our universe is expected to form at a finite but arbitrarily large time after the beginning of the inflationary state. In this picture our universe (our bubble) is only 12 billion years old, but the Universe as a whole (the entire bubble forming inflationary state) is of a finite but arbitrarily old age.

IV. VILENKIN'S TUNNELING UNIVERSE AND HARTLE-HAWKING'S NO-BOUNDARY PROPOSAL

But how to produce that initial spherical S^3 universe? Vilenkin [11] suggested that it could be formed from quantum tunneling. Consider the embedding diagram for de Sitter space: de Sitter space can be embedded as the surface $W^2 + X^2 + Y^2 + Z^2 - V^2 = r_0^2$ in a five-dimensional Minkowski space with metric $ds^2 = -dV^2 + dW^2 + dX^2 + dY^2 + dZ^2$. This can be seen as an S^3 cosmology with radius $a(t) = r_0 \cosh(t/r_0)$ where $V = r_0 \sinh(t/r_0)$ and $a^2 = W^2 + X^2 + Y^2 + Z^2$ gives the geometry of S^3 . This solution represents a classical trajectory with a turning point at $a = r_0$. But just as it reaches this turning point it could tunnel to $a = 0$ where the trajectory may be shown as a hemisphere of the Euclidean four-sphere $W^2 + X^2 + Y^2 + Z^2 + V^2 = r_0^2$ embedded in a flat Euclidean space with the metric $ds^2 = dV^2 + dW^2 + dX^2 + dY^2 + dZ^2$ and $a(t_E) = r_0 \cos(t_E/r_0)$ where $a^2 = W^2 + X^2 + Y^2 + Z^2$ and $V = r_0 \sin(t_E/r_0)$. The time-reversed version of this process would show tunneling from a point at ($V = -r_0, W=0, X=0, Y=0, Z=0$) to a three sphere at $V=0$ of radius r_0 which then expands with proper time like $a(t) = r_0 \cosh(t/r_0)$ giving a normal de Sitter space—thus Vilenkin's universe created from nothing is obtained [11].

Hawking has noted that in this case, in Hartle and Hawking's formulation, the point ($V = -r_0, W=0, X=0, Y=0, Z=0$) is not special, the curvature does not blow up there: it is like other points in the Euclidean hemispherical section [13]. However, this point is still the earliest point in Euclidean time since it is at the center of the hemisphere specified by the Euclidean boundary at $V=0$. So the beginning point in the Vilenkin model is indeed like the south pole of the Earth [13].

Vilenkin's tunneling universe was based on an analogy between quantum creation of universes and tunneling in ordinary quantum mechanics [11]. In ordinary quantum me-

chanics, a particle bounded in a well surrounded by a barrier has a finite probability to tunnel through the barrier to the outside if the height of the barrier is finite (as in the α -decay of radioactive nuclei [132–134]). The wave function outside the barrier is an outgoing wave, the wave function in the well is the superposition of an outgoing wave and an ingoing wave which is the reflection of the outgoing wave by the barrier. Because of the conservation of current, there is a net outgoing current in the well. The probability for the particle staying in the well is much greater than the probability for the particle running out of the barrier. The energy of the particle in the well *cannot* be zero, otherwise the uncertainty principle is violated. Thus there is always a finite zero-point-energy. The Vilenkin universe was supposed to be created from “nothing,” where according to Vilenkin “nothing” means “a state with no classical spacetime” [135]. Thus this is essentially different from tunneling in ordinary quantum mechanics since in ordinary quantum mechanics tunneling always takes place from one classically allowed region to another classically allowed region where the current and the probability are conserved. But creation from “nothing” is supposed to take place from a classically forbidden (Euclidean) region to a classically allowed (Lorentzian) region, so the conservation of current is obviously violated. Vilenkin obtained his tunneling universe by choosing a so-called “tunneling boundary condition” for the Wheeler-DeWitt equation [135,136]. His “tunneling from nothing” boundary condition demands that when the universe is big ($a^2 \Lambda/3 > 1$ where Λ is the cosmological constant and a is the scale factor of the universe) there is only an outgoing wave in the superspace [135,136]. If the probability and current are conserved (in fact there does exist a conserved current for the Wheeler-DeWitt equation [137], and a classically allowed solution with $a=0$ and zero “energy”), there must be a finite probability for the universe being in the state before tunneling (i.e., $a=0$) and this probability is much bigger than the probability for tunneling. This implies that there must be “something” instead of “nothing” before tunneling. This becomes more clear if matter fields are included in considering the creation of universes. In the case of a cosmological constant Λ and a conformally coupled scalar field ϕ (conformal fields are interesting not only for their simplicity but also because electromagnetic fields are conformally invariant) as the source terms in Einstein's equations, in the mini-superspace model (where the configurations are the scale factor a of the S^3 Robertson-Walker metric and a homogeneous conformally coupled scalar field ϕ) the Wheeler-DeWitt equation separates [12,138]

$$\frac{1}{2} \left(-\frac{d^2}{d\chi^2} + \chi^2 \right) \Phi(\chi) = E \Phi(\chi), \quad (1)$$

$$\frac{1}{2} \left[-\frac{1}{a^p} \frac{d}{da} \left(a^p \frac{d}{da} \right) + \left(a^2 - \frac{\Lambda}{3} a^4 \right) \right] \Psi(a) = E \Psi(a), \quad (2)$$

where $\Psi(a)\Phi(\chi)$ is the wave function of the universe [$\chi \equiv (4\pi/3)^{1/2} \phi a$], E is the “energy level” of the conformally coupled scalar field (we use quotes because for radiation the conserved quantity is $E = 4\pi\rho a^4/3$ instead of the energy $4\pi\rho a^3/3$ where ρ is the energy density), and p is a constant

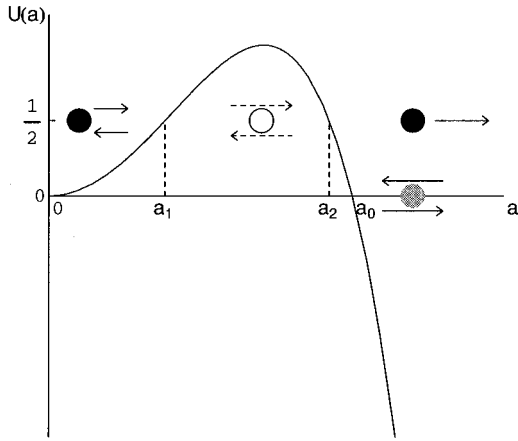


FIG. 2. The potential function in the Wheeler-DeWitt equation in the minisuperspace model. The horizontal axis is the scale factor of the universe. If the conformally coupled scalar field is in the ground state, it has a “zero-point-energy” $1/2$. If this “zero-point-energy” is considered, the quantum behavior of the universe is like a particle of unit mass with total energy $1/2$ moving in the potential $U(a)$. Regions $0 < a < a_1$ and $a > a_2$ are classically allowed; region $a_1 < a < a_2$ is classically forbidden. The left dark disk is a tiny radiation-dominated closed oscillating universe, which oscillates between big bangs and big crunches. The smoothness of the potential at $a=0$ may indicate that any big bang and big crunch singularities are removed by quantum theory. This tiny oscillating universe has a small but non-zero probability to tunnel through the barrier out to become a de Sitter-type inflating universe, which is represented by the dark disk on the right. The circle inside the barrier is a Euclidean bouncing space. If the “zero-point-energy” $1/2$ were neglected (as Hartle and Hawking did), the left classically allowed region would shrink to a point. The grey disk represents a contracting and re-expanding de Sitter universe. If the “zero-point-energy” is neglected, the Universe could start out at the metastable minimum as a point with $a=0$, tunneling through the barrier out to become a de Sitter universe. In this paper we argue that we have no reason to neglect the “zero-point-energy” so that it is the tiny oscillating universe initial state that applies.

determining the operator ordering. Equation (1) is just the Schrödinger equation of a harmonic oscillator with unit mass and unit frequency and energy E , the eigenvalues of E are $n + \frac{1}{2}$ where $n=0,1,2, \dots$. Equation (2) is equivalent to the Schrödinger equation for a unit mass particle with total energy $E = n + \frac{1}{2}$ in the one-dimensional potential

$$U(a) = \frac{1}{2} \left(a^2 - \frac{\Lambda}{3} a^4 \right). \quad (3)$$

It is clear that in the case of $n < \frac{1}{2}(3[4\Lambda] - 1)$, there exist one classically forbidden region $a_1 < a < a_2$ and two classically allowed regions $0 \leq a < a_1$ and $a > a_2$ where $a_{1,2} \equiv (3/2\Lambda)[1 \mp \sqrt{1 - \frac{4}{3}(2n+1)\Lambda}]$ (Fig. 2). Because $U(a)$ is regular at $a=0$, we expect that the wave function $\Psi(a)$ is also regular at $a=0$. If $\Lambda \ll 1$ and the conformally coupled scalar field is in the ground state with $n=0$, we have $a_1 \approx 1$, $a_2 \approx (3/\Lambda)^{1/2}$ and the potential in region $0 \leq a < a_1$ is $U(a) \approx \frac{1}{2}a^2$ like a harmonic oscillator. The quantum behav-

ior of the universe in region $0 \leq a < a_1$ is like a quantum harmonic oscillator. This may describe a quantum oscillating (Lorentzian) universe without big bang or big crunch singularities, which has a finite (but small) probability [$\approx \exp(-1/\Lambda)$] to tunnel through the barrier to form a de Sitter-type inflating universe. The existence of this tiny oscillating universe is due to the existence of a finite “zero-point-energy” ($1/2$) of a conformally coupled scalar field and this “zero-point-energy” is required by the uncertainty principle. Since a conformally coupled scalar field has an equation of state like that of radiation, the Friedmann equation for $k = +1$ is

$$\left(\frac{da}{dt} \right)^2 = \frac{C}{a^2} + \frac{\Lambda}{3} a^2 - 1, \quad (4)$$

where $C = 8\pi\rho a^4/3 = \text{const}$ and ρ is the energy density of the conformally coupled scalar field. Equation (4) is equivalent to the energy-conservation equation for a classical unit mass particle with zero total energy moving in the potential

$$V(a) = \frac{1}{2} \left(1 - \frac{\Lambda}{3} a^2 - \frac{C}{a^2} \right). \quad (5)$$

The difference between $U(a)$ and $V(a)$ is caused by the fact that in the integral of action the volume element contains a factor a^3 which is also varied when one makes the variation to obtain the dynamical equations. The potential $V(a)$ is singular at $a=0$ and near $a=0$ we have $V(a) \approx -C/2a^2$. For $\Lambda \ll 1$ and $n=0$ (we take $C=2E=2n+1$), the classical universe in region $0 \leq a < a_1$ is radiation dominated. This universe expands from a big bang singularity, reaches a maximum radius, then re-collapses to a big crunch singularity: $a=0$ is a singularity in the classical picture. But from the above discussion, the Wheeler-DeWitt equation gives a regular wave function at $a=0$. In such a case near $a=0$ the quantum behavior of the universe is different from classical behavior. This implies that, near $a=0$, classical general relativity breaks down and quantum gravity may remove singularities. This case is like that of a hydrogen atom where the classical instability (according to classical electrodynamics, an electron around a hydrogen nucleus will fall into the nucleus due to electromagnetic radiation) is cured by quantum mechanics. Anyway, it is *not* nothing at $a=0$. There is a small classically allowed, oscillating, radiation dominated, closed, quantum (by “quantum” we mean that its quantum behavior deviates significantly from its classical behavior) Friedmann universe near $a=0$, which has a small probability to tunnel through the barrier to form an inflationary universe. (If $\Lambda > 0.75$ there is no classically forbidden region and thus no tunneling.)

So in this model the universe did not come from a point (nothing) but from a tiny classically allowed, oscillating, quantum Friedmann universe whose radius is of order the Planck magnitude. But where did this oscillating universe come from? Because it has a finite probability to tunnel (each time it reaches maximum radius) to a de Sitter space, it has a finite “half-life” for decay into the de Sitter phase and cannot last forever. It could, of course, originate by tunneling

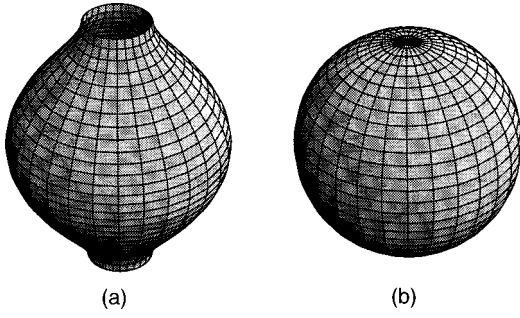


FIG. 3. (a) The solution of the Euclidean Einstein's equations representing the tunneling regime (open circle) in Fig. 2. This is a solution to the Euclidean Einstein's equations with a positive cosmological constant and a conformally coupled scalar field in its ground state. This is a Euclidean space bouncing between the state with maximum radius a_2 and the state with (non-zero) minimum radius a_1 . One "copy" of this Euclidean bouncing solution is shown in this diagram, which has two boundaries with minimum radius a_1 . (b) This is the case when the "zero-point-energy" of the conformally coupled scalar field is neglected, as Hartle and Hawking did. In this case the minimum radius is zero, and thus one copy of the bouncing Euclidean solution is a four-sphere. This four-sphere has no-boundary, which is the basis of Hawking's quantum cosmology. But we argue that since the "zero-point-energy" of the conformally coupled scalar field cannot be neglected, the true solution should be that given by diagram (3a), which does not enforce Hartle and Hawking's no-boundary proposal.

from a collapsing de Sitter phase (the time-reversed version of the creation of a de Sitter state from the oscillating state), but then we are back where we started. In fact, starting with a collapsing de Sitter phase one is more likely to obtain an expanding de Sitter phase by simply re-expanding at the classical turning point rather than tunneling into and then out of the tiny oscillating universe state. An alternative might be to have the original tiny oscillating universe created via a quantum fluctuation (since it has just the "zero-point-energy") but here we are basically returning to the idea of Tryon [139] that you could get an entire Friedmann universe of any size directly via quantum fluctuation. But quantum fluctuation of what? You have to have laws of physics and a potential etc.

Hartle and Hawking [12] made their no-boundary proposal and obtained a model of the universe similar to Vilenkin's tunneling universe. The no-boundary proposal is expressed in terms of a Euclidean path integral of the wave function of the universe

$$\Psi(h_{ab}, \phi_1, \partial M) = \sum_M \int \mathcal{D}g_{ab} \mathcal{D}\phi \times \exp[-I(g_{ab}, \phi, M)], \quad (6)$$

where the summation is over compact manifolds M with the prescribed boundary ∂M (being a compact three-manifold representing the shape of the universe at a given epoch) as the *only* boundary; g_{ab} is the Euclidean metric on the manifold M with induced three-metric h_{ab} on ∂M , ϕ is the matter field with induced value ϕ_1 on ∂M ; I is the Euclidean action

obtained from the Lorentzian action S via Wick rotation: $I = -iS(t \rightarrow -i\tau)$. In the mini-superspace model the configuration space is taken to include the $k=+1$ Robertson-Walker metric and a homogeneous matter field. In the WKB approximation the wave function is (up to a normalization factor)

$$\Psi \approx \sum_M B_M \exp[-I_{\text{cl}}(g_{ab}, \phi, M)], \quad (7)$$

where I_{cl} is the Euclidean action for the solutions of the Euclidean field equations (Einstein's equations and matter field equations). The factor B_M is the determinant of small fluctuations around solutions of the field equations [12]. If the matter field is a conformally coupled scalar field $\phi \equiv (3/4\pi)^{1/2} \chi/a$ (which is the case that Hartle and Hawking [12] discussed), ρa^4 is conserved where ρ is the energy density of ϕ satisfying the field equations. Then the Friedmann equation is given by Eq. (4). The corresponding Euclidean equation is obtained from Eq. (4) via $t \rightarrow -i\tau$

$$\left(\frac{da}{d\tau}\right)^2 = 1 - \frac{\Lambda}{3}a^2 - \frac{C}{a^2}. \quad (8)$$

The solution to Eq. (8) is (for the case $\frac{4}{3}\Lambda C < 1$)

$$a(\tau) = H^{-1} \left[\frac{1}{2} + \frac{1}{2}(1 - 4H^2C)^{1/2} \times \cos(2H\tau) \right]^{1/2}, \quad (9)$$

where $H = (\Lambda/3)^{1/2}$. This is a Euclidean bouncing space with a maximum radius $a_{\text{max}} = H^{-1}[\frac{1}{2} + \frac{1}{2}(1 - 4H^2C)^{1/2}]^{1/2}$ and a minimum radius $a_{\text{min}} = H^{-1}[\frac{1}{2} - \frac{1}{2}(1 - 4H^2C)^{1/2}]^{1/2}$ (Fig. 3). If $C=0$, we have $a_{\text{max}} = H^{-1}$, $a_{\text{min}} = 0$, and $a(\tau) = H^{-1} \cos(H\tau)$, one copy of this bouncing space is a four-sphere with the Euclidean de Sitter metric $ds^2 = d\tau^2 + H^{-2} \cos^2(H\tau)[d\chi^2 + \sin^2 \chi(d\theta^2 + \sin^2 \theta d\phi^2)]$ —which is just a four-sphere embedded in a five-dimensional Euclidean space (V, W, X, Y, Z) with metric $ds^2 = dV^2 + dW^2 + dX^2 + dY^2 + dZ^2$ —this is the solution that Hartle and Hawking used [12]. But, as we have argued above, according to Hartle and Hawking [12] and Hawking [138], the Wheeler-DeWitt equation for $\Phi(\chi)$ [Eq. (1)] gives rise to a "zero-point-energy" for the conformally coupled scalar field: $C_0 = 2E(n=0) = 1$ (the state with $C=0$ violates the uncertainty principle). One copy of this bouncing Euclidean space is *not* a compact four-dimensional manifold with no boundaries, but has two boundaries with $a = a_{\text{min}}$ (see Fig. 3). If $H \ll 1$ (i.e. $\Lambda \ll 1$), we have $a_{\text{max}} \approx H^{-1}$, $a_{\text{min}} \approx 1$.

Penrose [125] has criticized Hawking's no-boundary proposal and the model obtained by gluing a de Sitter space onto a four-sphere hemisphere by pointing out that there are only very few spaces for which one can glue a Euclidean and a Lorentzian solution together since it is required that they have both a Euclidean and a Lorentzian solution, but the generic case is certainly very far from that. Here "with a zero-point-energy" we have both a Euclidean solution and a

Lorentzian solution, and they can be glued together. But the Euclidean solution is not closed in any way; that is, it does not enforce the no-boundary proposal. Hartle and Hawking argued that there should be a constant ϵ_0 in E which arises from the renormalization of the matter field, i.e., E should be $n + \frac{1}{2} + \epsilon_0$ [12]. But there is *no* reason that ϵ_0 should be $-\frac{1}{2}$ to exactly cancel the “zero-point-energy” $\frac{1}{2}$. (As in the case of a quantum harmonic oscillator, we have no reason to neglect the zero-point-energy.) In fact, since ϵ_0 comes from the renormalization of the matter field (without quantization of gravity), it should be much less than the Planck magnitude, i.e., $\epsilon_0 \ll 1$, and thus ϵ_0 is negligible compared with $\frac{1}{2}$. In fact in [138] Hawking has dropped ϵ_0 .

In [12] Hartle and Hawking have realized that for excited states ($n > 0$), there are two kinds of classical solutions: one represents universes which expand from zero volume, to reach a maximum radius, and then re-collapse (like our tiny oscillating universe); the other represents the de Sitter-type state of continual expansion. There are probabilities for a universe to tunnel from one state to the other. Here we argue that for the ground state ($n = 0$), there are also two such kinds of Lorentzian universes. One is a tiny quantum oscillating universe (having a maximum radius with Planck magnitude). Here “quantum” just means that the classical description fails (so singularities might be removed). The other is a big de Sitter-type universe. These two universes can be joined to one another through a Euclidean section, which describes quantum tunneling from a tiny oscillating universe to an inflating universe (or from a contracting de Sitter-type universe to a tiny oscillating universe). During the tunneling, the radius of the universe makes a jump (from the Planck length to H^{-1} or *vice versa*).

As Hartle and Hawking [12] calculated the wave function of the universe for the ground state, they argued that, for the conformally coupled scalar field case, the path integral over a and $\chi = (4\pi/3)^{1/2} \phi a$ separates since “not only the action separates into a sum of a gravitational part and a matter part, but the boundary condition on the $a(\eta)$ and $\chi(\eta)$ summed over do not depend on one another” where η is the conformal time. The critical point for the variable’s separation in the path integral is that “the ground state boundary conditions imply that geometries in the sum are conformal to half of a Euclidean-Einstein static universe; i.e., the range of η is $(-\infty, 0)$. The boundary conditions at infinite η are that $\chi(\eta)$ and $a(\eta)$ vanish. The boundary conditions at $\eta = 0$ are that $a(0)$ and $\chi(0)$ match the arguments of the wave function a_0 and χ_0 ” [12]. But this holds only for some specific cases, such as de Sitter space. Our solution (9) does not obey Hartle and Hawking’s assumption that η ranges from $-\infty$ to 0. For a general $k = +1$ (Euclidean) Robertson-Walker metric, $\eta = \int (d\tau/a)$ is a functional of a , and the action of matter (an integral over η) is a functional of a . Therefore, the action *cannot* be separated into a sum of a gravitational part and a matter part as Hartle and Hawking did. The failure of Hartle and Hawking’s path integral calculation is also manifested in the fact that de Sitter space is *not* a solution of the Friedmann equation if the “zero-point-energy” of the conformally coupled scalar field is considered, whereas the semiclassical approximation implies that the principal contribution to the

path integral of the wave function comes from the configurations which solve Einstein’s equations. One may hope to overcome this difficulty by introducing a scalar field with a flat potential $V(\phi)$ (as in the inflation case). But this does not apply to the quantum cosmology case since as $a \rightarrow 0$ the universe always becomes radiation-dominated unless the energy density of radiation is exactly zero (but the uncertainty principle does not allow this case to occur).

V. CTCs AND THE CHRONOLOGY PROTECTION CONJECTURE

From the arguments in the last section, we find that the Universe does *not* seem to be created from nothing. On the other hand, if the Universe is created from *something*, that something could have been *itself*. Thus it is possible that the Universe is its own mother. In such a case, if we trace the history of the Universe backward, inevitably we will enter a region of CTCs. Therefore CTCs may play an important role in the creation of the Universe. It is interesting to note that Hawking and Penrose’s singularity theorems do not apply if the Universe has had CTCs. And, it has been shown that, if a compact Lorentzian spacetime undergoes topology changes, there must be CTCs in this spacetime [140,19,20]. [Basically there are two type of spacetimes with CTCs: for the first type, there are CTCs everywhere (Gödel space belongs to this type); for the second type, the CTCs are confined within some regions and there exists at least one region where there are no closed causal (timelike or null) curves, and the regions with CTCs are separated from the regions without closed causal curves by Cauchy horizons (Misner space belongs to this type). In this paper, with the word “spacetimes with CTCs” we always refer to the second type unless otherwise specified.]

While in classical general relativity there exist many solutions with CTCs, some calculations of vacuum polarization of quantum fields in spacetimes with CTCs indicated that the energy-momentum tensor (in this paper when we deal with quantum fields, with the word “the energy-momentum tensor” we always refer to “the renormalized energy-momentum tensor” because “the unrenormalized energy-momentum tensor” has no physical meaning) diverges as one approaches the Cauchy horizon separating the region with CTCs from the region without closed causal curves. This means that spacetimes with CTCs may be unstable against vacuum polarization since when the energy-momentum tensor is fed back to the semiclassical Einstein’s equations (i.e. Einstein’s equations with quantum corrections to the energy-momentum tensor of matter fields) the back-reaction may distort the spacetime geometry so strongly that a singularity may form and CTCs may be destroyed. Based on some of these calculations, Hawking [19,20] has proposed the chronology protection conjecture which states that the laws of physics do not allow the appearance of CTCs. (It should be mentioned that the chronology protection conjecture does *not* provide any restriction on spacetimes with CTCs but no Cauchy horizons since there is *no* indication that this type of spacetime is unstable against vacuum polarization. In the next section we will show a simple ex-

ample of a spacetime with CTCs but no Cauchy horizons, where the energy-momentum tensor is finite everywhere.)

But, on the other hand, Li, Xu, and Liu [22] have pointed out that even if the energy-momentum tensor of vacuum polarization diverges at the Cauchy horizon, it does *not* mean that CTCs must be prevented by physical laws because: (1) Einstein's equations are local equations and the energy-momentum tensor may diverge only at the Cauchy horizon (or at the polarized hypersurfaces) and be well-behaved elsewhere within the region with CTCs; (2) the divergence of the energy-momentum tensor at the Cauchy horizon does *not* mean that the Cauchy horizon must be destroyed by the back-reaction of vacuum polarization, *but* instead means that near the Cauchy horizon the usual quantum field theory on a prescribed classical spacetime background cannot be used and the quantum effect of gravity must be considered. (This is like the case that Hawking and Penrose's singularity theorems do *not* mean that the big bang cosmology is wrong but mean that near the big bang singularity quantum gravity effects become important [13].) When Hawking proposed his chronology protection conjecture, Hawking [20] and Kim and Thorne [16] had a controversy over whether quantum gravity can save CTCs. Kim and Thorne claimed that quantum gravitational effects would cut the divergence off when an observer's proper time from crossing the Cauchy horizon was the Planck time, and this would only give such a small perturbation on the metric that the Cauchy horizon could not be destroyed. But, Hawking [20] noted that one would expect the quantum gravitational cut-off to occur when the invariant distance from the Cauchy horizon was of order the Planck length, and this would give a very strong perturbation on the metric so that the Cauchy horizon would be destroyed. Since there does not exist a self-consistent quantum theory of gravity at present, we cannot judge who (Hawking or Kim and Thorne) is right. But in any case, these arguments imply that in the case of a spacetime with CTCs where the energy-momentum tensor of vacuum polarization diverges at the Cauchy horizon, quantum gravity effects should become important near the Cauchy horizon. Li, Xu, and Liu [22] have argued that if the effects of quantum gravity are considered, in a spacetime with CTCs the region with CTCs and the region without closed causal curves may be separated by a *quantum barrier* (e.g. a region where components of the metric have complex values) instead of a Cauchy horizon generated by closed null geodesics. By quantum processes, a time traveler may tunnel from the region without closed causal curves to the region with CTCs (or *vice versa*), and the spacetime itself can also tunnel from one side to the other side of the quantum barrier [22]. In classical general relativity, a region with CTCs and a region without closed causal curves must be separated by a Cauchy horizon (compactly generated or non-compactly generated) which usually contains closed null geodesics if it is compactly generated [20]. But if quantum gravity effects are considered (e.g. in quantum cosmology), they can be separated by a complex geometric region (as a quantum barrier) instead of a Cauchy horizon [22]. (In the path integral approach to quantum cosmology, complex geometries are *required* in order to make the path integral convergent and to overcome the difficulty

that in general situations a Euclidean space *cannot* be directly joined to a Lorentzian space [141].) And, using a simple example of a space with a region with CTCs separated from a region without closed causal curves by a complex geometric region, Li, Xu, and Liu [22] have shown that in such a space the energy-momentum tensor of vacuum polarization is finite everywhere and the chronology protection conjecture has been challenged.

Without appeal to quantum gravity, counter-examples to the chronology protection conjecture also exist. By introducing a spherical reflecting boundary between two mouths of a wormhole, Li [23] has shown that with some boundary conditions for geodesics (e.g. the reflection boundary condition) closed null *geodesics* [usually the "archcriminal" for the divergence of the energy-momentum tensor as the Cauchy horizon is approached (see e.g. [16])] may be removed from the Cauchy horizon separating the region with CTCs and the region without closed causal curves. In such a case the spacetime contains neither closed null *geodesics* nor closed timelike *geodesics*, though it contains both closed timelike *non-geodesic* curves and closed null *non-geodesic* curves. Li [23] has shown that in this spacetime the energy-momentum tensor is finite everywhere. Following Li [23], Low [24] has given another example of spacetime with CTCs but without closed causal *geodesics*.

Recently, with a very general argument, Li [26] has shown that the appearance of an absorber in a spacetime with CTCs may make the spacetime stable against vacuum polarization. Li [26] has given some examples to show that there exist many collision processes in high energy physics for which the total cross-sections increase (or tend to a constant) as the frequency of the incident waves increases. Based on these examples, Li [26] has argued that material will become opaque for waves (particles) with extremely high frequency or energy, since in such cases the absorption caused by various types of scattering processes becomes very important. Based on calculation of the renormalized energy-momentum tensor and the fluctuation in the metric, Li [26] has argued that if an absorbing material with appropriate density is introduced, vacuum polarization may be smoothed out near the Cauchy horizon so that the metric perturbation caused by vacuum fluctuations will be very small and a spacetime with CTCs can be stable against vacuum polarization.

Boulware [21] and Tanaka and Hiscock [25] have found that for sufficiently massive fields in Gott space [8,142] and Grant space [18] respectively, the energy-momentum tensor remains regular on the Cauchy horizon. Krasnikov [27] has found some two-dimensional spacetimes with CTCs for which the energy-momentum tensor of vacuum polarization is bounded on the Cauchy horizon. Sushkov [28] has found that for an automorphic complex scalar field in Misner space there is a vacuum state for which the energy-momentum tensor is zero everywhere. More recently, Cassidy [31] and Li and Gott [30] have independently found that for the real conformally coupled scalar field in Misner space there exists a quantum state for which the energy-momentum tensor is zero everywhere. Li and Gott [30] have found that this quantum state is the "adapted" Rindler vacuum (i.e. the usual Rindler vacuum with multiple images) and it is a self-

consistent vacuum state because it solves the semiclassical Einstein's equations exactly. Li and Gott [30] have also found that for this "adapted" Rindler vacuum in Misner space, an inertial particle detector perceives nothing. In this paper, we find that for a multiply connected de Sitter space there also exists a self-consistent vacuum state for a conformally coupled scalar field (see Sec. IX).

Thorne [9] has noted that, even if Hawking's argument that a quantum gravitational cut-off would occur when the geometric invariant distance from the Cauchy horizon is of order the Planck length is correct, by using two wormholes the metric fluctuations near the Cauchy horizon can be made arbitrarily small so a spacetime with CTCs created from two wormholes can be stable against vacuum polarization. Recently Visser [29] has generalized this result to the Roman ring case.

The above arguments indicate that the back-reaction of vacuum polarization may *not* destroy the Cauchy horizon in spacetimes with CTCs, and thus such spacetimes can be stable against vacuum polarization.

In a recent paper, Cassidy and Hawking [143] have admitted that "back-reaction does not enforce chronology protection." On the other hand, Cassidy and Hawking [143] have argued that the "number of states" may enforce the chronology protection conjecture since "this quantity will always tend to zero as one tries to introduce CTCs." Their arguments are based on the fact that for the particular spacetime with CTCs they constructed [which is the product of a multiply connected (via a boost) three-dimensional de Sitter space and S^1] the entropy of a massless scalar field diverges to minus infinity when the spacetime develops CTCs [143]. However, whether this conclusion holds for general spacetimes with CTCs remains an open question and further research is required. And, from ordinary statistical thermodynamics we know that entropy is always positive, so the physical meaning of a *negative* entropy is unclear. The number of states in phase space is given by $N = \Delta p \Delta q / (2\pi\hbar)^s$ where $\Delta q = \Delta q_1 \Delta q_2 \dots \Delta q_s$, $\Delta p = \Delta p_1 \Delta p_2 \dots \Delta p_s$, q_i ($i = 1, 2, \dots, s$) is a canonical coordinate, p_i is a canonical momentum, and s is the number of degrees of freedom. The uncertainty principle demands that $\Delta p_i \Delta q_i \geq 2\pi\hbar$ and thus we should always have $N \geq 1$. Thus the "fact" that the number of states tends to zero as one tries to develop CTCs (i.e. as one approaches the Cauchy horizon) may simply imply that near the Cauchy horizon quantum effects of gravity cannot be neglected, which is consistent with Li, Xu, and Liu's argument [22]. The entropy is defined by $k_B \ln N$ where N is the number of states and k_B is the Boltzmann constant. When N is small, quantization of the entropy becomes important (remember that the number of states N is always an integer). The entropy cannot *continuously* tend to negative infinity; it should *jump* from $k_B \ln 3$ to $k_B \ln 2$, *jump* from $k_B \ln 2$ to zero (but in Cassidy and Hawking's arguments [143] we have not seen such a jump), then the uncertainty principle demands that the entropy should stand on the zero value as one approaches the Cauchy horizon. On the other hand, ordinary continuous thermodynamics holds only for the case with $N \geq 1$. Thus, as one approaches the Cauchy horizon the thermodynamic limit has already been violated and ordinary

thermodynamics should be revised near the Cauchy horizon. In other words, Cassidy and Hawking's results [143] cannot be extended to the Cauchy horizon. Based on the fact that the effective action density diverges at the polarized hypersurfaces of spacetimes with CTCs [31], Cassidy and Hawking [143] have argued that the effective action "would provide new insight into issues of chronology protection." But we should note that the effective action is only a *tool* for computing some physical quantities (such as the energy-momentum tensor) and the effective action itself has not much physical meaning. The divergence of the effective action may imply that the effective action is not a good *tool* as the polarized hypersurfaces are approached. Our argument is supported by the fact that there exist many examples for which the energy-momentum tensor is finite everywhere, as mentioned above.

Recently, Kay, Radzikowski, and Wald [144] have proved two theorems which demonstrate that some fundamental quantities such as Hadamard functions and energy-momentum tensors must be ill-defined on a compactly generated Cauchy horizon in a spacetime with CTCs, as one extends the *usual* quantum field theory in a global hyperbolic spacetime to an acausal spacetime with a compactly generated Cauchy horizon. Basically speaking, their theorems imply that the *usual* quantum field theory cannot be *directly* extended to a spacetime with CTCs [144]. Their theorems tell us that serious difficulties arise when attempting to *define* quantum field theory on a spacetime with a compactly generated Cauchy horizon [144]. The ordinary quantum field theory must be significantly changed or some new approach must be introduced when one tries to do quantum field theory on a spacetime with CTCs. A candidate procedure for overcoming this difficulty is the Euclidean quantization proposed by Hawking [145,146]. Quantum field theory is well-defined in a Euclidean space because there are no CTCs in a Euclidean space [147]. In fact, even in simply connected Minkowski spacetime, quantum field theory is *not* well-defined since the path integral does not converge. To overcome this difficulty, the technique of Wick-rotation (which is essentially equivalent to Euclidean quantization) is used. Kay, Radzikowski, and Wald [144] have also argued that their results may be interpreted as indicating that in order to create CTCs it would be necessary to enter a regime where quantum effects of gravity will be dominant (see also the discussions of Visser [148,149]); this is also consistent with Li, Xu, and Liu's arguments [22]. Cramer and Kay [150,151] have shown that Kay, Radzikowski, and Wald's theorems [144] also apply to Misner space (for Sushkov's automorphic field case [28] and Krasnikov's two-dimensional case [27], respectively) where the Cauchy horizon is not compactly generated, in the sense that the energy-momentum tensor must be ill-defined on the Cauchy horizon itself. But we note that this only happens in a set of measure zero which does not make much sense in physics for if the renormalized energy-momentum tensor is zero everywhere except on a set of measure zero where it is formally ill-defined, then continuity would seem to require setting it to zero there also [30].

Perhaps a conclusion on the chronology protection conjecture can only be reached after we have a quantum theory

of gravity. However, we can conclude that the back-reaction of vacuum polarization does *not* enforce the chronology protection conjecture, a point Hawking himself also admits [143]. (Originally the back-reaction of vacuum polarization was supposed to be the strongest candidate for chronology protection [19,20].)

VI. MULTIPLY CONNECTED MINKOWSKI SPACETIMES WITH CTCs

A simple spacetime with CTCs is obtained from Minkowski spacetime by identifying points that are related by time translation. Minkowski spacetime is (R^4, η_{ab}) . In Cartesian coordinates (t, x, y, z) the Lorentzian metric η_{ab} is given by

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2. \quad (10)$$

Now we identify points (t, x, y, z) with points $(t + nt_0, x, y, z)$ where t_0 is a positive constant and n is any integer. Then we obtain a spacetime with topology $S^1 \times R^3$ and the Lorentzian metric. Such a spacetime is closed in the time direction and has no Cauchy horizon. All events in this spacetime are threaded by CTCs. (This is the only acausal spacetime without a Cauchy horizon considered in this paper.) Minkowski spacetime (R^4, η_{ab}) is the covering space of this spacetime.

Usually there is no well-defined quantum field theory in a spacetime with CTCs. (Kay-Radzickowski-Wald's theorems [144] enforce this claim, though they do not apply directly to an acausal spacetime without a Cauchy horizon.) However, in the case where a covering space exists, we can do it in the covering space with the method of images. In fact in most cases where the energy-momentum tensor in spacetimes with CTCs has been calculated, this method has been used (for the theoretical basis for the method of images see Ref. [15] and references cited therein). The method of images is sufficient for our purposes in this paper (computing the energy-momentum tensor and the response function of particle detectors). Thus in this paper we use this method to deal with quantum field theory in spacetimes with CTCs.

For any point (t, x, y, z) in $(S^1 \times R^3, \eta_{ab})$, there are an infinite number of images of points $(t + nt_0, x, y, z)$ in the covering space (R^4, η_{ab}) . For the Minkowski vacuum $|0_M\rangle$ of a conformally coupled scalar field (by ‘‘conformally coupled’’ we mean that the mass of the scalar field is zero and the coupling between the scalar field ϕ and the gravitational field is given by $\frac{1}{6}R\phi^2$ where R is the Ricci scalar curvature) in the Minkowski spacetime, the Hadamard function is

$$G_M^{(1)}(X, X') = \frac{1}{2\pi^2} \frac{1}{-(t-t')^2 + (x-x')^2 + (y-y')^2 + (z-z')^2}, \quad (11)$$

here $X = (t, x, y, z)$ and $X' = (t', x', y', z')$. With the method of images, the Hadamard function of the ‘‘adapted’’ Minkowski vacuum (which is the Minkowski vacuum with multiple images) in the spacetime $(S^1 \times R^3, \eta_{ab})$ is given by the summation of the Hadamard function in Eq. (11) for all images

$$G^{(1)}(X, X') = \frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{1}{-(t-t'+nt_0)^2 + (x-x')^2 + (y-y')^2 + (z-z')^2}. \quad (12)$$

The regularized Hadamard function is usually taken to be

$$G_{\text{reg}}^{(1)}(X, X') = G^{(1)}(X, X') - G_M^{(1)}(X, X') = \frac{1}{2\pi^2} \sum_{n \neq 0} \frac{1}{-(t-t'+nt_0)^2 + (x-x')^2 + (y-y')^2 + (z-z')^2}. \quad (13)$$

The renormalized energy-momentum tensor is given by [152,153]

$$\langle T_{ab} \rangle_{\text{ren}} = \frac{1}{2} \lim_{X' \rightarrow X} \left(\frac{2}{3} \nabla_a \nabla_{b'} - \frac{1}{3} \nabla_a \nabla_b - \frac{1}{6} \eta_{ab} \nabla_c \nabla^{c'} \right) G_{\text{reg}}^{(1)}. \quad (14)$$

Inserting Eq. (13) into Eq. (14) we get

$$\langle T_{\mu}^{\nu} \rangle_{\text{ren}} = \frac{\pi^2}{90t_0^4} \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (15)$$

We find that this energy-momentum tensor is constant and finite everywhere and has the form of radiation. Thus CTCs do not mean that the energy-momentum tensor must diverge.

Now let us consider a particle detector [153,154] moving

in this spacetime. The particle detector is coupled to the field ϕ by the interaction Lagrangian $cm(\tau)\phi[X(\tau)]$, where c is a small coupling constant, m is the detector's monopole moment, τ is the proper time of the detector's world line, and $X(\tau)$ is the trajectory of the particle detector [153]. Suppose initially the detector is in its ground state with energy E_0 and the field ϕ is in some quantum state $|\rangle$. Then the transition probability for the detector to all possible excited states with energy $E > E_0$ and the field ϕ to all possible quantum states is given by [153]

$$P = c^2 \sum_{E > E_0} |\langle E | m(0) | E_0 \rangle|^2 \mathcal{F}(\Delta E), \quad (16)$$

where $\Delta E = E - E_0 > 0$ and $\mathcal{F}(\Delta E)$ is the response function

$$\mathcal{F}(\Delta E) = \int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\tau' e^{-i\Delta E(\tau - \tau')} \times G^+(X(\tau), X(\tau')), \quad (17)$$

which is independent of the details of the particle detector and is determined by the positive frequency Wightman function $G^+(X, X') \equiv \langle |\phi(X)\phi(X')| \rangle$ [while the Hadamard function is defined by $G^{(1)}(X, X') \equiv \langle |\phi(X)\phi(X') + \phi(X')\phi(X)| \rangle$]. The response function represents the bath of particles that the detector effectively experiences [153]. The remaining factor in Eq. (16) represents the selectivity of the detector to the field and depends on the internal structure of the detector [153]. The Wightman function for the Minkowski vacuum is

$$G_M^+(X, X') = \frac{1}{4\pi^2} \frac{1}{-(t-t'-i\epsilon)^2 + (x-x')^2 + (y-y')^2 + (z-z')^2}, \quad (18)$$

where ϵ is an infinitesimal positive real number which is introduced to indicate that G^+ is the boundary value of a function which is analytic in the lower-half of the complex $\Delta t \equiv t - t'$ plane. For the adapted Minkowski vacuum in our spacetime $(S^1 \times R^3, \eta_{ab})$, the Wightman function is

$$G^+(X, X') = \frac{1}{4\pi^2} \sum_{n=-\infty}^{\infty} \frac{1}{-(t-t'+nt_0-i\epsilon)^2 + (x-x')^2 + (y-y')^2 + (z-z')^2}. \quad (19)$$

Assume that the detector moves along the geodesic $x = \beta t$ ($\beta < 1$), $y = z = 0$, then the proper time is $\tau = t/\zeta$ with $\zeta = 1/\sqrt{1-\beta^2}$. On the geodesic, the Wightman function is reduced to

$$\begin{aligned} G^+(\tau, \tau') &= \frac{1}{4\pi^2} \sum_{n=-\infty}^{\infty} \frac{1}{-(t-t'+nt_0-i\epsilon)^2 + \beta^2(t-t')^2} \\ &= -\frac{1}{4\pi^2\zeta^2} \sum_{n=-\infty}^{\infty} \frac{1}{(\tau-\tau'+nt_0/\zeta-i\epsilon/\zeta)^2 - \beta^2(\tau-\tau')^2}. \end{aligned} \quad (20)$$

Inserting Eq. (20) into Eq. (17), we obtain

$$\mathcal{F}(\Delta E) = -\frac{1}{4\pi^2\zeta^2} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} dT \int_{-\infty}^{\infty} d\Delta\tau e^{-i\Delta E\Delta\tau} \frac{1}{(\Delta\tau + nt_0/\zeta - i\epsilon/\zeta)^2 - \beta^2(\Delta\tau)^2}, \quad (21)$$

where $\Delta\tau = \tau - \tau'$ and $T = (\tau + \tau')/2$. The integration over $\Delta\tau$ is taken along a contour closed in the lower-half plane of complex $\Delta\tau$. Inspecting the poles of the integrand, we find that all poles are in the upper-half plane of complex $\Delta\tau$ (remember that $\beta < 1$). Therefore according to the residue theorem we have

$$\mathcal{F}(\Delta E) = 0. \quad (22)$$

Such a particle detector perceives no particles, though the renormalized energy-momentum tensor of the field has the form of radiation.

Another simple space with CTCs constructed from Minkowski space is Misner space [6]. In Cartesian coordi-

nates (t, x, y, z) in Minkowski spacetime, a boost transformation in the (t, x) plane (we can always adjust the coordinates so that the boost is in this plane) takes point (t, x, y, z) to point $(t \cosh b + x \sinh b, x \cosh b + t \sinh b, y, z)$ where b is the boost parameter. In Rindler coordinates (η, ξ, y, z) , defined by

$$t = \xi \sinh \eta, \quad (23)$$

$$x = \xi \cosh \eta,$$

$$y = y,$$

$$z = z,$$

the Minkowski metric can then be written in the Rindler form

$$ds^2 = -\xi^2 d\eta^2 + d\xi^2 + dy^2 + dz^2. \quad (24)$$

The Rindler coordinates (η, ξ, y, z) only cover the right quadrant of Minkowski space (i.e. the region R defined by $x > |t|$). By a reflection $(t, x, y, z) \rightarrow (-t, -x, y, z)$ [or $(\eta, \xi, y, z) \rightarrow (\eta, -\xi, y, z)$], the Rindler coordinates and the Rindler metric can be extended to the left quadrant (L, defined by $x < -|t|$). By the transformation

$$\eta \rightarrow \tilde{\xi} - i\frac{\pi}{2}, \quad \xi \rightarrow \pm i\tilde{\eta}, \quad y \rightarrow y, \quad z \rightarrow z, \quad (25)$$

the Rindler coordinates can be extended to the future quadrant (F, defined by $t > |x|$) and the past quadrant (P, defined by $t < -|x|$). In region L the Rindler metric has the same form as the metric in region R, which is given by Eq. (24). But in F and P the Rindler metric is extended to be

$$ds^2 = -d\tilde{\eta}^2 + \tilde{\eta}^2 d\tilde{\xi}^2 + dy^2 + dz^2. \quad (26)$$

Misner space is obtained by identifying (t, x, y, z) with $(t \cosh nb + x \sinh nb, x \cosh nb + t \sinh nb, y, z)$. Under such an identification, point (η, ξ, y, z) in R (or L) is identified with points $(\eta + nb, \xi, y, z)$ in R (or L), point $(\tilde{\eta}, \tilde{\xi}, y, z)$ in F (or P) is identified with points $(\tilde{\eta}, \tilde{\xi} + nb, y, z)$ in F (or P). Clearly there are CTCs in R and L but there are no closed causal curves in F and P, and these regions are separated by the Cauchy horizons $x = \pm t$, generated by closed null geodesics.

Misner space is not a manifold at the intersection of $x = t$ and $x = -t$. However, as Hawking and Ellis [40] have pointed out, if we consider the bundle of linear frames over Minkowski space, the corresponding induced bundle of linear frames over Misner space is a Hausdorff manifold and therefore well-behaved everywhere.

The energy-momentum tensor of a conformally coupled scalar field in Misner space has been studied in [14,30]. Hiscock and Konkowski [14] have calculated the energy-momentum tensor of the adapted Minkowski vacuum. In Rindler coordinates their results can be written as

$$\langle T_{\mu}^{\nu} \rangle_{\text{M,ren}} = \frac{A}{12\pi^2 \xi^4} \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (27)$$

where the constant A is

$$A = \sum_{n=1}^{\infty} \frac{2 + \cosh nb}{(\cosh nb - 1)^2}. \quad (28)$$

Equation (27) holds only in region R [because Rindler coordinates defined by Eq. (23) only cover R], but it can be analytically extended to other regions by writing $\langle T_{\mu}^{\nu} \rangle_{\text{M,ren}}$ in Cartesian coordinates or by the transformations mentioned

above. Obviously for any finite b , $\langle T_{\mu}^{\nu} \rangle_{\text{M,ren}}$ diverges as one approaches the Cauchy horizon ($\xi \rightarrow 0$). This divergence is coordinate independent since $\langle T^{\mu\nu} \rangle_{\text{M,ren}} \langle T_{\mu\nu} \rangle_{\text{M,ren}}$ also diverges as $\xi \rightarrow 0$. This indicates that though the Minkowski vacuum is a good and self-consistent vacuum for simply connected Minkowski space, the adapted Minkowski vacuum is *not* self-consistent for Misner space (i.e. it does not solve Einstein's equations given the Misner space geometry). This result has led Hawking [19,20] to conjecture that the laws of physics do not allow the appearance of CTCs (i.e., his chronology protection conjecture).

Li and Gott [30] have studied the adapted Rindler vacuum in Misner space. The Hadamard function for the Rindler vacuum is [155]

$$G_{\text{R}}^{(1)}(X, X') = \frac{1}{2\pi^2} \frac{\gamma}{\xi \xi' \sinh \gamma [-(\eta - \eta')^2 + \gamma^2]}, \quad (29)$$

where $X = (\eta, \xi, y, z)$, $X' = (\eta', \xi', y', z')$, and γ is defined by

$$\cosh \gamma = \frac{\xi^2 + \xi'^2 + (y - y')^2 + (z - z')^2}{2\xi\xi'}. \quad (30)$$

The Hadamard function for the adapted Rindler vacuum in Misner space is

$$G^{(1)}(X, X') = \frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{\gamma}{\xi \xi' \sinh \gamma [-(\eta - \eta' + nb)^2 + \gamma^2]}. \quad (31)$$

Though $G_{\text{R}}^{(1)}$ and $G^{(1)}$ given by Eq. (29) and Eq. (31) are defined only in region R, they can be analytically extended to regions L, F, and P in Minkowski and Misner space. The regularized Hadamard function for the adapted Rindler vacuum is $G_{\text{reg}}^{(1)}(X, X') = G^{(1)}(X, X') - G_{\text{M}}^{(1)}(X, X')$, where $G_{\text{M}}^{(1)}$ is the Hadamard function for the Minkowski vacuum given by Eq. (11). Inserting this together with Eq. (31) and Eq. (11) into Eq. (14), we obtain the energy-momentum tensor for a conformally coupled scalar field in the adapted Rindler vacuum [30]

$$\langle T_{\mu}^{\nu} \rangle_{\text{R,ren}} = \frac{1}{1440\pi^2 \xi^4} \left[\left(\frac{2\pi}{b} \right)^4 - 1 \right] \times \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (32)$$

which is expressed in Rindler coordinates and thus holds only in region R but can be analytically extended to other regions with the method mentioned above for the case of the adapted Minkowski vacuum. We [30] have found that unless $b = 2\pi$, $\langle T_{\mu}^{\nu} \rangle_{\text{R,ren}}$ blows up as one approaches the Cauchy horizon ($\xi \rightarrow 0$) (as also does $\langle T^{\mu\nu} \rangle_{\text{R,ren}} \langle T_{\mu\nu} \rangle_{\text{R,ren}}$). But, if $b = 2\pi$, we have

$$\langle T_{\mu}^{\nu} \rangle_{R,ren} = 0, \quad (33)$$

which is regular as one approaches the Cauchy horizon and can be regularly extended to the whole Misner space, where it is also zero. In such a case, the vacuum Einstein's equations without cosmological constant are automatically satisfied. Thus this is an example of a spacetime with CTCs at the semiclassical quantum gravity level. We [30] have called this vacuum the *self-consistent vacuum* for Misner space, and $b = 2\pi$ is the *self-consistent condition*. (Cassidy [31] has also independently proven that for a conformally coupled scalar field in Misner space there should exist a quantum state for which the energy-momentum tensor is zero everywhere. But he has not shown what quantum state it should be. We [30] have shown that it is the adapted Rindler vacuum.)

Another way to deal with quantum fields in spacetimes with CTCs is to do the quantum field theory in the Euclidean section and then analytically extend the results to the Lorentzian section [147]. For Misner space the Euclidean section is obtained by taking η and b to be $-i\bar{\eta}$ and $-i\bar{b}$. The resultant space is the Euclidean space with metric $ds^2 = \xi^2 d\bar{\eta}^2 + d\xi^2 + dy^2 + dz^2$ and $(\bar{\eta}, \xi, y, z)$ and $(\bar{\eta} + n\bar{b}, \xi, y, z)$ are identified where $(\bar{\eta}, \xi, y, z)$ are cylindrical polar coordinates with $\bar{\eta}$ the angular polar coordinate and ξ the radial polar coordinate. The geometry at the hypersurface $\xi = 0$ is conical singular unless $\bar{b} = 2\pi$. When extending that case to the Lorentzian section, we get $b = 2\pi$ which is just the self-consistent condition. This may be the geometrical explanation of the self-consistent condition. By doing quantum field theory in the Euclidean space, then analytically extending the results to the Lorentzian section, we obtain the renormalized energy-momentum tensor in region R (or L) of the Misner space. Then we can extend the renormalized energy-

momentum tensor in R (or L) to regions F (or P). The results are the same as that obtained with the method of images.

Let us consider a particle detector moving in Misner space with the adapted Rindler vacuum. Suppose the detector moves along a geodesic with $x = a$, $y = \beta t$, and $z = 0$ (a and β are constants and a is positive), which goes through the P, R, and F regions. The proper time of the detector is $\tau = t/\zeta$ with $\zeta = 1/\sqrt{1-\beta^2}$. On this geodesic, the Hadamard function in Eq. (31) is reduced to

$$G^{(1)}(t, t') = \frac{1}{2\pi^2} \frac{\gamma}{\sinh \gamma \sqrt{(a^2 - t^2)(a^2 - t'^2)}} \times \sum_{n=-\infty}^{\infty} \frac{1}{-(\eta - \eta' + nb)^2 + \gamma^2}, \quad (34)$$

where γ is given by

$$\cosh \gamma = \frac{2a^2 - t^2 - t'^2 + \beta^2(t - t')^2}{2\sqrt{(a^2 - t^2)(a^2 - t'^2)}}, \quad (35)$$

and $\eta - \eta'$ is given by

$$\sinh(\eta - \eta') = \frac{a(t - t')}{\sqrt{(a^2 - t^2)(a^2 - t'^2)}}. \quad (36)$$

Though this Hadamard function is originally defined only in R, it can be analytically extended to F, P, and L. The Wightman function is equal to 1/2 of the Hadamard function with t replaced by $t - i\epsilon/2$ and t' replaced by $t' + i\epsilon/2$, where ϵ is an infinitesimal positive real number. Then the response function is [30]

$$\mathcal{F}(E) = \frac{1}{4\pi^2} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} dT \int_{-\infty}^{\infty} d\Delta \tau \times \frac{\gamma^+ e^{-iE\Delta\tau}}{\sinh \gamma^+ \sqrt{\left[a^2 - \zeta^2 \left(T + \frac{\Delta\tau}{2} - \frac{i\epsilon}{2\zeta} \right)^2 \right] \left[a^2 - \zeta^2 \left(T - \frac{\Delta\tau}{2} + \frac{i\epsilon}{2\zeta} \right)^2 \right] \{ -[(\eta - \eta')^+ + nb]^2 + \gamma^{+2} \}}}, \quad (37)$$

where $T \equiv (\tau + \tau')/2$, $\Delta\tau \equiv \tau - \tau'$; γ^+ and $(\eta - \eta')^+$ are given by Eqs. (35) and (36) with t replaced by $t - i\epsilon/2$ and t' replaced by $t' + i\epsilon/2$. The integral over $\Delta\tau$ can be worked out by the residue theorem where we choose the integration contour to close in the lower-half complex- $\Delta\tau$ plane. The result is zero since there are no poles in the lower-half plane. Therefore such a detector cannot be excited and so it detects nothing [30]. We [30] have also calculated the response functions for detectors on world lines with constant ξ , y , and z and world lines with constant $\bar{\xi}$, y , and z —both are zero.

VII. VACUUM POLARIZATION IN VILENKIN'S TUNNELING UNIVERSE

In order to compare our model for the creation of the universe with Vilenkin's tunneling universe, in this section we calculate the vacuum fluctuation of a conformally coupled scalar field in Vilenkin's tunneling universe. The geometry of Vilenkin's tunneling universe has been described in Sec. IV. Such a universe is described by a Lorentzian de Sitter space joined to a Euclidean de Sitter

space [11]. The Lorentzian section has the topology $R^1 \times S^3$ and the metric

$$ds^2 = -d\tau^2 + r_0^2 \cosh^2 \frac{\tau}{r_0} [d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2)]. \quad (38)$$

The Euclidean section has the topology S^4 and the metric

$$ds^2 = d\tau^2 + r_0^2 \cos^2 \frac{\tau}{r_0} [d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2)]. \quad (39)$$

The Lorentzian section and the Euclidean section are joined at the boundary Σ defined by $\tau=0$. Σ is a three-sphere with the minimum radius in de Sitter space and the maximum radius in the Euclidean four-sphere. The boundary condition for a conformally coupled scalar field ϕ is [156,157]

$$\left. \frac{\partial \phi}{\partial \tau} \right|_{\Sigma} = 0, \quad (40)$$

which is a kind of Neumann boundary condition and indicates that the boundary Σ is like a kind of reflecting boundary. The Green functions (including both the Hadamard function and the Wightman function) should also satisfy this boundary condition

$$\left. \frac{\partial G(\tau, \chi, \theta, \phi; \tau', \chi', \theta', \phi')}{\partial \tau} \right|_{\Sigma} = 0. \quad (41)$$

The vacuum state of a conformally coupled scalar field in de Sitter space is usually taken to be that obtained from the Minkowski vacuum by the conformal transformation according to which de Sitter space is conformally flat. (The quantum state so obtained is usually called the conformal vacuum [153].) Such a vacuum is de Sitter invariant and we call it the conformal Minkowski vacuum. The Hadamard function for this de Sitter vacuum (i.e. the conformal Minkowski vacuum) is [158]

$$G_{\text{CM}}^{(1)}(X, X') = \frac{1}{4\pi^2 r_0^2} \frac{1}{1-Z(X, X')}, \quad (42)$$

where $X = (\tau, \chi, \theta, \phi)$, $X' = (\tau', \chi', \theta', \phi')$, and $Z(X, X')$ is defined by

$$\begin{aligned} Z(X, X') = & -\sinh \frac{\tau}{r_0} \sinh \frac{\tau'}{r_0} \\ & + \cosh \frac{\tau}{r_0} \cosh \frac{\tau'}{r_0} \{ \cos \chi \cos \chi' \\ & + \sin \chi \sin \chi' [\cos \theta \cos \theta' \\ & + \sin \theta \sin \theta' \cos(\phi - \phi')] \}. \end{aligned} \quad (43)$$

In Vilenkin's tunneling universe, the Hadamard function satisfying the boundary condition (41) is given by

$$\begin{aligned} G^{(1)}(X, X') &= G_{\text{CM}}^{(1)}(X, X') + G_{\text{CM}}^{(1)}(X^-, X') \\ &= \frac{1}{4\pi^2 r_0^2} \left[\frac{1}{1-Z(X, X')} \right. \\ &\quad \left. + \frac{1}{1-Z(X^-, X')} \right], \end{aligned} \quad (44)$$

where $X^- = (-\tau, \chi, \theta, \phi)$ is the image of $X = (\tau, \chi, \theta, \phi)$ with respect to the reflecting boundary Σ .

There are various schemes for obtaining the renormalized energy-momentum tensor for de Sitter space (e.g. [158,159]). They all are equivalent to subtracting from the Hadamard function a reference term $G_{\text{ref}}^{(1)}$ to obtain a regularized Hadamard function and then calculating the renormalized energy-momentum tensor by [152,153]

$$\langle T_{ab} \rangle_{\text{ren}} = \frac{1}{2} \lim_{X' \rightarrow X} \mathcal{D}_{ab'}(X, X') G_{\text{reg}}^{(1)}(X, X'). \quad (45)$$

For the conformally coupled scalar field, the differential operator $\mathcal{D}_{ab'}$ is

$$\begin{aligned} \mathcal{D}_{ab'} &= \frac{2}{3} \nabla_a \nabla_{b'} - \frac{1}{6} g_{ab'} g_{cd'} \nabla^c \nabla^{d'} \\ &\quad - \frac{1}{3} \nabla_{a'} \nabla_{b'} + \frac{1}{3} g_{ab'} \nabla_{d'} \nabla^{d'} \\ &\quad + \frac{1}{6} \left(R_{ab} - \frac{1}{2} R g_{ab} \right), \end{aligned} \quad (46)$$

where $g_{ab'}$, $g_{cd'}$ are geodesic parallel displacement bivectors [160]. [It is easy to show that if $R_{ab}=0$ Eq. (45) and Eq. (46) are reduced to Eq. (14).] The regularized Hadamard function for the adapted conformal Minkowski vacuum in Vilenkin's tunneling universe is

$$\begin{aligned} G_{\text{reg}}^{(1)}(X, X') &= G^{(1)}(X, X') - G_{\text{ref}}^{(1)}(X, X') \\ &= [G_{\text{CM}}^{(1)}(X, X') - G_{\text{ref}}^{(1)}] \\ &\quad + G_{\text{CM}}^{(1)}(X^-, X'). \end{aligned} \quad (47)$$

(In this paper the exact form of $G_{\text{ref}}^{(1)}$ is not important for us.) Substituting Eqs. (42)–(44) and Eq. (47) into Eq. (45), we find that $\lim_{X' \rightarrow X} \mathcal{D}_{ab'} G_{\text{CM}}^{(1)}(X^-, X') = 0$, which shows that the boundary condition (40) does not produce any renormalized energy-momentum tensor; but the action of $\mathcal{D}_{ab'}$ on $G_{\text{CM}}^{(1)}(X, X') - G_{\text{ref}}^{(1)}$ should give the energy-momentum tensor for the conformal Minkowski vacuum in an eternal de Sitter space [158,159]

$$\frac{1}{2} \lim_{X' \rightarrow X} \mathcal{D}_{ab'} [G_{\text{CM}}^{(1)}(X, X') - G_{\text{ref}}^{(1)}] = -\frac{1}{960\pi^2 r_0^4} g_{ab}. \quad (48)$$

Therefore, the energy-momentum tensor of a conformally coupled scalar field in the adapted Minkowski vacuum in Vilenkin's tunneling universe is

$$\langle T_{ab} \rangle_{\text{ren}} = -\frac{1}{960\pi^2 r_0^4} g_{ab}, \quad (49)$$

which is the same as that for an eternal de Sitter space.

Now consider a particle detector moving along a geodesic with $\chi, \theta, \phi = \text{const}$. The response function is given by Eq. (17) but with the integration over τ and τ' ranging from 0 to ∞ . The Wightman function is obtained from the corresponding Hadamard function by the relation

$$G^+(\tau, \chi, \theta, \phi; \tau', \chi', \theta', \phi') = \frac{1}{2} G^{(1)}(\tau - i\epsilon/2, \chi, \theta, \phi; \tau' + i\epsilon/2, \chi', \theta', \phi'), \quad (50)$$

where ϵ is an infinitesimal positive real number. Along the world line of the detector, we have

$$Z(\tau, \tau') = -\sinh \frac{\tau}{r_0} \sinh \frac{\tau'}{r_0} + \cosh \frac{\tau}{r_0} \cosh \frac{\tau'}{r_0} = \cosh \frac{\tau - \tau'}{r_0}, \quad (51)$$

$$Z(-\tau, \tau') = +\sinh \frac{\tau}{r_0} \sinh \frac{\tau'}{r_0} + \cosh \frac{\tau}{r_0} \cosh \frac{\tau'}{r_0} = \cosh \frac{\tau + \tau'}{r_0}, \quad (52)$$

and

$$G^+(X, X') = \frac{1}{8\pi^2 r_0^2} \left(\frac{1}{1 - \cosh \frac{\tau - \tau' - i\epsilon}{r_0}} + \frac{1}{1 - \cosh \frac{\tau + \tau'}{r_0}} \right). \quad (53)$$

Then the response function is

$$\mathcal{F}(\Delta E) = \frac{1}{8\pi^2} \int_0^\infty dT \int_{-\infty}^\infty d\Delta \tau e^{-i\Delta E r_0 \Delta \tau} \times \left[\frac{1}{1 - \cosh(\Delta \tau - i\epsilon)} + \frac{1}{1 - \cosh 2T} \right], \quad (54)$$

where $\Delta \tau = (\tau - \tau')/r_0$ and $T = (\tau + \tau')/2r_0$. It is easy to calculate the contour integral over $\Delta \tau$. We find that the integration of the second term is zero and therefore, the result is the same as that for an inertial particle detector in an eternal de Sitter space [100,153]. Thus we have

$$\frac{d\mathcal{F}}{dT} = \frac{r_0}{2\pi} \frac{\Delta E}{e^{2\pi r_0 \Delta E} - 1}, \quad (55)$$

which is just the response function for a detector in thermal radiation with the Gibbons-Hawking temperature [100]

$$T_{\text{G-H}} = \frac{1}{2\pi r_0}. \quad (56)$$

[The factor r_0 over 2π in Eq. (55) is due to the fact that by definition $T = (\tau + \tau')/2r_0$ is dimensionless.] Therefore such a detector perceives a thermal bath of radiation with the temperature $T_{\text{G-H}}$.

Though the boundary between the Lorentzian section and the Euclidean section behaves as a reflecting boundary, a particle detector cannot distinguish Vilenkin's tunneling universe from an eternal de Sitter space, and they have the same energy-momentum tensor for the conformally coupled scalar field.

VIII. A TIME-NONORIENTABLE de SITTER SPACE

A time-nonorientable de Sitter space can be constructed from de Sitter space by identifying antipodal points [161,40]. Under such an identification, point $X = (\tau, \chi, \theta, \phi)$ is identified with $-X = (-\tau, \pi - \chi, \pi - \theta, \pi + \phi)$. Friedman and Higuchi [162,163] have described this space as a ‘‘Lorentzian universe from nothing’’ (without any Euclidean section), although one could also describe it as always existing. Friedman and Higuchi have studied quantum field theory in this space but have not calculated the renormalized energy-momentum tensor [162].

de Sitter space is the covering space of this time-nonorientable model. Using the method of images, the Hadamard function of a conformally coupled scalar field in the time-nonorientable de Sitter space with the ‘‘adapted’’ conformal Minkowski vacuum can be constructed as

$$G^{(1)}(X, X') = G_{\text{CM}}^{(1)}(X, X') + G_{\text{CM}}^{(1)}(-X, X') \\ = \frac{1}{4\pi^2 r_0^2} \left[\frac{1}{1 - Z(X, X')} + \frac{1}{1 - Z(-X, X')} \right] \\ = \frac{1}{4\pi^2 r_0^2} \left[\frac{1}{1 - Z(X, X')} + \frac{1}{1 + Z(X, X')} \right]. \quad (57)$$

The regularized Hadamard function is

$$G_{\text{reg}}^{(1)}(X, X') = G^{(1)}(X, X') - G_{\text{ref}}^{(1)}(X, X') \\ = [G_{\text{CM}}^{(1)}(X, X') - G_{\text{ref}}^{(1)}(X, X')] + G_{\text{CM}}^{(1)}(-X, X'). \quad (58)$$

Inserting Eq. (57) and Eq. (58) into Eq. (45), we find that the contribution of $G_{\text{CM}}^{(1)}(-X, X')$ to the energy-momentum tensor is zero. Therefore the renormalized energy-momentum tensor is the same as that in an eternal de Sitter space, which is given by Eq. (49).

Suppose a particle detector moves along a world line with $\chi, \theta, \phi = \text{const}$. The response function is given by Eq. (17). The Wightman function is obtained from the Hadamard function through Eq. (50). On the world line of the particle detector, we have

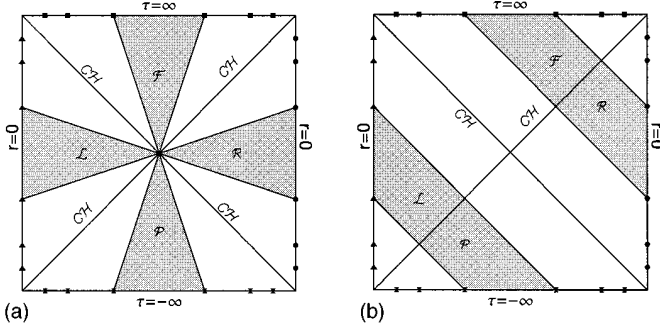


FIG. 4. Penrose diagrams of our multiply connected de Sitter space mapped onto its universal covering space (de Sitter space). Under a boost transformation, points with the same symbols (squares, disks, triangles, or double-triangles) are identified. Our multiply connected de Sitter space is divided into four regions \mathcal{R} , \mathcal{L} , \mathcal{F} , and \mathcal{P} , which are separated by Cauchy horizons \mathcal{CH} . The shaded regions represent fundamental cells of the multiply connected de Sitter space. [Figures 4(a) and 4(b) represent two different choices of the fundamental cells, but they are equivalent.] The fundamental cells \mathcal{R} and \mathcal{L} have a finite four-volume, whereas the fundamental cells \mathcal{F} and \mathcal{P} (which extend infinitely to the future and the past, respectively) have an infinite four-volume. In 4(a) the left and right boundaries of \mathcal{F} are identified, likewise for \mathcal{P} ; the upper and lower boundaries of \mathcal{R} are identified, likewise for \mathcal{L} . In 4(b) region $\mathcal{F}+\mathcal{R}$ is partially bounded by two null surfaces, the lower one is the future light cone of an event E , and the upper one is the future light cone of an event E' which is identified with E under the action of a boost. These two future light cones are identified creating a periodic boundary condition for the causally connected region $\mathcal{F}+\mathcal{R}$. \mathcal{R} and \mathcal{F} are separated by a Cauchy horizon \mathcal{CH} . Self-consistency (non-divergence of $T^{ab}T_{ab}$ as \mathcal{CH} is approached) requires retarded potentials in \mathcal{R} and \mathcal{F} . Region $\mathcal{P}+\mathcal{L}$ is partially bounded by the past light cone of an event F and the past light cone of an event F' which is identified with F under the action of a boost. These two surfaces are identified creating a periodic boundary condition for $\mathcal{P}+\mathcal{L}$, where self-consistency as \mathcal{CH} separating \mathcal{P} from \mathcal{L} is approached requires advanced potentials.

$$G^+(\tau, \tau') = \frac{1}{8\pi^2 r_0^2} \left(\frac{1}{1 - \cosh \frac{\tau - \tau' - i\epsilon}{r_0}} + \frac{1}{1 + \cosh \frac{\tau - \tau' - i\epsilon}{r_0}} \right). \quad (59)$$

Inserting this into Eq. (17) we get

$$\frac{d\mathcal{F}}{dT} = \frac{r_0}{2\pi} \frac{\Delta E}{e^{\pi r_0 \Delta E} - 1}, \quad (60)$$

which represents a thermal spectrum with a temperature equal to twice the Gibbons-Hawking temperature. Therefore a particle detector moving along such a geodesic in this time-nonorientable spacetime perceives thermal radiation with temperature $T=2T_{\text{G-H}}$.

For this time-nonorientable de Sitter space, the area of the event horizon is one half that of an eternal de Sitter space. This together with $T=2T_{\text{G-H}}$ tells us that the first thermodynamic law of event horizons $\delta M_c = T\delta A$ is preserved, where M_c is the mass within the horizon, and A is the area of the horizon [100].

IX. A MULTIPLY CONNECTED de SITTER SPACE WITH CTCs

A. Construction of a multiply connected de Sitter space

de Sitter space is a solution of the vacuum Einstein's equations with a positive cosmological constant Λ , which is one of the maximally symmetric spacetimes (the others being Minkowski space and anti-de Sitter space) [39,40]. de Sitter space can be represented by a timelike hyperbolic hypersurface

$$W^2 + X^2 + Y^2 + Z^2 - V^2 = r_0^2, \quad (61)$$

embedded in a five-dimensional Minkowski space (V, W, X, Y, Z) with the metric

$$ds^2 = -dV^2 + dW^2 + dX^2 + dY^2 + dZ^2, \quad (62)$$

where $r_0 = (3/\Lambda)^{1/2}$ [40,90]. de Sitter space has ten killing vectors—four of them are boosts, and the other six are rotations. The global coordinates $(\tau, \chi, \theta, \phi)$ have been described in previous sections. Static coordinates (t, r, θ, ϕ) on de Sitter space are defined by

$$\begin{aligned} V &= (r_0^2 - r^2)^{1/2} \sinh \frac{t}{r_0}, \\ W &= (r_0^2 - r^2)^{1/2} \cosh \frac{t}{r_0}, \\ X &= r \sin \theta \cos \phi, \\ Y &= r \sin \theta \sin \phi, \\ Z &= r \cos \theta, \end{aligned} \quad (63)$$

where $-\infty < t < \infty$, $0 \leq r < r_0$, $0 < \theta < \pi$, and $0 \leq \phi < 2\pi$. In these coordinates the de Sitter metric is written as

$$\begin{aligned} ds^2 &= - \left(1 - \frac{r^2}{r_0^2} \right) dt^2 + \left(1 - \frac{r^2}{r_0^2} \right)^{-1} dr^2 \\ &+ r^2 (d\theta^2 + \sin^2 \theta d\phi^2). \end{aligned} \quad (64)$$

We divide de Sitter space dS into four regions

$$\mathcal{R} \equiv \{p \in dS | W > |V|\}, \quad (65)$$

$$\mathcal{L} \equiv \{p \in dS | W < -|V|\}, \quad (66)$$

$$\mathcal{F} \equiv \{p \in dS | V > |W|\}, \quad (67)$$

$$\mathcal{P} \equiv \{p \in dS | V < -|W|\}, \quad (68)$$

which are separated by horizons where $W = \pm V$ and $X^2 + Y^2 + Z^2 = r_0^2$. (See Fig. 4.) It is obvious that the static coordinates defined by Eq. (63) only cover region \mathcal{R} . However, similar to the Rindler coordinates, these static coordinates can be extended to region \mathcal{F} by the complex transformation

$$t \rightarrow l - i \frac{\pi}{2} r_0, \quad r \rightarrow \tilde{t}, \quad \theta \rightarrow \theta, \quad \phi \rightarrow \phi, \quad (69)$$

where $-\infty < l < \infty$ and $\tilde{t} > 2r_0$. In region \mathcal{F} , with the coordinates $(\tilde{t}, l, \theta, \phi)$, the de Sitter metric can be written as

$$ds^2 = - \left(\frac{\tilde{t}^2}{r_0^2} - 1 \right)^{-1} d\tilde{t}^2 + \left(\frac{\tilde{t}^2}{r_0^2} - 1 \right) dl^2 + \tilde{t}^2 (d\theta^2 + \sin^2 \theta d\phi^2). \quad (70)$$

Transforming the coordinate \tilde{t} to the proper time τ by

$$\tilde{t} = r_0 \cosh \frac{\tau}{r_0}, \quad (71)$$

the de Sitter metric in \mathcal{F} is written as

$$ds^2 = -d\tau^2 + \sinh^2 \frac{\tau}{r_0} dl^2 + r_0^2 \cosh^2 \frac{\tau}{r_0} (d\theta^2 + \sin^2 \theta d\phi^2). \quad (72)$$

(See Fig. 4.) The coordinates (τ, l, θ, ϕ) are related to (V, W, X, Y, Z) by

$$\begin{aligned} V &= r_0 \sinh \frac{\tau}{r_0} \cosh \frac{l}{r_0}, \\ W &= r_0 \sinh \frac{\tau}{r_0} \sinh \frac{l}{r_0}, \\ X &= r_0 \cosh \frac{\tau}{r_0} \sin \theta \cos \phi, \\ Y &= r_0 \cosh \frac{\tau}{r_0} \sin \theta \sin \phi, \\ Z &= r_0 \cosh \frac{\tau}{r_0} \cos \theta. \end{aligned} \quad (73)$$

The universe with metric (72) is a type of Kantowski-Sachs universe [164]. Any hypersurface of $\tau = \text{const}$ has topology $R^1 \times S^2$ and has four killing vectors. Similarly, the static coordinates can also be extended to \mathcal{P} and \mathcal{L} .

Another coordinate system which will be used in this paper is the steady-state coordinate system (τ, x, y, z) , defined by

$$\begin{aligned} \tau &= r_0 \ln \frac{W+V}{r_0}, \\ x &= \frac{r_0 X}{W+V}, \\ y &= \frac{r_0 Y}{W+V}, \\ z &= \frac{r_0 Z}{W+V}. \end{aligned} \quad (74)$$

These coordinates cover regions $\mathcal{R} + \mathcal{F}$ and the horizon at $W = V > 0$. With these steady-state coordinates, the de Sitter metric can be written in the steady-state form

$$ds^2 = -d\tau^2 + e^{2\tau/r_0} (dx^2 + dy^2 + dz^2). \quad (75)$$

Introducing the conformal time

$$\bar{\eta} = -r_0 e^{-\tau/r_0} = -\frac{r_0^2}{W+V}, \quad (76)$$

and spherical coordinates (ρ, θ, ϕ) defined by $x = \rho \sin \theta \cos \phi$, $y = \rho \sin \theta \sin \phi$, and $z = \rho \cos \theta$, the de Sitter metric can be written as

$$ds^2 = \frac{r_0^2}{\bar{\eta}^2} [-d\bar{\eta}^2 + d\rho^2 + \rho^2 (d\theta^2 + \sin^2 \theta d\phi^2)]. \quad (77)$$

The de Sitter metric is invariant under the action of the de Sitter group. Because the boost group in de Sitter space is a sub-group of the de Sitter group, the de Sitter metric is also invariant under the action of the boost group. A boost transformation in the (V, W) plane in the embedding five-dimensional Minkowski space induces a boost transformation in the de Sitter space. Under such a transformation, point (V, W, X, Y, Z) is taken to $(V \cosh b + W \sinh b, W \cosh b + V \sinh b, X, Y, Z)$. In static coordinates in \mathcal{R} , point (t, r, θ, ϕ) is taken to $(t + \beta, r, \theta, \phi)$ where $\beta = br_0$. In coordinates $(\tilde{t}, l, \theta, \phi)$ in \mathcal{F} , point $(\tilde{t}, l, \theta, \phi)$ is taken to $(\tilde{t}, l + \beta, \theta, \phi)$. Similar to Misner space, our multiply connected de Sitter space is constructed by identifying points (V, W, X, Y, Z) with $(V \cosh nb + W \sinh nb, W \cosh nb + V \sinh nb, X, Y, Z)$ on de Sitter space dS . In regions \mathcal{R} , points (t, r, θ, ϕ) are identified with $(t + n\beta, r, \theta, \phi)$; in region \mathcal{F} , points $(\tilde{t}, l, \theta, \phi)$ are identified with $(\tilde{t}, l + n\beta, \theta, \phi)$. We denote the multiply connected de Sitter space so obtained by dS/B , where B denotes the boost group. Under the identification generated by the boost transformation, clearly dS/B has CTCs in regions \mathcal{R} and \mathcal{L} , but has no closed causal curves in regions \mathcal{F} and \mathcal{P} . The boundaries at $W = \pm V$ and $X^2 + Y^2 + Z^2 = r_0^2$ are the Cauchy horizons which separate the

causal regions \mathcal{F} and \mathcal{P} from the acausal regions \mathcal{R} and \mathcal{L} and are generated by closed null geodesics (Fig. 4).

Similar to the case of Misner space, dS/B is not a manifold at the two-sphere defined by $W=V=0$ and $X^2+Y^2+Z^2=r_0^2$. However, as in Hawking and Ellis's arguments for Misner space [40], the quotient of the bundle of linear frames over de Sitter space by the boost group is a Hausdorff manifold and thus is well-behaved everywhere. It may not be a serious problem in physics that dS/B is not a manifold at the two-sphere mentioned above since this is a set of measure zero.

B. Conformal relation between our multiply connected de Sitter space and Misner space

It is well known that de Sitter space is conformally flat. The de Sitter metric is related to the Minkowski metric by the conformal transformation

$$g_{ab} = \Omega^2 \eta_{ab}. \quad (78)$$

It is easy to show this relation by writing the steady-state de Sitter metric using conformal time [see Eq. (77)]. However, in this paper it is more convenient to show this conformal relation by writing the de Sitter metric in the static form and the Minkowski metric in the Rindler form, and using the transformation [165]

$$\begin{aligned} \eta &= \frac{t}{r_0}, \\ \xi &= \frac{\sqrt{1-r^2/r_0^2}}{1-r \cos \theta/r_0}, \\ y &= \frac{r \sin \theta \cos \phi/r_0}{1-r \cos \theta/r_0}, \\ z &= \frac{r \sin \theta \sin \phi/r_0}{1-r \cos \theta/r_0}, \end{aligned} \quad (79)$$

then the conformal factor Ω^2 is

$$\Omega^2 = r_0^2 (1 - r \cos \theta/r_0)^2. \quad (80)$$

The conformal relations given by Eq. (79) and Eq. (80) define a *conformal map* between the static de Sitter space and the Rindler space. The horizon at $r=r_0$ in the static de Sitter space coordinates corresponds to the horizon $\xi=0$ in Rindler space, and the world line $r=0$ in de Sitter space corresponds to the world line with $\xi=1$ and $y=z=0$ in Rindler space. This conformal relation can also be extended to region \mathcal{F} in de Sitter space and region F in Minkowski space, where we have

$$\begin{aligned} \tilde{\eta} &= \pm \frac{\sqrt{\tilde{t}^2/r_0^2 - 1}}{1 - \tilde{t} \cos \theta/r_0}, \\ \tilde{\xi} &= \frac{l}{r_0}, \\ y &= \frac{\tilde{t} \sin \theta \cos \phi/r_0}{1 - \tilde{t} \cos \theta/r_0}, \\ z &= \frac{\tilde{t} \sin \theta \sin \phi/r_0}{1 - \tilde{t} \cos \theta/r_0}, \end{aligned} \quad (81)$$

and

$$\Omega^2 = r_0^2 (1 - \tilde{t} \cos \theta/r_0)^2. \quad (82)$$

Equation (81) and Eq. (82) give a *locally* conformal map in the sense that in \mathcal{F} in de Sitter space, the map given by Eq. (81) and Eq. (82) with a “+” sign only covers $\theta_0 < \theta < \pi$, where $\theta_0 = \text{Arccos}(r_0/\tilde{t})$; the map given by Eq. (81) and Eq. (82) with a “-” sign only covers $0 < \theta < \theta_0$. (Remember that in F in Rindler space we have $\tilde{\eta} > 0$.) This conformal map is singular at $\theta = \theta_0$. However, since the hypersurfaces $\tilde{t} = \text{const}$ and $\tilde{\eta} = \text{const}$ are homogeneous, in a neighborhood of any point in region F, we can always adjust coordinates (θ, ϕ) so that Eq. (81) and Eq. (82) hold, except for the points lying in region O defined by $\tilde{\eta}^2 \geq 1 + y^2 + z^2$ (i.e. $t^2 - x^2 - y^2 - z^2 \geq 1$) in F; because as $\tilde{t} \rightarrow \infty$ we have $\tilde{\eta}^2 / (1 + y^2 + z^2) \rightarrow 1$. This means that there always exists a *locally* conformal map between \mathcal{F} and F-O (defined by $t^2 - x^2 - y^2 - z^2 < 1$ in F), and future infinity ($\tilde{t} \rightarrow \infty$) in \mathcal{F} corresponds to the hyperbola $\tilde{\eta}^2 = 1 + y^2 + z^2$ (i.e. $t^2 - x^2 - y^2 - z^2 = 1$) in F.

With the above conformal transformation, Misner space is naturally transformed to the multiply connected de Sitter space dS/B with

$$\beta = b r_0. \quad (83)$$

For a conformally coupled scalar field in a conformally flat spacetime, the Green function $G(X, X')$ of the conformal vacuum is related to the corresponding Green function $\bar{G}(X, X')$ in the flat spacetime by [153]

$$G(X, X') = \Omega^{-1}(X) \bar{G}(X, X') \Omega^{-1}(X'), \quad (84)$$

the renormalized energy-momentum tensors are related by [153]

$$\begin{aligned} \langle T_a^b \rangle_{\text{ren}} &= \Omega^{-4} \langle \bar{T}_a^b \rangle_{\text{ren}} + \frac{1}{16\pi^2} \\ &\times \left[\frac{1}{9} a_1 \quad {}^{(1)}H_a^b + 2a_3 \quad {}^{(3)}H_a^b \right], \end{aligned} \quad (85)$$

where

$${}^{(1)}H_{ab} = 2\nabla_a \nabla_b R - 2g_{ab} \nabla^c \nabla_c R - \frac{1}{2} R^2 g_{ab} + 2RR_{ab}, \quad (86)$$

$${}^{(3)}H_{ab} = R_a^c R_{cb} - \frac{2}{3} RR_{ab} - \frac{1}{2} R_{cd} R^{cd} g_{ab} + \frac{1}{4} R^2 g_{ab}, \quad (87)$$

and for a scalar field $a_1 = \frac{1}{120}$ and $a_3 = -\frac{1}{360}$ [153]. [The sign before $1/16\pi^2$ is positive here because we are using signature $(-, +, +, +)$.] For de Sitter space we have $R_{ab} = \Lambda g_{ab}$, $R = 4\Lambda$, and thus ${}^{(1)}H_{ab} = 0$, ${}^{(3)}H_{ab} = \frac{1}{3}\Lambda^2 g_{ab} = (3/r_0^4)g_{ab}$. Inserting them into Eq. (85), we have

$$\langle T_a^b \rangle_{\text{ren}} = \Omega^{-4} \langle \bar{T}_a^b \rangle_{\text{ren}} - \frac{1}{960\pi^2 r_0^4} \delta_a^b. \quad (88)$$

Since the renormalized energy-momentum tensor for Minkowski space in the Minkowski vacuum is zero, we have $\langle \bar{T}_a^b \rangle_{\text{ren}} = 0$, and thus for a conformally coupled scalar field in the conformal Minkowski vacuum in a simply connected de Sitter space dS

$$\langle T_{ab} \rangle_{\text{ren}} = -\frac{1}{960\pi^2 r_0^4} g_{ab}, \quad (89)$$

which is just the expected result [see Eq. (49)].

If we insert the energy-momentum tensor in Eq. (89) into the semiclassical Einstein's equations

$$G_{ab} + \Lambda g_{ab} = 8\pi \langle T_{ab} \rangle_{\text{ren}}, \quad (90)$$

and recall that for de Sitter space we have $G_{ab} = R_{ab} - \frac{1}{2} R g_{ab} = -(3/r_0^2)g_{ab}$, we find that the semiclassical Einstein's equations are satisfied if and only if

$$\Lambda - \frac{3}{r_0^2} + \frac{1}{120\pi r_0^4} = 0. \quad (91)$$

If $\Lambda = 0$, the solutions to Eq. (91) are $r_0 = (360\pi)^{-1/2}$ and $r_0 = \infty$ [89]. Gott [89] has called the vacuum state in de Sitter space with $r_0 = (360\pi)^{-1/2}$ the self-consistent vacuum state (it has a Gibbons-Hawking thermal temperature $T_{G-H} = 1/2\pi r_0$) [100]. In this self-consistent case, $\langle T_{ab} \rangle_{\text{ren}} = -g_{ab}/960\pi^2 r_0^4$ itself is the source term producing the de Sitter geometry [89]. This may give rise to inflation at the Planck scale [89]. (In a recent paper of Panagiotakopoulos and Tetradis [167], inflation at the Planck scale has been suggested to lead to homogeneous initial conditions for a second stage inflation at the GUT scale.) The second solution $r_0 = \infty$ corresponds to Minkowski space. These perhaps supply a possible reason that the effective cosmological constant is either of order unity in Planck units or exactly zero. That is interesting because we observe $\Lambda_{\text{eff}} = 0$ today and a high Λ_{eff} is needed for inflation. If $\Lambda \neq 0$, we find that the solutions to Eq. (91) are

$$r_0^2 = \frac{3}{2\Lambda} \left(1 \pm \sqrt{1 - \frac{\Lambda}{270\pi}} \right). \quad (92)$$

A de Sitter space with r_0 given by Eq. (92) automatically satisfies the semiclassical Einstein's equations (90). Such a de Sitter space and its corresponding vacuum are thus self-consistent.

C. Renormalized energy-momentum tensor in multiply connected de Sitter space

From Eq. (88) we find that if we know the energy-momentum tensor of a conformally coupled scalar field in some vacuum state in Misner space, we can get the energy-momentum tensor in the corresponding conformal vacuum in the multiply connected de Sitter space.

Two fundamental vacuums in Minkowski space are the Minkowski vacuum and the Rindler vacuum [153,166]. The energy-momentum tensor of the conformally coupled scalar field in the adapted Minkowski vacuum in Misner space has been worked out by Hiscock and Konkowski [14]; their results are given by Eq. (27). Inserting Eq. (27) into Eq. (88), and using Eqs. (79)–(83), we obtain the energy-momentum tensor of a conformally coupled scalar field in the adapted conformal Minkowski vacuum in our multiply connected de Sitter space dS/B . In static coordinates (t, r, θ, ϕ) , it is written as

$$\langle T_\mu^\nu \rangle_{\text{CM,ren}} = \frac{\tilde{A}}{12\pi^2 r_0^4 (1 - r^2/r_0^2)^2} \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \frac{1}{960\pi^2 r_0^4} \delta_\mu^\nu, \quad (93)$$

where

$$\tilde{A} = \sum_{n=1}^{\infty} \frac{2 + \cosh \frac{n\beta}{r_0}}{\left(\cosh \frac{n\beta}{r_0} - 1 \right)^2}. \quad (94)$$

This result is defined in region \mathcal{R} , but it can be extended to region \mathcal{F} through the transformation in Eq. (69), and can also be extended to region \mathcal{L} and \mathcal{P} through similar transformations. Similar to Misner space, this energy-momentum tensor diverges at the Cauchy horizon as $r \rightarrow r_0$ for any finite β ; and the divergence is coordinate independent since $\langle T^{\mu\nu} \rangle_{\text{CM,ren}} \langle T_{\mu\nu} \rangle_{\text{CM,ren}}$ also diverges there. Though the conformal Minkowski vacuum is a good vacuum for simply connected de Sitter space [158,159], it (in the adapted version) is not self-consistent for the multiply connected de Sitter space dS/B . (That is, it does not solve the semiclassical Einstein's equations.)

In the case of an eternal Schwarzschild black hole, there are the Boulware vacuum [168] and the Hartle-Hawking vacuum [169]. The globally defined Hartle-Hawking vacuum bears essentially the same relationship to the Boulware vacuum as the Minkowski vacuum does to the Rindler vacuum [170]. For the Boulware vacuum, the energy-

momentum tensor diverges at the event horizon of the Schwarzschild black hole, which means that this state is *not* a good vacuum for the Schwarzschild black hole because, when one inserts this energy-momentum tensor back into Einstein's equations, the back-reaction will seriously alter the Schwarzschild geometry near the event horizon. For the Hartle-Hawking vacuum, however, the energy-momentum tensor is finite everywhere and a static observer outside the horizon sees Hawking radiation [171]. People usually regard the Hartle-Hawking vacuum as the reasonable vacuum state for an eternal Schwarzschild black hole because, when its energy-momentum tensor is fed back into Einstein's equations, the Schwarzschild geometry is only altered slightly [172]. Therefore, in the case of Misner space, Li and Gott [30] have tried to find a vacuum which is also self-consistent and found that the adapted Rindler vacuum is such a vacuum if $b = 2\pi$.

Here we also try to find a self-consistent vacuum for our multiply connected de Sitter space. Let us consider the adapted conformal Rindler vacuum in dS/B . The energy-momentum tensor of a conformally coupled scalar field in the adapted Rindler vacuum in Misner space is given by Eq. (32). Inserting Eq. (32) into Eq. (88) and using Eqs. (79)–(83), we obtain the energy-momentum tensor for the adapted conformal Rindler vacuum of a conformally coupled scalar field in our multiply connected de Sitter space

$$\begin{aligned} \langle T_{\mu}^{\nu} \rangle_{\text{CR,ren}} &= \frac{1}{1440\pi^2 r_0^4 (1-r^2/r_0^2)^2} \\ &\times \left[\left(\frac{2\pi r_0}{\beta} \right)^4 - 1 \right] \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &- \frac{1}{960\pi^2 r_0^4} \delta_{\mu}^{\nu}, \end{aligned} \quad (95)$$

where the coordinate system is the static coordinate system (t, r, θ, ϕ) . Similarly, this result can also be analytically extended to the whole dS/B , though the static coordinates only cover region \mathcal{R} . We find that, if

$$\beta = 2\pi r_0, \quad (96)$$

this energy-momentum tensor is regular on the whole space. [Equation (96) corresponds to $b = 2\pi$ via Eq. (83).] Other-

wise both $\langle T_{\mu}^{\nu} \rangle_{\text{CR,ren}}$ and $\langle T^{\mu\nu} \rangle_{\text{CR,ren}} \langle T_{\mu\nu} \rangle_{\text{CR,ren}}$ diverge as the Cauchy horizon is approached. For the case $\beta = 2\pi r_0$, the energy-momentum tensor is

$$\langle T_{ab} \rangle_{\text{CR,ren}} = -\frac{1}{960\pi^2 r_0^4} g_{ab}, \quad (97)$$

which is the same as the energy-momentum tensor for the conformal Minkowski vacuum in the simply connected de Sitter space.

The Euclidean section of our multiply connected de Sitter space is a four-sphere S^4 embedded in a five dimensional flat Euclidean space with those points related by an azimuthal rotation with angle β/r_0 being identified. There are conical singularities unless $\beta/r_0 = 2\pi$. This may be regarded as a geometrical explanation of the self-consistent condition in Eq. (96).

Similarly, our multiply connected de Sitter space solves the semiclassical Einstein's equations with a cosmological constant Λ and the energy-momentum tensor in Eq. (97) (and thus it is self-consistent) if $r_0^2 = (3/2\Lambda)[1 \pm \sqrt{1 - (\Lambda/270\pi)}]$ (if $\Lambda = 0$, we have the two solutions $r_0^2 = 1/360\pi$ and $r_0 = \infty$ [89]).

D. Particle detectors in the multiply connected de Sitter space

It is well known that in the simply connected de Sitter space, an inertial particle detector perceives thermal radiation with the Gibbons-Hawking temperature [Eq. (56)] if the conformally coupled scalar field is in the conformal Minkowski vacuum [100,153]. Now we want to find what a particle detector perceives in the adapted conformal Rindler vacuum in our multiply connected de Sitter space.

The response function of the particle detector is still given by Eq. (17). The Wightman function is obtained from the corresponding Hadamard function by Eq. (50). The Hadamard function for the conformally coupled scalar field in multiply connected de Sitter space is related to that in Misner space via Eq. (84) [with $G(X, X')$ replaced by $G^{(1)}(X, X')$]. The Hadamard function for the adapted Rindler vacuum in Misner space is given by Eq. (31). Inserting Eq. (31) [as $\tilde{G}_R^{(1)}$] into Eq. (84) and using Eqs. (79)–(83), we obtain the Hadamard function for the adapted conformal Rindler vacuum of the conformally coupled scalar field in our multiply connected de Sitter space

$$G_{\text{CR}}^{(1)}(X, X') = \frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{\gamma}{\sinh \gamma \sqrt{(1-r^2/r_0^2)(1-r'^2/r_0^2)} [-(t-t'+n\beta)^2 + r_0^2 \gamma^2]}, \quad (98)$$

where $X = (t, r, \theta, \phi)$, $X' = (t', r', \theta', \phi')$, and γ is written in (t, r, θ, ϕ) as

$$\cosh \gamma = \frac{1}{\sqrt{(1-r^2/r_0^2)(1-r'^2/r_0^2)}} \left\{ 1 - \frac{rr'}{r_0^2} [\cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi')] \right\}. \quad (99)$$

The Wightman function is obtained from Eq. (98) via Eq. (50). The Hadamard function given by Eq. (98) and the Wightman function obtained from that are defined in region \mathcal{R} in the multiply connected de Sitter space, but they can be analytically extended to region \mathcal{F} via the transformation in Eq. (69). However, it should be noted that as we make the continuation from \mathcal{R} to \mathcal{F} , $\sqrt{(1-r^2/r_0^2)(1-r'^2/r_0^2)}$ should be continued to be $-\sqrt{(\tilde{t}^2/r_0^2-1)(\tilde{t}'^2/r_0^2-1)}$ instead of $+\sqrt{(\tilde{t}^2/r_0^2-1)(\tilde{t}'^2/r_0^2-1)}$. This is because if we take $\sqrt{1-z^2}=i\sqrt{z^2-1}$, we should also take $\sqrt{1-z'^2}=i\sqrt{z'^2-1}$ (instead of $-i\sqrt{z'^2-1}$) (z and z' should be continued along the same path), thus

$$\begin{aligned} \sqrt{(1-z^2)(1-z'^2)} &= (i\sqrt{z^2-1})(i\sqrt{z'^2-1}) \\ &= -\sqrt{(z^2-1)(z'^2-1)}. \end{aligned}$$

Using similar transformations, the results can also be continued to regions \mathcal{P} and \mathcal{L} (we do not write them out because we do not use them here).

We consider particle detectors moving along three kinds of world lines in our multiply connected de Sitter space:

(1) *A particle detector moving along a world line with $r, \theta, \phi = \text{constants}$ in \mathcal{R} .* In such a case, on the world line of the particle detector, γ is zero and the Hadamard function is reduced to

$$\begin{aligned} G_{\text{CR}}^{(1)}(\tau, \tau') &= -\frac{1}{2\pi^2(1-r^2/r_0^2)} \sum_{n=-\infty}^{\infty} \frac{1}{(t-t'+n\beta)^2} \\ &= -\frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{1}{(\tau-\tau'+n\beta\sqrt{1-r^2/r_0^2})^2}, \end{aligned} \quad (100)$$

where $\tau = t\sqrt{1-r^2/r_0^2}$ is the proper time of the particle detector. The corresponding Wightman function obtained from Eq. (50) is

$$G_{\text{CR}}^+(\tau, \tau') = -\frac{1}{4\pi^2} \sum_{n=-\infty}^{\infty} \frac{1}{(\tau-\tau'+n\beta\sqrt{1-r^2/r_0^2}-i\epsilon)^2}, \quad (101)$$

where ϵ is an infinitesimal positive real number. Inserting it into Eq. (17), obviously the integration over $\Delta\tau = \tau - \tau'$ is zero since all poles of the integrand are in the upper-half plane of complex $\Delta\tau$ while the integration contour is closed in the lower-half plane. Therefore the response function $\mathcal{F}(\Delta E)$ is zero and no particles are detected. All of these world lines are accelerated, except for the one at $r=0$.

(2) *A particle detector moving along a geodesic with $l, \theta, \phi = \text{const}$ in region \mathcal{F} .* In this region the Hadamard function is

$$G_{\text{CR}}^{(1)}(X, X') = -\frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{\tilde{\gamma}}{\sinh \tilde{\gamma} \sqrt{(\tilde{t}^2/r_0^2-1)(\tilde{t}'^2/r_0^2-1)} [- (l-l'+n\beta)^2 + r_0^2 \tilde{\gamma}^2]}, \quad (102)$$

where $\tilde{\gamma}$ is given by

$$\cosh \tilde{\gamma} = \frac{1}{\sqrt{(\tilde{t}^2/r_0^2-1)(\tilde{t}'^2/r_0^2-1)}} \left\{ -1 + \frac{\tilde{t} \tilde{t}'}{r_0^2} [\cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi')] \right\}. \quad (103)$$

[Equation (102) and Eq. (103) are obtained from Eq. (98) and Eq. (99) via the transformation in Eq. (69) respectively.] On the world line of the particle detector, the Hadamard function is reduced to

$$G_{\text{CR}}^{(1)}(\tilde{t}, \tilde{t}') = -\frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{\tilde{\gamma}}{\sinh \tilde{\gamma} \sqrt{(\tilde{t}^2/r_0^2-1)(\tilde{t}'^2/r_0^2-1)} (-n^2\beta^2 + r_0^2 \tilde{\gamma}^2)}, \quad (104)$$

and $\cosh \tilde{\gamma}$ is reduced to

$$\cosh \tilde{\gamma} = \frac{\tilde{t} \tilde{t}'/r_0^2 - 1}{\sqrt{(\tilde{t}^2/r_0^2-1)(\tilde{t}'^2/r_0^2-1)}}. \quad (105)$$

Using the proper time τ defined by Eq. (71), on the world line of the particle detector $\cosh \tilde{\gamma}$ and $G_{\text{CR}}^{(1)}$ can be written as

$$\cosh \tilde{\gamma} = \frac{\cosh 2T + \cosh \Delta\tau - 2}{\cosh 2T - \cosh \Delta\tau}, \quad (106)$$

and

$$G_{\text{CR}}^{(1)}(T, \Delta\tau) = \frac{1}{\pi^2 r_0^2} \sum_{n=-\infty}^{\infty} \frac{\tilde{\gamma}}{\sinh \tilde{\gamma} (\cosh 2T - \cosh \Delta\tau) (n^2 b^2 - \tilde{\gamma}^2)}, \quad (107)$$

where $\tau > 0$, $\tau' > 0$, $\Delta\tau = (\tau - \tau')/r_0$, $T = (\tau + \tau')/2r_0$, and $b = \beta/r_0$. The Wightman function is equal to one half of the Hadamard function with $\Delta\tau$ replaced by $\Delta\tau - i\epsilon$ [Eq. (50)]. Thus the response function is

$$\mathcal{F}(\Delta E) = \sum_{n=-\infty}^{\infty} \mathcal{F}_n(\Delta E), \quad (108)$$

where

$$\mathcal{F}_n(\Delta E) = \frac{1}{2\pi^2} \int_0^\infty dT \int_{-\infty}^\infty d\Delta\tau e^{-i\Delta E r_0 \Delta\tau} \left[\frac{\tilde{\gamma}}{\sinh \tilde{\gamma} (\cosh 2T - \cosh \Delta\tau) (n^2 b^2 - \tilde{\gamma}^2)} \right]_{\Delta\tau \rightarrow \Delta\tau - i\epsilon}. \quad (109)$$

Now we consider the poles in the complex $\Delta\tau$ plane of the integrand in the integral of $\mathcal{F}_n(\Delta E)$. The poles are given by the equation

$$\tilde{\gamma} = \pm nb. \quad (110)$$

(It is easy to check that $\cosh 2T = \cosh \Delta\tau$ does not give any poles.) From Eq. (110) and Eq. (106), we have (we neglect the term $i\epsilon$, and at the end of the calculation we return it back to the expressions)

$$\cosh 2T + \cosh \Delta\tau - 2 = \cosh nb (\cosh 2T - \cosh \Delta\tau). \quad (111)$$

Solutions to Eq. (111) are

$$\Delta\tau = \Delta\tau_n + i2m\pi \equiv \Delta\tau_{nm}, \quad (112)$$

where

$$\begin{aligned} \Delta\tau_n &= \pm \text{Arccosh} \frac{(\cosh nb - 1) \cosh 2T + 2}{\cosh nb + 1} \\ &= \pm 2 \text{Arcsinh} \left(\sinh T \tanh \frac{nb}{2} \right), \end{aligned} \quad (113)$$

where $\text{Arccosh}z$ is the principal value of $\text{arccosh}z$, and here it is real (similarly for $\text{Arcsinh}z$). We need to check if all $\Delta\tau_{nm}$ are roots of Eq. (110), because the number of roots might increase as we go from Eq. (110) to Eq. (111). [For example, for any integer m , $x_m = \pm 2 + im\pi$ solves the equation $\cosh(2x) = \cosh 4$; but, only $x_0 = +2$ solves the equation $2x = 4$.] $\Delta\tau_n$ is obviously a root of Eq. (110). The question is: as $\Delta\tau$ goes from $\Delta\tau_n$ to $\Delta\tau_n + i2m\pi$, does Eq. (106) give the same $\tilde{\gamma}$ which is a real value [$\pm nb$; see Eq. (110)]? (Remember that $\text{arccosh}z$ is a multi-valued complex function.) To answer this question, let $\Delta\tau = \Delta\tau_n + i\theta$ (where θ is real). Then from Eq. (106) we have

$$\begin{aligned} \tilde{\gamma} &= \text{arccosh} \frac{\cosh 2T + \cosh \Delta\tau - 2}{\cosh 2T - \cosh \Delta\tau} \\ &= \ln \frac{\sinh T + \sinh \frac{\Delta\tau}{2}}{\sinh T - \sinh \frac{\Delta\tau}{2}} \\ &= \ln \frac{\sinh T + \sinh \frac{\Delta\tau_n}{2} \cos \frac{\theta}{2} + i \cosh \frac{\Delta\tau_n}{2} \sin \frac{\theta}{2}}{\sinh T - \sinh \frac{\Delta\tau_n}{2} \cos \frac{\theta}{2} - i \cosh \frac{\Delta\tau_n}{2} \sin \frac{\theta}{2}} \\ &\equiv \ln \frac{z_1}{z_2}, \end{aligned} \quad (114)$$

where we have used $\text{arccosh}z = \ln(z + \sqrt{z^2 - 1})$. The real components of z_1 and z_2 are respectively

$$\Re(z_1) = \sinh T + \sinh \frac{\Delta\tau_n}{2} \cos \frac{\theta}{2}, \quad (115)$$

$$\Re(z_2) = \sinh T - \sinh \frac{\Delta\tau_n}{2} \cos \frac{\theta}{2}. \quad (116)$$

By Eq. (113), we find that $\Re(z_1)$ and $\Re(z_2)$ are always positive for any real θ . This means that as $\Delta\tau$ goes from $\Delta\tau_n$ to $\Delta\tau_n + i2m\pi$, the arguments (the argument of a complex number $z = |z|e^{i\alpha}$ is α) of z_1 and z_2 do not change, neither does the argument of z_1/z_2 . The value of $\tilde{\gamma}$ remains in the same branch of $\ln z$ as θ varies. Thus, for all $\Delta\tau_{nm} = \Delta\tau_n + i2m\pi$, we have $\tilde{\gamma} = \pm nb$ and Eq. (110) is satisfied. Therefore all $\Delta\tau_{nm}$ in Eq. (112) are poles.

The residues of the integrand in Eq. (109) at poles $\Delta\tau_{nm}$ are (here $i\epsilon$ is returned to the expressions)

$$\text{Res}(\Delta\tau = i2m\pi + i\epsilon, n=0) = \frac{iEr_0}{4\pi^2} e^{2m\pi\Delta Er_0}, \quad (117)$$

$$\begin{aligned} \text{Res}(\Delta\tau = \Delta\tau_n + i2m\pi + i\epsilon, n \neq 0) \\ = -\frac{1}{4\pi^2} \frac{e^{2m\pi\Delta Er_0 - i\Delta Er_0\Delta\tau_n}}{(\cosh nb + 1)\sinh \Delta\tau_n}. \end{aligned} \quad (118)$$

Then by the residue theorem (the contour for the integral is closed in the lower-half plane of complex $\Delta\tau$) we have

$$\frac{d\mathcal{F}_0}{dT} = \frac{r_0}{2\pi} \frac{\Delta E}{e^{2\pi\Delta Er_0} - 1}, \quad (119)$$

and

$$\begin{aligned} \frac{d\mathcal{F}_{n \neq 0}}{dT} &= \frac{\sin(\Delta Er_0|\Delta\tau_n|)}{\pi(\cosh nb + 1)\sinh|\Delta\tau_n|} \\ &\times \frac{1}{e^{2\pi\Delta Er_0} - 1}. \end{aligned} \quad (120)$$

The $\sin(\Delta Er_0|\Delta\tau_n|)$ factor in Eq. (120) indicates that the $n \neq 0$ terms' contribution can be both positive (absorption by the detector) and negative (emission from the detector). We see that the contribution of the $n=0$ term is just the Hawking radiation with the Gibbons-Hawking temperature $T_{\text{G-H}} = 1/2\pi r_0$ in the simply connected de Sitter space. The contribution of the $n \neq 0$ terms is a kind of ‘‘gray-body’’ Hawking radiation: the temperature is $T_{\text{G-H}}$, but its density or flux decreases as the universe expands ($|\Delta\tau_n|$ increases as the universe expands). The sum of all $n \neq 0$ contributions is

$$\begin{aligned} \sum_{n \neq 0} \frac{d\mathcal{F}_n}{dT} &= \frac{1}{\pi^2} \frac{1}{e^{2\pi r_0 \Delta E} - 1} \\ &\times \sum_{n \neq 0} \frac{\sin(\Delta Er_0|\Delta\tau_n|)}{(\cosh nb + 1)\sinh|\Delta\tau_n|}. \end{aligned} \quad (121)$$

In the case of $b=2\pi$ (the self-consistent case), we have $\cosh nb \approx \exp(|n|b)/2 \gg 1$ ($n \neq 0$) and thus $\Delta\tau_n \approx \pm 2T$. Then

$$\begin{aligned} \sum_{n \neq 0} \frac{d\mathcal{F}_n}{dT} &\approx \frac{1}{2\pi} \frac{A}{e^{2\pi r_0 \Delta E} - 1} \\ &\times \frac{\sin(2\Delta Er_0 T)}{\sinh 2T}, \end{aligned} \quad (122)$$

where $A = 4\sum_{n=1}^{\infty} (\cosh 2n\pi + 1)^{-1} \approx 0.015$. As $T \rightarrow \infty$, the contribution of all $n \neq 0$ terms decreases exponentially to

zero. Thus, at events far from the Cauchy horizon in \mathcal{F} , the particle detector perceives pure Hawking radiation given by the $n=0$ term. As $T \rightarrow 0$ (near the Cauchy horizon), we have

$$\sum_{n \neq 0} \frac{d\mathcal{F}_n}{dT} \approx \frac{Ar_0}{2\pi} \frac{\Delta E}{e^{2\pi r_0 \Delta E} - 1}. \quad (123)$$

This is a ‘‘gray-body’’ Hawking radiation with $A \approx 1.5\%$. Near the Cauchy horizon the total radiation is the sum of a pure Hawking radiation (given by the $n=0$ term) and a ‘‘gray-body’’ Hawking radiation (given by all $n \neq 0$ terms). The total intensity of the radiation near the Cauchy horizon is a factor of $\approx 101.5\%$ that of regular Hawking radiation, but its spectrum is the same as the usual Hawking radiation.

(3) *A particle detector moving along a co-moving world line in the steady-state coordinate system.* Suppose the detector moves along the geodesic $\rho, \theta, \phi = \text{constants}$ (such a world line is a timelike geodesic passing through \mathcal{R} and into \mathcal{F}) where $\rho \equiv (x^2 + y^2 + z^2)^{1/2}$ and the proper time τ are related to the static radius r by

$$r = -r_0 \rho / \bar{\eta} = \rho e^{\tau/r_0}. \quad (124)$$

The Cauchy horizon is at $r=r_0$, or $\rho = -\bar{\eta} = r_0 e^{-\tau/r_0}$. On the world line of the detector the Hadamard function is

$$\begin{aligned} G_{\text{CR}}^{(1)}(T, \Delta\tau) &= \frac{1}{2\pi^2 r_0^2} \frac{\gamma}{2L \sinh \frac{\Delta\tau}{2}} \\ &\times \sum_{n=-\infty}^{\infty} \frac{1}{\gamma^2 - \left(\frac{t-t'}{r_0} + nb\right)^2}, \end{aligned} \quad (125)$$

where $\Delta\tau = (\tau - \tau')/r_0$, $T = (\tau + \tau')/2r_0$, $L = \rho e^T/r_0 \equiv r(T)/r_0$, γ is given by

$$\cosh \gamma = \frac{1 - L^2}{\sqrt{1 + L^4 - 2L^2 \cosh \Delta\tau}}, \quad (126)$$

and $t-t'$ is related to T and $\Delta\tau$ by

$$\cosh \frac{t-t'}{r_0} = \frac{\cosh \Delta\tau - L^2}{\sqrt{1 + L^4 - 2L^2 \cosh \Delta\tau}}. \quad (127)$$

By analytical continuation, Eqs. (125)–(127) hold in the whole region covered by the steady-state coordinates in de Sitter space. The Wightman function G^+ is equal to one half of $G^{(1)}$ with $\Delta\tau$ replaced by $\Delta\tau - i\epsilon$ [Eq. (50)]. The response function is $\mathcal{F}(\Delta E) = \sum_{n=-\infty}^{\infty} \mathcal{F}_n(\Delta E)$ where

$$\mathcal{F}_n(\Delta E) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dT \int_{-\infty}^{\infty} d\Delta\tau e^{-i\Delta Er_0 \Delta\tau} \left\{ \frac{\gamma}{2L \sinh \frac{\Delta\tau}{2} \left[\gamma^2 - \left(\frac{t-t'}{r_0} + nb\right)^2 \right]} \right\}_{\Delta\tau \rightarrow \Delta\tau - i\epsilon}. \quad (128)$$

The poles of the integrand in the complex- $\Delta\tau$ plane are given by

$$\frac{t-t'}{r_0} + nb = \pm \gamma. \quad (129)$$

This together with Eq. (126) and Eq. (127) leads to

$$(\cosh \Delta\tau - L^2) \cosh nb + \sinh \Delta\tau \sinh nb = 1 - L^2. \quad (130)$$

The roots of Eq. (130) are

$$\Delta\tau = \Delta\tau_n^\pm + i2m\pi \equiv \Delta\tau_{nm}^\pm, \quad (131)$$

where

$$\Delta\tau_n^\pm = \ln(1 + 2\mu^2 \pm 2\mu\sqrt{1 + \mu^2}) - nb, \quad (132)$$

where $\mu \equiv \sinh(nb/2)$. By carefully checking $\Delta\tau_{nm}^\pm$ in Eq. (131), as we did in case (2), we find that: (1) For $L < 1$ (or $\rho e^T < r_0$, i.e., in region \mathcal{R}), only $\Delta\tau_{n0}^\pm = \Delta\tau_n^\pm$ solve Eq. (129); (2) for $L > 1$ (or $\rho e^T > r_0$, i.e., in region \mathcal{F}), only

$\Delta\tau_{nm}^+ = \Delta\tau_n^+ + i2m\pi$ solve Eq. (129). (Here it is assumed that $b > \ln 2$ and the self-consistent case with $b = 2\pi$ obviously satisfies this condition.) All other $\Delta\tau$'s in Eq. (131) are not roots of Eq. (129), though they solve Eq. (130). Therefore the poles are (where $i\epsilon$ is returned)

$$\Delta\tau = \begin{cases} \Delta\tau_n^+ + i\epsilon, & \text{in } \mathcal{R}; \\ \Delta\tau_n^+ + i2m\pi + i\epsilon, & \text{in } \mathcal{F}. \end{cases} \quad (133)$$

Obviously in region \mathcal{R} all poles are in the upper-half plane of complex $\Delta\tau$. Therefore

$$\frac{d\mathcal{F}}{dT} = 0, \quad (134)$$

when the particle detector is in region \mathcal{R} . So the particle detector sees nothing while it is in region \mathcal{R} .

In region \mathcal{F} , only the poles with $m < 0$ are in the lower-half plane of complex $\Delta\tau$. The residues of the integrand at poles $\Delta\tau_{nm}^+ + i\epsilon$ are

$$\text{Res}(\Delta\tau = i2m\pi + i\epsilon, n=0) = \frac{ir_0\Delta E}{4\pi^2} e^{2m\pi\Delta E r_0}, \quad (135)$$

$$\text{Res}(\Delta\tau = \Delta\tau_n^+ + i2m\pi + i\epsilon, n \neq 0) = \frac{1}{16\pi^2 L \sinh \frac{\Delta\tau_n^+}{2}} \frac{\alpha_1(1 + L^4 - 2L^2 \cosh \Delta\tau_n^+) e^{2m\pi\Delta E r_0 - i\Delta E r_0 \Delta\tau_n^+}}{-\alpha_1 L(L^2 - 1) \cosh \frac{\Delta\tau_n^+}{2} + (\alpha_2 + nb)(L^2 \cosh \Delta\tau_n^+ - 1)}, \quad (136)$$

where $\alpha_1 = \text{arccosh}[(L^2 - 1)/\sqrt{1 + L^4 - 2L^2 \cosh \Delta\tau_n^+}]$ and $\alpha_2 = \text{arccosh}[(L^2 - \cosh \Delta\tau_n^+)/\sqrt{1 + L^4 - 2L^2 \cosh \Delta\tau_n^+}]$. By the residue theorem, we have that $d\mathcal{F}_0/dT$ has the same value as that in Eq. (119), which represents Hawking radiation with the Gibbons-Hawking temperature; the contribution of all $n \neq 0$ terms (note that $\Delta\tau_n^+ = -\Delta\tau_{-n}^+$) is

$$\frac{d}{dT} \sum_{n \neq 0} \mathcal{F}_n = \frac{1}{4\pi^2 (e^{2\pi r_0 \Delta E} - 1)} \sum_{n=1}^{\infty} \frac{\sin(\Delta E r_0 \Delta\tau_n^+)}{L \sinh \frac{\Delta\tau_n^+}{2}} \frac{\alpha_1(1 + L^4 - 2L^2 \cosh \Delta\tau_n^+)}{\alpha_1 L(L^2 - 1) \cosh \frac{\Delta\tau_n^+}{2} - (\alpha_2 + nb)(L^2 \cosh \Delta\tau_n^+ - 1)}, \quad (137)$$

which represents a ‘‘gray-body’’ Hawking radiation. As $T \rightarrow \infty$ (or $L \rightarrow \infty$), $(d/dT) \sum_{n \neq 0} \mathcal{F}_n$ exponentially drops to zero; therefore, at events far from the Cauchy horizon in \mathcal{F} , the particle detector only perceives pure Hawking radiation [the same as that in case (2)]. As $L \rightarrow 1$ (approaching the Cauchy horizon), we also have $(d/dT) \sum_{n \neq 0} \mathcal{F}_n \rightarrow 0$. Thus near the Cauchy horizon, on the region \mathcal{F} side, the particle detector comoving in the steady-state coordinate system perceives pure Hawking radiation with Gibbons-Hawking temperature.

From the above discussion, we find that in our multiply connected de Sitter space with the adapted Rindler vacuum, region \mathcal{R} is cold (where the temperature is zero) but region \mathcal{F} is hot (where the temperature is $T_{\text{G-H}}$). Similarly, region \mathcal{L} is cold but \mathcal{P} is hot, the above results can be easily extended to

these regions. This gives rise to an arrow of increasing entropy, from a cold region to a hot region (Fig. 5).

E. Classical stability of the Cauchy horizon and the arrow of time

In classical electromagnetic theory, it is well known that both the retarded potential ϕ_{ret} and the advanced potential ϕ_{adv} (and any part-retarded-and-part-advanced potential $a\phi_{\text{ret}} + b\phi_{\text{adv}}$ with $a + b = 1$) are solutions of Maxwell’s equations. But from our experience, we know that all the electromagnetic perturbations we see are propagated only by the *retarded* potential. (For example, if at some time and some place, a light signal is emitted, it can only be received by a receiver at another place sometime *later*.) This indicates that there is an *arrow of time* in the solutions of Maxwell’s

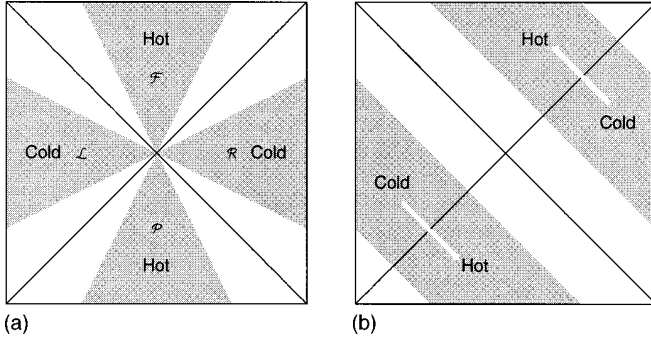


FIG. 5. With our adapted conformal Rindler vacuum, our multiply connected de Sitter space is cold (with zero temperature) in \mathcal{R} and \mathcal{L} , but hot (with the Gibbons-Hawking temperature) in \mathcal{F} and \mathcal{P} . The arrows indicate the direction of increasing entropy.

equations, though Maxwell's equations themselves are time-symmetric. This arrow of time is sometimes called the electromagnetic arrow of time, or the causal arrow of time. How this arrow of time arises is a mystery. Many people have tried to solve this problem by attributing it to a boundary condition of the Universe [60,173–175] (for review of the arrows of time, see [176,177]). In this subsection we argue that the principle of self-consistency [178,179] naturally gives rise to an arrow of time in our multiply connected de Sitter space.

First let us consider the arrow of time in Misner space. Suppose at an event E in region F in Misner space [by boost and translation, assume we have moved E to $(t=t_0, x=0, y=0, z=0)$], a spherical pulse of electromagnetic wave is created. If the potential is retarded [here ‘retarded’ and ‘advanced’ are defined relative to the direction of $(\partial/\partial t)^a$ (t is the time coordinate in the global Cartesian coordinates of the covering space—Minkowski space)], the pulse will propagate in the future direction as a light cone originating from E . At any point on the light cone, the energy-momentum tensor of the wave is

$$T^{ab} = \mu k^a k^b, \quad (138)$$

where $\mu \equiv \mu(t)$ is a scalar function and $k^a = k^0(\partial/\partial t)^a + k^1(\partial/\partial x)^a + k^2(\partial/\partial y)^a + k^3(\partial/\partial z)^a$ is a null vector tangent to the light cone, and the energy density measured by an observer with four-velocity vector $(\partial/\partial t)^a$ (whose ordinary three-velocity is zero) is

$$\rho = T_{ab} \left(\frac{\partial}{\partial t} \right)^a \left(\frac{\partial}{\partial t} \right)^b = \mu (k^0)^2. \quad (139)$$

(Thus μ measures the energy density of the electromagnetic wave.) By Einstein's equations, the back-reaction of T_{ab} on R and $R_{ab}R^{ab}$ (where R_{ab} is the Ricci tensor and $R = R_a^a$ is the Ricci scalar curvature) is $\delta R \sim T_a^a$, $\delta(R_{ab}R^{ab}) \sim T_{ab}T^{ab}$. The Riemann tensor can be decomposed as $R_{abcd} = C_{abcd} + Q_{abcd}$, where C_{abcd} is the Weyl tensor and Q_{abcd} is constructed entirely from the Ricci tensor

$$Q_{abcd} = g_{a[c}R_{d]b} - g_{b[c}R_{d]a} - \frac{1}{3}Rg_{a[c}g_{d]b}, \quad (140)$$

where square brackets denote antisymmetrization [180]. The Weyl tensor describes the part of the curvature that is due to pure gravitational field, whereas the Ricci tensor describes the part that, according to Einstein's equations, is directly due to the energy-momentum tensor of matter [59]. Therefore, in some sense, the values of T_a^a and $T_{ab}T^{ab}$ determine the influence of matter fields on the stability of the background spacetime. An infinite T_a^a or $T_{ab}T^{ab}$ implies that the spacetime is unstable against this perturbation and a singularity may form; on the other hand, if T_a^a and $T_{ab}T^{ab}$ are finite, the spacetime may be stable against this perturbation. Self-consistent solutions should require that T_a^a and $T_{ab}T^{ab}$ do not blow up. If they did, the starting geometry—on the basis of which T_a^a and $T_{ab}T^{ab}$ were calculated—would be greatly perturbed and the T_a^a and $T_{ab}T^{ab}$ calculation itself would be invalid, and thus it would not be a self-consistent solution. For electromagnetic fields we always have $T_a^a = 0$, so we need only consider $T_{ab}T^{ab}$. For T_{ab} in Eq. (138), we also have

$$T^{ab}T_{ab} = 0. \quad (141)$$

Thus significant perturbations (indicated by a non-vanishing $T_{ab}T^{ab}$) can only occur when the light cone ‘collides’ with its images under the boost transformation. At any point p on the intersection of the light cone L and its n th image L_n (suppose $n > 0$), the energy-momentum tensor is

$$T^{ab} = \mu k^a k^b + \tilde{\mu} \tilde{k}^a \tilde{k}^b, \quad (142)$$

where k^a is the null vector tangent to the light cone L at p , \tilde{k}^a is the null vector tangent to the light cone L_n at p ; μ measures the energy density in light cone L , $\tilde{\mu}$ measures the energy density in light cone L_n . From Eq. (142) we have

$$T^{ab}T_{ab} = [2\mu\tilde{\mu}(k^a\tilde{k}_a)^2]_p, \quad (143)$$

the index p denotes that the quantity is evaluated at the point p .

Since the point p on L_n is obtained from some point p' on L by boost transformation, p and p' must have the same timelike separation from the origin ($t=0, x=0, y=0, z=0$) (remember that p is on the intersection of L and L_n ; see Fig. 6a). If we take the \tilde{k}^a at p being transported from the k'^a at p' , we have $\tilde{\mu}_{p \in L_n} = \mu_{p' \in L}$. Because the light cone L is spherically symmetric, we have $t_p = t_{p'}$. Therefore we have $\mu_{p' \in L} = \mu_{p \in L}$ and at p we have $\tilde{\mu} = \mu$. Under the boost transformation B , we have

$$\begin{aligned} (\tilde{k}^a)_p &= B[(k'^a)_{p'}] \\ &= k'^0 \left[\cosh nb \left(\frac{\partial}{\partial t} \right)^a + \sinh nb \left(\frac{\partial}{\partial x} \right)^a \right] \\ &\quad + k'^1 \left[\cosh nb \left(\frac{\partial}{\partial x} \right)^a + \sinh nb \left(\frac{\partial}{\partial t} \right)^a \right] \\ &\quad + k'^2 \left(\frac{\partial}{\partial y} \right)^a + k'^3 \left(\frac{\partial}{\partial z} \right)^a, \end{aligned} \quad (144)$$

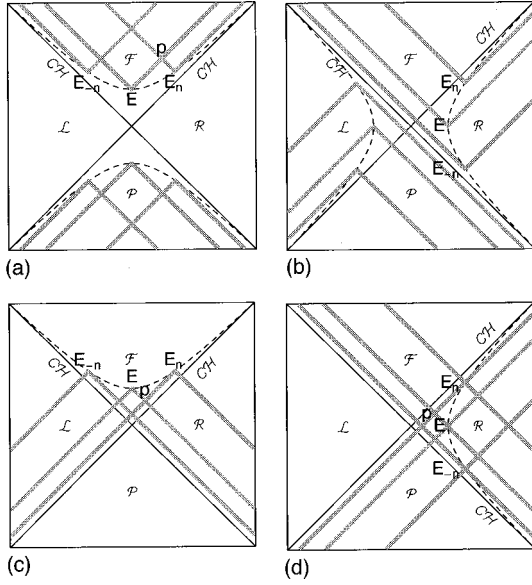


FIG. 6. Self-consistency near the Cauchy horizons in a space-time with CTCs naturally gives rise to an arrow of time. Grey thick lines represent light cones of electromagnetic waves or photons emitted from event E. (a) This diagram shows that in \mathcal{F} the retarded potential is self-consistent. The “collision” of an electromagnetic wave with its images cannot destroy the Cauchy horizon, since the proper time from the “collision” (event p) to the origin is always bigger than the proper time from E to the origin. Likewise the advanced potential in region \mathcal{P} would not destroy the Cauchy horizon. (b) This diagram shows that a retarded potential in \mathcal{R} and an advanced potential in \mathcal{L} (or *vice versa*) are self-consistent. But the potentials in \mathcal{R} and \mathcal{L} cannot be both retarded or both advanced, otherwise the “collision” of two waves from \mathcal{R} and \mathcal{L} respectively will destroy the Cauchy horizon. (c) This diagram shows that the advanced potential in \mathcal{F} is not self-consistent, since the collision of an electromagnetic wave with its images will destroy the Cauchy horizon. (As $n \rightarrow \pm\infty$, the collision event p approaches the Cauchy horizon.) (d) This diagram shows that a part-advanced-and-part-retarded potential in \mathcal{R} (or \mathcal{L}) is also not self-consistent, as $T^{ab}T_{ab}$ would also diverge as the Cauchy horizon is approached.

where $(k'^a)_{p'} = k'^0(\partial/\partial t)^a + k'^1(\partial/\partial x)^a + k'^2(\partial/\partial y)^a + k'^3(\partial/\partial z)^a$. Due to the spherical symmetry, we have $k'^0 = k^0$. Define (r, θ, ϕ) by $x = r \cos \theta$, $y = r \sin \theta \cos \phi$, and $z = r \sin \theta \sin \phi$. Then we have $r' = r$, $\theta' = \pi - \theta$, $\phi' = \phi$ (“’” means “at p' ”), and

$$\begin{aligned} k^1 &= k^0 \cos \theta, & k^2 &= k^0 \sin \theta \cos \phi, \\ k^3 &= k^0 \sin \theta \sin \phi, \end{aligned} \quad (145)$$

and

$$\begin{aligned} k'^1 &= k'^0 \cos \theta' = -k^0 \cos \theta = -k^1, \\ k'^2 &= k'^0 \sin \theta' \cos \phi' = k^0 \sin \theta \cos \phi = k^2, \\ k'^3 &= k'^0 \sin \theta' \sin \phi' = k^0 \sin \theta \sin \phi = k^3. \end{aligned} \quad (146)$$

Then

$$\begin{aligned} (k^a \tilde{k}_a)_p &= (k^0)^2 [-(1 + \cos^2 \theta) \cosh nb \\ &\quad + 2 \cos \theta \sinh nb + \sin^2 \theta], \end{aligned} \quad (147)$$

and

$$\begin{aligned} T^{ab}T_{ab} &= 2\rho^2(t_p) [-(1 + \cos^2 \theta) \cosh nb \\ &\quad + 2 \cos \theta \sinh nb + \sin^2 \theta]^2. \end{aligned} \quad (148)$$

It is easy to find that $T^{ab}T_{ab}$ reaches a maximum at $\theta=0$ and

$$(T^{ab}T_{ab})_{\max} = 8\rho^2(t_p)e^{-2nb}, \quad (149)$$

where $\rho(t_p)$ is the energy density from L as measured in a frame at event p with ordinary velocity $v_x = v_y = v_z = 0$. $(T^{ab}T_{ab})_{\max}$ is always finite [less than $8\rho^2(t_p)$] since n is positive. If $n < 0$ we have $(T^{ab}T_{ab})_{\max} = 8\rho^2(t_p)e^{2nb} < 8\rho^2(t_p)$. So if we have a retarded potential in region \mathcal{F} , even considering the infinite number of images, $T_{ab}T^{ab}$ is always finite.

If the potential is advanced however, the pulse wave will propagate backward in the past direction as a light cone originating from E. And, within a finite time, it will hit the Cauchy horizon. By an analysis similar to the above arguments, we find that in this case

$$\begin{aligned} T^{ab}T_{ab} &= 2\rho^2(t_p) [(1 + \cos^2 \theta) \cosh nb \\ &\quad + 2 \cos \theta \sinh nb - \sin^2 \theta]^2, \end{aligned} \quad (150)$$

which reaches a maximum at $\theta=0$ and

$$(T^{ab}T_{ab})_{\max} = 8\rho^2(t_p)e^{2|n|b}. \quad (151)$$

Since $\rho(t_p)$ is finite (the past light cone from E at $\theta=0$ hits the Cauchy horizon in a finite affine distance), thus $(T^{ab}T_{ab})_{\max} \rightarrow \infty$ as $n \rightarrow \pm\infty$. As $n \rightarrow \pm\infty$, L and L_n collide at the Cauchy horizon [as $n \rightarrow \pm\infty$ the point $p(\theta=0)$ approaches the Cauchy horizon] (see Fig. 6c). Thus $(T^{ab}T_{ab})_{\max}$ diverges as the Cauchy horizon is approached and the Cauchy horizon may be destroyed. Therefore the advanced potential is *not* self-consistent in region \mathcal{F} of Misner space. It is easy to see that any part-retarded-and-part-advanced potential is also *not* self-consistent in \mathcal{F} . The *only* self-consistent potential in region \mathcal{F} is the *retarded* potential.

Similarly, in region \mathcal{P} the only self-consistent potential is the *advanced* potential (see Fig. 6a). [Note that here “advanced” and “retarded” are defined relative to the global time direction in Minkowski spacetime (the covering space). An observer in \mathcal{P} will regard it as “retarded” relative to his own time direction.]

In region \mathcal{R} , by boost and translation, we can always move the event E (where a spherical pulse of electromagnetic waves is emitted) to $(t=0, x=x_0, y=0, z=0)$. Either pure retarded or pure advanced potentials are self-consistent in this region because the light cone never “collides” with the images of itself and thus we always have $T^{ab}T_{ab}=0$ (see Fig. 6b). But, for a part-retarded-and-part-advanced potential, the retarded light cone (L^+) propagates forward while the advanced light cone (L^-) propagates backward, both originating from E. The forward part of the light cone will

collide with images of the backward part of the light cone and *vice versa* (see Fig. 6d). We find that at a point p on the intersection of L^+ and L_n^- (or L^- and L_n^+)

$$T^{ab}T_{ab} = 2\rho(t)\rho(-t)[(1 + \cos^2 \theta)\cosh nb - 2 \cos \theta \sinh nb + \sin^2 \theta]^2, \quad (152)$$

where $\rho(t)$ is the energy density from L^+ observed in a frame on L^+ with time coordinate t and with ordinary velocity $v_x = v_y = v_z = 0$ and $\rho(-t)$ is the energy density from L^- seen in a frame on L^- with time coordinate $-t$ and with ordinary velocity $v_x = v_y = v_z = 0$. $T^{ab}T_{ab}$ reaches a maximum at $\theta = \pi$, and

$$(T^{ab}T_{ab})_{\max} = 8\rho(t)\rho(-t)e^{2|n|b}, \quad (153)$$

where t is the global time coordinate in the covering Minkowski space. As p approaches the Cauchy horizon, where $n \rightarrow \pm\infty$, $\rho(t)$ and $\rho(-t)$ are both finite, since in the $\theta = \pi$ direction the future and past light cones of E both hit the Cauchy horizon in a finite affine distance. Thus $(T^{ab}T_{ab})_{\max} \rightarrow \infty$ as p approaches the Cauchy horizon (where $n \rightarrow \pm\infty$). Therefore in region R both the retarded and the advanced potential are self-consistent, but the part-retarded-and-part-advanced potential is *not* self-consistent. This conclusion also holds for region L. Furthermore, there must be a correlation between time arrows in region L and region R: if we choose the retarded potential in R, we must choose the advanced potential in L (see Fig. 6b); if we choose the advanced potential in R, we must choose the retarded potential in L. Otherwise the collision of light cones from R and light cones from L will destroy the Cauchy horizon.

As another treatment for perturbations in Misner space, consider that at an event E in region F two photons are created [181] [we choose E to be at $(t = t_0, x = 0, y = 0, z = 0)$ as before]. One photon runs to the right along the $+x$ direction, the other photon runs to the left along the $-x$ direction. They have the same frequency (thus the same energy). The tangent vector of the null geodesic of the right-moving photon is chosen to be ${}^r k^a = (q/v)(\partial/\partial u)^a \equiv (\partial/\partial \lambda_r)^a$, where λ_r is an affine parameter of the geodesic, q is a constant and $u = t + x, v = t - x$. The tangent vector of the null geodesic of the left-moving photon is chosen to be ${}^l k^a = (q/u)(\partial/\partial v)^a \equiv (\partial/\partial \lambda_l)^a$, where λ_l is an affine parameter of that geodesic. The null vectors ${}^r k^a$ and ${}^l k^a$ are invariant under boost transformations. At any point where a photon with null wave-vector k^a is passing by, the frequency of the photon measured in a frame of reference passing by the same point with the four-velocity v^a is $\omega = -k^a v_a$. If $v^a = (\partial/\partial t)^a$ (i.e., the frame of reference has ordinary three-velocity $v_x = v_y = v_z = 0$) and $k^a = {}^r k^a$ or ${}^l k^a$, we have $\omega_r = \omega_l = q/2t_0 \equiv \omega_0$ (thus q measures the frequency of the photon). At any point where the n th image of the right-moving (left-moving) photon is passing by, using the boost transformation we can always find a frame of reference in which the frequency of the photon is ω_0 . But at a point p where the right-moving (left-moving) photon passes the n th image of the left-moving (right-moving) photon, we cannot find a frame of reference such that the two ‘‘colliding’’ photons both have frequency

ω_0 . In such a case we should analyze it in the center-of-momentum frame. The four-velocity of the center-of-momentum frame is $v^a = \gamma({}^r k^a + {}^l k^a)$ where $\gamma^2 = -[({}^r k^a + {}^l k^a)({}^r k_a + {}^l k_a)]^{-1} = uv/q^2 = \tilde{\eta}^2/q^2$ where $\tilde{\eta} = (t^2 - x^2)^{1/2}$ is the proper time separation of p from the origin ($t = 0, x = 0, y = 0, z = 0$). Therefore the total energy of the two oppositely directed photons in the center-of-momentum frame is

$$\mathcal{E} = \omega_1 + \omega_2 = -{}^r k^a v_a - {}^l k^a v_a = \frac{1}{\gamma} = \frac{2t_0}{\tilde{\eta}} \omega_0. \quad (154)$$

(For all other frames the total energy would be greater.) If the potential is retarded, so photons move in the future direction, all points where photons and their images ‘‘collide’’ are in the future of the hypersurface $t^2 - x^2 = t_0^2$. Therefore we have $\tilde{\eta} \geq \tilde{\eta}_0 = t_0$ and $\mathcal{E} \leq 2\omega_0$, so the total energy in the center-of-momentum frame is always bounded. But, if the potential is advanced, photons move in the past direction; thus all points where photons and oppositely directed image photons ‘‘collide’’ are in the past of the hypersurface $t^2 - x^2 = t_0^2$. In particular, the right-moving (left-moving) photon collides with the ∞ -th ($-\infty$ -th) image of the left-moving (right-moving) photon at the Cauchy horizon, where $\tilde{\eta} = 0$ and thus $\mathcal{E} \rightarrow \infty$. Thus, the Cauchy horizon may be destroyed by these photon pairs. Therefore in agreement with our earlier argument, the advanced potential is *not* self-consistent in region F. The *retarded* potential is self-consistent in region F. Similarly, the *advanced* potential is self-consistent in region P. In region R and region L, both the retarded potential and the advanced potential are self-consistent, because the photons and their images will not collide with each other and at any point a photon is passing by we can always find a frame for whom the frequency of this photon is ω_0 . And, the potentials in region R and region L must be correlated in the following way: If the potential in R is retarded, the potential in L must be advanced; if the potential in R is advanced, the potential in L must be retarded (we would call them ‘‘anti-correlated’’). Otherwise the photons from L and photons from R passing in opposite directions would be measured to have infinite energy in center-of-momentum frames as the Cauchy horizon is approached and this may similarly destroy the Cauchy horizon. These conclusions are consistent with those obtained from the analysis of the perturbation of a pulse wave discussed above.

Our multiply connected de Sitter space is conformally related to Misner space via Eqs. (78)–(83). Because light cones and chronological relations are conformally invariant [182] (thus regions $\mathcal{F}, \mathcal{P}, \mathcal{R}$, and \mathcal{L} in multiply connected de Sitter space correspond respectively to regions F, P, R, and L in Misner space under the conformal map, as discussed in Sec. IX B), Maxwell’s equations are also conformally invariant [180,182], so it is easy to generalize the results from Misner space to our multiply connected de Sitter space. Under the conformal transformation $g_{ab} \rightarrow \Omega^2 g_{ab}$, the energy-momentum tensor of the electromagnetic field is transformed as $T_a^b \rightarrow \Omega^{-4} T_a^b$ [180]. Thus $T^{ab}T_{ab}$ is transformed as $T^{ab}T_{ab} \rightarrow \Omega^{-8} T^{ab}T_{ab}$. From the above discussion of $T^{ab}T_{ab}$

in Misner space, we know that $T^{ab}T_{ab}$ is zero everywhere except at the intersection of two light cones. Thus, in multiply connected de Sitter space, $T^{ab}T_{ab}$ is also zero everywhere except at the intersection of two light cones. At the intersection of two light cones in multiply connected de Sitter space, it is easy to show that the maximum value of $T^{ab}T_{ab}$ is at the points with $\theta=0$ or $\theta=\pi$ on the intersection. From Eq. (80) and Eq. (82) we find that for $\theta=0$ or $\theta=\pi$, Ω^2 is non-zero except at the points with $\theta=0$ on the Cauchy horizon (where $r=r_0$ or $\tilde{t}=r_0$). Also because Ω^2 is finite everywhere on the Cauchy horizon (i.e. it is never infinite), we have that: (1) if $T^{ab}T_{ab}$ diverges on the Cauchy horizon in Misner space, the corresponding $T^{ab}T_{ab}$ also diverges on the Cauchy horizon in our multiply connected de Sitter space; (2) if $T^{ab}T_{ab}$ is finite in some region (except at the Cauchy horizon) in Misner space, the corresponding $T^{ab}T_{ab}$ is also finite in the corresponding region (not at the Cauchy horizon) in the multiply connected de Sitter space; (3) if $T^{ab}T_{ab}$ is zero in some region (not a single point) in Misner space, the corresponding $T^{ab}T_{ab}$ is also zero in the corresponding region in the multiply connected de Sitter space. Under the conformal transformation $g_{ab}\rightarrow\Omega^2g_{ab}$, the affine parameter of a null geodesic is transformed as $\lambda\rightarrow\tilde{\lambda}:d\tilde{\lambda}/d\lambda=C\Omega^2$ where C is a constant [180] and thus the null vector $k^a=(\partial/\partial\lambda)^a$ is transformed as $k^a\rightarrow C^{-1}\Omega^{-2}k^a$. Then $\gamma=[-(\cdot k^a + \cdot k^a)(\cdot k_a + \cdot k_a)]^{-1/2}$ is transformed as $\gamma\rightarrow C\Omega\gamma$ and the total energy of the photon pairs in the center-of-momentum frame is transformed as $\mathcal{E}\rightarrow C^{-1}\Omega^{-1}\mathcal{E}$ and the constant C^{-1} can be absorbed into ω_0 . Therefore, we can transplant the above results for Misner space directly to our multiply connected de Sitter space: *In region \mathcal{F} the only self-consistent potential is the retarded potential; in region \mathcal{P} the only self-consistent potential is the advanced potential; in regions \mathcal{R} and \mathcal{L} both the retarded potential and the advanced potential are self-consistent, but they must be anti-correlated* (Fig. 6).

The Cauchy horizon [40] separating a region with CTCs from that without closed causal curves is also called a chronology horizon [9]. A chronology horizon is called a *future* chronology horizon if the region with CTCs lies to the future of the region without closed causal curves; a chronology horizon is called a *past* chronology horizon if the region with CTCs is in the past of the region without closed causal curves. It is generally believed that a future chronology horizon is classically unstable unless there is some diverging effect near the horizon [9,30]. The argument says that a wave packet propagating in the future direction in this spacetime will pile up on the future chronology horizon and destroy the horizon due to the effect of the infinite blue-shift of the frequency (and thus the energy) seen by a timelike observer near a closed null geodesic on the horizon [181,9]. But if there is some diverging mechanism (like the diverging effect of a wormhole in a spacetime with CTCs constructed from a wormhole [7]) near the horizon, the amplitude of the wave packet will decrease with time due to this mechanism, and this may cancel the effect of the blue-shift of the frequency, making the energy finite and thus rendering the future chronology horizon classically stable. Unfortunately, in our mul-

tiple connected de Sitter spacetime (as also in Misner space) there is no such diverging mechanism. A light ray propagating in de Sitter space will focus rather than diverge. This can be seen from the focusing equation [183]

$$\frac{d^2\mathcal{A}^{1/2}}{d\lambda^2} = -\left(\sigma^2 + \frac{1}{2}R_{ab}k^ak^b\right)\mathcal{A}^{1/2}, \quad (155)$$

where \mathcal{A} is the cross-sectional area of the bundle of rays, λ is the affine parameter along the central ray, the null vector k^a is $k^a=(\partial/\partial\lambda)^a$, and σ is the magnitude of the shear of the rays. For de Sitter space we have $R_{ab}k^ak^b=\Lambda g_{ab}k^ak^b=0$ and thus we have $d^2\mathcal{A}^{1/2}/d\lambda^2\leq 0$, so the ray will never diverge. (In fact this always holds if the spacetime satisfies either the weak energy condition or the strong energy condition and it is called the focusing theorem [183].) Hawking [20] has given a general proof along the above lines that any *future* chronology horizon is classically unstable unless light rays are diverging when they propagate near the chronology horizon. You could cause this instability by shaking an electron in the vicinity of the future chronology horizon. The retarded wave would then propagate to the future causing the instability.

However, in Hawking's proof [20], if we replace a future chronology horizon with a *past* chronology horizon, then the proof breaks down because, in such a case, a wave packet propagating toward the future near the past chronology horizon will suffer a red-shift instead of a blue-shift. Therefore a *past* chronology horizon, according to Hawking's argument, is classically stable in a world with retarded potentials. If the universe started with a region of CTCs, but there are no CTCs now, that early region of CTCs would be bounded to the future by a past chronology horizon, and that horizon would be classically stable in a world with retarded potentials—which is what we want. In our multiply connected de Sitter space, this is realized, since the arrow of time in region \mathcal{F} is in the future direction and the arrow of time in region \mathcal{P} is in the past direction [here “future” and “past” are defined globally by the direction of $(\partial/\partial\tau)^a$, where τ is the time coordinate in the global coordinate system $(\tau, \chi, \theta, \phi)$ of the de Sitter covering space]. \mathcal{F} and \mathcal{R} can have retarded potentials, while \mathcal{P} and \mathcal{L} have advanced potentials, as we have noted. In this case the Cauchy horizons separating \mathcal{F} from \mathcal{R} and \mathcal{P} from \mathcal{L} are classically stable, as indicated by our detailed study of $T^{ab}T_{ab}$ as these Cauchy horizons are approached. What about the Cauchy horizons separating \mathcal{P} from \mathcal{R} and \mathcal{F} from \mathcal{L} ? In region \mathcal{P} , the potentials are advanced, so Hawking's instability does not arise as one approaches the Cauchy horizon separating it from \mathcal{R} . In region \mathcal{R} , the potentials are retarded, so by Hawking's argument, one might think that there would be an instability as the Cauchy horizon separating \mathcal{R} from \mathcal{P} is approached from the \mathcal{R} side. But, as we have shown, with retarded potentials in \mathcal{R} , $T^{ab}T_{ab}$ does not diverge as the Cauchy horizon separating \mathcal{R} from \mathcal{P} is approached from the \mathcal{R} side, indicating no instability. Why? Because one can always find frames where the passing photon energies are bounded as the Cauchy horizon is approached. Hawking's argument works

only if one can pick a particular frame like the frame of a timelike observer crossing the Cauchy horizon and observe the blow up of the energy in that frame. (Thus Hawking's approach is observer-dependent, while our approach with $T^{ab}T_{ab}$ is observer-independent.) Hawking's timelike observer would be killed by these photons. But, as we have shown, \mathcal{R} is in a pure vacuum state in our model, so there are no timelike observers in this region, and no preferred frame. If there were timelike particles of positive mass crossing from \mathcal{P} to \mathcal{R} through the Cauchy horizon, we have shown (Li and Gott [30]) that these would cause a classical instability; but there are none. There are, as we shall show in the next subsection, no real particles in regions \mathcal{L} and \mathcal{R} (because these are vacuum states) and no real particles in region \mathcal{F} and \mathcal{P} until the vacuum state there decays by forming bubbles at a timelike separation $|\tau| > \tau_0$ from the origin (τ_0 will be given in the next subsection). Thus, there are no particles crossing the Cauchy horizons separating \mathcal{P} from \mathcal{R} and \mathcal{F} from \mathcal{L} . Thus, there is no instability caused by particles crossing the Cauchy horizons; and since there are no timelike observers in region \mathcal{R} to be hit by photons as the Cauchy horizon separating \mathcal{R} from \mathcal{P} is approached, there is no instability, as indicated by the fact $T^{ab}T_{ab}$ does not blow up as that Cauchy horizon is approached. As indicated in Fig. 4b, region $\mathcal{F}+\mathcal{R}$ is one causally connected region which can be pictured as partially bounded to the future by the future light cone of an event E' and bounded to the past by the future light cone of an event E ; but E and E' are identified by the action of the boost, so these two light cones are identified, creating a periodic boundary condition for region $\mathcal{F}+\mathcal{R}$. As our treatment using $T_{ab}T^{ab}$ with images indicates, retarded photons created in $\mathcal{F}+\mathcal{R}$ cause no instability. Particles with timelike world lines crossing the Cauchy horizons separating $\mathcal{F}+\mathcal{R}$ from $\mathcal{P}+\mathcal{L}$ would cause instability by crossing an infinite number of times between the future light cones of E and E' , thus making an infinite number of passages through the region $\mathcal{F}+\mathcal{R}$ (also $\mathcal{P}+\mathcal{L}$) shown in Fig. 4b. However, as we have shown, there should be no such particles with timelike world lines crossing the Cauchy horizons separating $\mathcal{F}+\mathcal{R}$ from $\mathcal{P}+\mathcal{L}$, and no photons crossing these horizons either, since the potentials in $\mathcal{F}+\mathcal{R}$ are retarded, while the potentials in $\mathcal{P}+\mathcal{L}$ are advanced. Thus, we expect $\mathcal{F}+\mathcal{R}$ and $\mathcal{P}+\mathcal{L}$ to both be stable, and causally disconnected from each other. (See further discussion in the next subsection.)

Thus, the principle of self-consistency [178,179] produces classical stability of the Cauchy horizons and naturally gives rise to an arrow of time in our model of the Universe.

F. Bubble formation in the multiply connected de Sitter space

From the above discussion we find that in the multiply connected de Sitter space region \mathcal{F} and region \mathcal{P} are causally independent in physics: the self-consistent potential in \mathcal{F} is the retarded potential, while the self-consistent potential in \mathcal{P} is the advanced potential, thus an event in \mathcal{F} can never influence an event in \mathcal{P} , and *vice versa*. \mathcal{F} and \mathcal{P} are physically disconnected though they are mathematically connected. If we choose the potential in \mathcal{R} to be retarded, then the potential in \mathcal{L} must be advanced. (Note that here "advanced" and

"retarded" are defined relative to the global time direction in de Sitter space—the covering space of our multiply connected de Sitter space.) Then region $\mathcal{F}+\mathcal{R}$ (including the Cauchy horizon separating \mathcal{F} from \mathcal{R}) forms a causal unit, and region $\mathcal{P}+\mathcal{L}$ (including the Cauchy horizon separating \mathcal{P} from \mathcal{L}) forms another causal unit. (See Fig. 4b, where the two null surfaces partially bounding the gray $\mathcal{F}+\mathcal{R}$ region to the past and future are identified. Similarly for the null surfaces partially bounding the $\mathcal{P}+\mathcal{L}$ region.) An event in $\mathcal{F}+\mathcal{R}$ and an event in $\mathcal{P}+\mathcal{L}$ are always causally independent in physics: they can never physically influence each other though they may be mathematically connected by some causal curves (null curves or timelike curves). Though $\mathcal{F}+\mathcal{R}$ and $\mathcal{P}+\mathcal{L}$ are connected in mathematics, they are disconnected in physics. They are separated by a Cauchy horizon. When we consider physics in $\mathcal{F}+\mathcal{R}$, we can completely forget region $\mathcal{P}+\mathcal{L}$ (and *vice versa*). Though in such a case the Cauchy horizon separating $\mathcal{F}+\mathcal{R}$ from $\mathcal{P}+\mathcal{L}$ is a null spacetime boundary, we do not need any boundary condition on it because the topological multi-connectivity in $\mathcal{F}+\mathcal{R}$ has already given rise to a periodic boundary condition (which is a kind of self-consistent boundary condition). (In Fig. 4b this is shown by the fact that the null curves partially bounding $\mathcal{F}+\mathcal{R}$ to the past and future are identified.) This periodic boundary condition (the self-consistent condition) is sufficient to fix the solutions of the universe. For example, in our multiply connected de Sitter space model, the stability of the Cauchy horizon requires that the regions with CTCs (\mathcal{R} and \mathcal{L}) must be confined in the past and in these regions all quantum fields must be in vacuum states (as we have already remarked, the appearance of any real particles there seems to destroy the Cauchy horizon [30]). This gives rise to an arrow of time and an arrow of entropy in this model.

$\mathcal{F}+\mathcal{R}$ is a Hausdorff manifold with a null boundary, and thus $\mathcal{F}+\mathcal{R}$ is geodesically incomplete to the past. But, the geodesic incompleteness of $\mathcal{F}+\mathcal{R}$ may *not* be important in physics because in the inflationary scenario all real particles are created during the reheating process after inflation within bubbles created in region \mathcal{F} and these particles emit only retarded photons which never run off the spacetime because here the geodesic incompleteness takes place only in the past direction. On the other hand, we can smoothly extend $\mathcal{F}+\mathcal{R}$ to $\mathcal{P}+\mathcal{L}$ so that the total multiply connected de Sitter space dS/B is geodesically complete but at the price that it is not a manifold at a two-sphere (Sec. IX A). This model describes two physically disconnected but mathematically connected universes. [The analogy between the causal structures in region $\mathcal{F}+\mathcal{R}$ and region $\mathcal{P}+\mathcal{L}$ might motivate us to identify antipodal points in our multiply connected de Sitter space, as we did for the simply connected de Sitter space (Sec. VIII). The spacetime so obtained is a Hausdorff manifold everywhere. It is geodesically complete but not time orientable. For computing the energy-momentum tensor of vacuum polarization, we must take into account the images of antipodal points in addition to the images produced by the boost transformation. Further research is needed to find a self-consistent vacuum for this spacetime.]

Now we consider formation of bubbles in $\mathcal{F}+\mathcal{R}$ in multiply connected de Sitter space. [The results (and the argu-

ments for $\mathcal{F}+\mathcal{R}$ in the previous paragraph) also apply to region $\mathcal{P}+\mathcal{L}$, except that while in $\mathcal{F}+\mathcal{R}$ bubbles expand in the future direction, in $\mathcal{P}+\mathcal{L}$ they expand in the past direction; here ‘‘future’’ and ‘‘past’’ are defined with respect to $(\partial/\partial\tau)^a$ where τ is the time coordinate in the global coordinates of de Sitter space.] Region \mathcal{R} (for its fundamental cell see Fig. 4) which is multiply connected has a finite four-volume $V_I = \frac{4}{3}\pi b r_0^4$ (here $b = \beta/r_0$, β is the de Sitter boost parameter). If the probability of forming a bubble per volume r_0^4 in de Sitter space is ϵ , then the total probability of forming a bubble in V_I is $P_I = \frac{4}{3}\pi b \epsilon$.

Region \mathcal{F} (its fundamental cell is shown in Fig. 4) has an infinite four-volume and thus there should be an infinite number of bubble universes formed [89,98]. The metric in region \mathcal{F} is given by Eq. (72) with $0 < \tau < \infty$, $0 \leq l < \beta$, $0 < \theta < \pi$, and $0 \leq \phi < 2\pi$ (see Fig. 4a); it is multiply connected (periodic in l with period β). In order that the inflation proceeds and the bubbles (which expand to the future—as expected with the retarded potential in region \mathcal{F}) do not percolate, it is required that $\epsilon < \epsilon_{\text{cr}}$ where $5.8 \times 10^{-9} < \epsilon_{\text{cr}} < 0.24$ [95]. Gott and Statler [99] showed that in order that we on earth today should not have witnessed another bubble colliding with ours within our past light cone (with 95% confidence) ϵ must be less than 7.60×10^{-4} for $\Omega = 0.1$ (for $\Omega = 0.4$ Gott [88] found $\epsilon < 0.01$). In our multiply connected de Sitter space, for inflation to proceed, there should be the additional requirement that bubbles do not collide with images of themselves (producing percolation). A necessary condition for a bubble formed in \mathcal{F} not to collide with itself is that from time τ when the bubble forms to future infinity ($\tau \rightarrow \infty$) a light signal moving along the l direction [where τ and l are defined in Eq. (69) and Eq. (71)] propagates a co-moving distance less than $\beta/2$, which leads to the condition that $\tau > \tau_0 \equiv r_0 \ln[(e^{b/2} + 1)/(e^{b/2} - 1)]$. In fact this is also a sufficient condition, which can be shown by the conformal mapping between region \mathcal{F} in the multiply connected de Sitter space and region F-O in Misner space defined by Eqs. (81)–(83). If the collision of two light cones in F occurs beyond the hyperbola $t^2 - x^2 - y^2 - z^2 = 1$ ($t > 0$) in Misner space (i.e., in the region O), the corresponding two light cones (and thus the bubbles formed inside these light cones) in \mathcal{F} will never collide because $t^2 - x^2 - y^2 - z^2 = 1$ in F corresponds to $\tau \rightarrow \infty$ in \mathcal{F} . It is easy to show that the condition for a light cone not to collide with its images within F-O is that $e^b(t^2 - x^2) - y^2 - z^2 > 1$, where (t, x, y, z) is the event where the light cone originates. By Eq. (81) this condition corresponds to $e^b[(\tilde{t}/r_0)^2 - 1] > 1 + (\tilde{t}/r_0)^2 - 2(\tilde{t}/r_0)\cos\theta$. Since $\tilde{t} > r_0$ and $-1 \leq \cos\theta \leq 1$, a sufficient condition is $e^b[(\tilde{t}/r_0)^2 - 1] > 1 + (\tilde{t}/r_0)^2 + 2(\tilde{t}/r_0)$, i.e. $e^b(\tilde{t}/r_0 - 1) > \tilde{t}/r_0 + 1$ which is equivalent to $\tau > \tau_0 = r_0 \ln[(e^{b/2} + 1)/(e^{b/2} - 1)]$. Therefore all bubbles formed after the epoch τ_0 in \mathcal{F} in the multiply connected de Sitter space will never collide with themselves. The $0 < \tau < \tau_0$ part of the fundamental cell in \mathcal{F} has a finite four-volume $V_{II} = V_I[\cosh^3(\tau_0/r_0) - 1]$. The total probability of forming a bubble in V_{II} is $P_{II} = \frac{4}{3}\pi b \epsilon [\cosh^3(\tau_0/r_0) - 1]$. For $b = 2\pi$ we have $\tau_0 \approx 0.086r_0$, $V_{II} \approx 0.011V_I$, and thus $P_{II} \approx 0.011P_I$.

For the case of $b = 2\pi$, in order that there be less than a 5% chance that a bubble forms in V_I (and thus less than 0.05% chance in V_{II}), ϵ should be less than 2×10^{-3} . This should be no problem because we expect that this tunneling probability ϵ should be exponentially small. Thus it would not be surprising to find region \mathcal{R} and region \mathcal{F} for epochs $0 < \tau < \tau_0 = 0.086r_0$ clear of bubble formation events (and clear of real particles), which is all we require.

Also note that there may be two epochs of inflation, one at the Planck scale caused by $\langle T_{ab} \rangle_{\text{ren}} = -g_{ab}/960\pi^2 r_0^4$ [Eq. (97)] which later decays in region \mathcal{F} at $\tau \gg \tau_0$ into an inflationary metastable state at the GUT scale produced by a potential $V(\phi)$, which, still later, forms bubble universes.

X. BABY UNIVERSE MODELS

Inflationary universes can lead to the formation of baby universes in several different scenarios [64,32–36]. If one of these baby universes simply turns out to be the original universe that one started out with, we have a multiply connected solution in many ways similar to our multiply connected de Sitter space. There would be a multiply connected region of CTCs bounded by a past Cauchy horizon which would be stable because of the self-consistency requirement as in the previous section, and this would also engender pure retarded potentials. Thus, in a wide class of scenarios, the epoch of CTCs would be long over by now, as we would be one of the many later-formed bubble universes. Also, the model might either be geodesically complete to the past or not. This might not be a problem in physics since we would in any case have a periodic boundary condition; and because with its pure retarded potentials, no causal signals could be propagated to the past in any case. There are several different baby universe scenarios—any one of which could accommodate our type of model.

First, there is the Farhi, Guth, and Guven [32] method of creation of baby universes in the lab. At late times in an open universe, for example, an advanced civilization might implode a mass (interestingly, it does not have to be a large mass—a few kilograms will do) with enough energy to drive it up to the GUT energy scale, whereupon it might settle into a metastable vacuum, creating a small spherical bubble of false vacuum with a $V = \Lambda/8\pi$ metastable vacuum inside. This could be done either by just driving the region up over the potential barrier, or by going close to the barrier and tunneling through. The inside of this vacuum bubble would contain a positive cosmological constant with a positive energy density and a negative pressure. This bubble could be created with an initial kinetic energy of expansion with the bubble wall moving outward. But the negative pressure would pull it inward, and it would eventually reach a point of maximum expansion (a classical turning point), after which it would start to collapse and would form a black hole. But occasionally, (probability $P = 10^{-10^{18}}$ for typical GUT scales [32]) when it reaches its point of maximum expansion it tunnels to a state of equal energy but a different geometry, like a doorknob, crossing the Einstein-Rosen bridge [184]. The ‘‘knob’’ itself would be the interior of the bubble, containing the positive cosmological constant, and sitting in the

metastable vacuum state with $V = \Lambda/8\pi$. The “knob” consists of more than a hemisphere of an initially static S^3 closed de Sitter universe, where the bubble wall is a surface of constant “latitude” on this sphere. At the wall, the circumferential radius is thus decreasing as one moves outward toward the external spacetime. Just outside the wall is the Einstein-Rosen neck which reaches a minimum circumferential radius at $r = 2M$, and then the circumference increases to join the open external solution. This “doorknob” solution then evolves classically. The knob inflates to form a de Sitter space of eventually infinite size. It is connected to the original spacetime by the narrow Einstein-Rosen bridge. But an observer sitting at $r = 2M$ in the Einstein-Rosen bridge will shortly hit a singularity in the future, just as in the Schwarzschild solution. So the connection only lasts for a short time. The interior of the “knob” is hidden from an observer in the external spacetime by an event horizon at $r = 2M$. Eventually the black hole evaporates via Hawking radiation [185], leaving a flat external spacetime (actually part of an open big bang universe) with simply a coordinate singularity at $r = 0$ as seen from outside. (See Fig. 7.)

From the point of view of an observer sitting at the center of $V = \Lambda/8\pi$ bubble, he would see himself, just after the tunneling event, as sitting in a de Sitter space that was initially static but which starts to inflate. Centered on this observer’s antipodal point in de Sitter space, he would see a bubble of ordinary $V = 0$ vacuum surrounding a black hole of mass M . The observer sees his side of the Einstein-Rosen bridge and an event horizon at $r = 2M$ which hides the external spacetime at late times from him. From the point of view of the de Sitter observer, the black hole also evaporates by Hawking radiation, eventually leaving an empty $V = 0$ bubble in an ever-expanding de Sitter space. This infinitely expanding de Sitter space, which begins expanding at the tunneling event, is a perfect starting point (just like Vilenkin’s tunneling universe) for making an infinite number of bubble universes, as this de Sitter space has a finite beginning and then expands forever. Now suppose *one* of these open bubble universes simply turns out to be the *original* open universe where that advanced civilization made the baby de Sitter universe in the first place (Fig. 7). Now the model is multiply connected, with no earliest event. There is a Cauchy horizon (\mathcal{CH} , see Fig. 7) separating the region of CTCs from the later region that does not contain them. This Cauchy horizon is generated by ingoing closed null geodesics that represent signals that could be sent toward the black hole, which then tunnel across the Euclidean tunneling section jumping across the Einstein-Rosen bridge and then continuing as ingoing signals to enter the de Sitter space and reach the open single bubble in the de Sitter space (that turns out to be the original bubble in which the tunneling event occurs). A retarded photon traveling around one of those closed null geodesics will be red-shifted more and more on each cycle, thus not causing an instability. Another novel effect is that although these null generators are converging just before the tunneling event, they are diverging just after the tunneling event, having jumped to the other side of the Einstein-Rosen bridge. Thus, converging rays are turned into diverging rays (as in the wormhole solution) during the tun-

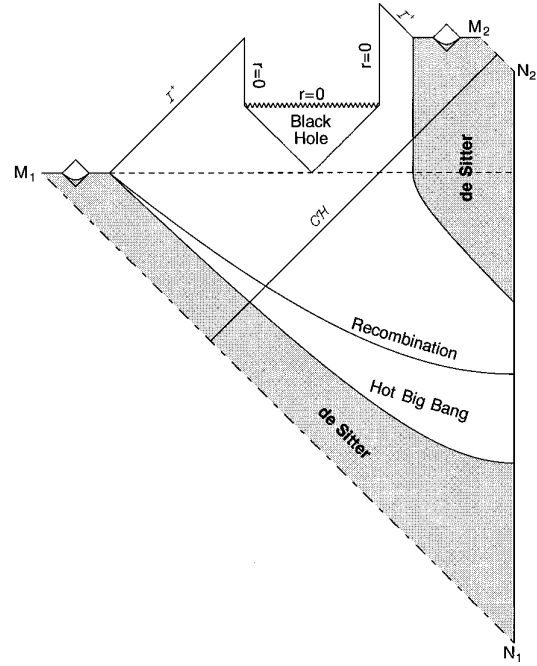


FIG. 7. A schematic Penrose diagram of a self-creating Universe based on the baby universe model of Farhi, Guth, and Guven. We identify M_1N_1 with M_2N_2 , to obtain a model of the Universe creating itself. (M_1N_1 is the future light cone of event N_1 .) In this model the closed null curves generating the Cauchy horizon (\mathcal{CH}) pass through a hot big bang region, where the dense absorber can make the Cauchy horizon stable against vacuum polarization effects. The metastable de Sitter phase is shown in grey. It decays along a hyperboloid H^3 near the bottom of the figure to form a single open bubble universe with a hot big bang phase and an epoch of recombination which is also shown. After recombination a supercivilization creates, at the right, an expanding bubble of de Sitter metastable vacuum. This reaches a point of maximum expansion at which point it tunnels to a doorknob-shaped configuration. The tunneling epoch is shown by the dashed line: just below the dashed line is how the spacetime appears just before tunneling, and just above the dashed line is how the spacetime appears just after the tunneling. Just after the tunneling, the geometry (just above the dashed line) from left to right goes from infinite radius (where future null infinity \mathcal{I}^+ meets the dashed line) to a minimum radius $r = 2M$ at the neck in the Einstein-Rosen bridge (where the inside and outside black hole event horizons meet just below the word “black hole”) then to a radius $r > 2M$ at the surface of the de Sitter bubble, reaching a maximum radius at the equator of the de Sitter bubble “knob” and finally decreasing to $r = 0$ at the center of the bubble at the extreme right. The de Sitter bubble expands forever. M_2 is at $t = \infty$. To the left of M_2 is another open bubble universe forming out of the metastable de Sitter vacuum. It is diamond-shaped—the bottom two lines representing the expanding bubble wall and the top two lines representing future null infinity for that bubble. Within this bubble the de Sitter vacuum decays to a hot big bang phase along a hyperboloid H^3 shown as a curved line crossing the diamond. Another open bubble universe forms to the right of M_1 . Recall, $M_1 = M_2$. These two bubble universes both form after the Cauchy horizon \mathcal{CH} as do an infinite number of others. The black hole singularity is shown, as well as the fact that the black hole evaporates. N_1N_2 is a CTC; CTCs occur on the N_1N_2 side of the Cauchy horizon \mathcal{CH} . After \mathcal{CH} , there are no CTCs.

neling event without violating the weak energy condition. These closed null geodesics need not be infinitely extendible in affine distance toward the past. It would seem that it can be arranged that the renormalized energy-momentum tensor does not blow up on this Cauchy horizon so that a self-consistent solution is possible. Using the method of images, note that the N -th image is from N cycles around the multiply connected spacetime. The path connecting an observer to the N -th image will have to travel N times through the hot big bang phase which occurs in the open bubble after the false vacuum with $V(\phi < \phi_0) = \Lambda/8\pi$ dumps its false vacuum energy into thermal radiation as it falls off the plateau and reaches the true vacuum $V(\phi = \phi_0) = 0$. Thus, to reach the N -th image one has to pass through the hot optically thick thermal radiation of the hot big bang N times. And this will cause the contribution of the N -th image to the renormalized energy-momentum tensor to be exponentially damped by a factor $e^{-N\tau}$ where $\tau \equiv nz\sigma_t \gg 1$ (where n is the number density of target particles, z is the thickness of hot material, σ_t is the total cross-section). Li [26] has calculated the renormalized energy-momentum tensor of vacuum polarization with the effect of absorption. Li [26] has estimated the fluctuation of the metric of the background spacetime caused by vacuum polarization with absorption, which is a small number in most cases. If the absorption is caused by electron-positron pair production by a photon in a photon-electron collision, the maximum value of the metric fluctuation is $(\delta g_{\mu\nu})_{\max} \sim l_p^2/(r_e L)$, where l_p is the Planck length, r_e is the classical radius of electron, L is the spatial distance between the identified points in the frame of rest relative to the absorber [26]. If we take L to be the Hubble radius at the recombination epoch ($\sim 10^{23}$ cm), we have $(\delta g_{\mu\nu})_{\max} \sim 10^{-76}$. Thus, we expect that the renormalized energy-momentum tensor will not blow up at the Cauchy horizon [26], so that a self-consistent solution is possible.

The tunneling event is shown as the epoch indicated by the dashed line in Fig. 7. During the tunneling event, the trajectory may be approximated as a classical space with four spacelike dimensions solving Einstein's equations, with the potential inverted, so that this Euclidean section bridges the gap between the two classical turning points. (In such a case, the concepts of CTCs and closed null curves should be generalized to contain a spacelike interval. Thus, there are neither closed null geodesics nor closed timelike geodesics with the traditional definitions. According to Li [23], this kind of spacetime can be stable against vacuum polarization.)

As Farhi, Guth, and Guven [32] note, the probability for forming such a universe is exponentially small, so an exponentially large number of trials would be required before an intelligent civilization would achieve this feat. If the metastable vacuum is at the Planck density, the number of trials required is expected to be not too large; but if it is at the GUT density which turns out to be many orders of magnitude lower than the Planck density, then the number of trials becomes truly formidable ($P \sim 10^{-10^{18}}$) [32]. Thus, Farhi, Guth, and Guven [32] guess that it is unlikely that the human race will ever succeed in making such a universe in the lab at the GUT scale. Gott [63], applying the Copernican principle

to estimate our future prospects, would come to similar conclusions. However, if our universe is open, it has an infinite number of galaxies, and it would likely have some super-civilizations powerful enough to succeed at such a creation event, or at least have so many super-civilizations (an infinite number) that even if they each tried only a few times, then some of them (again an infinite number) would succeed. In fact, if the probability for a civilization to form on a habitable planet like the Earth and eventually succeed at creating a universe in the lab is some finite number greater than zero (even if it is very low), then our universe (if it is an open bubble universe) should spawn an infinite number of such baby universes.

This notion has caused Harrison [33] to speculate that our universe was created in this way in the lab by some super-civilization in a previous universe. He noted correctly that if super-civilizations in a universe can create many baby universes, then baby universes created in this way should greatly outnumber the parent universes, and that you (being not special) are simply likely to live in one of the many baby universes, because there are so many more of them. Here he is using implicitly the formulation of Gott [63] that according to the Copernican principle, out of all the places for intelligent observers to be, there are, by definition, only a few special places and many non-special places, and you are simply more likely to be in one of the many non-special places. Thus, if there are many baby universes created by intelligent supercivilizations in an infinite open bubble universe, then you are likely to live in a baby universe created in this way. Harrison uses this idea to explain the strong anthropic principle. The strong anthropic principle as advanced by Carter [62] says that the laws of physics, in our universe at least, must be such as to allow the development of the intelligent life. Why? Because we are here. It is just a self-consistency argument. This might lead some to believe, particularly with inflationary cosmologies that are capable of producing an infinite number of bubble universes, that these different universes might develop with many different laws of physics, given a complicated, many-dimensional inflationary potential with many different minima, and many different low energy laws of physics. If some of these did not allow the development of intelligent life and some of these did, well, which type of universe would you expect to find yourself in?—one that allowed intelligent observers, of course. (By the same argument, you are not surprised to find yourself on a habitable planet—Earth—although such habitable planets may well be outnumbered by uninhabitable ones—Mercury, Venus, Pluto, etc.) Thus, there may be many more universes that have laws of physics that do not allow intelligent life—you just would not find yourself living there. It has been noticed that there are various coincidences in the physical constants—like the numerical value of the fine structure constant, or the ratio of the electron to proton mass, or the energy levels in the carbon nucleus—which, if they were very different, would make intelligent life either impossible, or much less likely. If we observe such a coincidence, according to Carter [62], it simply means that if it were otherwise, we would not be here. Harrison [33] has noted that if intelligent civilizations made baby universes they might

well, by intelligent choice, make universes that purposely had such coincidences in them in order to foster the development of intelligent life in the baby universes they created. If that were the case, then the majority of universes would have laws of physics conducive to the formation of intelligent life. In this case, the reason that we observe such coincidences is that a previous intelligent civilization made them that way. One might even speculate in this scenario that if they were smart enough, they could have left us a message of sorts in these dimensionless numbers (a theme that resonates, by the way, with part of Carl Sagan's thesis in *Contact*). However, it is unclear whether any super-civilization would be able to control the laws of physics in the universes they created. All, they might reasonably be able to do would be to drive the baby universe up into a particular metastable vacuum [32]. But then, such a metastable vacuum inflates in the knob, and an infinite number of bubble universes form later, with perhaps many different laws of physics depending on how they tunnel away from the metastable vacuum and which of the many potential minima they roll down into. Controlling these phase transitions would seem difficult. Thus, it would seem difficult for the super-civilization that made the metastable state that later gave rise to our universe to have been able to manipulate the physical constants in our universe. Harrison's model could occur in many generations, making it likely that we were produced as great, great, . . . , great grandchildren universes from a sequence of intelligent civilizations. Harrison [33] was able to explain all the universes by this mechanism except for the first one. For that, he had to rely on natural mechanisms. This seems to be an unfortunate gap. In our scenario, suppose that "first" universe simply turned out to be one of the infinite ones formed later by intelligent civilizations. Then the Universe—note capital U—would be multiply connected, and would have a region of CTCs; all of the individual universes would owe their birth to some intelligent civilization in particular in this picture.

All this may overestimate the importance of intelligent civilizations. It may be that bubbles of inflating metastable vacuum are simply produced at late times in any big bang cosmology by natural processes, and that baby universes produced by natural processes may vastly outnumber those produced by intelligent civilizations. Such a mechanism has been considered by Frolov, Markov, and Mukhanov [55]. They considered the hypothesis that spacetime curvature is limited by quantum mechanics and that as this limit is approached, the curvature approaches that of de Sitter space. Then, as any black hole collapses, the curvature increases as the singularity is approached; but before getting there it will convert into a collapsing de Sitter solution. This can be done in detail in the following way. Inside the horizon, but outside the collapsing star the geometry becomes Schwarzschild which is a radially collapsing but stretching cylinder. This can be matched onto a radially collapsing and radially shrinking cylinder in de Sitter space as described by the metric in Eq. (72) with the time τ being negative and the coordinate l being unbounded rather than periodic. Both surfaces are cylinders with identical intrinsic curvature, but with different extrinsic curvature. This mismatch is cured by intro-

ducing a shell of matter which converts the stretching of the Schwarzschild cylinder to collapsing as well which then matches onto the collapsing de Sitter solution. This phase transition may occur in segments which then merge as noted by Barabes and Frolov [56,57]. The de Sitter solution then bounces and becomes an expanding de Sitter solution which can in turn spawn an infinite number of open bubble universes. This all happens behind the event horizon of the black hole. Within the de Sitter phase, one finds a Cauchy horizon like the interior Cauchy horizon of the Reissner-Nordstrom solution, but this inner Cauchy horizon is not unstable because the curvature is bounded by the de Sitter value so the curvature is not allowed to blow up on the inner horizon. (This is an argument that one could also rely on to produce self-consistent multiply connected de Sitter phases with CTCs—if needed.) This model thus produces, inside the black hole, to the future, and behind the event horizon, an expanding de Sitter phase that has a beginning, just like Vilenkin's tunneling universe. If one of those bubble universes simply turns out to be the original one in which the black hole formed, then the solution is multiply connected with a region of CTCs. This would make every black hole produce an infinite number of universes. This would be the dominant mechanism for making new bubble universes, since the number of black holes in our universe would appear to greatly outnumber the number of baby universes ever produced by intelligent civilizations, since the tunneling probability for that process to succeed is exceedingly small.

Smolin [34,35] has proposed that this type of mechanism works and furthermore that the laws of physics (in the bubble universes) are like those in our own but with small variations. Then, there would be a Darwinian evolution of universes. Universes that produced many black holes would have more children that would inherit their characteristics—with some small variations. Soon, most universes would have laws of physics that were fine-tuned to produce the maximum number of black holes. Smolin [34,35] points out that this theory is testable, since we can calculate whether small changes in the physical constants would decrease the number of black holes formed. In this picture we should be near a global maximum in the black hole production rate. One problem is that the laws of physics that maximize the number of black holes and those that simply maximize the number of main sequence stars may be rather similar, and the laws that maximize the number of main sequence stars might well simply maximize the number of intelligent observers, and the anthropic principle alone would suggest a preference for us observing such laws, even if no baby universes were created in black holes. Another possible problem with this model, pointed out by Rothman and Ellis [186], is that if the density fluctuations in the early universe had been higher in amplitude, this would form many tiny primordial black holes (presumably more black holes per comoving volume than in our universe), so, we well might wonder why the density fluctuations in our universe were so small. One way out might be that tiny black holes do not form any baby universes, but this seems a bit forced since the de Sitter neck formed can be as small as the Planck scale or GUT scale and it would seem that even primordial black holes could be

large enough to produce an infinite number of open bubble universes.

Another possibility is the recycling universe of Garriga and Vilenkin [36]. In this model there is a metastable vacuum with cosmological constant Λ_1 , and a true lowest energy vacuum with a cosmological constant Λ_2 . Λ_1 is at the GUT or Planck energy scale, while Λ_2 is taken to be the present value of Λ (as might be the case in a flat- Λ model). As long as $\Lambda_2 > 0$, then Garriga and Vilenkin assert that there is a finite (but small) probability per unit four volume that the Λ_2 state could tunnel to form a bubble of Λ_1 state, which could therefore inflate, decaying into bubbles of Λ_2 vacuum, which could recycle forming Λ_1 bubbles, and so forth. They point out that depending on the coordinate system, a bubble of Λ_2 forming inside a Λ_1 universe could also be seen as a Λ_1 bubble forming inside of a Λ_2 universe. Take two de Sitter spaces, one with Λ_1 and one with Λ_2 , and cut each along a vertical slice ($W = W_0$) in the embedding space. They can then be joined along an appropriate hyperbola of one sheet representing a bubble wall, with the Λ_2 universe lying to the $W < W_0$ side and the Λ_1 universe lying to the $W > W_0$ side. Slicing along hyperplanes with $V + W = \text{const}$ gives a steady-state coordinate system for a Λ_1 universe in which a bubble of Λ_2 vacuum appears. Slicing along hyperplanes with $V - W = \text{const}$, however, gives a steady-state coordinate system for a Λ_2 universe in which a bubble of Λ_1 appears. So, one can find a steady-state coordinate system in which there is a Λ_1 universe, with bubbles of Λ_2 inside it, and bubbles of Λ_1 inside these Λ_2 bubbles, and so forth. If the roll down is slow, within the Λ_2 bubble as it forms, as in Gott's open bubble universe [89], then it will have at least 67 e -folds of inflation with $\Lambda \approx \Lambda_1$ before it falls off the plateau into the absolute minimum at Λ_2 , and this will be an acceptable big bang model which will have the usual big bang properties except that it will eventually be dominated by a lambda term Λ_2 . Being bubble universes, they will all be open with negative curvature as in Gott's model [89] but they will be asymptotically open de Sitter models at late times with $a(t) = r_0 \sinh(t/r_0)$ and $\Lambda = \Lambda_2$. Garriga and Vilenkin [36] wondered whether such a recycling model could be geodesically complete toward the past. Such a outcome, they pointed out, would violate no known theorems and should be investigated. They hoped to find such a geodesically-complete-to-the-past model so that it could be eternal without a need for a beginning. However, in the special case, where $\Lambda_1 = \Lambda_2$, one can show that the recycling steady state solution becomes a simple single de Sitter space geometry with Λ_1 , and the usual steady-state coordinate system in a single de Sitter space is not geodesically complete to the past.

Now take this recycling model where it turns out that one of the Λ_1 bubbles formed inside an Λ_2 bubble inside a Λ_1 region is, in fact, the Λ_1 region that one started out with. In this case, we would have a multiply connected model such as we are proposing which would include a region of CTCs (Fig. 8). (If $\Lambda_1 = \Lambda_2$, this model is just the multiply connected de Sitter space we have considered.) If our multiply connected model was geodesically complete to the past, so

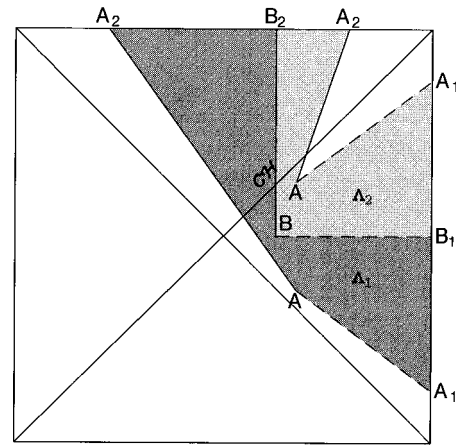


FIG. 8. A self-creating Universe model based on Garriga and Vilenkin's recycling Universe. In a region of cosmological constant Λ_1 , a bubble B of cosmological constant Λ_2 is formed by tunneling at the epoch BB_1 . The expanding bubble wall is represented by BB_2 . At a later time, within bubble B a bubble A forms at epoch AA_1 by tunneling. The expanding bubble wall is shown by AA_2 . Inside bubble A the cosmological constant is Λ_1 . In the limit where $\Lambda_1 = \Lambda_2$ we can plot this in a single de Sitter space. Now we identify the two hypersurfaces denoted by A_1AA_2 to obtain a model of the Universe creating itself. The Cauchy horizon bounding the region of CTCs is indicated by CH . After CH there are no CTCs. If $\Lambda_1 = \Lambda_2$, this reduces to our multiply connected de Sitter model $\mathcal{F} + \mathcal{R}$ shown in Fig. 4(b).

would the covering space (a simply connected Garriga-Vilenkin model) be. If our multiply connected model was geodesically incomplete to the past, so would the covering space (a simply connected Garriga-Vilenkin model) be also. In our model, there would be a strong self-consistency reason for pure retarded potential, whereas in the Garriga-Vilenkin recycling model, there would be no such strong reason for it. With pure retarded potentials throughout, the issue of whether the spacetime was geodesically complete to the past is less compelling, as we have argued above, and our model, having a periodic boundary condition, would not need further boundary conditions, unlike a simply connected recycling model that was geodesically incomplete to the past.

Thus, there are a number of models in which baby universes are created which can be converted into models in which the Universe creates itself, if one of those created baby universes turns out to be the original universe that one started with. Since these models are all ones in which there are an infinite number of baby universes created, this multiply connected outcome must occur unless the probability for a particular multiple connectivity to exist is exactly zero. In other words, it should occur, unless it is forbidden by the laws of physics. Given quantum mechanics, it would seem that such multiple connectivities would not be absolutely forbidden, particularly in the Planck foam era.

We should note here that, in principle, there might even be solutions that are simply connected in which there was an early region of CTCs bounded to the future by a Cauchy horizon followed by an inflationary region giving rise to an

infinite number of bubble universes. The models considered so far have all obeyed the weak energy condition, and these models have all been multiply connected; in other words, they have a genus of 1, like a donut, since one of the later baby universes is connected with the original one. Consider an asymptotically flat spacetime with two connected wormhole mouths that are widely separated. The existence of the wormhole connection increases the genus by one. Instead of a flat plane, it becomes a flat plane with a handle. To do this, the wormhole solution must violate the weak energy condition [7]. It must have some negative energy density material, for it is a diverging lens (converging light rays entering one wormhole mouth, diverge upon exiting the other mouth). For a compact two dimensional surface, the integrated Gaussian curvature over the surface divided by 4π is equal to 1 minus the genus. Thus, the integrated Gaussian curvature over a sphere (genus=0) is 4π , while the integrated Gaussian curvature over a donut (genus=1) is zero, and the integrated Gaussian curvature over a Figure 8 pretzel (genus=2) is -4π . Negative curvature is added each time the genus is increased. Conversely, positive curvature can be added to reduce the genus by 1. When a donut is cut, so that it resembles a letter ‘‘C,’’ the ends of the letter ‘‘C’’ are sealed with positive curvature (two spherical hemispherical caps would do the job, for example). Our solutions are already multiply connected, so they might in principle be made simply connected by the addition of some extra positive mass density, without violating the weak energy condition. An example of this is seen by comparing Grant space [18] with Gott’s two-string spacetime [8]. Grant space is multiply connected, has $T_{ab}=0$ everywhere, and includes CTCs. It can be pictured as a cylinder. Gott’s two-string spacetime is simply connected, but is identical to Grant space at large distances from the strings. It also contains CTCs. It can be pictured as a cardboard cylinder that has been stepped on and then stapled shut at one end, like an envelope. There are two corners at the closed end, representing the two strings, but the cylinder continues outward forever toward its open end (so it is like a test tube, a cylinder closed on one end). The two strings provide positive energy density (i.e. they do not violate the weak energy condition). CTCs that wrap around the two strings far out in the cylinder (which is identical to a part of Grant space; see Laurence [187]) can be shrunk to points by slipping them through the strings—but they become spacelike curves during this process. Thus, Gott space represents how a multiply connected spacetime with CTCs (Grant space) can be converted into a simply connected spacetime with CTCs by adding to the solution material that obeys the weak energy condition. A similar thing might in principle be possible with these cosmological models. Since our multiply connected versions already obey the weak energy condition, so would the associated simply connected versions.

XI. CONCLUSIONS

The question of first-cause has been a troubling one for cosmology. Often, this has been solved by postulating a universe that has existed forever in the past. Big bang models

supposed that the first-cause was a singularity, but questions about its almost, but not quite, uniformity remained. Besides, the big bang singularity just indicated a breakdown of classical general relativity, and with a proper theory-of-everything, one could perhaps push through to earlier times. Inflation has solved some of these problems, but Borde and Vilenkin have shown that if the initial inflationary state is metastable, then it must have had a finite beginning also. Ultimately, the problem seems to be how to create something out of nothing.

So far, the best attempt at this has been Vilenkin’s tunneling from nothing model and the similar Hartle-Hawking no-boundary proposal. Unfortunately, tunneling is, as the name suggests, usually a process that involves tunneling from *one* classical state to *another*, thus, with the Wheeler-DeWitt potential and ‘‘energy’’ $E=0$ that Hartle and Hawking adopted, the Universe, we argue, should really start not as nothing but as an S^3 universe of radius zero—a point. A point is as close to nothing as one can get, but it is not nothing. Also, how could a point include the laws of physics? In quantum cosmology, the wave function of the Universe is treated as the solution of a Schrödinger-like equation (the Wheeler-DeWitt equation), where the three-sphere S^3 radius a is the abscissa and there is a potential $U(a)$ with a metastable minimum at $U(a=0)=0$, and a barrier with $U(a)>0$ for $0<a<a_0$, and $U(a)<0$ for $a>a_0$. Thus, the evolution can be seen as a particle, representing the universe, starting as a point, $a=0$, at the bottom of the metastable potential well, with $E=0$. Then it tunnels through the barrier and emerges at $a=a_0$ with $E=0$, whereupon it becomes a classically inflating de Sitter solution. It can then decay via the formation of open single bubble universes [89,98]. The problem with this model is that it ignores the ‘‘zero-point-energy.’’ If there is a conformal scalar field ϕ , then the ‘‘energy’’ levels should be $E_n=n+\frac{1}{2}$. Even for $n=0$ there is a ‘‘zero-point-energy.’’ The potential makes the system behave like a harmonic oscillator in the potential well near $a=0$. A harmonic oscillator cannot sit at the bottom of the potential well—the uncertainty principle would not allow it. There must be some zero-point-energy and the particle must have some momentum, as it oscillates within the potential well when the field ϕ is included. Thus, when the ‘‘zero-point-energy’’ is considered, we see that the initial state is not a point but a tiny oscillating ($0\leq a\leq a_1$) big bang universe, that oscillates between big bangs and big crunches (though the singularities at the big bangs and big crunches might be smeared by quantum effects). This is the initial classical state from which the tunneling occurs. It is metastable, so this oscillating universe could not have existed forever: after a finite half-life, it is likely to decay. It reaches maximum radius a_1 , and then tunnels to a classical de Sitter state at minimum radius a_2 where $a_2<a_0$. The original oscillating universe could have formed by a similar tunneling process from a contracting de Sitter phase, but such a phase would have been much more likely to have simply classically bounced to an expanding de Sitter phase instead of tunneling into the oscillating metastable state at the origin. In this case, if one found oneself in an expanding de Sitter phase, it would be much more likely that it was the result of

classical bounce from a contracting de Sitter phase, rather than the result of a contracting de Sitter phase that had tunneled to an oscillating phase and then back out to an expanding de Sitter phase. Besides, a contracting de Sitter phase would be destroyed by the formation of bubbles which would percolate before the minimum radius was ever reached.

In this paper, we consider instead the notion that the Universe did not arise out of nothing, but rather created itself. One of the remarkable properties of the theory of general relativity is that in principle it allows solutions with CTCs. Why not apply this to the problem of the first-cause? Usually the beginning of the Universe is viewed like the south pole. Asking what is before that is like asking what is south of the south pole, it is said. But as we have seen, there remain unresolved problems with this model. If instead there were a region of CTCs in the early universe, then asking what was the earliest point in the Universe would be like asking what is the easternmost point on the Earth. There is no easternmost point—you can continue going east around and around the Earth, always traveling to the east. If the Universe contained an early region of CTCs, there would be no first-cause. Every event would have events to its past. And yet the Universe would not have existed eternally in the past (see Fig. 1). Thus, one of the most remarkable properties of general relativity—the ability in principle to allow CTCs—would be called upon to solve one of the most perplexing problems in cosmology. Such an early region of CTCs could well be over by now, being bounded to the future by a Cauchy horizon. We construct some examples to show that vacuum states can be found such that the renormalized energy-momentum tensor does not blow up as one approaches the Cauchy horizon. For such a model to work the Universe has to reproduce at some later time the same conditions that obtained at an earlier time. Inflation is particularly useful in this regard, for starting with a tiny piece of inflating state, at later times a huge volume of inflating state is produced, little pieces of which look just like the one we started with [64]. Many inflationary models allow creation of baby inflationary universes inside black holes, either by tunneling across the Einstein-Rosen bridge, or by formation as one approaches the singularity. If one of these baby universes simply turns out to be the universe we started with, then a multiply connected model with early CTCs bounded by a Cauchy horizon is produced. Since any closed null geodesics generating the Cauchy horizon must circulate through the optically thick region of the hot big bang phase of the universe after the inflation has stopped, the renormalized energy-momentum tensor should not blow up as the Cauchy horizon is approached.

As a particularly simple example we consider a multiply connected de Sitter solution where an event E_i is topologically identified with an event E'_i that lies inside its future light cone via a boost transformation. If the boost $b = 2\pi$, we show that we can find a Rindler-type vacuum where the renormalized energy-momentum tensor does not blow up as the Cauchy horizon is approached but rather produces a cosmological constant throughout the spacetime which self-consistently solves Einstein's equations for this geometry.

Thus, it is possible to find self-consistent solutions. When analyzing classical fields in this model, the only self-consistent solution without a blow up as the Cauchy horizon is approached occurs when there is a pure retarded potential in the causally connected region of the model. Thus, the multiply connected nature of this model and the possibility of waves running into themselves, ensure the creation of an arrow of time in this model. This is a remarkable property of this model. Interestingly, this model, although having no earliest event and having some timelike geodesics that are infinitely extendible to the past, is nevertheless geodesically incomplete to the past. This is not a property we should have thought desirable, but since pure retarded potentials are established automatically in this model, there are no waves propagating to the past and so there may be no problem in physics with this, since there are never any waves that run off the edge of the spacetime. The region of CTCs has a finite four-volume equal to $4\pi b r_0^4/3$ and should be in a pure vacuum state containing no real particles or Hawking radiation and no bubbles. After the Cauchy horizon for a certain amount of proper time (depending on the bubble formation probability per four volume r_0^4) no bubbles (or real particles) form, but eventually this model expands to infinite volume, creating an infinite number of open bubble universes, which do not percolate. At late times in the de Sitter phase a particle detector would find the usual Hawking radiation just as in the usual vacuum for de Sitter space.

There are a number of problems to be solved in this model. The chronology projection conjecture proposes that the laws of physics conspire so as to prevent the formation of CTCs. This conjecture was motivated by Hiscock and Konkowski's result that the energy-momentum tensor of the adapted Minkowski vacuum in Misner space diverges as the Cauchy horizon is approached. But as we have shown [30], the adapted Rindler vacuum for Misner space has $\langle T_{ab} \rangle_{\text{ren}} = 0$ throughout the space if $b = 2\pi$; thus, this is a self-consistent vacuum for this spacetime since it solves Einstein's equations for this geometry. It is true that $\langle T_{ab} \rangle_{\text{ren}}$ remains formally ill-defined on the Cauchy horizon itself [$\xi = 0$ in Eq. (32) with $b = 2\pi$], a set of measure zero. But it is not clear that this creates a problem for physics, since continuity might require that this formally ill-defined quantity be defined to be zero on this set of measure zero as well, since it is zero everywhere else. In fact, a treatment in the Euclidean section shows this is the case, for in the Euclidean section, if $b = 2\pi$, $\langle T_{ab} \rangle_{\text{ren}} = 0$ everywhere, including at $\xi = 0$. Other counter-examples to the chronology protection conjecture have also been found, as discussed in Sec. V. Hawking himself has also admitted that the back-reaction of vacuum polarization does not enforce the chronology protection conjecture.

One of the remarkable properties of general relativity is that it allows, in principle, the formation of event horizons. This appears to be realized in the case of black holes. Just as black hole theory introduced singularities at the end, standard big bang cosmology introduced singularities at the beginning of the universe. Now, with inflation, we see that event horizons should exist in the early universe as well [89]. Inflationary ideas prompt the suggestion that baby universes

may be born. If one of the baby universes simply turns out to be the one we started with, then we get a model with an epoch of CTCs that is over by now, bounded toward the future by a Cauchy horizon. We have argued that the divergence of the energy-momentum tensor as one approaches the Cauchy horizon does not necessarily occur, particularly when the Cauchy horizon crosses through a hot big bang phase where absorption occurs.

If the energy-momentum tensor does not diverge as the Cauchy horizon is approached, other problems must still be tackled. The classical instability of a Cauchy horizon to the future (a future chronology horizon) in a spacetime with CTCs is one. But this problem is solved in a world with retarded potentials for a Cauchy horizon that occurs to our past (a past chronology horizon) and which ends an epoch of CTCs. It thus seems easier to have a Cauchy horizon in the early universe. At the microscopic level, quantum mechanics appears to allow acausal behavior. Indeed the creation and annihilation of a virtual positron-electron pair can be viewed as creation of a small closed loop, where the electron traveling backward in time to complete the loop appears as a positron. So, why should the laws of physics forbid time travel globally? Indeed one of the most remarkable properties of the laws of physics is that although they are time (CPT) symmetric, the solutions we observe have an arrow of time and retarded potentials. Without this feature of the solutions, acausal behavior would be seen all the time. Interestingly, in our model, the multiply connected nature of the spacetime geometry forces an arrow of time and retarded potentials. Thus, it is the very presence of the initial region of CTCs that produces the strong causality that we observe later on. This is a very interesting and unexpected property. An entropy arrow of time is automatically produced as well, with the region of CTCs in the simplest models sitting automatically in a cold vacuum state, with the universe becoming heated after the Cauchy horizon. Recently, Cassidy and Hawking [143] have proposed yet another supposed difficulty for CTCs, in that the formally defined entropy appears to diverge to negative infinity as the Cauchy horizon is ap-

proached. Yet, in the early universe this may turn out to be an advantage, since to produce the ordinary entropy arrow of time we observe in the universe today, we must necessarily have some kind of natural low-entropy boundary condition in the early universe [58,59]. This could occur on the Cauchy horizon that ends the period of CTCs.

New objections to spacetimes with CTCs can continue to surface, as old problems are put to rest, so it might seem that disproving the chronology protection conjecture would be a tall order. But, proving that there are no exceptions to the chronology protection conjecture, ever, would seem an equally daunting task. This is particularly true since we currently do not have either a theory of quantum gravity or a theory-of-everything.

Perhaps the most obvious problem with the model we have proposed is that the simplest solutions we have obtained so far are not geodesically complete to the past. But we may need no boundary condition since we have a periodic boundary condition instead. This thus may not be a problem in physics if retarded potentials are the only ones allowed. Alternatively, as Garriga and Vilenkin have indicated, it would violate no known theorems for some type of recycling universe (making bubble universes within bubble universes *ad infinitum*) to exist that was geodesically complete to the past. If such solutions exist, it might be possible to find a solution in which there was an early epoch of CTCs that would be geodesically complete to the past as well by simply identifying an earlier bubble with a later one.

Thus, a number of important questions remain, and we would not minimize them. The models presented here, however, do have some interesting and attractive properties, suggesting that this *type* of model should be investigated further, and that we *should* ask the question:

Do the laws of physics prevent the Universe from being its own mother?

ACKNOWLEDGMENTS

This research was supported by NSF grant AST95-29120 and NASA grant NAG5-2759.

-
- [1] A. Einstein, Sitzungsber. Preuss. Akad. Wiss., p. 142 (1917).
 - [2] W. J. van Stockum, Proc. R. Soc. Edinburgh **57**, 135 (1937).
 - [3] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
 - [4] A. H. Taub, Ann. Math. **53**, 472 (1951).
 - [5] E. T. Newman, L. Tamburino, and T. J. Unti, J. Math. Phys. **4**, 915 (1963).
 - [6] C. W. Misner, in *Relativity Theory and Astrophysics I: Relativity and Cosmology*, edited by J. Ehlers, Lectures in Applied Mathematics Vol. 8 (American Mathematical Society, Providence, 1967), p. 160.
 - [7] M. S. Morris, K. S. Thorne, and U. Yurtsever, Phys. Rev. Lett. **61**, 1446 (1988).
 - [8] J. R. Gott, Phys. Rev. Lett. **66**, 1126 (1991).
 - [9] K. S. Thorne, in *General Relativity and Gravitation 1992: Proceedings of the 13th International Conference on General Relativity and Gravitation*, edited by R. J. Gleiser, C. N. Kozameh, and O. M. Moreschi (IOP Publishing, Bristol, 1993), p. 295.
 - [10] M. Visser, *Lorentzian Wormholes—from Einstein to Hawking* (AIP, New York, 1995).
 - [11] A. Vilenkin, Phys. Lett. **117B**, 25 (1982).
 - [12] J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).
 - [13] S. W. Hawking, *A Brief History of Time, From the Big Bang to Black Holes* (Bantam Books, New York, 1988).
 - [14] W. A. Hiscock and D. A. Konkowski, Phys. Rev. D **26**, 1225 (1982).
 - [15] V. P. Frolov, Phys. Rev. D **43**, 3878 (1991).
 - [16] S.-W. Kim and K. S. Thorne, Phys. Rev. D **43**, 3929 (1991).
 - [17] G. Klinkhammer, Phys. Rev. D **46**, 3388 (1992).
 - [18] J. D. E. Grant, Phys. Rev. D **47**, 2388 (1993).

- [19] S. W. Hawking, in *Proceedings of the 6-th Marcel Grossmann Meeting*, edited by H. Sato (World Scientific, Singapore, 1992), p. 3.
- [20] S. W. Hawking, *Phys. Rev. D* **46**, 603 (1992).
- [21] D. G. Boulware, *Phys. Rev. D* **46**, 4421 (1992).
- [22] L.-X. Li, J.-M. Xu, and L. Liu, *Phys. Rev. D* **48**, 4735 (1993).
- [23] L.-X. Li, *Phys. Rev. D* **50**, R6037 (1994).
- [24] R. J. Low, *Class. Quantum Grav.* **12**, L37 (1995).
- [25] T. Tanaka and W. A. Hiscock, *Phys. Rev. D* **52**, 4503 (1995).
- [26] L.-X. Li, *Class. Quantum Grav.* **13**, 2563 (1996).
- [27] S. V. Krasnikov, *Phys. Rev. D* **54**, 7322 (1996).
- [28] S. V. Sushkov, *Class. Quantum Grav.* **14**, 523 (1997).
- [29] M. Visser, *Phys. Rev. D* **55**, 5212 (1997).
- [30] L.-X. Li and J. R. Gott, *Phys. Rev. Lett.* **80**, 2980 (1998).
- [31] M. J. Cassidy, *Class. Quantum Grav.* **14**, 3031 (1997).
- [32] E. Farhi, A. H. Guth, and J. Guven, *Nucl. Phys.* **B339**, 417 (1990).
- [33] E. R. Harrison, *Q. J. R. Astron. Soc.* **36**, 193 (1995).
- [34] L. Smolin, *Class. Quantum Grav.* **9**, 173 (1992).
- [35] L. Smolin, *The Life of the Cosmos* (Oxford University Press, Oxford, 1997).
- [36] J. Garriga and A. Vilenkin, *Phys. Rev. D* **57**, 2230 (1998).
- [37] A. Friedmann, *Z. Phys.* **10**, 377 (1922).
- [38] E. Hubble, *Proc. Natl. Acad. Sci. USA* **15**, 168 (1929).
- [39] S. W. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).
- [40] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge University Press, Cambridge, England, 1973).
- [41] G. Gamow, *Phys. Rev.* **74**, 505 (1948).
- [42] G. Gamow, *Nature (London)* **162**, 680 (1948).
- [43] R. A. Alpher and R. Herman, *Nature (London)* **162**, 774 (1948).
- [44] A. A. Penzias and R. W. Wilson, *Astrophys. J.* **142**, 419 (1965).
- [45] J. C. Mather *et al.*, *Astrophys. J. Lett.* **354**, L37 (1990).
- [46] J. C. Mather *et al.*, *Astrophys. J.* **420**, 439 (1994).
- [47] G. F. Smoot *et al.*, *Astrophys. J. Lett.* **396**, L1 (1992).
- [48] S. W. Hawking, *Proc. R. Soc. London* **A300**, 187 (1967).
- [49] S. W. Hawking and R. Penrose, *Proc. R. Soc. London* **A314**, 529 (1970).
- [50] G. Lemaitre, *Ann. Soc. Sci. Bruxelles A.* **53**, 51 (1933).
- [51] J. D. Bekenstein, *Phys. Rev. D* **11**, 2072 (1975).
- [52] R. Durrer and J. Laukenmann, *Class. Quantum Grav.* **13**, 1069 (1996).
- [53] R. C. Tolman, *Relativity, Thermodynamics, and Cosmology* (Clarendon, Oxford, 1934).
- [54] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).
- [55] V. P. Frolov, M. A. Markov, and V. F. Mukhanov, *Phys. Rev. D* **41**, 383 (1990).
- [56] C. Barrabès and V. P. Frolov, *Helv. Phys. Acta* **69**, 253 (1996).
- [57] C. Barrabès and V. P. Frolov, *Phys. Rev. D* **53**, 3215 (1996).
- [58] R. Penrose, in *General Relativity, an Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979), p. 581.
- [59] R. Penrose, in *Fourteenth Texas Symposium on Relativistic Astrophysics*, *Annals of the New York Academy of Sciences* Vol. 571, edited by E. J. Fenyves (The New York Academy of Sciences, New York, 1989), p. 249.
- [60] J. A. Wheeler and R. P. Feynman, *Rev. Mod. Phys.* **17**, 157 (1945).
- [61] J. R. Gott, *Astrophys. J.* **187**, 1 (1974).
- [62] B. Carter, in *Confrontation of Cosmological Theories with Observations*, edited by M. Longair (Dordrecht, Reidel, 1974), p. 291.
- [63] J. R. Gott, *Nature (London)* **363**, 315 (1993).
- [64] A. Linde, *Particle Physics and Inflationary Cosmology* (Harwood Academic Publishers, Chur, Switzerland, 1990).
- [65] E. W. Kolb and M. S. Turner, *The Early Universe* (Addison-Wesley Publishing Company, Redwood City, CA, 1990).
- [66] Ya. B. Zeldovich, *Mon. Not. R. Astron. Soc.* **160**, 1P (1972).
- [67] P. J. E. Peebles and J. T. Yu, *Astrophys. J.* **162**, 815 (1970).
- [68] E. R. Harrison, *Phys. Rev. D* **1**, 2726 (1970).
- [69] J. Bardeen, P. J. Steinhardt, and M. S. Turner, *Phys. Rev. D* **28**, 679 (1983).
- [70] A. Guth and S. Y. Pi, *Phys. Rev. Lett.* **49**, 1110 (1982).
- [71] S. W. Hawking, *Phys. Lett.* **115B**, 295 (1982).
- [72] A. A. Starobinsky, *Phys. Lett.* **117B**, 175 (1982).
- [73] N. Bahcall and R. M. Soneira, *Astrophys. J.* **270**, 20 (1983).
- [74] V. de Lapparent, M. Geller, and J. Huchra, *Astrophys. J. Lett. Ed.* **302**, L1 (1986).
- [75] J. R. Gott, A. Melott, and M. Dickinson, *Astrophys. J.* **306**, 341 (1986).
- [76] J. R. Gott, D. N. Weinberg, and A. Melott, *Astrophys. J.* **321**, 2 (1987).
- [77] M. Geller and J. Huchra, *Science* **246**, 897 (1989).
- [78] C. Park, *Mon. Not. R. Astron. Soc.* **242**, 59P (1990).
- [79] C. Park, Ph.D. thesis, Princeton University, 1990.
- [80] C. Park and J. R. Gott, *Mon. Not. R. Astron. Soc.* **249**, 288 (1991).
- [81] J. R. Gott, B. Gao, and C. Park, *Astrophys. J.* **383**, 90 (1991).
- [82] C. Park, J. R. Gott, A. Melott, and I. D. Karachentsev, *Astrophys. J.* **387**, 1 (1992).
- [83] S. J. Maddox, G. Efstathiou, W. J. Sutherland, and J. Loveday, *Mon. Not. R. Astron. Soc.* **242**, 43P (1990).
- [84] W. Saunders *et al.*, *Nature (London)* **349**, 32 (1991).
- [85] C. Park, J. R. Gott, and L. N. da Costa, *Astrophys. J. Lett.* **392**, L51 (1992).
- [86] S. A. Shectman *et al.*, in *Wide Field Spectroscopy and the Distant Universe*, the 35th Herstmonceux Conference, edited by S. J. Maddox and A. Aragon-Salamanca (World Scientific, Singapore, 1995), p. 98.
- [87] M. S. Vogeley, C. Park, M. J. Geller, J. P. Huchra, and J. R. Gott, *Astrophys. J.* **420**, 525 (1994).
- [88] J. R. Gott, in *Critical Dialogues in Cosmology*, edited by N. Turok (World Scientific, Singapore, 1997), p. 519.
- [89] J. R. Gott, *Nature (London)* **295**, 304 (1982).
- [90] E. Schrödinger, *Expanding Universes* (Cambridge University Press, London, 1956).
- [91] E. A. Milne, *Nature (London)* **130**, 9 (1932).
- [92] S. Coleman, *Phys. Rev. D* **15**, 2929 (1977).
- [93] S. Coleman and F. de Luccia, *Phys. Rev. D* **21**, 3305 (1980).
- [94] S. W. Hawking, I. G. Moss, and J. M. Steward, *Phys. Rev. D* **26**, 2681 (1982).
- [95] A. H. Guth and E. J. Weinberg, *Nucl. Phys.* **B212**, 321 (1983).

- [96] A. D. Linde, Phys. Lett. **108B**, 389 (1982).
- [97] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
- [98] J. R. Gott, in *Inner Space/Outer Space*, edited by E. W. Kolb *et al.* (The University of Chicago Press, Chicago, 1986), p. 362.
- [99] J. R. Gott and T. S. Statler, Phys. Lett. **136B**, 157 (1984).
- [100] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
- [101] D. N. Page, Phys. Rev. D **25**, 1499 (1982).
- [102] B. Ratra and P. J. E. Peebles, Astrophys. J. Lett. **432**, L5 (1994).
- [103] B. Ratra and P. J. E. Peebles, Phys. Rev. D **52**, 1837 (1995).
- [104] M. Bucher, A. S. Goldhaber, and N. Turok, Phys. Rev. D **52**, 3314 (1995).
- [105] M. Bucher, A. S. Goldhaber, and N. Turok, Phys. Rev. D **52**, 5538 (1995).
- [106] K. Yamamoto, M. Sasaki, and T. Tanaka, Astrophys. J. **455**, 412 (1995).
- [107] A. D. Linde, Phys. Lett. B **351**, 99 (1995).
- [108] A. D. Linde and A. Mezhlumian, Phys. Rev. D **52**, 6789 (1995).
- [109] A. D. Linde, Phys. Lett. **129B**, 177 (1983).
- [110] J. R. Gott and M. J. Rees, Astron. Astrophys. **45**, 365 (1975).
- [111] A. G. Riess, W. H. Press, and R. P. Kirshner, Astrophys. J. Lett. **438**, L17 (1995).
- [112] J. Mould and W. Freedman, Nature (London) **381**, 555 (1996).
- [113] A. Sandage *et al.*, Nature (London) **381**, 555 (1996).
- [114] T. Kundic *et al.*, Astrophys. J. **482**, 75 (1997).
- [115] M. Bolte and C. J. Hogan, Nature (London) **376**, 399 (1995).
- [116] J. R. Gott and E. L. Turner, Astrophys. J. **213**, 309 (1977).
- [117] R. Cen, J. R. Gott, J. P. Ostriker, and E. L. Turner, Astrophys. J. **423**, 1 (1994).
- [118] J. P. Ostriker and P. J. Steinhardt, Nature (London) **377**, 600 (1995).
- [119] E. L. Turner, Astrophys. J. Lett. **365**, L43 (1990).
- [120] M. Fukugita, T. Futamase, and M. Kasai, Mon. Not. R. Astron. Soc. **246**, 24P (1990).
- [121] C. S. Kochanek, CfA Report No. 4258 (1996).
- [122] B. Ratra, A. J. Banday, K. M. Gorski, and N. Sugiyama, astro-ph/9512148 (1995).
- [123] B. Ratra and N. Sugiyama, astro-ph/9512157 (1995).
- [124] C. Park, W. N. Colley, J. R. Gott, B. Ratra, D. N. Spergel, and N. Sugiyama, astro-ph/9711057 (1997).
- [125] S. Hawking and R. Penrose, *The Nature of Space and Time* (Princeton University Press, Princeton, NJ, 1996).
- [126] A. Borde and A. Vilenkin, in *Relativistic Astrophysics: The Proceedings of the Eighth Yukawa Symposium*, edited by M. Sasaki (University of Academic Press, Japan, 1995); gr-qc/9403004 (1994).
- [127] A. Borde and A. Vilenkin, Int. J. Mod. Phys. D **5**, 813 (1996).
- [128] A. Vilenkin, Phys. Rev. D **27**, 2848 (1983).
- [129] A. A. Starobinsky, in *Field Theory, Quantum Gravity, and Strings*, Lectures in Physics, edited M. J. de Vega and N. Sánchez (Springer-Verlag, Berlin, 1986).
- [130] A. D. Linde, Phys. Lett. B **175**, 395 (1986).
- [131] A. S. Gonchanov, A. D. Linde, and V. F. Mukhanov, Int. J. Mod. Phys. A **2**, 561 (1987).
- [132] G. Gamow, Z. Phys. **51**, 204 (1928).
- [133] G. Gamow, Z. Phys. **52**, 510 (1928).
- [134] R. W. Gurney and E. U. Condon, Phys. Rev. **33**, 127 (1929).
- [135] A. Vilenkin, Phys. Rev. D **30**, 509 (1984).
- [136] A. Vilenkin, Phys. Rev. D **37**, 888 (1988).
- [137] J. J. Halliwell, in *Quantum Cosmology and Baby Universes*, edited by S. Coleman, J. B. Hartle, T. Piran, and S. Weinberg (World Scientific, Singapore, 1991), p. 159.
- [138] S. W. Hawking, Nucl. Phys. **B239**, 257 (1984).
- [139] E. P. Tryon, Nature (London) **246**, 396 (1973).
- [140] R. P. Geroch, J. Math. Phys. **8**, 782 (1967).
- [141] J. J. Halliwell and J. B. Hartle, Phys. Rev. D **41**, 1815 (1990).
- [142] M. P. Headrick and J. R. Gott, Phys. Rev. D **50**, 7244 (1994).
- [143] M. J. Cassidy and S. W. Hawking, Phys. Rev. D **57**, 2372 (1998).
- [144] B. S. Kay, M. J. Radzikowski, and R. M. Wald, Commun. Math. Phys. **183**, 533 (1997).
- [145] S. W. Hawking, in *Recent Developments in Gravitation*, Cargese Lectures, edited by M. Levey and S. Deser (Plenum Press, New York, 1978), p. 145.
- [146] 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.
- [147] S. W. Hawking, Phys. Rev. D **52**, 5681 (1995).
- [148] M. Visser, Phys. Lett. B **415**, 8 (1997).
- [149] M. Visser, gr-qc/9710020 (1997).
- [150] C. R. Cramer and B. S. Kay, Class. Quantum Grav. **13**, L143 (1996).
- [151] C. R. Cramer and B. S. Kay, Phys. Rev. D **57**, 1052 (1998).
- [152] R. M. Wald, Phys. Rev. D **17**, 1477 (1978).
- [153] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England, 1982).
- [154] W. G. Unruh, Phys. Rev. D **14**, 870 (1976).
- [155] J. S. Dowker, Phys. Rev. D **18**, 1856 (1978).
- [156] S. A. Hayward, Class. Quantum Grav. **9**, 1851 (1992); **9**, 2453(E) (1992).
- [157] S. A. Hayward, “Junction Conditions for Signature Change,” gr-qc/9303034 (1993).
- [158] T. S. Bunch and P. C. W. Davies, Proc. R. Soc. London **A360**, 117 (1978).
- [159] D. Bernard and A. Folacci, Phys. Rev. D **34**, 2286 (1986).
- [160] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, Science Publishers, New York, 1965).
- [161] E. Calabi and L. Marcus, Ann. Math. **75**, 63 (1962).
- [162] J. L. Friedman and A. Higuchi, Phys. Rev. D **52**, 5687 (1995).
- [163] J. L. Friedman, in *The Sixth Canadian Conference on General Relativity and Relativistic Astrophysics*, edited by S. P. Braham, J. D. Gegenberg, and R. J. Mc Keller (American Mathematical Society, Providence, RI, 1997), p. 43.
- [164] R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).
- [165] P. Candelas and J. S. Dowker, Phys. Rev. D **19**, 2902 (1979).
- [166] R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics* (The University of Chicago Press, Chicago, 1994).

- [167] C. Panagiotakopoulos and N. Tetradis, hep-th/9710526 (1997).
- [168] D. G. Boulware, Phys. Rev. D **11**, 1404 (1975).
- [169] J. B. Hartle and S. W. Hawking, Phys. Rev. D **13**, 2188 (1976).
- [170] D. W. Sciama, P. Candelas, and D. Deutsch, Adv. Phys. **30**, 327 (1981).
- [171] P. Candelas, Phys. Rev. D **21**, 2185 (1980).
- [172] J. W. York, Phys. Rev. D **31**, 775 (1985).
- [173] T. Gold, Am. J. Phys. **30**, 403 (1962).
- [174] J. E. Hogarth, Proc. R. Soc. London **A267**, 365 (1962).
- [175] F. Hoyle and J. V. Narlikar, Proc. R. Soc. London **A227**, 1 (1964).
- [176] H. D. Zeh, *The Physical Basis of the Direction of Time* (Springer-Verlag, Berlin, 1992).
- [177] L. S. Schulman, *Time's Arrows and Quantum Measurement* (Cambridge University Press, Cambridge, England, 1997).
- [178] I. D. Novikov, *Evolution of the Universe* (Cambridge University Press, Cambridge, England, 1983).
- [179] J. Friedman, M. S. Morris, I. D. Novikov, F. Echeverria, G. Klinkhammer, K. S. Thorne, and U. Yurtsever, Phys. Rev. D **42**, 1915 (1990).
- [180] R. M. Wald, *General Relativity* (The University of Chicago Press, Chicago, 1984).
- [181] C. W. Misner and A. H. Taub, Sov. Phys. JETP **28**, 122 (1969) [*Zh. Éksp. Teor. Fiz* **55**, 233 (1968)].
- [182] R. K. Sachs and H. Wu, *General Relativity for Mathematicians* (Springer-Verlag, Berlin, 1977).
- [183] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman and Company, San Francisco, 1973).
- [184] A. Einstein and N. Rosen, Phys. Rev. **48**, 73 (1935).
- [185] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
- [186] T. Rothman and G. F. R. Ellis, Q. J. R. Astron. Soc. **34**, 201 (1993).
- [187] D. Laurence, Phys. Rev. D **50**, 4957 (1994).