Did the Universe have a beginning?

Alexander Vilenkin*
California Institute of Technology, Pasadena, California 91125
(Received 20 March 1992; revised manuscript received 3 August 1992)

It is argued that "eternal inflation" must have a beginning in time. Conditions are formulated for a spacetime to describe an eternally inflating universe without a beginning, and it is shown that these conditions cannot be satisfied. A rigorous proof is given for a two-dimensional spacetime, and a plausibility argument for four dimensions.

PACS number(s): 98.80.Bp

I. INTRODUCTION

Inflation is a period of rapid (quasiexponential) expansion in the early history of the Universe [1]. During this period, regions initially within the causal horizon are blown up to sizes much greater than the present Hubble radius, and the observable part of the Universe comes very close to being homogeneous, isotropic, and flat. The inflationary expansion is driven by the potential energy of a scalar field φ, while the field slowly "rolls down" its potential $V(\varphi)$. When φ reaches the minimum of the potential this vacuum energy thermalizes, and inflation is followed by the usual radiation-dominated expansion. The evolution of the field φ is influenced by quantum fluctuations, and as a result thermalization does not occur simultaneously in different parts of the Universe. Fluctuations in the thermalization time give rise to small density fluctuations on the observable scales, but result in large deviations from homogeneity and isotropy on scales much greater than the present horizon. In fact, it can be shown [2–9] that at any time there are parts of the universe which are still in the inflationary phase. Once started, inflation never ends completely. Inflating regions constantly undergo thermalization, but the exponential expansion of the remaining regions more than compensates for the loss.

A model in which the inflationary phase has no end and continually produces new islands of thermalization naturally leads to the following question: can this model be also extended to the infinite past, avoiding in this way the problem of the initial singularity? The Universe would then be in a steady state of eternal inflation which does not have a beginning. This possibility was pointed out by Linde [5] immediately after Steinhardt [4] suggested that inflation may have no end. However, it was soon realized by Linde himself [6] and by others [2,4] that this idea could not be implemented in the simplest model in which the inflating Universe is described by a de Sitter space. In a more general case, the situation remained unclear [8].

The purpose of this paper is to investigate the possibility of eternal inflation, without a beginning, in greater detail. To simplify the terminology, the word "eternal" will be used for the case when inflation is infinite in both time directions, and "semieternal" when it is infinite only in the future. (Note that this terminology is different from that used by Linde [7,8].) In the following sections, some necessary conditions will be formulated for a spacetime to describe an eternally inflating universe. Then it will be shown that under rather general assumptions these conditions cannot be satisfied. Hence, the Universe must have a beginning, and eternal inflation (in the above sense) is impossible.

Before analyzing the most general case, we shall clarify the ideas using simplified models. In the next section the possibility of eternal inflation is discussed in the context of "old" inflationary scenario [10], in which the vacuum energy is strictly constant and vacuum decay occurs through bubble nucleation. Old inflation is known to be semieternal; bubbles cannot fill the entire Universe, since the space between them is expanding so fast. In the thermalized parts of the Universe the distribution of matter produced by colliding bubble walls is grossly inhomogeneous, making old inflation unsuitable as a realistic cosmological model. Here we shall disregard this aspect of the problem and will only be concerned with the question of whether or not old inflation can be continued back to the infinite past. We shall see that the answer to this question is "no". Exponential expansion of the false vacuum is unavoidably preceded by an exponential contraction. If vacuum decays by bubble nucleation, the bubbles have no trouble filling the Universe during the contracting phase, so that the whole Universe thermalizes and collapses to a singularity, without ever making it to the expanding phase.

Guided by this discussion we shall formulate two conditions that a spacetime describing an eternally inflating universe should satisfy (Sec. III). In Sec. IV it will be shown that these conditions cannot be satisfied in a nonsingular two-dimensional spacetime. The realistic case of a four-dimensional inflating universe will be studied in Sec. V, and the conclusions will be briefly discussed in Sec. VI.

---

*On leave from Tufts University.

II. CAN OLD INFLATION BE ETERNAL?

In the old inflationary scenario, the false vacuum has the energy-momentum tensor

$$T_{\mu\nu}=\rho g_{\mu\nu} , \tag{2.1}$$

with $\rho=$const. The homogeneous and isotropic solution of Einstein's equations with $T_{\mu\nu}$ from (2.1) is the de Sitter space, which can be represented in the form

$$ds^2=dt^2-e^{2Ht}d\mathbf{x}^2 , \tag{2.2}$$

where

$$H^2=8\pi G\rho/3 . \tag{2.3}$$

This space has a horizon of radius H^{-1}; observers separated by a greater distance cannot communicate. We shall first assume that the inflating part of the Universe is described by the metric (2.2). The general case will be considered afterward.

Let us first recall the arguments showing that old inflation is semieternal [10]. Bubbles nucleating in false vacuum expand, rapidly approaching the speed of light. In de Sitter space this corresponds to having asymptotically static boundaries in comoving coordinates. The physical radius of a bubble formed at time t_1 is [for $H(t-t_1)\gg 1$]

$$r(t,t_1)\approx H^{-1}e^{H(t-t_1)} . \tag{2.4}$$

An expanding bubble can affect the geometry of the outside region only within a distance of $\sim H^{-1}$ from its boundary. Hence, although bubbles carve large volumes out of de Sitter space, the geometry of the remaining regions is practically unchanged. We shall first assume that inflation starts at some time t_0 and later consider the limit $t_0\to-\infty$.

Bubble nucleation is a stochastic process with a constant probability λ per unit spacetime volume. The probability for no bubbles to be formed in a four-volume Ω is [10,11]

$$\mathcal{P}_\Omega=e^{-\lambda\Omega} . \tag{2.5}$$

Here is a quick derivation. Let $\mathcal{P}(A)$ be the probability for no bubbles to nucleate in spacetime region A. Then, for nonoverlapping regions A and B, $\mathcal{P}(A\cup B)=\mathcal{P}(A)\mathcal{P}(B)$, and for an infinitesimal volume $d\Omega$, $\mathcal{P}\approx 1-\lambda\, d\Omega$. The only function $\mathcal{P}(\Omega)$ with these properties is (2.5).

The probability for a given spacetime point $x=(t,\mathbf{x})$ to be in the inflationary phase is given by (2.5) with Ω being the volume occupied by false vacuum in the past light cone of x. For $H(t-t_0)\gg 1$, this is

$$\Omega\equiv\frac{4\pi}{3H^3}(t-t_0) . \tag{2.6}$$

(This equation is easily understood if we note that null geodesics continued to large negative values of t_0 asymptotically approach the horizon, which is a sphere of radius H^{-1}.) From (2.5) and (2.6), the fraction of space that is still inflating at time t is

$$f(t)\equiv\exp\left[-\frac{4\pi\lambda}{3H^3}(t-t_0)\right] . \tag{2.7}$$

It decreases with time and vanishes as $t\to+\infty$. However, the physical volume of inflating regions, $V(t)\propto e^{3Ht}f(t)$, grows with time. The reason is very simple: for sufficiently small λ the rate of expansion of false vacuum regions is greater than the rate of their decay.

An argument similar to that in Ref. [9] shows that inflating regions form a self-similar fractal of dimension $d<3$. This fractal dimension can be found from

$$f(t)=\left[\frac{H^{-1}}{R}\right]^{3-d} , \tag{2.8}$$

where H^{-1} is the size of the smallest bubbles and $R\approx H^{-1}\exp[H(t-t_0)]$ is the size of the largest bubbles (formed at $t\approx t_0$). A comparison of (2.7) and (2.8) gives

$$d=3-\frac{4\pi\lambda}{3H^4} . \tag{2.9}$$

The meaning of the fractal dimension is easy to understand. Consider a sphere of radius r centered on a point in an inflating region. As r is increased, the volume occupied by false vacuum inside the sphere grows (on average) like $V\propto r^d$. The deviation of d from 3 can be attributed to the fact that, as the sphere becomes larger, it is likely to include larger and larger bubbles. Of course, the inflating regions have a fractal nature only on scales $H^{-1}<r<R$. For $r>R$, $V\propto r^3$.

Let us now ask what happens if we remove the beginning of inflation to the infinite past, $t_0\to-\infty$. As I already mentioned in the Introduction we will not be concerned with the question of whether or not this model is realistic (it is not). The question is whether or not, by letting $t_0\to-\infty$, one obtains a consistent model of eternal inflation.

As $t_0\to-\infty$, the upper cutoff on the size of the bubbles is removed, $R\to\infty$. At the same time the probability (2.5) for a point to be in the inflationary phase and the fraction of space f occupied by the false vacuum both vanish. We note, however, that for a point $(\mathbf{x},t)$ in an inflating region there is a finite probability that inflation will continue for any given time interval Δt. This probability is given by

$$\mathcal{P}_{\Delta\Omega}=e^{-\lambda\Delta\Omega} , \tag{2.10}$$

where $\Delta\Omega$ is the four-volume between the past light cones originating at $(\mathbf{x},t)$ and $(\mathbf{x},t+\Delta t)$:

$$\Delta\Omega=\frac{4\pi}{3H^3}\Delta t . \tag{2.11}$$

The vanishing of f in Eq. (2.7) simply expresses the fact that an object of fractal dimension $d<3$ cannot fill the three-dimensional space; a randomly chosen point is most likely to be inside an infinitely large bubble. However, the physical volume occupied by the false vacuum is still increasing with time, and it may appear that we have a model of eternal inflation.

The trouble with this model is that the metric (2.2) does not cover the whole de Sitter space [12]. The spacetime described by this metric is geodesically incomplete. If a timelike geodesic is followed back in time, it reaches $t=-\infty$ in a finite proper time, indicating that the spacetime has an "edge" and can be continued. The full de Sitter space is covered by the metric

$$ds^2=d\tilde{t}^2-H^{-2}\cosh^2(H\tilde{t})d\Omega_3^2 , \qquad (2.12)$$

where $d\Omega_3^2$ is the metric on a unit three-sphere. This shows that the phase of exponential expansion at $\tilde{t}>0$ is preceded by a phase of exponential contraction at $\tilde{t}<0$. Of course, the contracting phase does not describe an inflating universe. If such a contracting universe were filled by a false vacuum, the nucleating bubbles would rapidly fill the space. The whole universe would thermalize and collapse to a singularity, without ever getting to the expanding phase.

The situation can be illustrated using a Penrose diagram. To make things simpler let us consider a $(1+1)$-dimensional version of (2.12);

$$ds^2=d\tilde{t}^2-H^{-2}\cosh^2(H\tilde{t})d\theta^2 . \qquad (2.13)$$

With a new time variable η, such that $d\eta=H\,d\tilde{t}/\cosh(H\tilde{t})$, the metric takes the form

$$ds^2=\frac{1}{H^2\cos^2\eta}(d\eta^2-d\theta^2) . \qquad (2.14)$$

Both variables η and θ have a finite range;

$$-\pi\le\theta\le\pi, \quad -\pi/2\le\eta\le\pi/2 , \qquad (2.15)$$

with $\theta=-\pi$ and $\theta=\pi$ identified. The corresponding Penrose diagram is shown in Fig. 1. Since the metric (2.14) is conformally flat, the null geodesics are $\theta=\pm\eta+\text{const}$ and are represented in the diagram by straight lines at 45° to the horizontal. A coordinate transformation

$$t=H^{-1}\ln\left[\frac{\sin\eta+\cos\theta}{\cos\eta}\right] , \quad x=H^{-1}\frac{\sin\theta}{\sin\eta+\cos\theta} , \qquad (2.16)$$

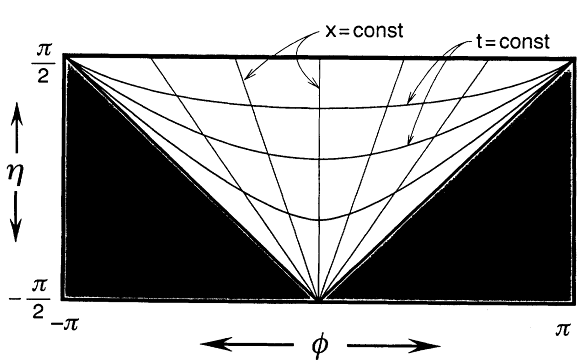

FIG. 1. Penrose diagram for de Sitter space. The lines of constant x and t in coordinates (2.17) are indicated. The shaded region is not covered by this coordinate system.

brings the metric (2.14) to a form similar to (2.2):

$$ds^2=dt^2-e^{2Ht}dx^2 . \qquad (2.17)$$

We see from (2.16) that $t\to-\infty$ corresponds to $\sin\eta+\cos\theta=0$, that is, $\eta=\pm\theta-\pi/2$. The spacetime region not covered by (2.17) is shaded in Fig. 1. The boundary of this region is the future light cone of a point at past timelike infinity.

The trajectories of expanding bubbles in de Sitter space are illustrated in Fig. 2. Bubbles nucleating at late times do not collide. The corresponding lines in the diagram hit the future infinity $\mathcal{J}^+(\eta=+\pi/2)$ before they intersect (see, e.g., the bubbles originating at points P and Q in the diagram). However, there is an infinite number of bubbles originating at past infinity $\mathcal{J}^-$ in the vicinity of $\eta=-\pi/2$ (since the spacetime volume of this region is infinite). These bubbles have no trouble colliding and rapidly fill the entire Universe, as illustrated in the figure.

In terms of the probability (2.10) for inflation to persist, a nonzero answer was obtained only because the two light cones bounding the volume $\Delta\Omega$ were cut off at the surface $t=-\infty$. In the full de Sitter space (2.14), $\Delta\Omega=\infty$ and $\mathcal{P}_{\Delta\Omega}=0$.

In our discussion so far we assumed that the part of spacetime occupied by false vacuum is exactly de Sitter. It has been conjectured [13] that, roughly speaking, all solutions of Einstein's equations with an energy-momentum tensor (2.1) asymptotically approach de Sitter space. This is the so-called cosmic no-hair conjecture. It is usually applied to the asymptotic future, but it should be equally valid for the asymptotic past. Particular versions of this conjecture with different assumptions about the initial state have been proved in Refs. [14–18]. In the most general case, the no-hair conjecture is certainly false. A simple counterexample is given by a small closed universe filled by gravitational waves which recollapses before the vacuum energy (2.1) can take over. A version of the conjecture that would be suitable for our purposes is that all solutions of Einstein's equations with $T_{\mu\nu}$ of the form (2.1) which are geodesically complete to the past are asymptotically de Sitter. The condition of geodesic completeness excludes singular spacetimes such as a closed collapsing universe. We shall not continue this line of argument, since, as we shall see in the following sections, eternal inflation can be ruled out without relying on spacetime being asymptotically de Sitter.

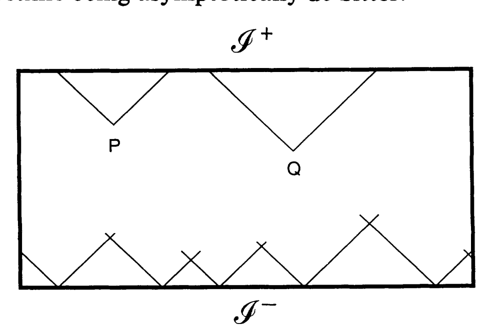

FIG. 2. Expanding bubbles in de Sitter space. Bubbles nucleating at points P and Q never collide, but the space is readily filled by the bubbles nucleating at $\mathcal{J}^-$.

III. CONDITIONS FOR ETERNAL INFLATION

The analysis in the previous section cannot be directly applied to realistic inflationary scenarios. In realistic models the false vacuum energy ρ is replaced by the scalar field potential $V(\varphi)$ which can vary in space and time. The field φ is usually assumed to be slowly varying and the spacetime to be locally close to de Sitter, but the global structure of spacetime can be quite different. Moreover, quantum nucleation of bubbles is replaced by a quantum random walk of the field φ, which is followed by thermalization when φ gets close to the bottom of the potential. One can still find the probability for inflation to persist at a given point for a specified period of time, but now this probability depends on the initial value of φ at that point. As a result, the locations of thermalization regions are strongly correlated (unlike the bubble nucleation sites). Finally, if the magnitude of $V(\varphi)$ gets near the Planck scale, the gravitational action may get significant quantum corrections, and classical Einstein equations can no longer be used.

In this section I shall try to formulate some necessary conditions that a spacetime should satisfy in order to describe an eternally inflating universe. I shall try to reduce to the minimum any assumptions about the dynamical laws that govern the evolution of geometry and of the scalar field φ. However, to make the discussion meaningful, we will have to assume that, to a reasonable approximation, the spacetime can be treated as a classical Riemannian manifold. Although eternal inflation is sometimes described as occurring at Planck scale, the description invariably relies on classical spacetime concepts such as "causality," "beginning," "end," etc. We shall also assume the spacetime to be causal, that is, to contain no closed nonspacelike curves.

In the previous section we saw that the spacetime (2.2) describing an inflating universe is geodesically incomplete. If eternal inflation is possible, one should be able to construct a complete spacetime which has the necessary properties of (2.2). It should be clear from the preceeding discussion that the essential property required for eternal inflation is a nonzero probability for inflation to continue at a given point for a specified interval of time. In the old inflationary scenario this probability is given by (2.10) and its nonzero value is guaranteed by the finiteness of the four-volume $\Delta\Omega$ in (2.11). To formulate the corresponding requirement in the general case, we shall assume that the boundaries of thermalized regions expand at a speed approaching the speed of light, like the walls of bubbles expanding in a false vacuum. More precisely, it will be assumed that a spacetime point P can be in an inflating region only if its past light cone contains no thermalized regions. I do not have any solid facts to justify this assumption, except that it is hard to imagine a different type of behavior in a relativistic theory. In addition, it will be assumed that the probability of forming thermalized regions does not vanish in the infinite past. Otherwise, it could be possible for the false vacuum to survive an infinitely long contraction phase. This possibility, however, is against the spirit of eternal inflation, which assumes a "steady-state" picture of the Universe. With this assumption, the probability of having no thermalized regions in an infinite volume vanishes, and we arrive at the following condition.

Condition 1. Let P and Q be two points in an inflating region, with Q to the future of P. Then the volume of the spacetime region between the past light cones of P and Q satisfies

$$\Delta\Omega < \infty \ . \tag{3.1}$$

The definition of $\Delta\Omega$ is somewhat ambiguous in cases when the light cones of P and Q intersect. In the general case, $\Delta\Omega$ should be understood as the volume of the difference between the pasts of P and Q.

I mentioned at the beginning of this section that a spacetime describing an inflating Universe is expected locally to be close to de Sitter space. This means that the Riemann tensor is approximately given by

$$R_{\mu\nu\sigma\tau} \approx H^2(x)(g_{\mu\sigma}g_{\nu\tau} - g_{\nu\sigma}g_{\mu\tau}) \ , \tag{3.2}$$

where $H(x)$ is a slowly varying function:

$$|\nabla H| \ll H^2 \ . \tag{3.3}$$

Corrections to (3.2) are expected to be of the order $(HL)^{-1}$, where L is the characteristic scale of variation of H. The magnitude of H at a given point is determined by the local value of the scalar field potential. For a classical picture of spactime to be valid, H must be bounded from above by some value $H_{\max}$ smaller than the Planck mass. Since inflation ends when φ gets sufficiently close to the minimum of $V(\varphi)$, it is clear that the magnitude of H in inflating regions should also be bounded from below. This leads to a condition 2.

Condition 2. The Riemann tensor in inflating regions is approximately given by (3.2) with

$$H_{\max} > H(x) > H_0 \ , \tag{3.4}$$

where H_0 and $H_{\max}$ are positive constants.

The question we would like to address in the following sections is whether or not a spacetime can be geodesically complete to the past and satisfy conditions 1 and 2. The analysis is much simpler in the two-dimensional case which is discussed in the next section. The realistic case of a four-dimensional spacetime in studied in Sec. V.

IV. NO ETERNAL INFLATION IN TWO DIMENSIONS

In a two-dimensional spacetime the metric can always be brought to a conformally flat form

$$ds^2 = \tilde{C}(x,t)(dt^2 - dx^2) = C(u,v)du\,dv \ , \tag{4.1}$$

where $u = t - x$, $v = t + x$. The null geodesics in metric (4.1) are lines of constant u and v. This makes it particularly convenient for the analysis of condition 1. Consider the light cones originating at points (u_0, v_0) and $(u_0 + \Delta, v_0 + \Delta)$ with an infinitesimal Δ. The two-dimensional version of condition 2 requires that the spacetime area between the two light cones be finite:

$$\Delta \int^{v_0} C(u_0, v)dv + \Delta \int^{u_0} C(u, v_0)du < \infty \ . \tag{4.2}$$

The lower bounds of integration cannot, in general, be set to $-\infty$ since the variables u and v can have a finite range. [See, for example, the two-dimensional de Sitter metric (2.14).]

Null geodesics have vanishing length, but the role similar to length is played by the affine parameter, p, which is defined up to multiplication by an arbitrary constant [12]. In terms of this parameter, the null geodesic equation has the standard form

$$x^{\mu''}+\Gamma^{\mu}_{\nu\sigma}x^{\nu'}x^{\sigma'}=0 \ , \tag{4.3}$$

where primes stand for differentiation with respect to p. For the geodesic $u=u_0$, $v=v(p)$ in metric (4.1) this reduces to

$$[C(u_0,v)v']'=0 \ , \tag{4.4}$$

and thus

$$dp=C(u_0,v)dv \ . \tag{4.5}$$

Similarly, for the geodesic $v=v_0$, $dp=C(u,v_0)du$. Now, if the geodesics are indefinitely continued to the past, null geodesic completeness requires that $p\to-\infty$. This implies that the integrals in (4.2) are divergent, and thus condition 1 cannot be satisfied.

In the above analysis I have implicitly assumed that the null geodesics $u=u_0$ and $v=v_0$ do not intersect again in the past of the point (u_0,v_0). If they did, then the integration in (4.2) would have to be cut off at the point of intersection, and the resulting area would be finite. A simple example of this sort is a flat spacetime (4.1) with $C=1$ and topology of $S_1\times R$, so that points (x,t) and $(x+L,t)$ are identified and $t=$const sections are circles of circumstance L. It is clear, however, that this "counterexample" has nothing to do with eternal inflation: since light rays can run around the Universe and come back, the whole Universe can be thermalized by a single "bubble" in a finite time.

Turning now to condition 2, it will be shown that this condition can be satisfied only if the Universe has an initial contracting phase. Let us consider a congruence of past-directed timelike geodesics emanating from some point P in an inflating region. We can construct a "synchronous" coordinate system by choosing these geodesics as lines of constant x and choose the time coordinate t to be the proper time along the geodesics with $t=0$ at P (and $t<0$ to the past of P). The constant time lines will then be the curves perpendicular to the congruence and the metric will have the form

$$ds^2=dt^2-a^2(x,t)dx^2 \ . \tag{4.6}$$

It will not be necessary to assume that this coordinate system covers the whole interior of the light cone, and we will not be concerned about the range of the coordinate x.

The coordinates (4.6) are singular at $t=0$, where all geodesics meet and

$$a(x,0)=0 \ . \tag{4.7}$$

Since the geodesics are converging at point P we must also have

$$\dot{a}(x,0)<0 \ , \tag{4.8}$$

where an overdot represents differentiation with respect to t. Similar singularities could generally occur in the past of P if some of the geodesics crossed again, but we will see that if condition 2 is satisfied, this never happens and the coordinates (4.6) can be extended to the infinite past.

The Riemann tensor in two dimensions is exactly of the form (3.2) with

$$H^2=\ddot{a}/a \ , \tag{4.9}$$

and the condition (3.4) reduces to

$$\ddot{a}>H_0^2 a \ . \tag{4.10}$$

Now, from Eqs. (4.10), (4.7), and (4.8) it is not difficult to show that, for $t<0$,

$$a(x,t)>H_0^{-1}\dot{a}(x,0)\sinh(H_0 t) \ . \tag{4.11}$$

Equation (4.11) shows that the particles represented by the geodesics spread arbitrarily far apart as $t\to-\infty$. They spread faster than they would in flat spacetime, where we would have $a(t)\propto|t|$, indicating that at early times the Universe is contracting from an infinite size. It also follows from (4.11) that $a(x,t)$ is always positive at $t<0$, and thus the geodesics never cross.

We conclude that if a two-dimensional universe is geodesically complete to the past, then (i) condition 1 cannot be satisfied and (ii) condition 2 can be satisfied only if the Universe has an initial contracting phase.

V. FOUR DIMENSIONS

Let us now turn to the realistic case of inflation in four dimensions. The analysis here is more complicated, and I will only be able to give a plausibility argument, rather than a proof, that $\Delta\Omega=\infty$. Starting with condition 1, we consider the spacetime volume between the past light cones of two points P and Q with a small timelike separation. As before, we begin by setting up a convenient coordinate system. If γ is a timelike curve connecting P and Q, we can choose one of the coordinates, u, so that the surfaces of constant u are past light cones of points on γ. Each light cone is a congruence of null geodesics, and we can choose the remaining coordinates x^1, x^2, and v so that $x^i=$const on the geodesics and v is an affine parameter along the geodesics. This determines v up to a transformation

$$v\to f(u,x^i)v+g(u,x^i) \ . \tag{5.1}$$

Noting that the null geodesics are normal to the surfaces $u=$const, we can write the metric in the form

$$ds^2=g_{uu}du^2+2g_{uv}du\,dv+2g_{ui}du\,dx^i+g_{ij}dx^i dx^j \ . \tag{5.2}$$

Using the null geodesic equation and the fact that v is an affine parameter, it is easily verified that

$$\frac{\partial g_{uv}}{\partial v}=0 \ . \tag{5.3}$$

Finally, we can use the freedom of transformation (5.1) to set

$$g_{uv}=1 \ . \tag{5.4}$$

The determinant of the metric (5.2), (5.4) is

$$g = -{}^{(2)}g \ , \tag{5.5}$$

where ${}^{(2)}g = \det(g_{ij})$, and the spacetime volume between the light cones $u = u_0$ and $u = u_0 + \Delta$ is

$$\Delta\Omega = \Delta \int_{u=u_0} dv\, dx^1 dx^2 \sqrt{-g} = \Delta \int dv \mathcal{A} \ . \tag{5.6}$$

Here,

$$\mathcal{A}(u_0, v) = \int \sqrt{{}^{(2)}g}\; d^2x \tag{5.7}$$

is the area of the surface of constant v on the light cone $u = u_0$. Now, since v is an affine parameter and we assume the spacetime to be geodesically complete to the past, v must have an infinite range. Hence, for $\Delta\Omega$ to be finite, the wave-front area $\mathcal{A}$ must vanish [19] faster than $|v|^{-1}$ as $v \to -\infty$.

The vector $\delta x^\mu(v)$ connecting points with the same values of v on a pair of nearby geodesics on the light cone satisfies the geodesic deviation equation,

$$\frac{D^2}{dv^2}\delta x^\mu = R^\mu_{\ \nu\sigma\rho} \frac{dx^\nu}{dv}\frac{dx^\sigma}{dv}\delta x^\rho \ . \tag{5.8}$$

The initial conditions for $\delta x^\mu(v)$ at $v = 0$ are

$$\delta x^\mu(0) = 0, \quad \frac{d}{dv}\delta x^\mu(0) \equiv \kappa^\mu \neq 0 \ . \tag{5.9}$$

With the Riemann tensor from (3.2), the right-hand side of (5.8) vanishes, and the geodesics diverge linearly with v, $\delta x^\mu = \kappa v$. This suggests that the wave-front area in (5.7) is quadratically divergent, $\mathcal{A} \propto v^2$, and that $\Delta\Omega$ in (5.6) is infinite.

It should be noted, however, that the form (3.2) of the Riemann tensor cannot be exact for $H(x) \neq$const. Small corrections to this form can become significant over a large distance along the geodesics, and can in principle modify the character of geodesic divergence. Unfortunately, I was unable to prove that this does not happen in a spacetime satisfying condition 2. Compared to the two-dimensional case, the situation is complicated by the fact that null geodesics can cross and form caustics, and the coordinate system (5.2) cannot generally be extended to the asymptotic region. (In two dimensions, the light cone at each point consists of two null rays travelling in opposite directions, and null geodesics pointing in the same direction never cross).

The behavior of timelike geodesics in a spacetime with Riemann tensor (3.2) can be studied as was done in Sec. IV for the two-dimensional case. One finds that the separation of nearby geodesics diverges exponentially with the proper time τ as $\tau \to -\infty$. Details are given in Appendix A. Faster-than-linear divergence of geodesics indicates an initial contracting phase. Since past-directed geodesics originating at point P are all contained within the past light cone of P, the rapid divergence of geodesics suggests that the wave-front area $\mathcal{A}$ should also diverge. This argument, however, falls short of a proof, since one can imagine a sequence of three-surfaces of increasing three-volume, which are bounded by two-surfaces of decreasing area. (A lower-dimensional analogy would be a balloon of increasing radius with an opening of decreasing circumference.) A proof of $\Delta\Omega = \infty$ at the same level of rigor as in Sec. 4 for the two-dimensional case remains a problem for future research.

VI. DISCUSSION

I have presented arguments indicating that the Universe cannot be in a state of eternal inflation, with no beginning and no end. It appears that inflation, when continued to the past, is necessarily preceded by a period of contraction, during which regions where false vacuum energy is thermalized would merge and fill the entire Universe. No false vacuum would then survive, and the Universe would never start inflating.

Thus, inflation does not help to avoid the problem of initial singularity. If an inflating Universe is followed back in time, we must arrive at some hypersurface Σ which, from the point of view of classical spacetime, represents the beginning of the Universe. What could be the nature of this initial hypersurface? It could be the surface where the curvature becomes Planckian, and the classical concepts of space and time are no longer applicable. Alternatively, the origin of the Universe could be a quantum tunneling event described by an instanton with a curvature well below Planckian. Then the surface Σ corresponds to the state of the Universe at nucleation and can be determined from the instanton. Variations of this idea have been discussed in Refs. [21–23]. In both cases Σ should not be pictured as a mathematical surface, but rather as a somewhat fuzzy boundary region corresponding to a transition from classical to quantum description of spacetime.

Although eternal inflation with no beginning appears to be impossible, I would like to emphasize that inflation can be, and typically is semieternal, that is, eternal only to the future. We thus have a peculiar situation where the classical spacetime of the Universe must have a beginning, but probably has no end. As inflation continues, new regions of thermalization are formed, and we live in one of these regions. Our region is likely to be at a very large (but finite) distance from the hypersurface Σ.

APPENDIX

In this appendix we shall discuss the divergence of timelike geodesics in a spacetime with Riemann tensor of the form (3.2). Let us consider a congruence of past-directed timelike geodesics originating at some point P. The geodesics can be parametrized by the proper time τ with $\tau=0$ at P and $\tau<0$ to the past of P.

The vector $\delta x^\mu(\tau)$ connecting points with the same values of τ on a pair of nearby geodesics satisfies the geodesic deviation equation,

$$\frac{\mathcal{D}^2}{d\tau^2}\delta x^\mu = R^\mu_{\ \nu\sigma\rho}\frac{dx^\nu}{d\tau}\frac{dx^\sigma}{d\tau}\delta x^\rho \ , \tag{A1}$$

where $\mathcal{D}/d\tau$ stands for a covariant derivative. Since $\mathcal{D}^2 x^\mu/d\tau^2 = 0$, it follows from (A1) that

$$\frac{\mathcal{D}^2}{d\tau^2}\left[\frac{dx_\mu}{d\tau}\delta x^\mu\right] = 0 \ , \tag{A2}$$

In the case under consideration $(dx^\mu/d\tau)\delta x^\mu$ and its first derivative are both equal to zero at $\tau=0$, and Eq. (A2) ensures that δx^μ remains orthogonal to $dx^\mu/d\tau$ at all τ,

$$\frac{dx^\mu}{d\tau}\delta x_\mu = 0 . \tag{A3}$$

Now, with the Riemann tensor from (3.2), Eqs. (A1) and (A3) reduce to

$$\frac{\mathcal{D}^2}{d\tau^2}\delta x^\mu = H^2(\tau)\delta x^\mu , \tag{A4}$$

where $H(\tau)\equiv H(x(\tau))$. This equation is solved by the ansatz

$$\delta x^\mu(\tau) = F(\tau)\delta x_0^\mu(\tau) , \tag{A5}$$

where the function $F(\tau)$ satisfies the equation

$$F''(\tau) = H^2(\tau)F(\tau) , \tag{A6}$$

the primes denote derivatives with respect to τ, and the vector $\delta x_0^\mu(\tau)$ is parallel transported along the geodesic $x(\tau)$,

$$\frac{\mathcal{D}}{d\tau}\delta x_0^\mu(\tau) = 0 . \tag{A7}$$

Since parallel transport preserves the norm of a vector, $|\delta x_0^\mu(\tau)| = \text{const}$ and $F(\tau)$ describes the change in the magnitude of $\delta x^\mu(\tau)$. The initial conditions for Eq. (A6) at $\tau=0$ are

$$F(0)=0 , \quad F'(0)<0 , \tag{A8}$$

where the second condition follows from the fact that the geodesics are converging at $\tau=0$.

Comparing Eqs. (A6), (A8) with (4.7), (4.8), (4.10), we see that they are formally equivalent, and the same argument as in Sec. IV yields

$$F(\tau) > H_0^{-1}F'(0)\sinh(H_0\tau) . \tag{A9}$$

For a slowly-varying function $H(\tau)$ with $|H'| \ll H^2$ we can go further and solve Eq. (A6) using a WKB-type approximation,

$$F(\tau) \approx H_0^{-1}F'(0)\sinh\left[\int_0^\tau H(\xi)d\xi\right] . \tag{A10}$$

It is also clear from Eqs. (A9) and (A10) that nearby geodesics exponentially diverge toward the past. Faster-than-linear divergence of the geodesics indicates that the Universe is contracting at early times.

As already noted, small corrections to the form (3.2) of the Riemann tensor can become significant over a large distance along the geodesics, and their effect on the geodesic divergence should be investigated.

ACKNOWLEDGMENTS

It is a pleasure to thank John Preskill and the High Energy Physics group at Caltech for hospitality during the period when this work was being done. I also benefited from discussions with Jaume Garriga, Stephen Hawking, Gary Horowitz, and Esteban Calzetta. This work was supported in part by U.S. Department of Energy under Contract No. DEAC-03-81ER40050 and by the National Science Foundation under Contract No. 8605578.

[1] For a review of inflation see, e.g., E. W. Kolb and M. S. Turner, *The Early Universe* (Addison-Wesley, New York, 1990); S. K. Blau and A. H. Guth, in *300 Years of Gravitation,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1987); A. D. Linde, *Particle Physics and Inflationary Cosmology* (Harwood Academic, Chur, Switzerland, 1990).

[2] A. Vilenkin, Phys. Rev. D **27**, 2848 (1983).

[3] A. A. Starobinsky, in *Field Theory, Quantum Gravity and Strings,* Proceedings of the Seminar Series, Meudon and Paris, France, 1984–1985, edited by M. J. de Vega and N. Sanchez, Lecture Notes in Physics Vol. 246 (Springer-Verlag, New York, 1986).

[4] P. J. Steinhardt, in *The Very Early Universe,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University Press, Cambridge, England, 1983).

[5] A. D. Linde, Cambridge University report, 1982 (unpublished).

[6] A. D. Linde, in *The Very Early Universe* [4].

[7] A. D. Linde, Mod. Phys. Lett. A **1**, 81 (1986); Phys. Lett. B **175**, 395 (1986).

[8] A. S. Goncharov and A. D. Linde, Zh. Eksp. Teor. Fiz. **92**, 1137 (1987) [Sov. Phys. JETP **65**, 635 (1987)]; A. S. Goncharov, A. D. Linde, and V. F. Mukhanov, Int. J. Mod. Phys. A **2**, 561 (1987); see also A. D. Linde in Ref. [1].

[9] M. Aryal and A. Vilenkin, Phys. Lett. B **199**, 351 (1987).

[10] A. H. Guth, Phys. Rev. D **23**, 347 (1981).

[11] A. H. Guth and E. J. Weinberg, Phys. Rev. D **23**, 876 (1981).

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

[13] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).

[14] W. Boucher and G. W. Gibbons, in *The Very Early Universe* [4].

[15] J. D. Barrow, in *The Very Early Universe* [4].

[16] A. A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. **37**, 55 (1983) [Sov. Phys. JETP Lett. **37**, 66 (1983)].

[17] R. M. Wald, Phys. Rev. D **28**, 2118 (1983).

[18] L. G. Jensen and J. A. Stein-Schabes, Phys. Rev. D **35**, 1146 (1986).

[19] Note that in the incomplete spacetime (2.2), which served as a prototype for the idea of eternal inflation, the light rays asymptotically approach the horizon and $\mathcal{A}\to 4\pi H^{-2}$ as $t\to -\infty$.

[20] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[21] A. Vilenkin, Phys. Lett. **117B**, 25 (1982); Phys. Rev. D **30**, 509 (1984).

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

[23] A. D. Linde, Nuovo Cimento Lett. **39**, 401 (1984).

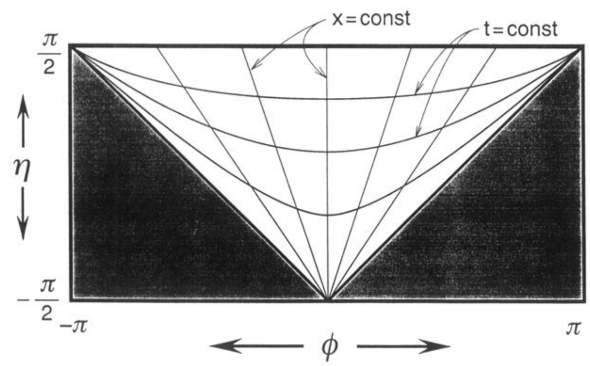

FIG. 1. Penrose diagram for de Sitter space. The lines of constant x and t in coordinates (2.17) are indicated. The shaded region is not covered by this coordinate system.