Inflationary initial data for generic spatial topology

Jonathan Morrow-Jones
5191 San Simeon Drive, Santa Barbara, California 93111

Donald M. Witt
Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1
(Received 8 March 1993)

Initial data sets for Einstein's equations with a positive cosmological constant which guarantee inflation to the future are constructed. The construction is local, allowing pieces to be sewn together to create inflationary initial data with a generic spatial topology. A class of the initial data evolves to have the de Sitter metric locally, but the resulting spacetime *cannot* be constructed by performing identifications on de Sitter spacetime or subsets of de Sitter spacetime. These solutions provide a new class of spacetimes which can be used to study the global effects of inflation. The space of Kantowski-Sachs solutions, which is the evolution of another special class of the initial data, is also considered.

PACS number(s): 98.80.Cq, 02.40.Pc, 04.20.Jb

I. BACKGROUND

Observations of our local region of the Universe present puzzles for cosmology. Although there may be structure, matter on large scales is not tremendously clumped together, which gravity tends to do. Also, the Universe is spatially quite flat, showing almost no dominant curvature. And it appears to be very old. Perhaps most perplexing of all, the 3 K background microwave radiation is exceptionally isotropic. Because of the "clumping" nature of gravity, standard big band cosmology would dictate that the early Universe must have been remarkably flat, homogeneous, and isotropic.

Inflationary cosmology attempts to resolve these mysteries by conjecturing an early phase of the Universe whose evolution was dominated by a large positive cosmological constant [1]. The precise mechanism leading to a temporary effective cosmological constant is as yet unsettled. Generally, some high-energy symmetry was spontaneously broken as the Universe was cooling from expansion. During this time, the symmetry-breaking field supercooled, yielding an effective positive cosmological constant.

Whenever this effective cosmological constant arose, the character of the Einstein equations must have changed dramatically. Viewed as an initial data problem in general relativity, the inflation period inherited the spatial character of the Universe at the time of the phase transition, no matter how locally curved or convoluted its topology. Once inherited, Cauchy development of the initial data dictated that the spatial topology could not change, since time evolution is a diffeomorphism from one spatial hypersurface to another. For inflationary cosmology to be appropriate, a large class of initial data must evolve toward exponential and isotropic expansion, under the influence of a large, positive cosmological constant. In this paper we consider the cosmological constant as so dominant; other matter terms in the stress-energy tensor are negligible.

Integration of initial data in general relativity is hampered by the strongly nonlinear nature of its equations. For pure cosmological constant solutions, Wald was able to show that all non-type-IX Bianchi cosmologies, and some type-IX Bianchi cosmologies, inflate and approach exponential and isotropic expansion [2]. Not all Bianchi type-IX solutions inflate—in fact, some collapse [3]. There are two other candidates for spatially homogeneous cosmologies: Kantowski-Sachs type, and $\mathbb{R}\times\mathbb{H}^2$, where $\mathbb{H}^2$ is the hyperbolic plane (constant negative curvature). The space $\mathbb{R}\times\mathbb{H}^2$ does not admit any solutions to the Einstein equations for a positive cosmological constant. However, like Bianchi type IX, some Kantowski-Sachs spaces inflate, while others collapse—these spaces will be discussed later in detail.

For inhomogeneous spacetimes, in the presence of matter, the prospects for proving any particular behavior of solutions is daunting, to say the least. Nonetheless, at least one attempt has been made to generalize Wald's technique to inhomogeneous spacetimes with matter [4]. However, this theorem requires an unphysically motivated assumption: Nowhere during the evolution of the initial data can the spatial scalar curvature be greater than zero. This is a restriction not only upon the initial data, but also upon its evolution. For example, Bianchi type-IX spaces admit initial data which have zero spatial scalar curvature and expand isotropically, but nonetheless halt expansion and collapse [3]. Thus the theorem can only be applied by integrating the initial data, and that defeats its usefulness.

An alternative approach to directly integrating Einstein's equation is to identify "attractor" solutions. These are solutions toward which initial data evolve, asymptotically. Attractor solutions in general relativity are believed to be stationary. Assuming that the initial data in general relativity evolve to be stationary, inflationary cosmology requires that there be a unique, stationary solution, with exponential and isotopic expansion in its future. The statement of this uniqueness is the "cosmic no-hair conjecture" [5]. As originally stated, the conjecture is false, since there exists a solution to the Einstein equations with a positive cosmological constant which is static and does not expand isotropically—the Nariai spacetime [6]. This solution presents no problem for inflation, however, since it is not stable [7] and therefore is not an attractor solution. Some rephrasings of the conjecture have led to proofs assuming properties of the conformal structure of the spacetime [8]. Unfortunately, this assumes a global structure the Universe may very well not possess. Indeed, in this paper, we present initial data which do inflate, but are not conformally well behaved, even in their covering space. Nonetheless, recent work shows that the only stable, locally static solution is a locally de Sitter spacetime and therefore expands exponentially and isotropically to the future [9].

In the absence of matter, the initial data in general relativity consist of the spatial metric and the extrinsic curvature, which describes the embedding of the spatial hypersurface within spacetime. Since general relativity is a gauge theory, initial data cannot be specified freely, but rather must satisfy constraint equations. In this paper we first specify locally spherically symmetric spatial metrics on generic three-topologies in Sec. II. In Sec. III, we solve the constraint equations locally to determine the extrinsic curvature, thereby showing that the topologies of Sec. II admits initial data for the Einstein equations with a positive cosmological constant. In Sec. IV we prove that the evolution of locally spherically symmetric initial data must evolve to be an isometric immersion, or local embedding, in de Sitter, Schwarzschild–de Sitter, or Nariai spacetime. We then show that the topologies of Sec. II admit initial data which evolve to immerse isometrically in de Sitter spacetime. Therefore the de Sitter attractor solution exists for spaces with generic spatial topology. As an interesting consequence of our construction, we create examples of spacetimes which locally have the de Sitter metric, but which cannot be generated by making identifications on either all de Sitter spacetime or a subset of de Sitter spacetime. Finally, in Sec. V we take a special class of spherically symmetric solutions to the Einstein equations with a positive cosmological constant: Kantowski-Sachs spaces. There, we show the entire space of solutions and interpret each trajectory in phase space in terms of the well-known Schwarzschild–de Sitter [5] family of solutions and the lesser-known Nariai solution [6]. By examining this space carefully, we shall see the evolution of each trajectory: those that inflate, those that collapse, and those that end on Cauchy horizons.

II. CONSTRUCTION OF LOCALLY SPHERICALLY SYMMETRIC THREE-SPACES

During inflation, the cosmological constant is very large and positive compared to other matter sources. As a first approximation, other matter terms can be considered negligible compared to the cosmological constant. Assuming that the only source is a cosmological constant, the Einstein field equations are

$$R_{ab}=\Lambda g_{ab} , \tag{1}$$

where Λ is the cosmological constant and R_{ab} denotes the Ricci curvature of the four-metric, g_{ab}.

Since the goal is to describe the evolution of the early Universe, a notion of time is needed. By defining a global time t, four-dimensional spacetime is split into three-dimensional hypersurfaces Σ^3 and time $\mathbb{R}$. The global time is necessary for defining the initial value problem, which in turn restricts topologies of "globally hyperbolic" spacetimes to split, $M^4=\mathbb{R}\times\Sigma^3$. The spacetime metric also splits as a consequence of the time slicing of spacetime,

$$g_{ab}=-n_a n_b+h_{ab} , \tag{2}$$

where n_a is the future-directed normal to the hypersurface of constant time, Σ^3 and h_{ab} is the hypersurface metric. The initial data in general relativity are determined by fixing the spatial metric h_{ab} and its normal derivative

$$K_{ab}=\tfrac{1}{2}\pounds_n h_{ab} , \tag{3}$$

called the extrinsic curvature. The pair (h_{ab},K_{ab}) cannot be specified freely on a hypersurface Σ^3 due to the gauge constraints of general relativity. We defer further discussion of the extrinsic curvature and the constraints of general relativity until the next section.

In this section we show the existence of locally spherically symmetric three-metrics h_{ab} on three-dimensional hypersurfaces Σ^3 with generic three-topology. We show this by explicit construction. Before proceeding with the construction, we define precisely what we mean by local spherical symmetry. Given any Lie group G, a local symmetry with respect to this group is defined.

Definition. A tensor field $T_{ab\cdots}{}^{cd\cdots}$, on a manifold M is *locally symmetric with respect to G* iff every point in M has an open neighborhood U such that the following conditions are satisfied: (i) There is a finite set of vectors fields $\{\xi_I{}^a\}$ on U which generate a faithful representation of the Lie algebra of G; (ii) $\pounds_{\xi_I}T_{ab\cdots}{}^{cd\cdots}|_U=0$ for these vectors.

If the tensor with local symmetry is the metric on the manifold, then the vectors are local Killing vectors. The particular symmetry of interest in the present work is a local spherical symmetry.

Definition. A tensor field $T_{ab\cdots}{}^{cd\cdots}$ on a manifold M is *locally spherically symmetric* iff it is locally symmetric with respect to SO(3) and the orbits of the vector fields are two dimensional.

Definition. A manifold M with metric g_{ab} is considered *locally spherically symmetric* if the metric is locally spherically symmetric.

As a natural extension of these definitions, the initial data are locally spherically symmetric if both h_{ab} and K_{ab} are locally spherically tensors on the initial hypersurface Σ^3. The above abstract conditions allow one to construct local spherical coordinates. Further, the metric and extrinsic curvature will only depend on a radial coordinate.

It follows from the definition that any globally spherically symmetric space must also be locally spherically symmetric. In general, the converse is not true. As an example, take $\mathbb{R}^3$ with the Euclidean metric. It is globally spherically symmetric. Using three linearly independent generators of the translational isometry, we identify the space periodically. The result is a three-torus $T^3 \equiv S^1 \times S^1 \times S^1$. Locally, the metric is the same, and therefore it is locally spherically symmetric, but it is no longer globally spherically symmetric.

We use "surgery" to construct spaces with local spherical symmetry. The fundamental pieces are the round three-sphere, with a line element,

$$ds^2 = a^2[d\psi^2 + \sin^2\psi (d\Omega_2)^2] ; \quad (4)$$

flat $\mathbb{R}^3$, with a line element,

$$ds^2 = a^2[d\psi^2 + \psi^2 (d\Omega_2)^2] ; \quad (5)$$

the hyperbolic plane $\mathbb{H}^3$, with a line element,

$$ds^2 = a^2[d\psi^2 + \sinh^2\psi (d\Omega_2)^2] ; \quad (6)$$

and the handle $\mathbb{R} \times S^2$, with a line element,

$$ds^2 = a^2[d\psi^2 + (d\Omega_2)^2] . \quad (7)$$

The overall factor a merely sets the length scale, and the term $(d\Omega_2)^2 = d\theta^2 + \sin^2\theta \, d\phi^2$ is the round metric on the unit two-sphere—these are the orbits of the local spherical symmetry. All these metrics possess symmetries beyond the local SO(3): The metric given by Eq. (4) has the isometry group SO(4), the metric given by Eq. (5) has the isometry group ISO(3), the metric given by Eq. (6) has the isometry group SO(3,1), and the last metric has the isometry group $\mathbb{R} \times$ SO(3).

A generalized form of these metrics can be written

$$ds^2 = a^2[d\psi^2 + f^2(\psi)(d\Omega_2)^2] . \quad (8)$$

For a general $f(\psi)$, the metric given by Eq. (8) possesses only an SO(3) isometry.

To sew spaces together while preserving local spherical symmetry, we remove a three-ball from the sphere, the hyperbolic plane, or flat space. The boundary of what remains is topologically S^2 and is an orbit of an SO(3) isometry. We want a smooth transition from one space to another while maintaining the general metric form [Eq. (8)]. Specifically, we need the function $f(\psi)$ to be smooth, but not analytic.

We start with the smooth, nonanalytic function

$$p(\psi) = \begin{cases} 0 & \text{if } \psi \le 0, \\ \exp(-1/\psi^2) & \text{if } \psi > 0 \end{cases} .$$

Then define

$$u_\epsilon(\psi) = \frac{\int_{-\infty}^{\psi} dx \, p(x-\epsilon) p(\epsilon - x)}{\int_{-\infty}^{+\infty} dx \, p(x-\epsilon) p(\epsilon - x)} . \quad (9)$$

The function $u_\epsilon(\psi)$ has the property that it is 0 on the interval $(-\infty, -\epsilon]$, it is 1 on the interval $[\epsilon, +\infty)$, and it is a monotonically increasing $C^\infty(\mathbb{R})$ function on the interval $(-\epsilon, \epsilon)$.

Using the function $u_\epsilon(\psi)$, we can now sew various spaces together. If, for the general three-metric of the form Eq. (8) we take

$$f(\psi) = [1 - u_\epsilon(\psi - \psi_1)] \sin(\psi) + u_\epsilon(\psi - \psi_1)(\psi - \psi_2) ,$$

then the resulting metric represents the sewing of a round S^3 into a flat $\mathbb{R}^3$. We show this pictorially in Fig. 1.

Because we sew with a smooth, nonanalytic function, the sewing of the sphere into $\mathbb{R}^3$ has no impact outside the matching region. We can also sew the handle [metric in Eq. (7)] into $\mathbb{R}^3$ or the handle into $\mathbb{H}^3$, etc. Also, before sewing the spaces together, we can create quotient spaces which still admit a locally spherically symmetric metric. An example of this is the space $\mathbb{R}P^3 \equiv S^3/\{\pm I\}$. To preserve the metric, we quotient a space by a discrete subgroup of its isometries. As a result, this procedure can be carried out endlessly—making quotient spaces, cutting, and sewing—until we obtain something like Fig. 2. Topologically, removing a three-ball from two manifolds, followed by sewing the two manifolds together across the resulting boundaries, is called "the connected sum" and is denoted # [10]. The construction we have just given generates locally spherically symmetric metrics on three-spaces with topologies of

$$\frac{S^3}{\Gamma_{1,SO(4)}} \# \frac{S^3}{\Gamma_{2,SO(4)}} \# \cdots \# \frac{S^3}{\Gamma_{k,SO(4)}}$$
$$\# \frac{\mathbb{R}^3}{\Gamma_{1,ISO(3)}} \# \frac{\mathbb{R}^3}{\Gamma_{2,ISO(3)}} \# \cdots \# \frac{\mathbb{R}^3}{\Gamma_{l,ISO(3)}}$$
$$\# \frac{\mathbb{H}^3}{\Gamma_{1,SO(3,1)}} \# \frac{\mathbb{H}^3}{\Gamma_{2,SO(3,1)}} \# \cdots \# \frac{\mathbb{H}^3}{\Gamma_{m,SO(3,1)}}$$
$$\# \frac{\mathbb{R}\times S^2}{\Gamma_{1,\mathbb{R}}} \# \frac{\mathbb{R}\times S^2}{\Gamma_{2,\mathbb{R}}} \# \cdots \# \frac{\mathbb{R}\times S^2}{\Gamma_{n,\mathbb{R}}} , \quad (10)$$

where for the Lie group G, $\Gamma_{i,G}$ is a discrete subgroup of G, possibly trivial (containing only the identity).

Having just shown the existence of an infinite number of examples of three-manifolds that admit a locally spherically symmetric metric, it is natural to wonder how large the above set compares to the set of all three-manifolds. In order to understand this more fully, some important mathematics must be presented.

The work of Thurston [11] on three-manifolds implies that most three-manifolds admit hyperbolic metrics, i.e., a geodesically complete Riemannian metric with constant

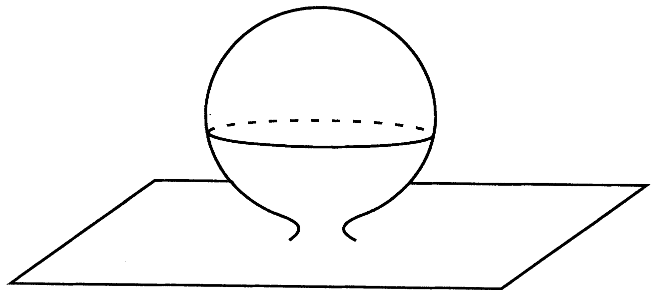

FIG. 1. Example of three-sphere smoothly sewn into flat three-dimensional plane. Sewing preserves local spherical symmetry.

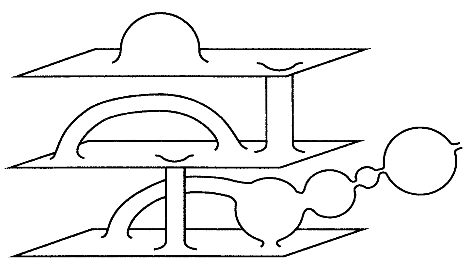

FIG. 2. Example of convoluted initial data hypersurface. Because sewing is local, local spherical symmetry is preserved.

negative sectional curvature. Topologically, all three-manifolds which admit a hyperbolic metrics are obtained from identifying points in $\mathbb{R}^3$ via the action of a discrete subgroup of SO(3,1).

In order to understand Thurston's work, the following background material is needed. First, a *knot* is the continuous embedding of a circle in a three-manifold and a *link* is a finite number of disjoint knots. Next, one performs *Dehn surgery* along a link in a manifold by removing tubular neighborhoods of each knot and then gluing back the removed solid tori differently. More precisely, one identifies the boundary of each hole left by identifying it with the boundary of another solid torus via a homeomorphism of the boundary different from the one defined by the inclusion of the removed torus in the manifold. Lickorish [19] proved that every closed orientable three-manifold can be obtained from Dehn surgery on the three-sphere. More recently, Thurston has proven that every closed orientable three-manifold is obtainable from the three-sphere S^3 by Dehn surgery along a restricted class of links, called hyperbolic links [11]. A link L is a *hyperbolic link* if and only if the three-sphere S^3 minus L admits a hyperbolic metric of finite volume. Furthermore, given any fixed hyperbolic link L, there are an infinite number of distinct closed three-manifolds obtained from surgeries along L which admit hyperbolic metrics and only finite number of closed three-manifolds which do not admit hyperbolic metrics. Hence, given a fixed hyperbolic link L, most three-manifolds obtained from Dehn surgery along L admit a hyperbolic metric. Therefore it follows that most three-manifolds constructed from L have the topology of $\mathbb{R}^3/\Gamma$, where Γ is a discrete subgroup of SO(3,1). Since all closed three-manifolds are obtained from Dehn surgeries along hyperbolic links, it is fair to conclude that in general most three-manifolds admit hyperbolic metrics and, more importantly, have the topology $\mathbb{R}^3/\Gamma$.

Although Thurston's work applies to closed three-manifolds, one can still count open manifolds which have compactifications which are manifolds by applying these results to their compactifications. Combining the above results with the fact that locally spherically symmetric metrics exist on connected sums of hyperbolic three-manifolds, handles, and other three-manifolds, it is fair to conclude that a locally spherically metric occurs on generic topologies.

The greater challenge now is to find *any* extrinsic curvature which satisfies the constraint equations for the initial data in general relativity, much less *locally spherically symmetric* extrinsic curvature. As an example, consider the case of a zero cosmological constant. Because of the positive mass theorem [12], there does not exist any smooth sewing of a nontrivial topological space into Minkowski space, which only alters a compact subset of Minkowski space. This occurs since the sewing of an nontrivial space into Minkowski space introduces a mass into Minkowski space, which impacts the geometry of the space out to infinity. So solving the constraint equations for the extrinsic curvature is not a simple matter. We shall see that the positive cosmological constant allows us to escape the restrictions of the positive mass theorem for a zero cosmological constant.

III. CONSTRUCTION OF THE INITIAL DATA

Recall that in initial value formulations the spacetime metric is split:

$$g_{ab} = -n_a n_b + h_{ab} \ ,$$

where n_a is the future-directed normal to the hypersurface of constant time Σ^3 and h_{ab} is the hypersurface metric. Defining the extrinsic curvature

$$K_{ab} = \tfrac{1}{2}\pounds_n h_{ab} \ ,$$

the Einstein equations for a positive cosmological constant [Eq. (1)] are rewritten

$$-\tfrac{1}{2}\,{}^{(3)}R - \tfrac{1}{2}(K^2 - K^a{}_b K^b{}_a) = -\Lambda \ , \qquad (11)$$

$$D_a K^a{}_b - D_b K = 0 \ , \qquad (12)$$

$$
\begin{aligned}
{}^{(3)}R^a{}_b - \tfrac{1}{2}\,{}^3R h^a{}_b + [\pounds_n(K^a{}_b - h^a{}_b K) \\
- KK^a{}_b + \tfrac{1}{2}h^a{}_b K^2 - \tfrac{1}{2}h^a{}_b(K^c{}_d K^d{}_c)] = -\Lambda h^a{}_b \ , \qquad (13)
\end{aligned}
$$

where D_a is the covariant derivative with respect to the hypersurface metric h_{ab}. The first two equations represent relations between h_{ab} and $K^a{}_b = h^{ac}K_{cb}$. These are the constraint equations for the initial data. Equation (13) along with the definition of $K^a{}_b$ evolves the initial data (h_{ab}, K_{ab}) to form a (unique) spacetime [13]. If Eqs. (11) and (12) are satisfied on the initial hypersurface, then the evolution equations guarantee that the constraints will be obeyed under evolution.

In Sec. II we showed that generic topologies, characterized by the connected sums of Eq. (10), admit locally spherically symmetric metrics in the form of Eq. (8). We must now find $K^a{}_b$, which, along with the spatial metric of Eq. (8), satisfy the constraint Eqs. (11) and (12). Specifically, we seek a $K^a{}_b$ which is locally spherically symmetric. Within the coordinate chart used for Eq. (8), $K^a{}_b$ must be only a function of the coordinate ψ.

Recall the general locally spherically symmetric metric of Eq. (8):

$$ds^2 = a^2[d\psi^2 + f^2(\psi)(d\Omega_2)^2] \ .$$

This metric has a scalar curvature

$$^{(3)}R[f(\psi)]=\frac{1}{a^2f^2}[-4ff''-2(f')^2+2]\ , \tag{14}$$

where $f'=df/d\psi$. The constraint in Eq. (12) becomes

$$K^{\theta}{}_{\theta}=K^{\phi}{}_{\phi} \tag{15}$$

and

$$\left[\partial_\psi+\frac{f'}{f}\right]K^{\phi}{}_{\phi}=\frac{f'}{f}K^{\psi}{}_{\psi}\ . \tag{16}$$

Combining Eqs. (15) and (16) into Eq. (11),

$$\frac{^{(3)}R[f]}{2}+\left[\frac{f}{f'}\partial_\psi(K^{\phi}{}_{\phi})^2+3(K^{\phi}{}_{\phi})^2\right]=\Lambda\ , \tag{17}$$

where $^{(3)}R[f]$ is given by Eq. (14).

We solve Eq. (17) for $(K^{\phi}{}_{\phi})$ with an arbitrary $f(\psi)$ by explicitly integrating from an initial ψ value ψ_0,

$$(K^{\phi}{}_{\phi})^2(\psi)=\frac{\Lambda}{3}+\left[(K^{\phi}{}_{\phi})^2(\psi_0)-\frac{\Lambda}{3}\right]\left[\frac{f(\psi_0)}{f(\psi)}\right]^3+\frac{[f'(\psi)]^2-1}{a^2f^2(\psi)}-f(\psi_0)\frac{[f'(\psi_0)]^2-1}{a^2f^3(\psi)}\ . \tag{18}$$

By Eq. (18), the value of $(K^{\phi}{}_{\phi})^2$ at ψ will be affected by values of $(K^{\phi}{}_{\phi})^2$ and f at ψ_0. Such "far-field" effects are intimately related to the concept of mass. Far-field effects would obstruct placing locally spherically symmetric initial data on spaces with generic topology: The far-field effects of one sewing region would break local spherical symmetry for another sewing region. As a result, choose

$$(K^{\phi}{}_{\phi})^2(\psi_0)=\frac{\Lambda}{3}+\frac{[f'(\psi_0)]^2-1}{a^2f^2(\psi_0)}\ ,$$

and so Eq. (18) becomes

$$(K^{\phi}{}_{\phi})^2(\psi)=\frac{\Lambda}{3}+\frac{[f'(\psi)]^2-1}{a^2f^2(\psi)}\ . \tag{19}$$

With Eq. (19), $K^{\phi}{}_{\phi}$ is determined completely in terms of the local three geometry, $f(\psi)$.

Substituting Eq. (19) into the constraint given by Eq. (16),

$$K^{\psi}{}_{\psi}=\frac{f}{f'}\partial_\psi K^{\phi}{}_{\phi}+K^{\phi}{}_{\phi}\ . \tag{20}$$

The solution is smooth in regions where $f(\psi)\neq 0$. If $f(\psi)\to 0$ (a "cap"), smoothness of $(K^{\phi}{}_{\phi})^2$ requires

$$\lim_{f(\psi)\to 0}|f'(\psi)|=1\ .$$

Also, to keep $^{(3)}R$ nonsingular when $f(\psi)\to 0$,

$$\lim_{f(\psi)\to 0}\left[\frac{f''}{f}\right]\to \text{finite}\ .$$

We have the freedom to choose either one of two roots for $K^{\phi}{}_{\phi}=\pm[(K^{\phi}{}_{\phi})^2]^{1/2}$. As usual, if the inflation epoch was brought about by adiabatic expansion cooling of the Universe, then the initial data must be expanding, which, in our conventions, corresponds to

$$K=K^{\psi}{}_{\psi}+2K^{\phi}{}_{\phi}>0\ .$$

Consequently, we generally take the non-negative root $K^{\phi}{}_{\phi}=[(K^{\phi}{}_{\phi})^2]^{1/2}$.

Consider sewing the three-sphere [Eq. (4)] into one of the other locally spherically symmetric metrics, including another three-sphere. For convenience, let $\psi=0$ be the north pole of the sphere, and so $f(\psi)=\sin(\psi)$ for $\psi<\psi_1-\epsilon$. An interesting case occurs when $\psi_1-\epsilon>\pi/2$, which implies $f'(\psi_1-\epsilon)<0$. Then matching the three-sphere requires that $f(\psi)$ go from decreasing to nondecreasing, which means that there must be a value of ψ in the matching region for which $f'(\psi)=0$. This is precisely the location of the narrowest part of the neck in the matching region, $f_{\min}$. Demanding that the right side of Eq. (19) remain non-negative becomes a restriction on the smallest allowed value of $f_{\min}$:

$$af_{\min}\geq\alpha\ , \tag{21}$$

where the length scale α is given by

$$\alpha^2=3/\Lambda\ . \tag{22}$$

Smaller necks cannot be used to match the three-sphere into another locally spherically symmetric space.

In fact, in order to keep the sewings local, whenever the function $f(\psi)$ passes through a minimum, it must satisfy the inequality of Eq. (21). Any smaller necks must induce far-field effects. This suggests that *all* smaller necks carry a "mass." In Sec. V we show this is precisely the case.

Using the general locally spherically symmetric metric given by Eq. (8) and the extrinsic curvature given by Eqs. (19), (15), and (20), we have just constructed locally spherically symmetric initial data for the topologies of Eq. (10). These initial data are smooth for the following reasons: The locally spherically symmetric metrics were already constructed to be smooth. The expression (19) can only be singular at $f(\psi)=0$; however, that will never happen in a matching region between the different topologies. Finally, (20) is nonsingular for $f'(\psi)\neq 0$. In order to show that it is nonsingular for $f'(\psi)=0$, rewrite it in the form

$$K^{\psi}{}_{\psi}=\frac{f}{2f'K^{\phi}{}_{\phi}}\partial_\psi(K^{\phi}{}_{\phi})^2+K^{\phi}{}_{\phi}\ . \tag{23}$$

By explicitly differentiating the $(K^{\phi}{}_{\phi})^2$ term in (23) and using (19), one can check that the f' cancels out and expression (23) is nonsingular for our choices. Since the initial data are nonsingular and constructed from algebraic expressions involving smooth functions and derivatives of smooth functions, the initial data are smooth. We must now determine what spacetimes evolve from these initial data.

IV. EVOLUTION OF THE INITIAL DATA

In Secs. II and III, we showed that three-manifolds with generic topology admit locally spherically symmetric initial data. This high degree of symmetry of our initial data allows analytical integration into a spacetime.

As a first step toward integrating our initial data for generic topology, we evolve the initial data for the building blocks of our construction: the three-sphere [Eq. (4)], flat $\mathbb{R}^3$ [Eq. (5)], the hyperbolic plane [Eq. (6)], and the handle [Eq. (7)]. The "integration" is performed by showing that our locally spherically symmetric initial data are induced through an isometric embedding in de Sitter spacetime. Since a given set of initial data uniquely evolves into a spacetime, the building blocks must evolve spacetime metrics which are locally isometric to the de Sitter metric.

de Sitter spacetime is the Lorentzian version of a sphere, sometimes called a "pseudosphere" [14]. Given five-dimensional Minkowski space, with a metric,

$$ds^2 = -dz_0^2 + dz_1^2 + dz_2^2 + dz_3^2 + dz_4^2 , \quad (24)$$

de Sitter spacetime is the four-dimensional manifold which satisfies

$$\alpha^2 = -z_0^2 + z_1^2 + z_2^2 + z_3^2 + z_4^2 , \quad (25)$$

where $\alpha^2 = 3/\Lambda$. The Minkowski metric [Eq. (24)] has the isometry group ISO(4,1). de Sitter spacetime inherits the isometry group SO(4,1) < ISO(4,1). Topologically, de Sitter spacetime is $\mathbb{R} \times S^3$ and is therefore simply connected. de Sitter spacetime is geodesically complete and has a constant sectional curvature (hence a spaceform). It is globally hyperbolic, meaning that there exists a hypersurface from which the initial data integrate to give the entire spacetime. Any such hypersurface *must* be diffeomorphic to S^3. For the initial data to evolve into a geodesically complete spacetime with a constant sectional curvature, the universal cover of the initial data hypersurface must be topologically S^3, and the universal cover of the resulting spacetime must be isometric to de Sitter spacetime [14].

Here we show that our initial data for generic spatial topology evolve to be locally isometric to de Sitter spacetime. In general, the universal cover of these topologies will not be diffeomorphic to S^3, and therefore the initial data *cannot* evolve into a geodesically complete spacetime: There will be some time to the past, future, or possibly both, beyond which the initial data cannot be evolved. Physically, this does not present a problem. We are interested in inflationary spacetimes. Times before the initial data hypersurface predate the inflationary epoch, in which spacetime evolved according to different (and as yet unspecified) physics. We will choose expanding initial data, which evolve infinitely far to the future. As we shall see, these spaces do possess noncrushing singularities and/or Cauchy horizons to the past.

Consider now the following parametrization of de Sitter spacetime [15]:

$$z_0 = \alpha \sinh(t/\alpha) ,$$
$$z_1 = \alpha \cosh(t/\alpha)\sin(\psi)\sin(\theta)\cos(\phi) ,$$
$$z_2 = \alpha \cosh(t/\alpha)\sin(\psi)\sin(\theta)\sin(\phi) ,$$
$$z_3 = \alpha \cosh(t/\alpha)\sin(\psi)\cos(\theta) ,$$
$$z_4 = \alpha \cosh(t/\alpha)\cos(\psi) ,$$

creating the de Sitter metric

$$ds^2 = -dt^2 + \alpha^2\cosh^2(t/\alpha)[d\psi^2 + \sin^2(\psi)(d\Omega_2)^2] .$$

The t=const hypersurfaces are round three-spheres with a spatial metric [Eq. (4)]. These t=const hypersurfaces are intersections of de Sitter spacetime [Eq. (25)] with spacelike planes in Minkowski $\mathbb{R}^5$, parametrized by

$$z_0 = c, \quad z_1 = \lambda ,$$

where c=const and λ is a free parameter. For expanding three-spheres, choose $c > 0$. The extrinsic curvature of the hypersurfaces is

$$K_{ab} = \tfrac{1}{2}\pounds_n h_{ab} = \tfrac{1}{2}\pounds_t h_{ab} = \frac{1}{\alpha}\tanh(t/\alpha) h_{ab} ,$$

which is the extrinsic curvature given by Eqs. (19), (15), and (20), with $a = \alpha\cosh(t/\alpha)$ and $f(\psi) = \sin(\psi)$. Thus our locally spherically symmetric initial data for the three-sphere evolve to be locally de Sitter spacetime, within its domain of dependence.

If we now consider a different parametrization of de Sitter spacetime,

$$z_0 = \alpha \sinh(t/\alpha) + \tfrac{1}{2}\alpha e^{t/\alpha}\psi^2 ,$$
$$z_1 = \alpha \cosh(t/\alpha) - \tfrac{1}{2}\alpha e^{t/\alpha}\psi^2 ,$$
$$z_2 = \alpha e^{t/\alpha}\psi \sin(\theta)\cos(\phi) ,$$
$$z_3 = \alpha e^{t/\alpha}\psi \sin(\theta)\sin(\phi) ,$$
$$z_4 = \alpha e^{t/\alpha}\psi \cos(\theta) ,$$

we obtain the de Sitter metric

$$ds^2 = -dt^2 + \alpha^2 e^{2t/\alpha}[d\psi^2 + \psi^2(d\Omega_2)^2] .$$

The t=const hypersurfaces are flat $\mathbb{R}^3$'s with a spatial metric [Eq. (5)]. These t=const hypersurfaces are intersections of de Sitter spacetime [Eq. (25)] with null planes in Minkowski $\mathbb{R}^5$, parametrized by

$$z_0 = c + \lambda, \quad z_1 = \lambda ,$$

where c=const and λ is a free parameter. For expanding $\mathbb{R}^3$'s, choose $c > 0$. The extrinsic curvature of the hypersurfaces is

$$K_{ab} = \frac{1}{\alpha} h_{ab} ,$$

which is the extrinsic curvature given by Eqs. (19), (15), and (20), with $a = \alpha e^{t/\alpha}$ and $f(\psi) = \psi$. Thus our locally spherically symmetric initial data for a flat $\mathbb{R}^3$ evolve to be locally de Sitter spacetime, within its domain of dependence.

Choosing yet another parametrization of de Sitter spacetime,

$$z_0 = \alpha \sinh(t/\alpha)\cosh(\psi) ,$$
$$z_1 = \alpha \cosh(t/\alpha) ,$$
$$z_2 = \alpha \sinh(t/\alpha)\sinh(\psi)\sin(\theta)\cos(\phi) ,$$
$$z_3 = \alpha \sinh(t/\alpha)\sinh(\psi)\sin(\theta)\sin(\phi) ,$$
$$z_4 = \alpha \sinh(t/\alpha)\sinh(\psi)\cos(\theta) ,$$

we obtain the de Sitter metric

$$ds^2=-dt^2+\alpha^2\sinh^2(t/\alpha)[d\psi^2+\sinh^2(\psi)(d\Omega_2)^2] .$$

The t=const hypersurfaces are hyperbolic planes $\mathbb{H}^3$ with a spatial metric [Eq. (6)]. These t=const hypersurfaces are intersections of de Sitter spacetime [Eq. (25)] with timelike planes in Minkowski $\mathbb{R}^5$, parametrized by

$$z_0=\lambda, \quad z_1=c ,$$

where $c=\text{const}>\alpha$ and λ is a free parameter. The extrinsic curvature of the hypersurfaces is

$$K_{ab}=\frac{1}{\alpha}\coth(t/\alpha)h_{ab} ,$$

which is the extrinsic curvature given by Eqs. (19), (15), and (20), with $a=\alpha\sinh(t/\alpha)$ and $f(\psi)=\sinh(\psi)$. Thus our locally spherically symmetric initial data for $\mathbb{H}^3$ evolve to be locally de Sitter spacetime, within its domain of dependence.

Finally, considering the parametrization of de Sitter spacetime,

$$z_0=\alpha\sinh(t/\alpha)\cosh(\psi) ,$$

$$z_1=\alpha\sinh(t/\alpha)\sinh(\psi) ,$$

$$z_2=\alpha\cosh(t/\alpha)\sin(\theta)\cos(\phi) ,$$

$$z_3=\alpha\cosh(t/\alpha)\sin(\theta)\sin(\phi) ,$$

$$z_4=\alpha\cosh(t/\alpha)\cos(\theta) ,$$

we obtain the de Sitter metric

$$ds^2=-dt^2+\alpha^2[\sinh^2(t/\alpha)d\psi^2+\cosh^2(t/\alpha)(d\Omega_2)^2] , \tag{26}$$

which is a Kantowski-Sachs metric. The t=const hypersurfaces are handles $\mathbb{R}\times S^2$ with a spatial metric [Eq. (7)], where the t-constant parameter $\sinh(t/\alpha)$ is absorbed in rescaling ψ. These t=const hypersurfaces are intersections of de Sitter spacetime [Eq. (25)] with bent planes in Minkowski $\mathbb{R}^5$, parametrized by

$$z_0=c\cosh(\lambda/c), \quad z_1=c\sinh(\lambda/c) ,$$

where c=const and λ is a free parameter. The extrinsic curvature of the hypersurfaces is

$$K_{\psi\psi}=\frac{1}{\alpha}\coth(t/\alpha)h_{\psi\psi} ,$$

$$K_{\phi\phi}=\frac{1}{\alpha}\tanh(t/\alpha)h_{\phi\phi} ,$$

$$K_{\theta\theta}=\frac{1}{\alpha}\tanh(t/\alpha)h_{\theta\theta} ,$$

which is the extrinsic curvature given by Eqs. (19), (15), and (20), with $a=\alpha\cosh(t/\alpha)$ and $f(\psi)=1$. Thus our locally spherically symmetric initial data for handles evolve to be locally de Sitter spacetime, within its domain of dependence. The Killing field $(\partial/\partial\psi)^a$ is the restriction of a Lorentz boost field in Minkowski $\mathbb{R}^5$ to de Sitter spacetime in regions where it is spacelike. In regions where the boost field is timelike, we get the well-known static slicing of de Sitter spacetime:

$$ds^2=-\cos^2(\psi)dt^2+\alpha^2(d\Omega_3)^2 .$$

The Kantowski-Sachs slicing of de Sitter spacetime is analogous to the interior solution of a Schwarzschild black hole (which is also Kantowski-Sachs type). We will see this in greater detail in Sec. V.

We have now seen that for the fundamental building blocks of our initial data sets, the initial data evolve to be locally de Sitter spacetime, within the domain of dependence of each piece. This, however, says nothing of the evolution of the matching regions.

In order to discuss the evolution of the initial data in the matching regions, we need the following result.

Proposition. Let (h_{ab},K_{ab}) be an initial data set on Σ^3 which satisfies the constraints given by Eqs. (11) and (12) and is locally invariant under the Lie group G. Then the initial data set evolves into a spacetime such that for some open neighborhood of Σ^3 the spacetime metric is locally invariant under G.

The initial data set on a hypersurface Σ^3 evolves into a spacetime with the topology $\mathbb{R}\times\Sigma^3$ and metric g_{ab}. To evolve the initial data, we introduce a time function t and a vector flow of time t^a such that $t^a\nabla_a t=1$. The time flow need not be normal to the surfaces of constant t. In general, the time flow will have a projection onto the normal n^a and a projection onto the hypersurface:

$$t^a=Nn^a+N^a .$$

The metric split [Eq. (2)] becomes

$$g_{ab}=-N^2\nabla_a t\nabla_b t+h_{ab} .$$

A *synchronous gauge* results from choosing $N=1$ and $N^a=0$, in which case the flow of time is normal to the hypersurfaces of constant t. Locally, a synchronous gauge is always a valid gauge choice for a spacetime. Using the synchronous gauge, the spacetime metric can be written as

$$ds^2=-dt^2+h_{ij}dx^idx^j ,$$

where the coordinates x^i are spatial coordinates.

Let $\{\xi_l^a\}$ be the set of Lie algebra generators of the group G. Using the above slicing of $M\equiv\mathbb{R}\times\Sigma^3$, each of the vectors fields ξ_l^a can be trivially extended to M by requiring that each have no normal component and that they be time independent. The extensions of the vector fields $\{\xi_l^a\}$ to M will also be denoted by $\{\xi_l^a\}$. In the above coordinates, the explicit vectors are $\xi_l^m(x,t)=(0,\xi_l^i(x))$, where $m=0,1,2,3$ and $i=1,2,3$. Therefore there is a set of vector fields $\{\xi_l^a\}$ on an open subset of M which generates the Lie algebra of G.

The flows of the vector fields $\{\xi_l^a\}$ generate a family of diffeomorphisms $\{\Psi_l\}$ of an open subset of M. Observe that these diffeomorphisms only act on the spatial coordinates and always map the coordinate t to itself because the vector fields have vanishing t components. Let $h_{ab}(t)$ and $K_{ab}(t)$ be the time evolutions of the three-metric and extrinsic curvature with $h_{ab}(0)=h_{ab}$ and $K_{ab}(0)=K_{ab}$. Now $\Psi_l h_{ab}(t)$ and $\Psi_l K_{ab}(t)$ are also a solution to the Einstein equations because of coordinate invariance. However, the initial data for both solutions are the same because $\Psi_l h_{ab}(0)=h_{ab}(0)$ and $\Psi_l K_{ab}(0)=K_{ab}(0)$ by construction. Suppose that, for some t, $\Psi_l h_{ab}(t)\neq h_{ab}(t)$ and/or $\Psi_l K_{ab}(t)\neq K_{ab}(t)$; then, $\Psi_l h_{ab}(t)$ and $\Psi_l K_{ab}(t)$ are a different solution to the initial value problem with the initial data h_{ab} and K_{ab}. However, this cannot be true because the initial value problem has a unique evolution. Therefore $\Psi_l h_{ab}(t)=h_{ab}(t)$ and $\Psi_l K_{ab}(t)=K_{ab}(t)$.

The above proof can be applied to any matter sources coupled to gravity provided that the sources also have the same local symmetry G and have a unique initial value problem. Also, the proof applies to discrete symmetries.

An important corollary of the proposition is that locally spherically symmetric initial data for Eq. (1) evolve, for some time into the future, to be a locally spherically symmetric spacetime. In order to prove this, apply the above proposition with $G=\mathrm{SO}(3)$ and use the fact that the time flow generates a diffeomorphism, which means the orbits of SO(3) do not change dimension under time evolution.

The above results imply that all the locally spherically symmetric initial data sets constructed must evolve into locally spherically symmetric spacetimes. Even though the topologies are very general, the local spacetime metrics for spacetimes satisfying Eq. (1) with local spherical symmetry are very restricted. This is demonstrated by the following theorem which is a generalization of Birkhoff's theorem.

Theorem (cosmic Birkhoff). The only locally spherically symmetric solutions to Eq. (1) are locally isometric either to one of the Schwarzschild–de Sitter family of solutions or Nariai spacetime [6]. Further, these solutions are real analytic within their coordinate charts.

Before discussing the proof the theorem, let us recall the properties of the Nariai solution. The Nariai spacetime metric is

$$ds^2=\alpha^2[-dt^2+\tfrac{1}{3}\cosh^2(\sqrt{3}t)d\psi^2+\tfrac{1}{3}(d\Omega_2)^2] . \quad (27)$$

It is the Cartesian product of two-dimensional de Sitter spacetime and constant round two-spheres with the product metric. It is homogeneous with the isometry group $\mathrm{SO}(2,1)\times\mathrm{SO}(3)$.

Birkhoff's theorem is usually given for vacuum Einstein equations, [16] but the proof generalizes readily for cosmological constant [3]. In particular, a spherically symmetric metric can be written

$$ds^2=-\frac{dt^2}{F^2(t,r)}+X^2(t,r)dr^2+Y^2(t,r)(d\Omega_2)^2 . \quad (28)$$

By denoting time derivatives by $\dot F$, $\dot X$, and $\dot Y$ and radial derivatives by F', X', and Y', Eq. (1) becomes

$$\frac{\dot Y'}{Y}-\frac{\dot X Y'}{XY}+\frac{\dot Y F'}{YF}=0 , \quad (29)$$

$$\frac{1}{Y^2}+\frac{2}{X}\left[-\frac{Y'}{XY}\right]'-3\left[\frac{Y'}{XY}\right]^2+2F^2\frac{\dot X\dot Y}{XY}+F^2\left[\frac{\dot Y}{Y}\right]^2=\Lambda , \quad (30)$$

$$\frac{1}{Y^2}+2F\left[F\frac{\dot Y}{Y}\right]^{\cdot}+3\left[\frac{\dot Y}{Y}\right]F^2+\frac{2}{X^2}\frac{Y'F'}{YF}-\left[\frac{X'}{XY}\right]^2=\Lambda . \quad (31)$$

First, we address the degenerate case, which occurs when all the S^2 orbits of the SO(3) isometry have the same radius. With respect to the metric given by Eq. (28), $\nabla_a Y\nabla^a Y=0$ implies S^2 orbits of constant radius, which is equivalent to requiring

$$\frac{Y'}{X}=F\dot Y . \quad (32)$$

Equation (32) is then substituted into Eqs. (29) and (30). These are, in turn, both solved for $\dot Y'$, giving two different expressions. Consistency implies

$$Y=\left[\frac{1}{\Lambda}\right]^{1/2} . \quad (33)$$

So, finally,

$$F^2=X^2=\frac{1}{1-\Lambda r^2} . \quad (34)$$

Equations (33) and (34) specify the Nariai solution to Eq. (1).

When $\nabla_a Y\nabla^a Y<0$, surfaces of constant Y are timelike, and so Y can be used to define a radial coordinate $r\equiv Y$. One can find a new t for which $\nabla t\cdot\nabla r=0$. Then Eq. (29) implies $\dot X=0$. Further, Eq. (31) leads to $(F'/F)^{\cdot}=0$, and so, with a proper choice of t, $\dot F=0$. Therefore the solution is necessarily static (in this region). Equation (30) becomes

$$1-\left[\frac{r}{X^2}\right]'=\Lambda r^2 ,$$

which integrates to

$$X^2=\left[1-\frac{2M}{r}-\frac{\Lambda}{3}r^2\right]^{-1} , \quad (35)$$

which is the Schwarzschild–de Sitter family of solutions, parametrized by M.

The case of $\nabla_a Y\nabla^a Y>0$ is dealt with similarly. Therefore the only spherically symmetric solutions to Eq. (1) are Schwarzschild–de Sitter and Nariai spacetimes. Analyticity follows from the restrictive form of the Einstein equations for spherical symmetric solutions. Because of analyticity, if somewhere in a coordinate chart a solution is in the Schwarzschild–de Sitter family with mass parameter M, then it must remain in the Schwarzschild–de Sitter family with the same mass parameter throughout the chart. Similarly, if a solution is Nariai type somewhere within a chart, then it must be Nariai type throughout the chart. If the spacetime is covered by these charts, then it is either everywhere locally Nariai type or everywhere locally Schwarzschild–de Sitter type with fixed mass parameter M.

As we saw in our proposition, our locally spherically symmetric initial data evolve to be a locally spherically symmetric spacetime. Using Birkhoff's theorem, we know the initial data evolve to be everywhere locally Nariai type or everywhere locally Schwarzschild–de Sitter type with fixed parameter M. For the initial data for generic spatial topologies, we saw that the building blocks evolved to be de Sitter spacetime ($M=0$). *Therefore our locally spherically symmetric initial data for generic spatial topology evolve to be everywhere locally de Sitter spacetime.*

One might question whether these solutions are merely unusual slicings of de Sitter spacetime, with identifications (quotient spaces). If this is true, then the universal cover of the initial data hypersurface must isometrically embed *globally* in de Sitter spacetime as a spacelike hypersurface. While such embeddings occur for special cases, they will not occur generically. We show this by counterexample.

Consider Fig. 3, in which a series of flat planes are joined, one to the other, by $\mathbb{R}\times S^2$ handles. This space is simply connected, and so it represents its own universal cover. We show that this space cannot isometrically embed in de Sitter spacetime.

Each of the flat $\mathbb{R}^3$ planes in Fig. 3 is obtained by intersecting a null plane in Minkowski $\mathbb{R}^3$ with the de Sitter hyperboloid. If two null planes are parallel, then one plane lies to the future of the other. Therefore the two cannot both be part of an everywhere spatial and path-connected hypersurface. So each of the flat $\mathbb{R}^3$ planes of Fig. 3 must represent the intersection of the de Sitter hyperboloid with a different null plane, no two of which are parallel. Of course, in a flat space, any two nonparallel planes with codimension 1 must intersect. The question is whether any two nonparallel null planes have a nonempty intersection with the de Sitter hyperboloid.

Recall that the de Sitter hyperboloid is isometrically embedded in Minkowski $\mathbb{R}^5$ and is the solution to

$$\alpha^2=-z_0^2+z_1^2+z_2^2+z_3^2+z_4^2 ,$$

where $\alpha^2=3/\Lambda$. This hyperboloid is invariant under SO(4,1). One of the two null planes is given parametrically:

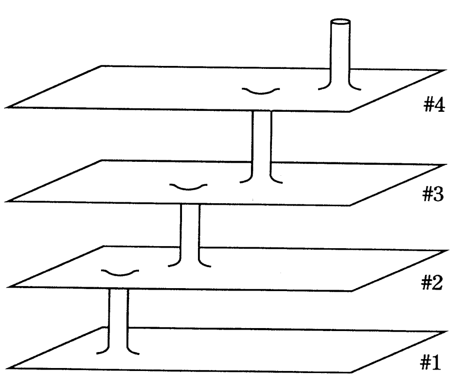

FIG. 3. Example of locally spherically symmetric initial data which is simply connected, but does not embed in de Sitter spacetime.

$$z_0=c_1+\lambda, \quad z_1=\lambda , \tag{36}$$

where $c_1=\text{const}$ and λ is a free parameter. Since we are interested in expanding solutions, $c_1>0$. The other nonparallel null plane is given by

$$z_0=c_2+\tau, \quad \sum_{i=1}^{4} w_i z_i=\tau, \quad \sum_{i=1}^{4} w_i^2=1 ,$$

where $c_2=\text{const}>0$ and τ is a free parameter. By choosing an element of $SO(3)<SO(4,1)$ which acts only upon z_2, z_3, and z_4, we write the second null plane:

$$z_0=c_2+\tau, \quad w_1z_1+w_2z_2=\tau, \quad w_1^2+w_2^2=1 , \tag{37}$$

without loss of generality. There are two cases to consider: $w_2\neq 0$ with $|w_1|<1$ or $w_2=0$ with $w_1=-1$.

If $w_2\neq 0$, then

$$\begin{aligned} &\tau=\lambda+c_1-c_2, \quad z_0=\lambda+c_1, \\ &z_1=\lambda, \quad |z_2|=\frac{|(1-w_1)\lambda+c_1-c_2|}{(1-w_1^2)^{1/2}} . \end{aligned} \tag{38}$$

Using Eq. (38) in Eq. (25), we get

$$\left[\frac{1-w_1}{1+w_1}\right]\left[\lambda-\frac{w_1c_1+c_2}{1-w_1}\right]^2+z_3^2+z_4^2 =\alpha^2+\frac{2}{1-w_1}c_1c_2>0 .$$

This equation always has a solution for a topological S^2. Therefore the two nonparallel null planes have nonempty intersection in the de Sitter hyperboloid.

When $w_2=0$, then

$$\sum_{i=2}^{4} z_i^2=\alpha^2+2c_1c_2>0 .$$

which again has a solution for a topological S^2. So any two nonparallel null planes with $c_1>0$ and $c_2>0$ (necessary for inflation) must have a nonempty intersection with the de Sitter hyperboloid.

Returning to Fig. 3, recall that no two planes can be parallel for an initial data surface. But we have just seen that any two nonparallel planes must have a topological S^2 intersection. Plane No. 1 intersects plane No. 2 in a topological S^2. Plane No. 2 intersects plane No. 3 in a topological S^2. But planes Nos. 1 and 3 do not intersect. Therefore the initial data on the hypersurface represented by Fig. 3 *cannot* be induced by a global, isometric embedding in de Sitter spacetime.

We have shown in this section that locally spherically symmetric initial data for Eq. (1) must evolve to be everywhere locally isometric to Schwarzschild–de Sitter spacetime with a fixed mass parameter M or everywhere locally isometric to Nariai spacetime. In order to avoid far-field effects from the sewing procedure, we found that the initial data for generic spatial topologies must evolve to be everywhere locally isometric to de Sitter spacetime. Further, we showed that this construction is nontrivial: The universal cover of the initial data does not generically embed globally in de Sitter spacetime.

V. SPACE OF KANTOWSKI-SACHS SOLUTIONS

We have demonstrated that three-surfaces with generic topology admit initial data which evolve to be everywhere locally de Sitter type, and therefore inflate. However, when we constructed the initial data for generic spatial topologies in Sec. III, we found that the narrowest allowed necks $\mathbb{R}\times S^2$ in matching regions are bounded by

$$af_{\min}\geq \alpha \ ,$$

where the length scale $\alpha^2=3/\Lambda$. Smaller necks will create far-field effects away from a matching region. In this section we show that the far-field effects are induced by the smaller necks because they carry "mass."

Recall that the locally spherically symmetric three-metric is

$$ds^2=a^2[d\psi^2+f^2(\psi)(d\Omega_2)^2] \ ,$$

within a coordinate chart. The narrowest neck of a matching region occurs at a point where $f'(\psi)=0$. Here we concentrate on regions for which $f'(\psi)=0$. In such regions the initial data possess an $\mathbb{R}\times\mathrm{SO(3)}$ local symmetry. From Sec. IV we know that the evolution of the initial data preserves its spatial symmetries within the domain of dependence of the initial data. Spacetime metrics with an $\mathbb{R}\times\mathrm{SO(3)}$ isometry are called Kantowski-Sachs [17] metrics and take the form

$$ds^2=\alpha^2[-dt^2+b^2(t)d\psi^2+c^2(t)(d\Omega_2)^2] \ , \qquad (39)$$

where we have removed a constant length scale $\alpha^2=3/\Lambda$. The Einstein equations for a positive cosmological constant [Eq. (1)] take the form

$$2\left[\frac{\dot b}{b}\right]\left[\frac{\dot c}{c}\right]+\left[\frac{\dot c}{c}\right]^2=3-\frac{1}{c^2} \ , \qquad (40)$$

$$\frac{d}{dt}(b\dot b)+2(b\dot b)\left[\frac{\dot c}{c}\right]-\dot b^2=3b^2 \ , \qquad (41)$$

$$\frac{d}{dt}(c\dot c)+(c\dot c)\left[\frac{\dot b}{b}\right]+1=3c^2 \ . \qquad (42)$$

The first equation is a constraint on the initial data, arising from Eq. (11). The second two equations evolve the initial data and come from Eq. (13). Naively, the initial data on a time surface $t=t_0$ would be $(b_0,\dot b_0,c_0,\dot c_0)$, satisfying Eq. (40). However, Eqs. (40)–(42) are independent $\beta\equiv\ln b$, depending only upon its time derivatives. This occurs since b_0 contains no geometric information, being absorbable in a rescaling of ψ in the metric given by Eq. (39). Therefore the initial data are given by $(\dot b_0/b_0,c_0,\dot c_0)$, subject to the constraint given by Eq. (40).

Now consider the constraint given by Eq. (40). When $dc/dt\neq 0$, Eq. (40) can be solved for $\dot b/b$. As a result, *when* $dc/dt\neq 0$, *the space of Kantowski-Sachs solutions is given by* $(c,\dot c)$. However, when $dc/dt=0$, Eq. (40) dictates that $c=1/\sqrt{3}$—this is the Nariai solution. So when $dc/dt=0$, all the dynamical information is contained in $\dot b/b$. We now have a picture of the space of Kantowski-Sachs solutions (Fig. 4). It consists of two half-planes connected by a point on their boundaries, where the point is part of a line coming out of the figure:

space of Kantowski-Sachs initial data

$$=\left\{\left[\frac{\dot b}{b},c,\frac{dc}{dt}\right]\middle|\frac{\dot b}{b}\text{ given by Eq. (40) } c>0,\text{ and }\frac{dc}{dt}>0\right\}$$
$$\cup\left\{\left[\frac{\dot b}{b},c,\frac{dc}{dt}\right]\middle|\frac{\dot b}{b}\text{ given by Eq. (40) } c>0,\text{ and }\frac{dc}{dt}<0\right\}$$
$$\cup\left\{\left[\frac{\dot b}{b},c,\frac{dc}{dt}\right]\middle|c=1/\sqrt{3}\text{ and }\frac{dc}{dt}=0\right\}$$

Kantowski-Sachs solutions, like locally spherically symmetric solutions in general, are restricted by the (cosmic) Birkhoff theorem. They must be either Nariai spacetime or in the Schwarzschild–de Sitter family with a fixed mass parameter M. Therefore all trajectories in Fig. 4 can be identified with a Schwarzschild–de Sitter solution or Nariai solution.

We have already seen that Nariai spacetime is the fixed point joining the upper half-plane to the lower half-plane. Moreover, the Nariai fixed point is unstable under Kantowski-Sachs perturbations [7]. When perturbed, Nariai spacetime either expands, approaching the de Sitter trajectory, or it collapses, approaching a Schwarzschild–de Sitter black hole singularity. The dynamics of the Nariai solution are not shown in Fig. 4, since its evolution occurs along the $\dot b/b$ dimension, coming out of the figure.

In Sec. IV we showed that de Sitter spacetime ($M=0$) has a Kantowski-Sachs form

$$ds^2=-dt^2+\alpha^2[\sinh^2(t/\alpha)d\psi^2+\cosh^2(t/\alpha)(d\Omega_2)^2] \ .$$

In Fig. 4 the de Sitter trajectory appears as the curve that "crosses" $\dot c=0$ at $c=1$. Of course, the trajectory does not actually cross the $\dot c=0$ axis, since it is not part of the phase space.

To understand the Kantowski-Sachs form of Schwarzschild–de Sitter solutions, for arbitrary M, consider the static chart of Schwarzschild–de Sitter [5] solutions:

$$ds^2=-\left[1-\frac{\Lambda}{3}r^2-\frac{2M}{r}\right]dt^2+\frac{dr^2}{1-(\Lambda/3)r^2-2M/r}+r^2(d\Omega_2)^2 \ .$$

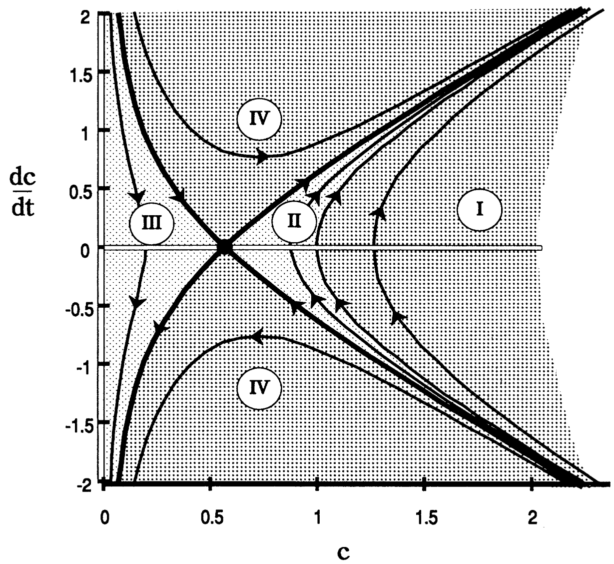

FIG. 4. Phase space for space of Kantowski-Sachs solutions. It consists of two half-planes (upper and lower) joined at the point of the separatrix. The unstable fixed point is Nariai type. All other trajectories are Schwarzschild–de Sitter type.

There are values of r for which the rational function $(r-r^3/\alpha^2-2M)/r$ is not positive for any value of the parameter M/α. If we take this rational function to be negative, the coordinate r becomes a time coordinate and the coordinate t becomes a spatial coordinate. For this case we write

$$ds^2=\alpha^2\left[-\frac{dT^2}{T^2+2M/\alpha T-1}+\left[T^2+\frac{2M}{\alpha T}-1\right]d\psi^2+T^2(d\Omega_2)^2\right] . \quad (43)$$

If we define a new time coordinate

$$dt=\frac{dT}{(T^2+2M/\alpha T-1)^{1/2}} ,$$

Eq. (43) takes the form of Eq. (39). In Fig. 4 the trajectory traced out by a Schwarzschild–de Sitter solution is given parametrically:

$$c=T, \quad \frac{dc}{dt}=\pm\left[T^2+\frac{2M}{\alpha T}-1\right]^{1/2} . \quad (44)$$

The plus sign gives the upper half-plane in Fig. 4, and the minus sign gives the lower half-plane (the time reversal of the upper half-plane). The trajectories in Eq. (44) are characterized by the number of real roots of the rational function:

$$\mathcal{R}(T)\equiv\left[T^2+\frac{2M}{\alpha T}-1\right] , \quad (45)$$

which corresponds to the number of "crossings" of the $\dot{c}=0$ axis. The possible cases are

$$M/\alpha\leq 0 \Longrightarrow \mathcal{R}(T) \text{ has one positive, real root ,} \quad (46)$$

$$0<M/\alpha<\frac{1}{3\sqrt{3}} \Longrightarrow \mathcal{R}(T) \text{ has two positive, real roots ,} \quad (47)$$

$$M/\alpha=\frac{1}{3\sqrt{3}} \Longrightarrow \mathcal{R}(T) \text{ has one positive, real, double root ,} \quad (48)$$

$$M/\alpha>\frac{1}{3\sqrt{3}} \Longrightarrow \mathcal{R}(T) \text{ has no positive real roots .} \quad (49)$$

The number of roots is directly related to the number of event horizons. For the Schwarzschild–de Sitter family of solutions, there are two kinds of potential event horizons: cosmic, due to the rapid expansion brought about by Λ, and black holes, brought about by the mass parameter M.

When the mass parameter falls in the range of Eq. (46), there is only one crossing of the $\dot{c}=0$ axis (the cosmic event horizon). These are the trajectories of region I in Fig. 4, and it includes de Sitter spacetime ($M=0$). When the mass parameter falls in the range of Eq. (47), there are two crossings of the $\dot{c}=0$ axis (both cosmic and black hole event horizons). These are the trajectories of regions II and III in Fig. 4. Region II is connected to the cosmic event horizon, and region III is connected to the black hole event horizon. When the mass parameter falls in the range of Eq. (48), there is only one crossing of the $\dot{c}=0$ axis. In this case the cosmic and black hole horizons *coincide*. This is the trajectory which forms the separatrix of Fig. 4, crossing at the Nariai fixed point. This solution has sometimes been associated with the Nariai solution, [18] but it is clearly distinct. Unlike the Nariai solution, it is not homogeneous, and it has no geodesic completion, since it possesses a crushing singularity. Although Fig. 4 shows the $M/\alpha=1/3\sqrt{3}$ trajectory crossing at the Nariai fixed point, it actually takes an infinite amount of comoving time, $dt=dT/(T^2+2M/\alpha T-1)^{1/2}$, to reach the fixed point. The other trajectories of Fig. 4 reach the $\dot{c}$ axis within finite comoving time. The difference for $M/\alpha=1/3\sqrt{3}$ is the double root in the definition of the comoving time. Finally, when $M/\alpha>1/3\sqrt{3}$, there are no real roots of $\mathcal{R}(T)$. In this case there are *no event horizons*. In fact, these spaces are covered by the Kantowski-Sachs charts. They are represented by region IV of Fig. 4. For $\dot{c}<0$ these solutions start at a large size in their past and evolve into crushing singularities; $\dot{c}>0$ is the time reversal.

Other than $M/\alpha=1/3\sqrt{3}$, all Schwarzschild–de Sitter trajectories reach the $\dot{c}=0$ axis in finite comoving time. But we know that the axis is not part of the phase space. To understand what happens here, we illustrate the situation for the crossing of the de Sitter trajectory. Figure 5 shows the conformal diagram for de Sitter spacetime. Recall that de Sitter spacetime inherits an SO(4,1) isometry from the Minkowski $\mathbb{R}^5$. The figure shows the orbit of a Lorentz boost, SO(1,1)<SO(4,1). In regions

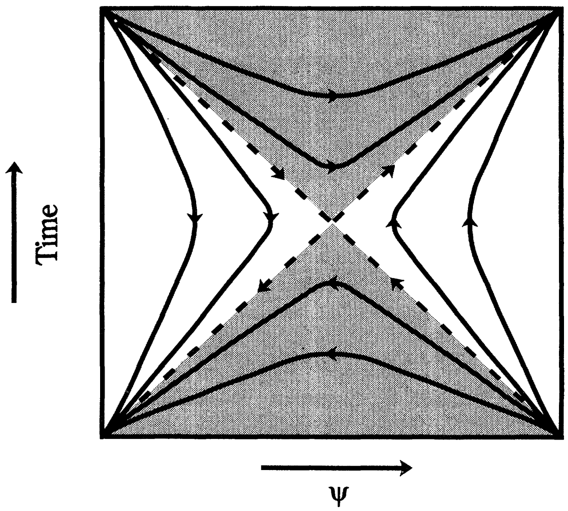

FIG. 5. Conformal diagram for de Sitter spacetime. Curves represent orbits of SO(1,1) Lorentz boost. Lower triangle and upper triangle are covered by Kantowski-Sachs metrics.

where it is timelike, it provides a definition for time, which creates static coordinate charts. In regions where the Lorentz boost is spacelike, it becomes a spatial isometry, and we get the Kantowski-Sachs slicing. The lower Kantowski-Sachs slicing of de Sitter spacetime (Fig. 5) corresponds to the $\dot{c}<0$ part of the de Sitter trajectory in Fig. 4. Likewise, the upper portion of the Kantowski-Sachs slicing of de Sitter spacetime in Fig. 5 corresponds to the $\dot{c}>0$ portion of the de Sitter trajectory in Fig. 4. Thus, other than $M/\alpha=1/3\sqrt{3}$, the crossing of the $\dot{c}=0$ axis in Fig. 4 corresponds to a Cauchy horizon. If the ψ coordinate is periodically identified, the solutions are inextendible beyond these Cauchy horizons.

We now see from Fig. 4 that *all* solutions with $c<1$ correspond to massive solutions. This is why matching regions having necks with radii smaller than Eq. (21) always induced far-field effects. In fact, an interesting example of this occurs for sewing a narrow neck $af_{\min}<\alpha$ into the three-sphere. Not only does the neck introduce far-field effects away from the matching region, but it induces an extrinsic curvature singularity at the antipodal point of the three-sphere. Within the constraint of spherical symmetry, the only way to remove the extrinsic curvature singularity is to introduce another "massive" neck at the antipodal point. This is analogous to trying to put an electric field monopole on a sphere.

In this section we have seen the space of Kantowski-Sachs solutions. From Fig. 4 we can see solutions which expand forever, approaching the de Sitter trajectory, solutions which collapse in crushing singularities, and solutions which end on Cauchy horizons. We have also seen that "small neck" solutions do indeed carry mass, which explains the far-field effects we saw in Sec. III.

VI. CONCLUSIONS

We have now constructed the locally spherically symmetric initial data for the Einstein equations with a positive cosmological constant [Eq. (1)]. By cutting and sewing, we have created locally spherically symmetric initial data for generic spatial topologies. The main restriction in constructing the initial data was ensuring that the matching regions never had a neck with radius smaller than Eq. (21). Smaller necks in a matching region would induce far-field effects. We found that this occurs because smaller necks carry mass. The locally spherically symmetric initial data evolve into a locally spherically symmetric spacetime. By suitable generalizations of Birkhoff's theorem, locally spherically symmetric initial data must evolve to be everywhere locally Schwarzschild–de Sitter type with a fixed mass parameter M or everywhere locally Nariai type. For the initial data on hypersurfaces with generic spatial topology, there could be no far-field effects from one matching region without breaking local spherical symmetry for another matching region. This restriction implied that locally spherically symmetric initial data on hypersurfaces with generic spatial topology evolve to be everywhere locally de Sitter type. We showed that despite being locally de Sitter type, neither these spaces nor their universal covers embed isometrically in de Sitter type. Therefore the construction is nontrivial.

Since there is only one geodesically complete and simply connected spaceform (space of constant curvature), we know the evolution of the initial data for generic spatial topology must have Cauchy horizons (inextendible in general) to their past, future, or both. By choosing expanding initial data, we place these horizons in the past. This is not merely for mathematical convenience. If inflation was brought about through a change of physics due to a cooling, expanding universe, then expanding data are appropriate. Evolving the Einstein equations [Eq. (1)] backwards in time is meaningless, since the preinflationary physics would have been different, and therefore the character of the Einstein equations would have been different.

If de Sitter spacetime is in fact an attractor solution, such as seen in Fig. 4 for Kantowski-Sachs spaces, then we have shown that this attractor solution exists for generic spatial topologies. This is important for inflation to be a viable concept. Since classical general relativity preserves spatial topology through evolution, it is important to know that inflation is not excluded from a Universe on purely topological grounds. Furthermore, when doing any global calculations involving inflation, these the generic topologies should be taken into account. It is not enough to only consider slices of the de Sitter spacetime with identifications.

ACKNOWLEDGMENTS

The authors are deeply grateful to Marc Firestone, Gary Horowitz, Jane Morrow-Jones, and Kristin Schleich for comments and discussions. The work of D.M.W. was supported by CITA, NSERC, and the Department of Physics at the University of British Columbia.

[1] L. Abbott and S.-Y. Pi, *Inflationary Cosmology* (World Scientific, Singapore, 1986).
[2] R. Wald, Phys. Rev. D **28**, 2118 (1983).
[3] J. Morrow-Jones, Ph.D. dissertation, University of California, Santa Barbara, 1988.
[4] L. Jensen and J. Stein-Schabes, Phys. Rev. D **35**, 1146 (1987).
[5] G. Gibbons and S. Hawking, Phys. Rev. D **15**, 2738 (1977).
[6] H. Nariai, Sci. Rep. Tohoku Univ., Ser. 1 **34**, 160 (1950); **35**, 62 (1951).
[7] V. Sahni and L. Kofman, Phys. Lett. A **117**, 275 (1986).
[8] W. Boucher, in *Classical General Relativity,* edited by W. Bonner, J. Islam, and M. MacCallum (Cambridge University Press, Cambridge, England, 1984); H. Friedrich, J. Geom. Phys. **3**, 101 (1986).
[9] J. Morrow-Jones and D. Witt, UCSB Report No. UCSBTH-88-08, 1988 (unpublished).
[10] J. Hemple, *3-Manifolds* (Princeton University Press, Princeton, NJ, 1976).
[11] W. Thurston, Bull. Am. Math. Soc. **6**, 357 (1982).
[12] R. Schoen and S-T. Yau, Commun. Math. Phys. **79**, 231 (1981); E. Witten, *ibid.* **80**, 381 (1981).
[13] Most proofs of the local evolution of initial data of Einstein's equations do not treat the case of $R_{ab} = \Lambda g_{ab}$ because it fails to satisfy the strong energy condition; however, the following reference does treat this case in detail: D. DeTurck, Compos. Math. **48**, 327 (1983).
[14] J. A. Wolf, *Spaces of Constant Curvature* (McGraw-Hill, New York, 1967).
[15] N. Birrell and P. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England, 1982).
[16] S. Hawking and G. Ellis, *The Large Scale Structure of Space-time* (Cambridge University Press, Cambridge, England, 1973).
[17] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).
[18] W. Boucher, G. Gibbons, and G. Horowitz, Phys. Rev. D **30**, 2447 (1984).
[19] W. B. R. Lickorish, Ann. Math. **76**, 531 (1962).

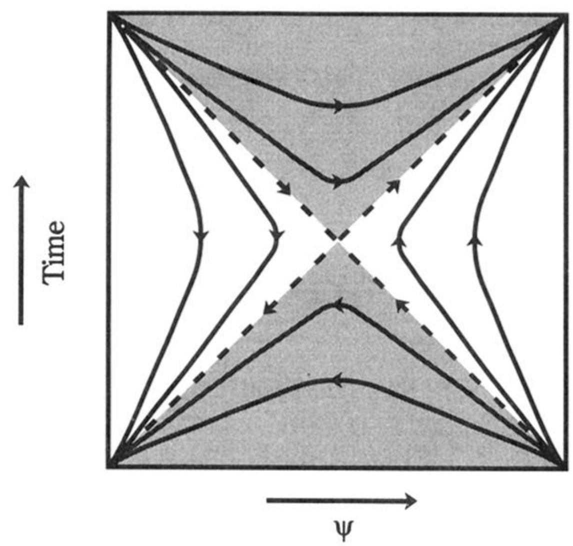

FIG. 5. Conformal diagram for de Sitter spacetime. Curves represent orbits of SO(1,1) Lorentz boost. Lower triangle and upper triangle are covered by Kantowski-Sachs metrics.