Initial conditions for bubble universes

Brett McInnes*
Department of Mathematics, National University of Singapore, 2 Science Drive 2, Singapore 117543, Republic of Singapore
(Received 6 March 2008; published 23 June 2008)

The "bubble universes" of Coleman and De Luccia play a crucial role in string cosmology. Since our own Universe is supposed to be of this kind, bubble cosmology should supply definite answers to the long-standing questions regarding cosmological initial conditions. In particular, it must explain how an initial singularity is avoided, and also how the initial conditions for inflation were established. I argue that the simplest nonanthropic approach to these problems involves a requirement that the spatial sections defined by distinguished bubble observers should not be allowed to have arbitrarily small volumes. Casimir energy is a popular candidate for a quantum effect which can ensure this, but (because it violates energy conditions) there is a danger that it could lead to nonperturbative instabilities in string theory. I make a simple proposal for the initial conditions of a bubble universe, and show that my proposal ensures that the system is nonperturbatively stable. Thus, low-entropy conditions can be established at the beginning of a bubble universe without violating the second law of thermodynamics and without leading to instability in string theory. These conditions are *inherited* from the ambient spacetime.

DOI: 10.1103/PhysRevD.77.123530 PACS numbers: 98.80.Cq, 04.20.Gz

I. GETTING INFLATION STARTED IN A BUBBLE

In string theory, the leading approach to the problem of the cosmological constant is given by the *landscape* [1]. String theory gives a consistent account of a set of possible universes which are so numerous— 10^{500} is the standard estimate—and have values of the cosmological constant spaced in such a way, that the value we actually observe ceases to seem surprising. Instead we conclude that our Universe corresponds to a point in the landscape.

The mathematical consistency of landscape universes does not suffice to solve the cosmological constant problem: one needs to explain how such a vast array of possible worlds actually comes into existence. This is achieved by means of the nucleation of *Coleman–De Luccia bubbles* [2–4]. These are bubbles of "true" vacuum which spontaneously arise within a larger spacetime containing a scalar field which is initially in a "false" vacuum state. With a suitable potential for the scalar, and with the usual assumptions ("potential domination") regarding the initial conditions for the inflaton, such bubbles can be made compatible with the standard inflationary account of the evolution of a universe like ours. This is the "open inflation" scenario [5], which works quite well in bubble universes—*provided, of course, that inflation can actually begin inside a bubble*: something which is by no means obvious, since the precise nature of inflationary initial conditions remains to be fully understood.

Indeed, if bubble nucleation is to be taken seriously as an account of the origin of our Universe, then it must be expected to answer all of the long-standing questions regarding cosmological initial conditions. In particular, it should supply answers to the following fundamental questions:

(i) Was the beginning singular? If not, how are the singularity theorems evaded?

(ii) The second law of thermodynamics dictates that the Universe began in an extremely low-entropy state. How was that arranged? Particularly: how does one enforce the *very special* conditions needed for inflation to start [6–8]?

The first of these questions requires no elaboration. The second question concerns the "specialness" of the initial conditions of *our* Universe. This specialness (or "nongenericity") is still manifested, even after the passage of more than 13×10^9 years, as an *arrow of time* [9–12]. The point is that a truly generic initial state would be dominated by black holes[1] [16]. But such an initial state would not evolve to a Universe like ours, with its extremely strong past/future asymmetry. A crucial instance of this is that inflation cannot begin with such initial conditions; in fact, as Albrecht [7] and others have stressed, inflation can only begin if the inflaton is itself initially in a very specific state, in which extremely few of the scalar field degrees of freedom have yet been excited. If we cannot produce a theory which *necessarily entails* such extraordinarily nongeneric initial conditions for at least some universes, then we will not be able to find a universe, even in the landscape, which remotely resembles our own. This was discussed at

*matmcinn@nus.edu.sg

[1]It has been argued [13] that this is not true of a universe which begins along a noncompact spatial hypersurface. Here I shall avoid this controversial question by postulating that the *original* Universe was created from "nothing" [14,15] along a *compact* hypersurface. In this case, a singularity-dominated beginning would indeed have been generic.

length in [17]. (See [13,18–22] for various theories of the origin of the arrow.)

I stress that settling this question is no mere technicality. Recently it has become clear that uncertainty as to the precise nature of the inflationary initial conditions has concrete consequences even for the interpretation of future *observational* data. Inflation can lose its ability to predict certain observational signatures if one weakens the usual assumptions regarding the preinflationary spacetime geometry [23].

The bubble universe theory addresses the first of my questions in a surprising way. As is well known, inflation is usually not past eternal [24]: an inflating region of spacetime must be geodesically incomplete to the past if its average rate of expansion is positive. A cautious interpretation of this fact [25] is that any inflationary history must be preceded by some unspecified but radically different state. This is the appropriate statement of the conclusion, because geodesic incompleteness can have a variety of physical interpretations. The most familiar interpretation is that incompleteness signals a singularity, but this is not necessarily the correct interpretation in the case of bubble universes.

Aguirre and Gratton have examined this question in the context of their theory of the arrow. (See [26] for a survey.) In fact, their analysis applies quite straightforwardly also to thin-walled Coleman–De Luccia bubbles. Their general argument implies that these bubbles are geodesically incomplete to the past, in the manner dictated by the Borde-Guth-Vilenkin results, simply because the description of the bubble interior given by the *distinguished cosmological observers* inside the bubble cannot be extended arbitrarily far into the full spacetime. In the case where the bubble wall is infinitely thin, the explanation for this is simple: the spacelike surfaces defined by these observers can approach arbitrarily close to a null surface, so their volumes (or rather the volumes of compact sets they contain) *shrink towards zero* at a finite proper time to the past of any event inside the bubble. We shall see later that this shrinkage to zero size actually persists even when the bubble wall is not infinitely thin: it is a generic property of bubble interiors under certain very mild conditions. Normally, such a situation would entail the existence of a singularity; but it turns out that the equation of state of the scalar field is such that zero volume *does not* imply infinite energy density. Thus, in this case, the incompleteness signals that the bubble spacetime is extensible, *not* singular; it can be extended, via the bubble wall, into the ambient spacetime. This is how the bubble universe proposal deals with the singularity problem.

This solution of the singularity problem can be questioned: clearly it depends on very strong assumptions about the exact matter content of the prenucleation spacetime. If, for example, the ambient spacetime contains other fields or objects, one will need to investigate their effects if they are absorbed by the bubble and encounter the zero-volume spatial slice; also, quantum effects (such as the Casimir effect) may alter the classical picture of the matter content in a decisive way. In the case of an infinitely thin bubble wall, incursions by external objects could be disastrous. The most dramatic example of such incursions involves the collision of a bubble with *another* bubble. The problem of understanding these questions in that context is currently the subject of intensive investigation; see the detailed discussions given in [27,28].

Leaving these complications aside for the moment, I can summarize by saying that the simplest versions of bubble universes offer an approach to the singularity problem simply by arguing that the earliest form of matter (necessarily) had an unusual equation of state, such that its energy density was not related to spatial volume in the familiar way. This permits an interpretation of the Borde-Guth-Vilenkin results in a way that does *not* involve singularities.

An answer to the first of the questions should set the scene for an answer to the second: since the zero-volume state[2] at the beginning of the bubble universe is *not* singular, *there is no obstruction to relating the thermodynamic conditions in the early bubble universe to conditions in the ambient spacetime*.

Thus, the problem of cosmological initial conditions can only be addressed, in the bubble universe context, by applying the second law of thermodynamics to the bubble nucleation process. In this work I argue that this suggests a small but significant modification of the usual approach to bubble nucleation theory. The idea that even exponentially suppressed corrections to the Coleman–De Luccia instanton can be important has been advocated by Buniy, Hsu, and Zee [29]; here I consider much less drastic modifications, which alter the bubble geometry only at the very earliest (bubble) times.

General aspects of applying the second law to bubble nucleation are explained in Sec. II. The key point here is that, for a bubble universe to resemble our own, its initial total entropy must be low as seen by the *distinguished observers* inside the bubble—the observers to whom the spatial geometry appears to be isotropic. But this is very difficult to arrange, because these same observers are the ones whose spatial sections shrink to zero volume as they probe backwards in time, and small spatial sections have a *very strong tendency to be anisotropic*. This key point will be reviewed in some detail. I argue that the most natural—though perhaps not the only—way of avoiding this problem is to find some means of avoiding a zero-volume ''initial'' state for a bubble universe.

[2]Strictly, I should say that the spatial sections have volumes which can be made *arbitrarily small*, not exactly zero; the distinction is however not important here.

In Sec. III, I discuss in detail the way in which a standard Coleman–De Luccia bubble universe avoids being (initially) singular and develops an arrow of time. I focus first on the case of negative vacuum energy inside the bubble, since the points I am making can be seen most clearly in that case (which does occur in the landscape, in the form of ''terminal vacua''). Using singularity theory, I show that the zero-volume spatial section will also be present in the case of positive vacuum energy. It can only be avoided by modifying the bubble universe in a way that violates the *null Ricci condition* or NRC. This is the statement that the Ricci tensor satisfies

$$R_{\mu\nu}n^{\mu}n^{\nu} \geq 0 \qquad (1)$$

at all points in spacetime and for all null vectors n^{μ}; it is equivalent to the *null energy condition* or NEC in cases where corrections to the Einstein equations can be neglected. If such corrections are important then the NRC can be violated even if the NEC is satisfied; in such a case we may speak of *effective* violations of the NEC.[3] My conclusion is that the required modification violates the NEC, though perhaps only effectively.

Real and effective NEC violations in string cosmology have been discussed in [31–33], and have recently attracted much more interest [34,35]. The perennial concern with regard to NEC violation is the possibility that it might lead to some kind of fatal instability [36]. Arkani-Hamed *et al.* [37,38] argue that NEC violation is not acceptable in string theory *except* when it is global and quantum mechanical, as in the case of the *Casimir effect* [39], or in other very special conditions (such as those associated with orbifold planes). However, even in these cases one must also take into account certain *nonperturbative* string effects, because it has been shown that these frequently do lead to problems when NEC violation occurs. In particular, we have to take into account the brane-antibrane pair-production instability analyzed by Seiberg and Witten [40] and subsequently by Maldacena and Maoz [41] and by Kleban *et al.* [42]. This instability means that NEC violation—even if it is only *effective*—is *not* always physically acceptable even in the cases where it does not lead to problems at the perturbative level.

In Sec. IV, I examine a particular model in which the interior of a bubble universe begins, with the aid of the Casimir effect, along a surface of nonzero minimal volume. I am able to show that, despite the violation of the NEC entailed by Casimir energies, the spacetime narrowly avoids becoming unstable in the Seiberg-Witten sense. Thus I have a toy model of a bubble universe which has satisfactory initial conditions for inflation; it is able to inherit an arrow of time.

I stress that the metric I find is asymptotic to one of the metrics normally used to describe bubble interiors, and differs substantially from such a metric only for an extremely short time. Thus my conclusions do not invalidate the large recent literature on eternal inflation in any way; nor, of course, am I suggesting that there is anything erroneous in the original Coleman–De Luccia analysis. The objective is simply to show that the bubble universes that populate the landscape can in fact have initial conditions similar to those of our own Universe.

II. BUBBLE NUCLEATION RESPECTS THE SECOND LAW

In this section I construct a very general argument to the effect that the second law of thermodynamics has specific consequences for the spatial geometry of the very earliest phase of a bubble universe.

First, note that, unlike the baby universes considered by Farhi and Guth [43] (see also [44]), a Coleman–De Luccia bubble is *not* isolated from the original spacetime: on the contrary, the bubble expands into the ambient universe and is permanently exposed to signals from it. Indeed, to a family of observers inside a bubble which nucleates in an approximately Minkowski spacetime, *the entire exterior spacetime lies to the past*. The second law of thermodynamics now has the following major consequence: *we cannot simply ''reset'' the initial conditions inside the bubble to suit ourselves.* The initial thermodynamic state of a bubble is set by the outside conditions and by what happens as one moves through the wall. In this connection, one should not expect the bubble wall to preserve all highly ordered structures it encounters—let alone generate them. That is, passage through the wall could lead to a dramatic increase of certain kinds of entropy. This is consistent with Coleman and De Luccia's description of passing into such a bubble as ''the ultimate ecological catastrophe.''

It follows from these simple observations that, if the bubble interior has extremely low initial entropy, this can only be a result of *inheriting* that condition from the ambient spacetime. Answering my question then amounts to establishing the following two statements:

(i) The ambient spacetime had extremely low entropy.

(ii) The inevitable increase in the entropy caused by bubble nucleation does not appear to be large as seen by an internal observer: low entropy is heritable.

One way of approaching the first point was proposed in [18]; in [19] I addressed it in a different way, by arguing that bubble universes nucleate in a ''mother universe'' which itself is the result of ''creation from nothing,'' after the manner of Vilenkin [14] and Ooguri *et al.* [15]. With a suitable spatial topology, one can use deep theorems from global differential geometry to argue that the original uni-

[3] That is, violation of the NRC amounts to violating the NEC for the effective stress-energy-momentum tensor obtained by absorbing the corrections into the physical stress-energy-momentum tensor. My main example (the Casimir effect) involves true NEC violation, but the distinction being made here is important, and should be borne in mind; see for example [30].

verse necessarily had a perfectly (locally) isotropic initial spatial section. This means that the initial *gravitational* entropy[4] was (necessarily) as low as possible, and indeed this is precisely why the *total* entropy of this initial universe was low [9,10]: extreme isotropy rules out black holes, which would otherwise strongly dominate the entropy accounting in a spatially compact universe. The gravitational entropy then increases due to the usual inflationary fluctuations, which mar the perfect geometric regularity of the very earliest spatial sections—if only to a microscopic degree.

This brings me to the second point. The second law dictates that the gravitational entropy cannot decrease during the bubble nucleation process. The question now is: what form will the increase take, *as seen by interior observers*? It is important to understand here that the bubble interior differs, in one crucial particular, from Minkowski or (anti) de Sitter spacetime. These latter spacetimes have very large (in fact, maximal) isometry groups, corresponding to their extremely simple matter contents. It follows that they do *not* have distinguished families of observers, as a generic Friedmann-Robertson-Walker (FRW) cosmological model does. Thus, for example, (regions of) de Sitter spacetime can be foliated in many different ways by spacelike surfaces having a variety of intrinsic geometries, and none of these foliations has a preferred status; for all of them correspond to observers who *see the same thing*, namely, isotropic dark energy with a particular invariant energy density. By sharp contrast, the surfaces of approximately constant scalar energy density inside the bubble *do* distinguish a special class of observers. *These* observers are the ones who, using whatever coarse graining they find appropriate, must deduce very low-entropy conditions in *their* earliest history, if the bubble universe is to have an arrow of time. This is another sense—apart from the ''ultimate ecological catastrophe'' aspect—in which the interior of a bubble universe is *not* analogous to (say) the interior of a forward light cone in de Sitter spacetime. The bubble universe contains distinguished observers whose (coarse-grained) observations are what we have to explain.

I begin my investigation of this question by noting that, for inflation to start, what is really needed is low *gravitational* entropy: if the spatial sections are too irregular, this will not be consistent with the required initial conditions for the inflaton. *Other* forms of entropy, such as the entropy of the Gibbons-Hawking radiation [47] associated with a cosmological horizon, will actually increase substantially during bubble nucleation, but this will not interfere with the inflaton initial conditions. (Nor, however, will it help to establish the particular form one needs for these conditions.) This observation refines my question considerably.

Now as we have seen, the characteristic property of the earliest spatial sections inside the bubble is that their volume scales are arbitrarily small; this is the proper interpretation of the Borde-Guth-Vilenkin results [26]. *But one does not expect "small" spatial sections to correspond to low gravitational entropy.* This can be explained as follows. Just as inflation leads an observer to think that his spatial sections have become smoother,[5] so also an unlimited *contraction* of a spatial section will make any irregularities more and more apparent. To put this another way, suppose that we consider the history of a small spatial patch in the present Universe. As we trace it back in time, we will see it becoming less and less isotropic around a generic point.

To see this, one needs to study the effect of including anisotropy in the spacetime dynamics. The anisotropy contributes a term of the form $C/a(t)^6$ to the field equations, where $a(t)$ is the scale factor and C is a constant. (This is explained extremely clearly in [49], which should be consulted for the details.[6]) This means that, as we go back in time, the anisotropy grows much more rapidly than the energy density of ordinary matter and radiation (or of any kind of dark energy), so it dominates the dynamics if the sections become sufficiently small, at least in the absence of very exotic forms of matter.

This discussion explains why one does not expect a *realistic* zero-volume state, such as that of a big crunch or a black hole, to be geometrically regular, though of course the intrinsic geometry of the late spatial sections is very regular in *idealized* FRW or Schwarzschild spacetimes. *Generically, spatial sections with very small volume scales correspond to high gravitational entropy.* It follows that observers inside a bubble universe, having deduced from observations that inflation took place, will have to conclude that the initial conditions for inflation were made possible by an *infinite* fine-tuning at the zero-volume state—infinite in the sense that the $C/a(t)^6$ term can only be ignored, if $a(t)$ really vanishes, if C is set *exactly* equal to zero. If they are aware that they live in a bubble, they will be forced to conclude that they owe their existence to a massive violation[7] of the second law. (They will realize that the Gibbons-Hawking entropy has increased, but since anisotropy generically completely dominates the dark en-

[4]The concept of gravitational entropy has not yet been made entirely precise: see, for example, [45,46]. That gravitational systems behave consistently with the second law is nevertheless not in doubt, and this is all we need here.

[5]See [48] for the precise statement of ''cosmic baldness,'' and [17] for a discussion.

[6]Generically these irregularities are of the kind originally discussed by Belinsky, Khalatnikov, and Lifschitz—see [50,51] for recent detailed discussions.

[7]The second law, being statistical, can of course be ''violated''; but the dire consequences of assuming that the current status of our Universe can be explained in that way are well known: see [52–55] for a survey.

ergy density, they will not be able to explain the situation by using this fact.)

There are basically four ways to deal with this problem. The first is to postulate the presence of some kind of matter with an energy density that grows even more rapidly, as volumes shrink towards zero, than the anisotropy; for example, a scalar field with a very negative potential [49,56]. One might then try to arrange for the growth of the entropy to be diverted away from the spatial geometry and into the scalar field. While this idea deserves (and requires) further development, it does not appear to be compatible with the landscape picture, and I shall not consider it further.

The second approach is the usual one: we simply ignore the effects of its environment on the bubble, and use idealized models of the geometry. Recently, however, it has been recognized [28,57–61] that *collisions* of bubbles are of the utmost importance, since, even if inflation is able to start in the aftermath, a permanent "memory" of the collision may be retained by the bubbles. Furthermore, the collisions release radiation into the ambient spacetime [61]. However, bubble collisions are just the most dramatic way in which the ambient spacetime can affect the initial conditions of a bubble universe. On a vastly smaller scale, the inflaton field in the ambient spacetime will suffer scalar and tensor perturbations; these may be tiny, but they must have *some* effect on the geometry of the bubble interior if they strike or are absorbed by the bubble. To suppose otherwise would, once again, amount to a violation of the second law of thermodynamics. These developments render obsolete the picture of a bubble existing in splendid isolation; we now have to think in terms of a bubble expanding in an environment where *it is constantly subjected to a bombardment of external signals of greater or lesser degrees of intensity.* Since the bubble initial conditions are "fine-tuned" (in the sense discussed earlier), it is hard to see how to justify ignoring these signals.

In a third approach, one might accept that external signals have these effects at a *generic* point on the earliest spatial slices of the bubble universe, but try to argue that there will always be *some* extremely atypical regions which remain undisturbed.[8] The observed Universe might have evolved from a tiny patch of this sort. This amounts to an invocation of the anthropic principle. Rather than become involved in the anthropic debate, I note instead that it is generally accepted that alternatives to that approach should always be fully investigated.

The fourth approach is the one to be explored here: one can try to *prevent* the spatial sections inside the bubble from ever being too small. This has to be done by considering small modifications of the Coleman–De Luccia analysis, taking into account effects previously neglected. By considering subleading corrections to the Coleman–De Luccia instanton, as advocated by Buniy *et al.* [29], one can study tunnelling processes not described purely by an analytic continuation of a single Euclidean solution. This opens the way to including global quantum effects, similar to the Casimir effect [39]. These will *not* change the standard large-scale picture of the bubble spacetime; they are only important when the bubble interior is very small. One can think of this procedure in terms of allowing other forms of energy, in addition to that of the scalar field, to act on the spacetime geometry.

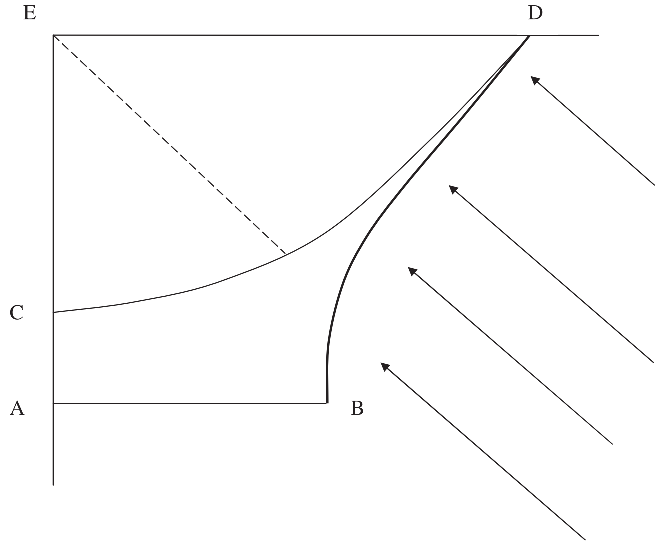

FIG. 1. Positive-vacuum-energy bubble in positive-vacuum-energy background.

A schematic Penrose diagram of the nucleation of a positive-vacuum-energy bubble residing in a larger spacetime (which itself has positive vacuum energy) is given in Fig. 1. The bubble nucleates along AB and the outer surface of its wall is represented by BD. Notice that the entire initial spatial section inside the bubble is exposed to outside influences. A gravitational wave (say) in the outside world can reach *any* point in the initial spatial section of the bubble even if it originates from a point deep inside the de Sitter "bulk"; this is symbolized by the arrows in the diagram.

Following Aguirre and Gratton [18], I think of semiclassical bubble nucleation as a three-stage process: the ambient spacetime and the bubble interior (ECD in the diagram) can be described more or less accurately by classical geometry, but the transition region (ACDB in the diagram) is a predominantly quantum domain. The idea is that, in that domain, quantum effects prevent the characteristic geodesic focussing associated with classical gravity, thus ruling out anything analogous to a shrinking of spatial sections to zero size. This is a reasonable expectation, because it is known that this "quantum defocusing" is precisely what happens in the course of the Hawking

[8]An interesting variant of this argument, to the effect that the region near to the center of the bubble is particularly favored, will be mentioned in the next section.

evaporation of a black hole—see [62] for a particularly clear discussion of this.

While this fourth proposal seems to be the simplest way of explaining how a bubble universe can inherit the geometric regularity of the ambient spacetime, it forces one to confront a basic issue: what, exactly, *are* the initial conditions for the bubble universe? That is, if one consider the spatial section along which a semiclassical description first becomes appropriate, the surface CD in Fig. 1, one needs to know something about the conditions imposed on this section by the quantum domain. Without this information, one will of course be unable to predict the subsequent evolution of the bubble interior.

The appropriate initial (and boundary) conditions for matter fields will be discussed in Sec. IV; for the moment, I shall focus on the initial conditions for the spacetime geometry of the bubble. I propose that the correct initial condition for the interior semiclassical spacetime is that the initial spatial section CD is of *minimal* but nonzero volume [specifically, that it is a spacelike surface of (approximately) *vanishing extrinsic curvature*].

There are four reasons for thinking that this is the right procedure. First, one can argue that, in string theory, it is not reasonable for *any* cosmological model with compact spatial sections to have spatial volumes much below the cube of the string length scale; so there should be a spatial section of minimal volume or zero extrinsic curvature, and that spatial section is the natural locus for a semiclassical description to be appropriate [17]. (Strictly speaking, the spatial sections of a bubble universe are infinite in extent, but in the "holographic" interpretation I adopt here [63] they are *effectively* finite. This will be discussed later.)

Second, a connection between zero extrinsic curvature and low entropy is suggested by Verlinde's [64] observation that Cardy's formula for the entropy of a conformal field theory can reproduce the Friedmann equation. Minimal volume is then naturally associated with low "holographic entropy," because the latter is related [65,66] to the extrinsic curvature of spatial sections.

Third, the Borde-Guth-Vilenkin theorem implies that the only way an inflating spacetime can avoid having zero-volume spatial sections is to have a longer history of contraction than of expansion. I therefore need to use *part of* a "bouncing" cosmological model [67], that is, part of a spacetime which does have a spacelike surface of zero extrinsic curvature. This surface is the only distinguished one in the spacetime, and so it is natural to use the part which begins along this surface. (True "bounce" cosmologies, including the contracting part, are interesting [68,69], but they encounter notorious entropic difficulties of precisely the kind I hope to resolve here, and the most recent work [70] only serves to reinforce these doubts regarding the thermodynamics of "bounces.") To put it another way: if the surface CD in Fig. 1 does *not* have vanishing extrinsic curvature, then this nonzero object would define a new fundamental time scale. [In the special case of FRW cosmology, the extrinsic curvature is given by $-\dot{a}(t)/a(t)$, where $a(t)$ is the scale factor and the dot denotes a *time* derivative.] It is hard to see how such a scale could arise in string theory.

Finally, my picture of the origin of the arrow of time in the ambient spacetime [19] supposes that the latter emerges from a state with no classical description—that is, from nothing [14,15]—along a surface of zero extrinsic curvature; so, to be consistent, I should assume that a similar principle applies when classicality emerges inside the bubble.

With a concrete proposal for the initial conditions of a bubble universe, one can explore the structure of the spacetime in the early history of such universes. Before doing so, however, let us see more concretely how all of these observations apply to the usual description of bubble universes.

III. THE ARROW OF TIME IN STANDARD BUBBLE UNIVERSES

The original examples of bubble universes were those studied by Coleman and De Luccia [2], who showed that they arise in the interior of bubbles of true vacuum nucleating in a false vacuum defined by a local minimum of a scalar field potential $V(\varphi)$. The scalar is assumed to be the *only* form of matter present in the spacetime. Let us consider in detail how an arrow can arise, under this assumption. Let us begin with a bubble of *negative* vacuum energy nucleating in a Minkowskian background; that is, following Coleman and De Luccia, I represent the scalar field inside the bubble by a negative cosmological constant, and I treat the wall as being infinitely thin. The conformal geometry of the bubble and its environment is depicted in Fig. 2.

The original Minkowski space is represented by the triangle BCF, and the bubble wall is the null surface AE. In a more realistic version, the bubble wall would be timelike; but then the wall would accelerate, so the curve representing it would not be geodesic, and so it is still able to terminate on future null infinity. This is important, because it means that the bubble is exposed to signals from the entire exterior spacetime, whether the wall is thin or not. The region FAEG represents the interior of the bubble; it is a part of the maximally symmetric simply connected four-dimensional spacetime of negative vacuum energy density $-1/8\pi L^2$, the anti–de Sitter spacetime AdS_4.

The timelike conformal boundary of this part of AdS_4 is represented by EG; as usual this causes the surface DE to have a future *Cauchy horizon*, EF, and also a past "Cauchy horizon," AE. (That is, AE is a Cauchy horizon in the original AdS_4.) The region FAE can be covered by coordinates such that DE is $t = 0$; in these coordinates the metric in FAE takes the form

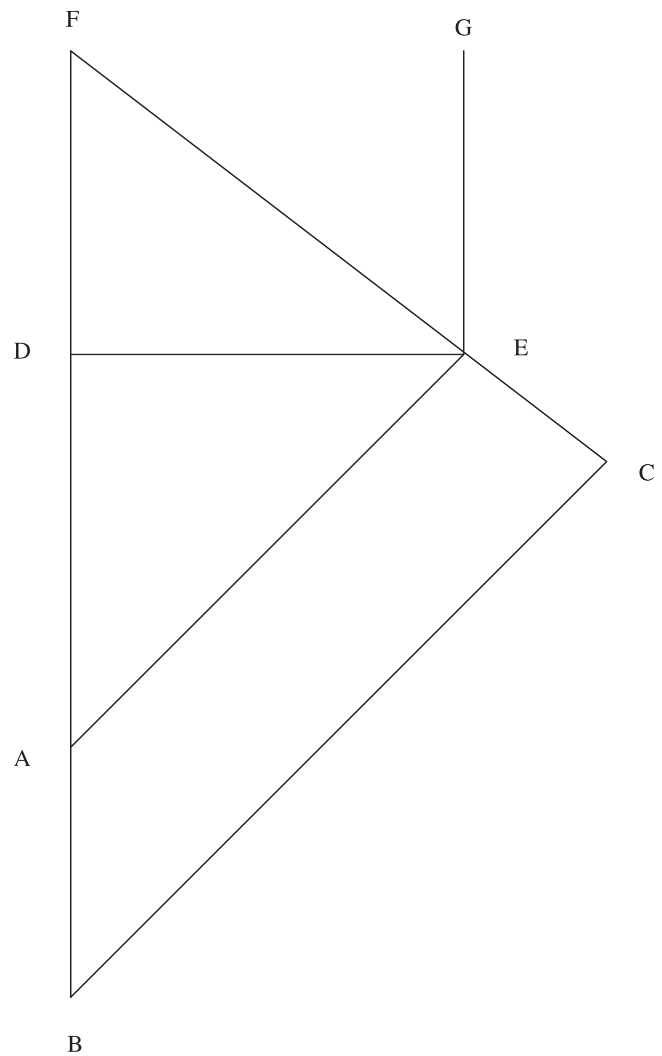

FIG. 2. Negative vacuum energy bubble in Minkowski spacetime.

$$g(\mathrm{AdS}_4) = -dt^2 + \cos^2(t/L) \times [dr^2 + L^2 \sinh^2(r/L)\{d\theta^2 + \sin^2(\theta)d\phi^2\}]. \quad (2)$$

Notice that the spatial sections are, at least locally,[9] copies of the hyperbolic space, H^3. Notice too that the spacelike hypersurfaces near to $t = \mp\pi L/2$ have volumes which are tending to *zero*. These regions are, respectively, the past and future Cauchy horizons of DE, that is, they correspond to AE and EF in the diagram. They are *not* singularities in the sense of having divergent curvature, despite the fact that their volume vanishes, because the equation of state of vacuum energy—the only form of energy in pure AdS_4—has a particular form: the energy density is constant, and cannot diverge under any circumstances. Notice finally that the spacetime is apparently time dependent; this is due to the fact that the corresponding inertial observers are not the Killing observers: but there is in fact a timelike Killing vector in this spacetime. Time *does not pass* in this bubble; vacuum energy cannot "age," and the spacetime itself is static.

Of course, this model is unrealistic in several ways: the bubble has a thin wall, the scalar field is treated as if it were exactly equivalent to vacuum energy, and the computation assumes strict semiclassical dominance of tunnelling amplitudes [29]; in particular, no allowance has been made for any kind of perturbation impinging on the bubble wall from outside.

In fact, a first step towards greater realism was taken by Coleman and De Luccia themselves, who gave a beautifully simple discussion of the consequences taking into account the first-order corrections to the thin-wall approximation. The essential point is that while the rotational symmetry of their instanton continues to enforce the initial vanishing of the time derivative of φ, it cannot force φ itself to vanish exactly. The results can best be pictured in the following way. I remarked above that the metric in Eq. (2) *appears* to represent a spacetime, with H^3 spatial sections, which is dynamic. This is not in fact the case. Coleman and De Luccia find, however, that the slightest perturbation away from the exact thin-wall conditions produces a spacetime which *really is* dynamic: the Killing vector is lost. This paves the way for an arrow to be established.

There is a crucial point here, however: making the spacetime more realistic can be expected to do away with the Cauchy horizons of the exact AdS_4 spacetime, since these horizons are typically unstable. But the Cauchy horizons do not simply disappear: generically, *they turn into spacelike surfaces of zero volume.* This aspect of AdS_4 was understood long ago: see, for example, the discussion on page 172 of the review article of Tipler, Clarke, and Ellis [71]. One says that the Cauchy horizons are replaced by *crushing singularities*, though these need not be "singular" in the sense I use here; see [71], page 166, for a definition.

With all this in mind, I proceed to the usual description of *realistic* versions of bubble universes with negative vacuum energy; it runs as follows. Once one recognizes that the spacetime is dynamic, one must expect the scalar field to fluctuate and to transfer energy to any other field to which it may be coupled, as happens in inflationary reheating. By the time the bubble interior nears the zero-volume spatial section which replaces the Cauchy horizon EF in Fig. 2, then, it will contain forms of matter with *conventional* equations of state, such that the energy density *does* diverge at late times. In short, a realistic version of a negative-energy bubble terminates in a true crunch. These are the "terminal vacua" in the landscape.

But if I grant that one Cauchy horizon becomes singular, why does that not happen along the other Cauchy horizon,

[9] That is, ignoring topological questions.

AE in Fig. 2? To see how this works, I consider the situation described in [3], where the scalar field inside the bubble is no longer represented by a simple vacuum energy.

The metric is a FRW metric with spatial sections of constant negative curvature; the O(4) symmetry group of the Coleman–De Luccia instanton becomes the O(1,3) group of (*local*) symmetries of three-dimensional hyperbolic space. Regularity of the Euclidean instanton, which has a characteristic length scale L, guarantees that the Lorentzian metric has the following general form:

$$g(\text{Bubble}) = -dt^2 + a(t/L)^2[dr^2 + L^2\sinh^2(r/L) \times \{d\theta^2 + \sin^2(\theta)d\phi^2\}], \tag{3}$$

where, if I choose zero as the origin of time,

$$a(t/L) = t/L + \mathbf{O}(t^3/L^3). \tag{4}$$

In a FRW cosmology with negatively curved spatial sections (with curvature proportional to $-1/L^2$), one can show straightforwardly that the pressure is given by

$$8\pi p = \frac{1/L^2 - 2a\ddot{a} - \dot{a}^2}{a^2}, \tag{5}$$

where the dot denotes a proper time derivative, while the energy density is

$$8\pi\rho = 3\frac{\dot{a}^2 - 1/L^2}{a^2}. \tag{6}$$

Applying this to the case at hand, I find, if I set $a(t/L) \approx t/L + \alpha t^3/L^3$, where α is a constant ([which is negative if the energy density is negative at small t), that

$$8\pi p \approx \frac{-18\alpha - 21\alpha^2 t^2/L^2}{L^2(1 + \alpha t^2/L^2)^2} \tag{7}$$

and

$$8\pi\rho \approx \frac{18\alpha + 27\alpha^2 t^2/L^2}{L^2(1 + \alpha t^2/L^2)^2}. \tag{8}$$

Thus I see that, even though the spatial sections shrink to zero size as t approaches zero, neither the pressure nor the density diverges in this limit, as would be the case for any normal form of matter or radiation. The scalar field pressure and density are related by a somewhat bizarre equation of state (obtained by eliminating t in the above relations for p and ρ), and this is what allows this field to avoid causing a singularity at $t = 0$.

Obviously this bubble universe has a very definite arrow of time: it begins in a perfectly smooth nonsingular state and ends in a (no doubt highly irregular) crunch singularity. In particular, it is clear that the gravitational entropy is initially low and finally very large. But the origin of this arrow is all too clear: *I built it in*, by *assuming* that the tunnelling originated in perfectly smooth Minkowski space, which justifies the description of the tunnelling by a perfectly smooth, exactly O(4)-invariant Euclidean instanton. I have in fact been guilty of practising Price's [10] "*double standard*": I made assumptions about the beginning of the bubble universe that I would never apply to its "generically" singular end. If I had allowed for perturbations in the ambient spacetime propagating into the bubble and—in accord with the second law—disturbing the geometry there, then, as I discussed in Sec. II, the *arbitrarily* small spatial slices near to $t = 0$ would not be perfectly smooth, and the scalar field might not be in a sufficiently low-entropy initial state.

It is true that, even in this case, the entropy of the initial state could still be *somewhat* lower than that of the final state, so the bubble might have an arrow of a sort. This argument has particular force when one considers the region of spacetime near to the initial nucleation event. Recall that Aguirre and Gratton [26] argued that the geodesic incompleteness of a (thin-walled) bubble universe is due to the fact that the spatial sections defined by distinguished observers have a tendency to become null. This tendency is less marked near the center of the bubble, and so one might hope that the growth of anisotropies as one moves back in time could be controlled in that region. If this is true, then it might pave the way towards dealing with the notorious problems associated with the infinite extent of the bubble spatial sections, since only a relatively small region of the bubble could come to resemble our Universe.

Against this, however, one has to bear in mind that the only Universe we have observed does not just have "low" initial entropy: its initial entropy is *fantastically* lower than it might have been, as Penrose [9] has shown by a well-known calculation. Thus my task is not just to show how bubble universes can have *some* kind of arrow—rather, I have to show how they can have an arrow *of the kind we observe*. Again, it is hard to see how such delicate initial conditions can be maintained in the face of anisotropies which, as I have discussed, grow extremely rapidly if spatial sections are allowed to become arbitrarily small. This is a matter which can be settled only by means of a detailed calculation, which I shall not attempt here.

Because it has both a beginning and an end, the negative-energy bubble is particularly suited to a discussion of the arrow, but the problem persists even in the more directly interesting case of a bubble with positive vacuum energy. An arrow of time will emerge in this case too, provided that the scalar field is in a sufficiently low-entropy state initially. But in this case too I will find a geometry like the one given in Eqs. (3) and (4), with a zero-volume initial state. Again, the pressure and density [given by (7) and (8), but with positive α] do not diverge even at zero volume, but the problem of large initial gravitational entropy persists. For that problem is associated with zero volume, not with the question as to whether a singularity is present.

As mentioned above, one way to deal with this problem would be to investigate, in detail, whether the growth of anisotropies can be controlled in the favorable region near the center of the bubble. Here, instead, I shall postulate that the scalar field is *not* the only important contributor to the total energy density inside the bubble: there must be another contribution due to quantum effects which "defocus" geodesics, so that there is never any surface of zero volume. The question now is: what is the mathematical description of this quantum contribution to the energy density?

Since the geometry of the earliest spatial sections inside a bubble universe is precisely the issue here, I must not base my arguments on FRW spacetime geometries. It will be useful, however, to begin by reminding ourselves of the reasons for the fact that FRW models with negatively curved spatial sections tend to be geodesically incomplete. The relevant singularity theorem is the one due to Penrose; it may be stated as follows. (See [72], page 239 for the theorem and the relevant concepts).

THEOREM (Penrose): Let M_4 be a spacetime satisfying the Einstein equations and the following conditions:

(a) The Null Ricci Condition (NRC) holds.

(b) M_4 is globally hyperbolic and contains a noncompact Cauchy surface.

(c) M_4 contains a trapped surface.

Then there is at least one incomplete future-directed null geodesic orthogonal to the trapped surface.

With regard to condition (a), recall the discussion of the NRC in Sec. I; with regard to (b), note that "noncompact Cauchy surface" can be weakened to "Cauchy surface with a noncompact universal cover." Thus, compactifying the hyperbolic spatial sections of a bubble[10] does not *in itself* allow me to avoid geodesic incompleteness here—but see Sec. IV, below.

Assuming that the NRC is not violated, the only condition of this theorem which needs to be verified in the case of FRW cosmologies with negatively curved spatial sections (whether compactified or not) is the last. Take the metric given in Eq. (3) and consider a 2-sphere with radial coordinate r at time t; its area is $4\pi L^2 a(t)^2 \sinh^2(r/L)$. The orthogonal *outward*-directed set of past-pointing null geodesics intersect the surface $t = t + dt$ (with negative dt) at radial coordinate $r - dt/a(t)$, and so the change in the area of the sphere as r increases is

$$dA = 8\pi L^2 a(t)\left[\sinh^2(r/L)da - \frac{dt}{L}\sinh(r/L) \times \cosh(r/L)\right] = 8\pi L \sinh^2(r/L) a(t) dt[L\dot{a} - \coth(r/L)]. \quad (9)$$

For a trapped surface to exist in this spacetime, one must be able to choose r and t in such a manner that dA is negative, that is, has the same sign as dt. Now suppose that there is at least one value of t, say t^*, such that $L\dot{a}(t^*) > 1$ at that time; if I assume the validity of the Einstein equations, then I see from Eq. (6) that this is precisely equivalent to assuming the existence of at least one spatial section on which *the total energy density is strictly positive*. Then the spacelike hypersurface $t = t^*$ contains a trapped surface, because if r is chosen sufficiently large then $\coth(r/L)$ (which of course is always greater than unity, but approaches it as r tends to infinity) becomes smaller than $L\dot{a}(t^*)$.

For example, in the case where a bubble universe contains nothing but pure positive vacuum energy, I obtain the version of de Sitter spacetime with hyperbolic spatial sections, with the "bubble de Sitter" metric:

$$g(\mathrm{BdS}) = -dt^2 + \sinh^2(t/L)[dr^2 + L^2\sinh^2(r/L) \times \{d\theta^2 + \sin^2(\theta)d\phi^2\}]; \quad (10)$$

in this case I have

$$dA = 8\pi L \sinh^2(r/L)\sinh(t/L)dt[\cosh(t/L) - \coth(r/L)]. \quad (11)$$

Notice that, in this case, it becomes steadily easier to keep the expression in square brackets positive as time progresses (in the sense that one need not take particularly large values of r in order to ensure this). The existence of trapped surfaces is in this sense a local question at late stages of inflation. Clearly, the spatial volume does vanish at $t = 0$.

The Penrose theorem now explains why FRW spacetimes with negatively curved spatial sections tend to be geodesically incomplete in the past (since I am applying the theorem to past-directed null geodesics). All I needed were the very mild conditions that the NRC should be satisfied—recall that this is equivalent to assuming the NEC if the Einstein equations hold—and that there should be at least one spatial section containing a trapped set. Note that both of these conditions are satisfied by the version of de Sitter spacetime with *spherical* spatial sections. Thus, contrary perhaps to intuition, what saves spatially spherical de Sitter spacetime from being geodesically incomplete is *not* "gravitational repulsion" (that is, violation of the strong energy condition, which is not assumed in the Penrose theorem) but rather the fact that the spatial sections do not have a noncompact universal cover.

I conclude that the only way to avoid having a zero-volume spacelike surface in a FRW bubble spacetime is to violate the NEC, at least effectively.

This discussion used FRW geometry, so this result is not surprising; but the advantage of using the Penrose theorem is that the argument can be adapted to show that similar conclusions follow if an *inflating* bubble is perturbed, even to a large extent. Bubble interiors do have spatial sections with noncompact universal covers, and this *topological*

[10]That is, projecting to a compact quotient of hyperbolic space by a discrete freely acting group of isometries.

statement is robust against perturbations. If one assumes that inflation occurs at late times, then the existence of spatial surfaces with positive total energy density at those times is only to be expected, since the positive energy density of the inflaton will dominate; this will lead to the existence of trapped surfaces. From another perspective: the existence of trapped surfaces seems to be inevitable, since it is a local question at late times, even in the case where the bubble has been perturbed extensively at *early* times. Thus, I expect to be able to apply the Penrose theorem even to bubbles which are *not* close to a FRW form. It follows quite generally (even for strongly perturbed spacetimes with highly irregular spatial geometries) that *an inflating bubble universe can only have an initial spacelike surface with vanishing extrinsic curvature and nonzero volume if the NEC is violated inside the bubble.*

A completely rigorous theory supporting this physical argument has been given by Andersson and Galloway [73], who prove a theorem (Theorem 4.1) to the following effect. Suppose that I take a globally hyperbolic asymptotically de Sitter spacetime satisfying the NRC, and assume that the Cauchy surfaces (or their universal covers) are *not compact*. Suppose now that I try to avoid having any spacelike surface with zero volume, by having a bounce. (''Asymptotically de Sitter'' is then assumed to hold both to the past and to the future.) Then Andersson and Galloway show that some future-directed null geodesic must fail to reach future infinity. (The spacetime must also satisfy a certain genericity condition, which essentially states that all spatial dimensions take part in the accelerated expansion; see [74] for further discussion, and see [75] for another application of results like this.) Since these spacetimes are supposed to evolve to a de Sitter-like (inflationary) state (in which all future-directed null geodesics do reach future infinity) I can conclude that the NEC must indeed be violated by all spacetimes of the kind in which I am interested here.

I now have an answer to my question as to how the matter content of a bubble universe must be modified in order to avoid spacelike sections of zero volume. The answer is simply that the NRC must be violated inside the bubble, by some effect which is normally ignored in discussions of bubble universes. This will involve either modifying the Einstein equations so that the NRC can be violated without violating the NEC, or directly violating the NEC itself. In the next section, I explore the second option.

IV. CASIMIR BUBBLES

My proposal is that the correct initial condition for a bubble interior as it emerges from the quantum domain is that of vanishing extrinsic curvature: this applies to the spacelike surface CD in Fig. 1. Formally, *but not physically*, the geometry here is like that of a bounce cosmology [67]; the great difference is that, in my case, the initial conditions for the semiclassical spacetime are *not* prepared by an earlier period of contraction. I stress that this is just a (natural) *proposal*: I have to verify that it makes sense physically, within the context of string theory.

The idea that the *Casimir effect* might play a crucial role in cosmology has often been suggested: see, for example, [39,76–79] and references therein. It has recently been raised in connection with the ''standard model landscape'' [37]. As is well known, the Casimir effect naturally leads to negative energy densities and pressures, violating the NEC. This is of great interest in string theory, because *all* currently known modulus stabilization schemes violate the NEC in one way or another. (Furthermore, it seems likely [80] that NEC violation of some kind is a fairly generic feature of theories involving higher dimensions.) Subsequently [38] it was found that by no means all forms of NEC violation are acceptable in string theory; Casimir effects are of great interest precisely because they belong to the ''acceptable'' class (outside the ''clock and rod'' sector). If I wish to embed my discussion in string theory, then ''Casimir cosmology'' is a particularly natural—though surely not the only—way to proceed.

The Casimir effect essentially arises from certain kinds of *boundary conditions* which one might find it physically appropriate to impose. In the case of a bubble universe, one has to ask: what kinds of boundary conditions are appropriate for fields inside the bubble, and how can they be enforced?

This brings me directly to attempts (see particularly [63] and references therein) to extend black hole complementarity to cosmology. Recall that black hole complementarity resolves the puzzles concerning Hawking radiation by declaring that one can describe black hole radiation by taking either *but not both* of two points of view (following the star as it collapses or using the observations of an observer who stays far away from it). Either perspective is postulated to give a complete description; paradoxes only arise if one tries to take a ''global'' point of view.

In the cosmological context, attention is focussed on *causal diamonds*, the entire region of a spacetime which is causally connected to the worldline of a single observer. The remainder of the global spacetime is then regarded as a set of redundant descriptions of the same data, and, once again, paradoxes arise if one attempts a global perspective. Now, in the case at hand, I wish to apply this philosophy to the bubble universe portrayed in Fig. 1. Take the observer whose worldline corresponds to the vertical left-hand boundary of the diagram. The relevant part of the corresponding causal diamond is represented by the dotted line. This line intersects any spatial section (such as CD) at a *finite* distance from the observer. From the point of view of complementarity, then, the spatial sections inside the bubble are effectively finite; regarding them as infinite means taking the global point of view of the bubble universe, and this is precisely what complementarity forbids.

The problem of deciding how to implement this insight mathematically is a difficult one. In order to proceed, I shall suggest a simple ansatz, which is not intended to be fully realistic but which will allow me to proceed in a quantitative way. My suggestion is prompted by the ideas discussed in Refs. [81,82], in which the authors discuss cosmological spacetimes with negatively curved spatial sections. As is well known, it is possible to perform *periodic identifications* of domains in ordinary hyperbolic space H^3, so that the quotient is compact. In classical general relativity this makes no difference if the domain involved is very large, but in [81,82] it is argued that string theory is sensitive to such identifications, and that the periodic structure has profound physical implications.[11]

Motivated by this, I propose to implement observer complementarity in the following simple manner: when studying quantum-mechanical aspects of the interior of a bubble universe, I should enforce *periodic boundary conditions* on all fields. Concretely, what this idea means is that I should reject all fluctuations of fields beyond a finite limit. Doing so will lead to a Casimir effect, which will however be significant only in the very earliest era of the bubble universe. I can now try to construct an internally consistent model of an inflating bubble with spatial sections which, with the help of Casimir energy, are able to avoid shrinking to size zero at any time. Since I am interested in the very earliest history of the bubble, where the inflaton is assumed to be rolling extremely slowly, I can approximate the energy density of the inflaton by that of a positive cosmological constant with characteristic length scale L; the negative Casimir energy density is superimposed on this.

Casimir energies can depend sensitively on the kinds of matter fields involved and whether the effects of higher dimensions are to be taken into account, and so on; but let me continue to proceed in the simplest possible manner, and assume as usual [79] that, for a four-dimensional FRW spacetime with effectively compact spatial sections, the Casimir density depends on the inverse fourth power of the scale factor. The total energy density is then a combination of the background vacuum density $+3/8\pi L^2$ with the Casimir energy; so the Friedmann equation takes the form

$$L^2\dot{a}^2 = \frac{8\pi}{3}L^2a^2\left[\frac{3}{8\pi L^2} - \frac{6}{8\pi L^2 a^4}\right] + 1. \quad (12)$$

Here the coefficient of the Casimir term has been fixed by requiring the surface of zero extrinsic curvature to correspond to a scale factor equal to unity. The solution for the "bubble de Sitter plus Casimir" metric is remarkably simple:

$$g(\mathrm{BdS}+\mathrm{C}) = -dt^2 + [1 + 3\sinh^2(t/L)] \times [dr^2 + L^2\sinh^2(r/L)\{d\theta^2 + \sin^2(\theta)d\phi^2\}]. \quad (13)$$

Notice that this is asymptotic, as t tends to infinity, to bubble de Sitter spacetime [Eq. (10); the factor of 3 can be absorbed in the limit], but it has a spatial surface of zero extrinsic curvature at $t = 0$. If I simply postulate that the semiclassical bubble history begins at that time, then we have a picture of the bubble interior in which the Casimir effect is significant for a very brief period, which is succeeded (as the Casimir energy rapidly dilutes but the inflaton energy does not) by an ordinary accelerated expansion. The Casimir effect allows the low-entropy conditions in the exterior to establish, via the surface $t = 0$, similar conditions in the interior; having done this duty, it rapidly disappears, and the usual description of a bubble interior becomes valid.

The Casimir effect is completely harmless at the perturbative level, but it is far from clear that this remains true nonperturbatively, particularly when it plays such an important role in fixing the spacetime geometry. In fact, it is known that such effects can lead to serious consequences [33], as follows. Seiberg and Witten [40] observed that branes, being extended objects, can be extremely sensitive to the geometry of the spaces in which they propagate. If the geometry takes certain forms, it can actually lead to a situation which Maldacena and Maoz [41] (see also [42]) describe as a *pair-production instability* for branes.

To be specific: suppose that a given spacetime has a Euclidean version which is conformally compactifiable; that is, it is conformal to the interior of a compact manifold-with-boundary. Such manifolds are said to be *asymptotically hyperbolic*: that is, the geometry comes to resemble that of hyperbolic space[12] at sufficiently large distances. For Euclidean Bogomol'nyi-Prasad-Sommerfield branes in four dimensions, the brane action consists of two terms: a positive one proportional to the (three-dimensional) area of the brane, and a negative one proportional to the volume enclosed by it. So I have, in four dimensions,

$$S = \Theta\left(A - \frac{3}{L}V\right), \quad (14)$$

where Θ is the tension, A is the area, V the volume enclosed, and L is the background asymptotic curvature radius. If at any point the volume term is larger than the

[11]Notice that the *local* isometry group of a compactified negatively curved space is the same as that of ordinary hyperbolic space H^3, namely, O(1,3), since the local metric is completely unaffected by the compactification. Therefore, the usual argument, whereby the O(4) symmetry of the Euclidean instanton becomes the O(1,3) symmetry of the spatial sections of the bubble, is unaffected.

[12]In my case this space will be four-dimensional; it should not be confused with the (also hyperbolic) three-dimensional transverse slices of the Lorentzian version.

area term, it will be possible to reduce the action of the system by creating brane-antibrane pairs and moving them to the appropriate positions, as described by Maldacena and Maoz [41]. Thus a severe nonperturbative instability will arise. In this way I obtain a powerful criterion for the acceptability of specific geometries from a stringy point of view: powerful because it applies even when the NEC is only violated *effectively*.

To see how this works in the present case, let me proceed as follows. I begin by constructing the asymptotically hyperbolic version of bubble de Sitter spacetime, with metric given in Eq. (10). I simply complexify both t and L, but not r. Relabelling the latter as $L\chi$, I obtain

$$g(\mathrm{AHBdS}) = dt^2 + L^2\sinh^2(t/L) \times [d\chi^2 + \sin^2(\chi)\{d\theta^2 + \sin^2(\theta)d\phi^2\}]; \quad (15)$$

this "asymptotically hyperbolic bubble de Sitter" metric is in fact the metric of four-dimensional hyperbolic space, foliated by 2-spheres. Note that the sign of the curvature has been reversed by the complexification of L. (In order to obtain *anti–de Sitter* spacetime from H^4, one chooses a quite different foliation, with negatively curved slices, and of course one does not complexify L; see [83] for the details.) Notice that this foliation makes it obvious that the conformal boundary is positively curved; this is important for establishing nonperturbative stability at *large* values of t, as was shown by Seiberg and Witten [40]. In fact, the brane action in this case can be evaluated explicitly: from Eq. (14) I have

$$S[\mathrm{BdS}](t) = 2\pi^2\Theta L^3[\sinh^3(t/L) - \tfrac{1}{4}\cosh(3t/L) + \tfrac{9}{4}\cosh(t/L) - 2]. \quad (16)$$

This function is actually non-negative at *all* positive values of t, large or small, so bubble de Sitter spacetime is *completely stable* against this particular nonperturbative effect. Actually, the function increases monotonically with t; this is characteristic of spatially flat or negatively curved asymptotically de Sitter spacetimes which satisfy the NEC. When the NEC is violated, there are grounds for serious concern that the action will not behave so benignly.

Applying this same complexification to the metric in Eq. (13), I have the asymptotically hyperbolic version of $g(\mathrm{BdS} + \mathrm{C})$:

$$g(\mathrm{AHBdS} + \mathrm{C}) = dt^2 + L^2[1 + 3\sinh^2(t/L)] \times [d\chi^2 + \sin^2(\chi)\{d\theta^2 + \sin^2(\theta)d\phi^2\}]. \quad (17)$$

Note that t/L is not complexified, so I can still interpret it as a dimensionless measure of time in this case. If I truncate this space at $t = T$, then the brane action for $t \geq T$ is

$$S[\mathrm{BdS} + \mathrm{C}]_T(t) = 2\pi^2\Theta L^3\left[(1 + 3\sinh^2(t/L))^{3/2} - \frac{3}{L}\int_T^t (1 + 3\sinh^2(\tau/L))^{3/2}d\tau\right], \quad (18)$$

where Θ is the tension, as in Eq. (14).

If $T = 0$, this function begins at $t = 0$ with a positive value equal to $2\pi^2\Theta L^3$ and then immediately *declines* as t increases. This decrease is characteristic of NEC-violating spacetimes, as was shown in the case of *flat* compact spatial sections in [33]; it is the reason for the fact that NEC-violating spacetimes are in danger of being nonperturbatively unstable. The positive curvature of the $t =$ constant sections in this case means that—as in the case of bubble de Sitter space—there is no such instability at *large* values of t, but there might be a problem at *small* values of t if the NEC violation causes the action to fall too low. (Maldacena and Maoz [41] discuss examples where this happens.) Since the Seiberg-Witten argument shows that the action *is* positive at large t, its decline must be halted at some point. The question is whether it is halted in time to prevent the action from becoming negative. The graph of the action function has a unique minimum (for all T) at $t = (\ln(\sqrt{3}))L$. In the case where I cut off the space at its neck (so that the spatial sections inside the bubble never contract, which is what I am supposing here), I have $T = 0$, and a simple numerical investigation shows that the action decreases from a positive value at $t = 0$ down to $S[\mathrm{BdS} + \mathrm{C}]_0((\ln(\sqrt{3}))L)$, which is still *positive* (for all L). The brane action subsequently increases indefinitely as the area term decisively overcomes the volume term as the Casimir energy is diluted (so that the action function comes to resemble that of bubble de Sitter). Thus the action is positive everywhere. Given this, it is easy to see that the same statement holds true for any $T \geq 0$: *there is no Seiberg-Witten instability in this system, as long as the spatial sections never contract.* The graph[13] of the action for $T = 0$ is given in Fig. 3; notice that the system escapes from being unstable despite the initial decrease of the action.

Allowing the spatial sections to contract means taking T to be negative. In this case, the initial value of $S[\mathrm{BdS} + \mathrm{C}]_T(t)$ becomes a larger positive number; but on the other hand the function decreases for a longer time, so it is not obvious that it remains positive everywhere. A numerical investigation shows that, as T is modified downwards, $S[\mathrm{BdS} + \mathrm{C}]_T((\ln(\sqrt{3}))L)$, the minimum value of the action, stays non-negative only down to a value of T that is very close to zero, $T \approx -0.0928L$. The scale factor at that value of t is given approximately by $a(-0.0928L) \approx 1.006\,44$. Clearly there is essentially no contraction in

[13]The horizontal axis is t/L, the vertical axis is $S[\mathrm{BdS} + \mathrm{C}]_0 \times (t)/2\pi^2\Theta L^3$.

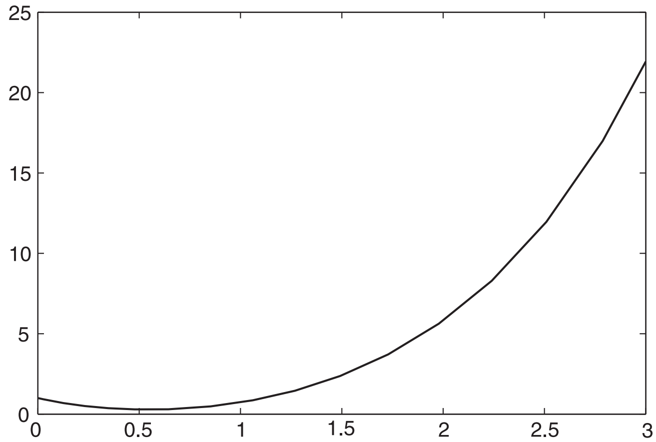

FIG. 3. The action $S[\mathrm{BdS} + \mathrm{C}]_0(t)$.

this case (the minimum value of the scale factor being unity).

I interpret this last result as strong evidence in favor of my postulate that the bubble history begins on or near to the surface of vanishing extrinsic curvature: if it tries to begin earlier, the system becomes violently unstable. The spacetime geometry is *not* like that of a bounce spacetime: there is little or no contraction as seen by the distinguished bubble observers.

Of course, the example I have considered here is a very special one: it is motivated by a desire to present a fully explicit metric. In fact, Casimir effects are not the only way to achieve NEC violation (or "effective" NEC violation—see [33,84]). However, numerical experiments lead us to believe that if the NEC is violated, effectively or otherwise, in ways that are compatible with the ideas of Arkani-Hamed *et al.* [38], then one will be led to a picture similar to the one presented in detail here: that is, the requirement of nonperturbative string stability will prohibit any more than a negligible amount of contraction inside a bubble universe.

In summary, it is very difficult for a bubble universe to resemble our world, because to do so it needs to begin with very special and delicate properties; but it may be possible if NEC violation is indeed compatible with, *yet constrained by*, stringy considerations.

V. CONCLUSION: BUILDING A LANDSCAPE

In the stringy picture of "creation from nothing" (or the "emergence of time") [15], the original "mother of all universes" is born along a spatial section that is as smooth as it can be, up to quantum fluctuations [17,19]. This allows inflation to start in the mother universe. The latter may however subsequently nucleate bubble universes of the kind I have been considering in this work. The arrow in these bubbles, if any, must be *inherited* from the mother universe; the arrow can then be handed down to subsequent generations. In this way one obtains an explanation of the observed arrow that does not involve wildly improbable or rare fluctuations into lower-entropy states. In this work, I have suggested a way of ensuring that this process of "inheritance" does occur.

However, the argument in favor of "arrow inheritance" in the NEC-violating case does depend on the "causal diamond" or "observer complementarity" philosophy. I needed this principle to justify the compactification of the bubble's spatial sections—or "periodic boundary conditions"—used in the previous section. While this idea is well motivated by black hole complementarity (and by ideas from string theory [81,82]), the extrapolation to cosmological horizons is not entirely secure. I should therefore ask: what would be the consequences for the landscape if this extrapolation had to be abandoned?

In that case, I would be led to conclude that *we are in the original universe*, the one presented to us directly by creation from nothing. For this original universe does have an arrow of time, such as we in fact observe; whereas no bubble universe would have this remarkable property. This would drastically change the role of bubble universes: far from seeding new life, they would merely destroy any ordered structure with which they collided in the original universe. This phenomenon might have to be taken into account in discussions of the nature of observers at very late times.

If bubble universes are unable to inherit an arrow, then we must find another way of building a landscape—that is, of actually constructing universes which realize the full set of string vacuum solutions. A way of doing so which automatically gives rise to spacetimes with low initial entropy is suggested by the work of Gibbons and Hartle [85], who raised the interesting question as to whether a universe created from nothing must be topologically *connected.* This is not at all obvious, because a compact manifold-with-boundary can, and generically will, have a boundary which breaks up into disconnected pieces; this idea is familiar from *cobordism* theory [86]. Gibbons and Hartle gave an elegant proof that the boundary must indeed be connected in the Hartle-Hawking case if all eigenvalues of the Ricci curvature of the Euclidean space are positive and bounded away from zero. This condition is certainly not satisfied by the spaces used in the work of Ooguri *et al.*, however, and so the question remains open. If indeed the relevant Euclidean space has multiple boundary components of zero extrinsic curvature, then potentially large numbers of spacetimes can be born from a single Euclidean ancestor; those born from a boundary component with a suitable (toral) topology will have an arrow, as explained in [19]. The question then, of course, will be whether these universes have suitably spaced values of the cosmological constant. Perhaps the methods of Dijkgraaf *et al.* [87] can be adapted to study this.

The Lorentzian spacetimes so created would be completely mutually inaccessible. However, it might be possible to find indirect evidence of the existence of the other universes in our own past, since all universes originate from a common Euclidean space.

ACKNOWLEDGMENTS

The author is grateful to Dr. Soon Wanmei for the diagrams He would also like to thank Matthew Kleban for patient and extremely helpful correspondence and the referee whose questions and suggestions have helped to clarify the argument at many points.

[1] Leonard Susskind, arXiv:hep-th/0302219.
[2] Sidney R. Coleman and Frank De Luccia, Phys. Rev. D **21**, 3305 (1980).
[3] Ben Freivogel, Matthew Kleban, Maria Rodriguez Martinez, and Leonard Susskind, J. High Energy Phys. 03 (2006) 039.
[4] Puneet Batra and Matthew Kleban, Phys. Rev. D **76**, 103510 (2007).
[5] Andrei Linde, Misao Sasaki, and Takahiro Tanaka, Phys. Rev. D **59**, 123522 (1999).
[6] Tanmay Vachaspati and Mark Trodden, Phys. Rev. D **61**, 023502 (1999).
[7] Andreas Albrecht, in *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity*, edited by J.D. Barrow, P.C.W. Davies, and C.L. Harper (Cambridge University Press, Cambridge, England, 2004).
[8] D.H. Coule, arXiv:0706.0205.
[9] R. Penrose, in *General Relativity: An Einstein Centenary Survey*, edited by S.W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).
[10] Huw Price, in *Time's Arrows Today*, edited by S. Savitt (Cambridge University Press, Cambridge, England, 1994); arXiv:physics/0402040.
[11] Lisa Dyson, Matthew Kleban, and Leonard Susskind, J. High Energy Phys. 10 (2002) 011.
[12] Robert M. Wald, arXiv:gr-qc/0507094.
[13] Sean M. Carroll and Jennifer Chen, arXiv:hep-th/0410270.
[14] A. Vilenkin, Phys. Lett. B **117**, 25 (1982).
[15] Hirosi Ooguri, Cumrun Vafa, and Erik Verlinde, Lett. Math. Phys. **74**, 311 (2005).
[16] Paul Frampton, Stephen D.H. Hsu, Thomas W. Kephart, and David Reeb, arXiv:0801.1847.
[17] Brett McInnes, arXiv:0711.1656.
[18] Anthony Aguirre and Steven Gratton, Phys. Rev. D **67**, 083515 (2003).
[19] Brett McInnes, Nucl. Phys. **B782**, 1 (2007).
[20] Claus Kiefer, Braz. J. Phys. **35**, 296 (2005); R. Holman and L. Mersini-Houghton, Phys. Rev. D **74**, 123510 (2006); A. Gorsky, Phys. Lett. B **646**, 183 (2007); Matias Aiello, Mario Castagnino, and Olimpia Lombardi, Found. Phys. **38**, 257 (2008); Laura Mersini-Houghton, arXiv:gr-qc/0609006.
[21] T. Banks, arXiv:hep-th/0701146.
[22] Katherine Freese, Matthew G. Brown, and William H. Kinney, arXiv:0802.2583.
[23] Cyril Pitrou, Thiago S. Pereira, and Jean-Philippe Uzan, J. Cosmol. Astropart. Phys. 04 (2008) 004.
[24] Arvind Borde and Alex Vilenkin, Phys. Rev. D **56**, 717 (1997); Arvind Borde, Alan H. Guth, and Alexander Vilenkin, Phys. Rev. Lett. **90**, 151301 (2003).
[25] Alan H. Guth, J. Phys. A **40**, 6811 (2007).
[26] Anthony Aguirre, arXiv:0712.0571.
[27] Anthony Aguirre and Steven Gratton, Phys. Rev. D **65**, 083507 (2002).
[28] Anthony Aguirre and Matthew C. Johnson, arXiv:0712.3038.
[29] Roman V. Buniy, Stephen D.H. Hsu, and A. Zee, Phys. Lett. B **660**, 382 (2008).
[30] Martin Bojowald and Reza Tavakol, arXiv:0802.4274v1.
[31] Brett McInnes, Nucl. Phys. **B718**, 55 (2005).
[32] Brett McInnes, Nucl. Phys. **B730**, 50 (2005).
[33] Brett McInnes, Nucl. Phys. **B748**, 309 (2006).
[34] Evgeny I. Buchbinder, Justin Khoury, and Burt A. Ovrut, Phys. Rev. D **76**, 123503 (2007); J. High Energy Phys. 11 (2007) 076.
[35] Paolo Creminelli and Leonardo Senatore, J. Cosmol. Astropart. Phys. 11 (2007) 010.
[36] Sean M. Carroll, Mark Hoffman, and Mark Trodden, Phys. Rev. D **68**, 023509 (2003).
[37] Nima Arkani-Hamed, Sergei Dubovsky, Alberto Nicolis, and Giovanni Villadoro, J. High Energy Phys. 06 (2007) 078.
[38] Nima Arkani-Hamed, Sergei Dubovsky, Alberto Nicolis, Enrico Trincherini, and Giovanni Villadoro, J. High Energy Phys. 05 (2007) 055.
[39] D.H. Coule, Classical Quantum Gravity **22**, R125 (2005).
[40] Nathan Seiberg and Edward Witten, J. High Energy Phys. 04 (1999) 017.
[41] Juan Maldacena and Liat Maoz, J. High Energy Phys. 02 (2004) 053.
[42] M. Kleban, M. Porrati, and R. Rabadan, J. High Energy Phys. 08 (2005) 016.
[43] Edward Farhi and Alan H. Guth, Phys. Lett. B **183**, 149 (1987).
[44] Anthony Aguirre and Matthew C. Johnson, Phys. Rev. D **73**, 123529 (2006).
[45] Filipe C. Mena and Paul Tod, Classical Quantum Gravity **24**, 1733 (2007).
[46] Ram Brustein, arXiv:hep-th/0702108.
[47] G.W. Gibbons and S.W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[48] G.W. Gibbons and S.N. Solodukhin, Phys. Lett. B **652**, 103 (2007).
[49] Joel K. Erickson, Daniel H. Wesley, Paul J. Steinhardt, and Neil Turok, Phys. Rev. D **69**, 063514 (2004).

[50] J. Mark Heinzle, Claes Uggla, and Niklas Rohr, arXiv:gr-qc/0702141.
[51] Giovanni Montani, Marco Valerio Battisti, Riccardo Benini, and Giovanni Imponente, arXiv:0712.3008.
[52] Andrei Linde, J. Cosmol. Astropart. Phys. 01 (2007) 022.
[53] Don N. Page, arXiv:hep-th/0612137.
[54] S. Carlip, J. Cosmol. Astropart. Phys. 06 (2007) 001.
[55] J. Richard Gott III, arXiv:0802.0233.
[56] D. Podolsky, arXiv:0704.0354.
[57] Jaume Garriga, Alan H. Guth, and Alexander Vilenkin, Phys. Rev. D **76**, 123512 (2007).
[58] Ben Freivogel, Gary T. Horowitz, and Stephen Shenker, J. High Energy Phys. 05 (2007) 090.
[59] Anthony Aguirre, Matthew C. Johnson, and Assaf Shomer, Phys. Rev. D **76**, 063509 (2007).
[60] Willy Fischler, Chethan Krishnan, Sonia Paban, and Marija Zanic, J. High Energy Phys. 05 (2008) 041.
[61] Spencer Chang, Matthew Kleban, and Thomas S. Levi, J. Cosmol. Astropart. Phys. 04 (2008) 034.
[62] Andrew Strominger and David Thompson, Phys. Rev. D **70**, 044007 (2004).
[63] Raphael Bousso and Ben Freivogel, J. High Energy Phys. 06 (2007) 018.
[64] Erik Verlinde, arXiv:hep-th/0008140.
[65] Qing-Guo Huang, Nucl. Phys. **B777**, 253 (2007).
[66] Qing-Guo Huang, arXiv:hep-th/0512004.
[67] M. Novello, and S. E. Perez Bergliaffa, arXiv:0802.1634.
[68] Yi-Fu Cai, Taotao Qiu, Yun-Song Piao, Mingzhe Li, and Xinmin Zhang, J. High Energy Phys. 10 (2007) 071.
[69] Felipe T. Falciano, Marc Lilley, and Patrick Peter, arXiv:0802.1196.
[70] D. H. Coule, arXiv:0802.1867.
[71] F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis, *General Relativity and Gravitation* (Plenum, New York, 1980), Vol. 2.
[72] Robert M. Wald, *General Relativity* (Chicago University Press, Chicago, 1984).
[73] L. Andersson and G. J. Galloway, Adv. Theor. Math. Phys. **6**, 307 (2003).
[74] Brett McInnes, J. High Energy Phys. 09 (2003) 009.
[75] Don N. Page, arXiv:0707.2081.
[76] Wlodzimierz Godlowski, Marek Szydlowski, and Zong-Hong Zhu, Gravitation and Cosmology **14**, 17 (2008).
[77] Marek Szydlowski and Wlodzimierz Godlowski, Int. J. Mod. Phys. D **17**, 343 (2008).
[78] A. A. Saharian and M. R. Setare, Phys. Lett. B **659**, 367 (2008).
[79] Brian R. Greene and Janna Levin, J. High Energy Phys. 11 (2007) 096.
[80] Daniel H. Wesley, arXiv:0802.3214.
[81] John McGreevy, Eva Silverstein, and David Starr, Phys. Rev. D **75**, 044025 (2007).
[82] Daniel Green, Albion Lawrence, John McGreevy, David R. Morrison, and Eva Silverstein, Phys. Rev. D **76**, 066004 (2007).
[83] Brett McInnes, J. High Energy Phys. 10 (2006) 029.
[84] Tomasz Stachowiak and Marek Szydlowski, Phys. Lett. B **646**, 209 (2007).
[85] G. W. Gibbons and J. B. Hartle, Phys. Rev. D **42**, 2458 (1990).
[86] H. Blaine Lawson and Marie-Louise Michelsohn, *Spin Geometry* (Princeton University Press, Princeton, 1990).
[87] Robbert Dijkgraaf, Rajesh Gopakumar, Hirosi Ooguri, and Cumrun Vafa, Phys. Rev. D **73**, 066002 (2006).