Lecture Notes on Classical de Sitter Space

Thomas Hartman

Cornell University

Please email corrections and suggestions to: hartman@cornell.edu

Contents

1 de Sitter Basics

de Sitter spacetime is the maximally symmetric spacetime of constant positive curvature. It is a solution of the vacuum Einstein equations with a positive cosmological constant. It is directly relevant for observation, in two (as fas as we know unrelated!) ways. First, there is evidence that the very early universe had a period of rapid expansion, 'inflation', well approximated by de Sitter spacetime. Second, our tiny present-day cosmological constant currently accounts for about 68% of the energy density of the universe, and this fraction is growing as the universe continues to expand. This means we are entering a second de Sitter phase. The early, inflationary de Sitter phase had a large cosmological constant and correspondingly tiny radius of curvature. The future, dark energy de Sitter has an energy set by today's cosmological constant, and enormous radius of curvature close to today's Hubble scale.

Some references
There is basic introduction to de Sitter in Carroll's book, and a more in depth introduction in Hawking and Ellis. For various coordinate systems, as well as discussion of de Sitter black holes and the first law, see Strominger's Les Houches lectures [hep-th/0110007], as well as Anninos [arXiv: 1205.3855]. For applications to inflation, including the production of cosmological perturbations, see the early chapters of the textbook by Baumann and McAllister (or related lecture notes by the authors available online). For a discussion of entropy bounds in this context, as well as a nice discussion of de Sitter perturbations and instabilities, see Bousso [hep-th/0205177] and references therein.

1.1 de Sitter as an embedding

Usually in GR we just define manifolds and curvatures intrinsically, not by 'embedding' them in a higher dimensional spacetime (which isn't even possible in general). But for de Sitter, the embedding is actually easy and intuitive so it's a great place to start: D-dimensional de Sitter spacetime can be viewed as a timelike hyperbola, embedded in $D + 1$-dimensional Minkowski spacetime $R^{1,D}$. The metric in the embedding space

2

$R^{1,D}$ is

$$ds^2_{embed} = -dX_0^2 + dX_1^2 + \cdots dX_D^2 \ . \tag{1.1}$$

de Sitter, with radius ℓ, is the hypersurface defined by the equation

$$X_\mu X^\mu = \ell^2 \ . \tag{1.2}$$

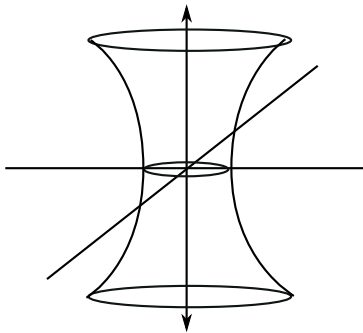This is just an hyperbola. It's easy to draw dS_2, embedded in $R^{1,2}$. See figure 1.

Figure 1: dS_2 viewed as a hyperbola in $R^{1,2}$. The vertical direction is time, X_0. The cross sections are S^1's. In the case of higher-dimensional de Sitter, this picture is still useful but somewhat impressionistic; the slices are S^{D-1}'s.

1.2 Global coordinates

The cross-sections of the dS_2 hyperbola (with the 'straight across' cutting of the figure) are S^1's. More generally, the cross sections of dS_D, sliced at constant-X_0, are clearly S^{D-1}'s. We can put down **global coordinates** on the hyperbola by defining

$$X_0 = \ell \sinh(t/\ell) \ , \quad X_i = \ell \cosh(t/\ell) z_i \ , \tag{1.3}$$

where $i = 1 \ldots D$ and the the z_i are constrained to the unit sphere, $\vec{z}^2 = 1$. This parameterization is designed so that plugging into (1.2) automatically solves the hyperbola equation. We can then put coordinates on the $(D-1)$-sphere $\vec{z}^2 = 1$ in the usual way.[*] To find the metric on the hyperbola we then just plug everything into (1.1). The result is

$$ds^2 = -dt^2 + \ell^2 \cosh^2(t/\ell) d\Omega^2_{D-1} \ . \tag{1.4}$$

[*] For example, in dS$_3$, we set $z_1 = \cos\theta$, $z_2 = \sin\theta\cos\phi$, $z_3 = \sin\theta\sin\phi$.

These are called global coordinates because they cover the entire hyperboloid.[*] Often we will foliate the spatial $(D-1)$-sphere by $(D-2)$-spheres,

$$ds^2 = -dt^2 + \ell^2 \cosh^2(t/\ell)(d\theta^2 + \sin^2\theta d\Omega_{D-2}^2) \qquad (1.5)$$

Relation to FRW

Global coordinates (1.4) are also called the **closed slicing** of de Sitter, especially in the context of cosmology. To explain this terminology, recall that the FRW metric for any homogeneous, isotropic universe is

$$ds_{FRW}^2 = -dt^2 + a(t)^2 d\Sigma^2 \qquad (1.6)$$

where the spatial slices Σ are either open (hyperbolas), flat (R^{D-1}), or closed (spheres). Clearly de Sitter is an example of a closed FRW universe. The scale factor $a(t) = \ell \cosh(t/\ell)$ is exponentially decreasing, reaches a minimum at $t = 0$, $a(0) = \ell$, then expands exponentially into the future.

Careful with the picture

Note that we should not take the picture in figure 1 too literally in the higher-dimensional case. For example, let's call the rightmost point of the hyperbola the 'north pole' of the sphere, and label it by the angle $\theta = 0$. Then the 'equator' of an S^1 consists of two distinct points, at $\theta = \pm\frac{\pi}{2}$. You can see these two points on the picture of dS$_2$, on opposites sides of the S^1. However in higher dimensions the equator is connected. For example, in dS_3, the global-time slices are S^2's so the equator at $\theta = \pi/2$ is the usual equator of a 2-sphere.

More generally, please be cautious with the embedding diagram. There is no $R^{1,D}$ – it's just a trick, a figment of your imagination. Perturbations of de Sitter cannot even be viewed as embedded surfaces, so ultimately we only care about the intrinsic geometry.

---

[*]This is a slight abuse of language: it is impossible to cover both poles of the $(D-1)$-sphere with a single coordinate patch for $D > 2$. So the coordinates are not quite global, but they are global up to the easily-dealt-with isolated points at the poles.

Analytic continuation to the sphere

dS$_D$ is a Minkowski-signature version of the Euclidean sphere, S^D. To see this very explicitly, note that under

$$X_0 \to iX_{D+1}, \qquad t \to i\tau \tag{1.7}$$

the entire discussion so far becomes a discussion of the D-sphere embedded in Euclidean R^{D+1}. Under this analytic continuation, the global coordinates of dS$_D$ become coordinates on S^D where we slice at constant polar angle

$$\Theta = \frac{\tau}{\ell} + \frac{\pi}{2} \ . \tag{1.8}$$

The shift by $\pi/2$ is there so that the metric on the sphere is written in the usual way, with a sine instead of cosine:

$$ds^2_{D-sphere} = \ell^2(d\Theta^2 + \sin^2\Theta d\Omega^2_{D-1}) \ . \tag{1.9}$$

Note that the time-slice at $t = 0$ can be viewed either as a slice of dS$_D$, or as the equator of S^D.

Curvatures

Now that we have the metric we can calculate the curvature of de Sitter. (It also follows easily by continuation from the sphere.) The Ricci tensor and Ricci curvature are

$$R_{\mu\nu} = \frac{D-1}{\ell^2}g_{\mu\nu} \ , \qquad R = \frac{D(D-1)}{\ell^2} \ . \tag{1.10}$$

It follows that dS_D solves the vacuum Einstein equations with positive cosmological constant related to the dS radius by

$$\Lambda = \frac{(D-1)(D-2)}{2\ell^2} \ . \tag{1.11}$$

1.3 Isometries

The embedding view gives an easy way to find isometries of de Sitter spacetime: they are inherited from isometries of $R^{1,D}$ that preserve the hyperboloid. This is just like the derivation of the $SO(3)$ isometries of a 2-sphere by viewing them as rotations in R^3. By analogy to the sphere, the isometries of $R^{1,D}$ that preserve the hyperboloid are rotations:

$$J_{ij} = X_i \partial_j - X_j \partial_i \tag{1.12}$$

and boosts:

$$K_i = X_0 \partial_i + X_i \partial_0 \ . \tag{1.13}$$

Together, these generate (under Lie brackets) the isometry group $SO(D,1)$ of dS$_D$. Convince yourself that there are

$$\frac{D(D-1)}{2} \text{ rotations} \quad + \quad D \text{ boosts} = \frac{D(D+1)}{2} \text{ generators} \ , \tag{1.14}$$

which meets the criterion for a *maximally symmetric* D-dimensional spacetime.

Exercise: Lie algebra
Writing the isometries of dS_2 in embedding space language, and check that under Lie brackets they obey $SO(1,1) \sim SL(2,R)$.

Of course we often need the *intrinsic* description of the isometries, and the vector fields above are written in the embedding space. It is easy to find the intrinsic coordinate representations of these vector fields as follows. First we imagine foliating $R^{1,D}$ by de Sitter hyperbolas of radius r. In other words, replace $\ell \to r$ in the coordinate definitions (1.3). Second, change coordinates from (1.12), (1.13) to these de-Sitter-adapted coordinates. Since we've only written isometries that preserve the hyperboloid, you will find that all the ∂_r components vanish. Finally, set $r = \ell$, and the remaining components give you a vector field on de Sitter in global coordinates.

Of course, the rotations will just give the usual Killing vectors of the $(D-1)$-sphere, so only the boosts are particularly interesting.

Exercise: induced isometries

Follow this procedure to find the boost isometries of dS_3, in global coordinates.

One of the answers, which is the same for all D, is

$$K_1 = X_0\partial_1 + X_1\partial_0 = \ell\cos\theta\partial_t - \sin\theta\tanh(t/\ell)\partial_\theta \ , \tag{1.15}$$

where θ is the polar angle on the spatial $(D-1)$-sphere.

1.4 Penrose diagram

Now we want to draw an intrinsic picture of de Sitter (ie, without embedding into a higher-dimensional spacetime). Global de Sitter can be viewed as a contracting-then-expanding $(D-1)$-sphere. (Assume for now that $D > 2$.) Let's foliate this $(D-1)$-sphere by $(D-2)$-spheres,

$$ds^2 = -dt^2 + \ell^2\cosh^2(t/\ell)(d\theta^2 + \sin^2\theta d\Omega^2_{D-2}) \tag{1.16}$$

The **north pole** $\theta = 0$ and **south pole** $\theta = \pi$ are just *points*, but any other value of $\theta \in (0,\pi)$ includes a whole $(D-2)$-sphere. For now we will focus on just the (t,θ)-directions. Often this is sufficient; for example we can use rotational symmetries of the $(D-2)$-sphere to put geodesics at a fixed point of the $(D-2)$-sphere without loss of generality.

In the coordinates (t,θ), null geodesics do not travel at $45°$. To draw a Penrose diagram (aka conformal diagram) we need to design a coordinate system that (i) covers the whole spacetime, (ii) makes massless particles travel at $45°$, and (iii) fits on the page. This

procedure is not unique, but that's fine; one way to proceed is as follows. The first condition is already accomplished by our choice of global coordinates. To impose the second condition, we change coordinates from $t \to \sigma(t)$, and require:

$$- dt^2 + \ell^2 \cosh^2(t/\ell)d\theta^2 = \Omega(\sigma)^2 \left[-d\sigma^2 + d\theta^2\right] \ . \tag{1.17}$$

The right-hand side is conformally flat; the conformal factor Ω does not affect null geodesics, so this will give us a nice causal diagram with $45°$ null geodesics. From this equation we read off $\Omega = \ell \cosh(t/\ell)$ and $dt = \Omega d\sigma$, so

$$\int \frac{dt}{\ell \cosh(t/\ell)} = \int d\sigma \tag{1.18}$$

Integrating,

$$\tan(\frac{\sigma}{2}) = \tanh(\frac{t}{2\ell}) \ . \tag{1.19}$$

Note that $t \in (-\infty, \infty)$ corresponds to

$$\sigma \in (-\pi/2, \pi/2), \tag{1.20}$$

so we've automatically satisfied the third condition, that our picture fits on a finite-size page. The metric is now

$$ds^2 = \frac{\ell^2}{\cos^2 \sigma} \left[-d\sigma^2 + d\theta^2 + \sin^2 \theta d\Omega_{D-2}^2\right] \ . \tag{1.21}$$

These are the coordinates used to draw the Penrose diagram in figure 2.

Some comments:

- It's a square in the (σ, θ) plane. This is because both σ and θ have coordinate range π.

- The straight-across (fixed σ) spatial slices of the Penrose diagram are simply the $(D-1)$-spheres on constant-X_0 slices of the embedding picture in figure 1.

- There are past and future conformal boundaries I^- and I^+. These are edges, in the conformal sense — everyone approaches them in the infinite past and infinite future. It takes infinite affine-time to reach them.
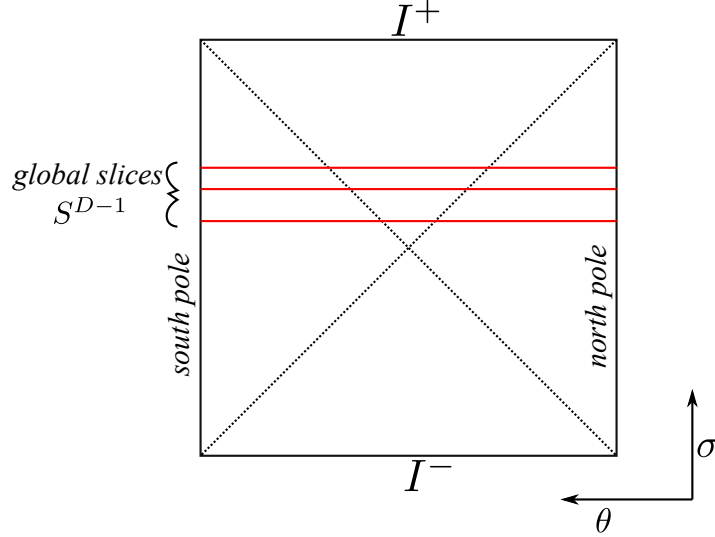
Figure 2: Penrose diagram for dS$_D$ with $D > 2$. Each pixel in the interior is a $(D-2)$-sphere; the radius of this $(D-2)$ sphere goes to zero at the north and south poles of the spatial slice, ie the left and right edges of the diagram. $\theta \in (0, \pi)$ is the polar angle on the spatial $(D-1)$-sphere, and σ is the conformally rescaled global time, running from $(-\frac{\pi}{2}, \frac{\pi}{2})$.

- The apparent 'edges' on the left and right sides of the diagram are not real boundaries, they are just the poles of the $(D-1)$-sphere at each global time-slice. It is important to note that each 'pixel' in the interior of the Penrose diagram represents a $(D-2)$-sphere, while the 'pixels' along these left and right edges are truly *points*, since they are at the poles of the $(D-1)$-sphere.

- Null geodesics restricted to the (t, θ)-plane travel at $45°$ on the Penrose diagram. Null geodesics with nontrivial motion along the S^{D-2} will appear on the Penrose diagram at less than $45°$.

- Throughout these lectures I will put the north pole on the right, and when I talk about static observers I imagine they sit along the right edge of the Penrose diagram. Evidently this is because I grew up on the east coast of the United States, and on on the west coast they put observers on the left.*

- There is an observer-dependent horizon, called the 'cosmological horizon'. This is the null surface beyond which the observer can never receive a signal. We'll have a lot more to say about this below.

---

*And I've never met anyone traveling back in time from a low-entropy future, but I imagine they put I^- at the top of their Penrose diagrams.

1.5 dS$_2$

$D = 2$ is a special case, since 1-sphere's are quite different from higher dimensional spheres. On the 1-sphere, the angle runs over $\theta \in (-\pi, \pi)$, so the Penrose diagram is twice as wide:

$$(1.22)$$
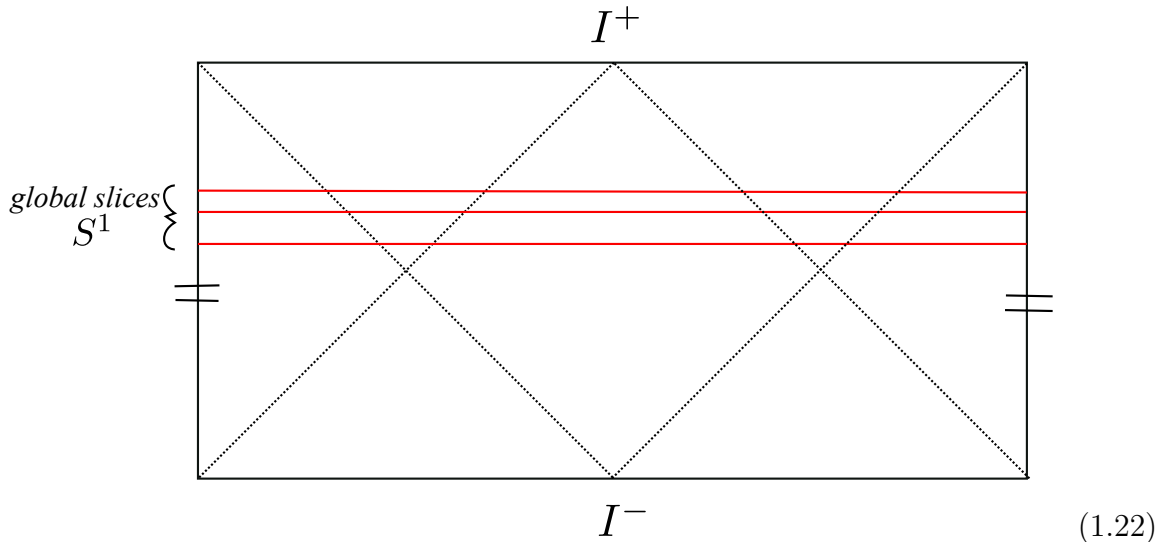
The left and right edges are identified, $\theta \sim \theta + 2\pi$.

2 Other coordinates on de Sitter

Now that we've covered the global properties of de Sitter, we'll talk about two other very useful coordinate systems, covering just parts of the spacetime: the *static patch* and the *flat slicing*. If this feels like a step backward, consider black holes: it is very useful to know the Schwarzschild coordinates, which cover the outside accessible to an asymptotic observer, even though the Kruskal coordinates cover a larger region. The analogue of the "Schwarzschild" coordinates for de Sitter are static patch coordinates. But we'll start with the flat slicing, which are somewhat analogous to the ingoing Eddington-Finkelstein coordinates for a black hole. They play a special role in cosmology.

2.1 Flat slicing

In terms of the embedding space $R^{1,D}$, the flat slicing of dS_D is

$$X_0 = \ell \sinh(t/\ell) + \frac{r^2}{2\ell}e^{t/\ell}, \quad X_1 = \ell \cosh(t/\ell) - \frac{r^2}{2\ell}e^{t/\ell}, \quad X_i = e^{t/\ell}y_i \qquad (2.1)$$

where $i = 2 \dots D$ and $r^2 \equiv \vec{y}^2$. Plugging into the line element (1.1) gives de Sitter in the flat slicing:

$$ds^2 = -dt^2 + e^{2t/\ell}d\vec{y}^2 \qquad (2.2)$$

where $d\vec{y}^2$ is the flat metric on R^{D-1}. Note that this t is different from the global time t.

From (2.1) we see that these coordinates only cover

$$X_0 + X_1 > 0 . \qquad (2.3)$$

If we draw a diagonal plane through the embedding diagram, this is the 'upper triangle.' Similar coordinates can be chosen to cover just the 'lower triangle.'

Flat vs Global coordinate transformation
Equating (2.1) and (1.3) gives the relation between flat and global coordinates. To write this in a useful way, let's first put down spherical coordinates on the flat slices:

$$ds^2 = -dt^2 + e^{2t/\ell}(dr^2 + r^2 d\Omega_{D-2}^2) \qquad (2.4)$$

and we will compare to (1.5). The $d\Omega_{D-2}$'s in these two expressions are the same, so we just need to worry about mapping $(t_g, \theta) \leftrightarrow (t, r)$. ($t_g$ is the global time coordinate, now with a subscript to distinguish it from flat-slicing-time t.) Proceeding via the embedding space we eventually find the simple coordinate change

$$e^{t/\ell} = \cos\theta\cosh(t_g/\ell) + \sinh(t_g/\ell), \qquad \frac{\ell}{r}e^{t/\ell} = \sin\theta\cosh(t_g/\ell) \qquad (2.5)$$

Note that at late times, the time coordinates become the same,

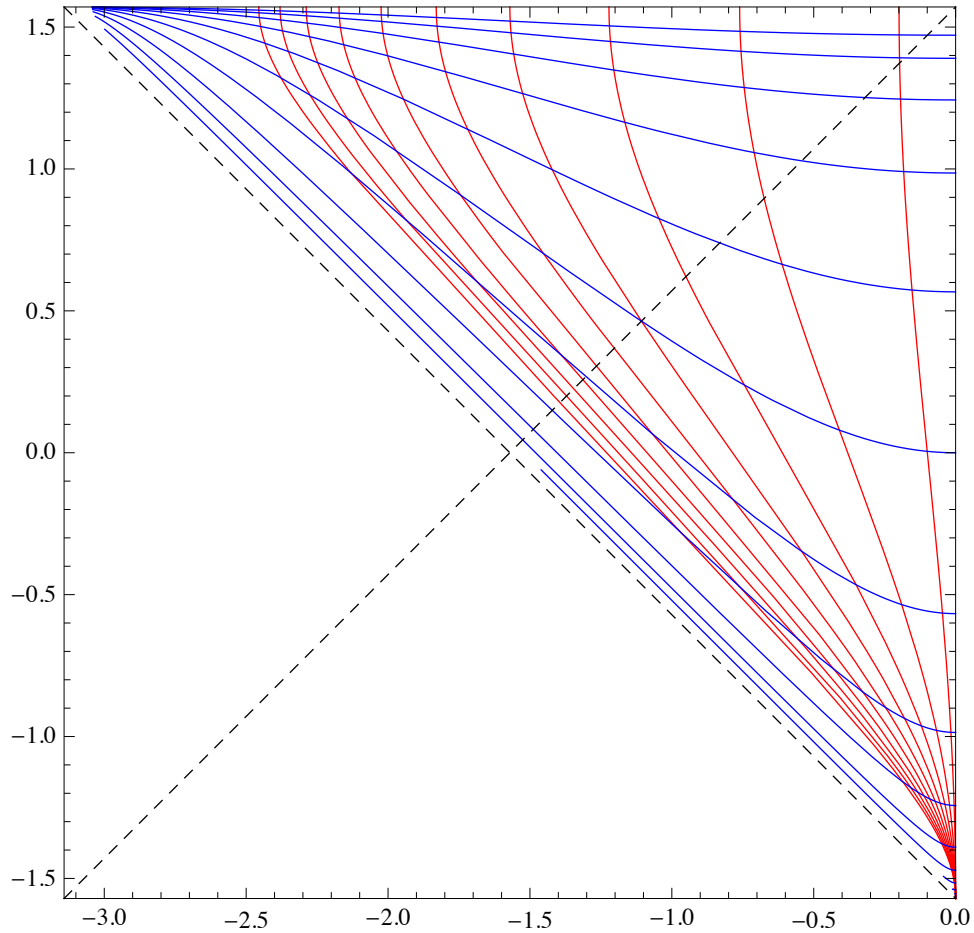$$t \sim t_g \quad \text{as} \quad t \to \infty . \qquad (2.6)$$

11

Figure 3: Flat slicing of de Sitter, drawn on the Penrose diagram. Blue curves are constant-t flat slices, and red curves are the surfaces of constant-r. Intersections of the blue curves with the dashed line are the cross-sections of the cosmological horizon; see discussion below around (2.15).

We can also use (2.5) to draw the flat slices on the Penrose diagram, and see which portion of de Sitter is covered by this coordinate patch. The result is in figure 3.

Exercise: getting to know the flat slicing

Derive (2.5) and reproduce figure 3.

The flat slicing covers only half the Penrose diagram, so this metric by itself is past-geodesically incomplete. Timelike worldlines, unless they are specially chosen to sit at the North pole, will exit the flat slicing in the past, in finite affine time.

On the other hand, once we have initial conditions on some full flat slice, it completely determines the future evolution of de Sitter.

Relation to cosmology

This is called the flat slicing because, comparing to (1.6), we see that it is a flat FRW universe with an exponentially growing scale factor,

$$a(t) = e^{t/\ell} . \tag{2.7}$$

Recall that in FRW cosmology, we assume that matter — galaxies, dust, etc — has a preferred rest frame. On average, matter is assumed to be at rest in the 'comoving' $\vec{y}$-coordinate. Therefore this describes an exponentially growing universe, in which matter is sitting at fixed comoving coordinates but speeding apart in proper distance. The matter worldlines are the red timelike curves in figure 3.

Of course we have not included any matter in our Einstein equation, but this is still a good approximation as long as the energy density from the cosmological constant dominates over the other components. And, the matter is being exponentially diluted, whereas the cosmological constant is, well, constant. So once Λ takes over, everything else quickly ceases to matter and we get exponential expansion. This will go on forever, unless, for some reason, the cosmological constant 'turns off' by some dynamical mechanism. This is believed to be exactly what happened in inflation.

Note that the de Sitter spacetime, by itself, has no preferred coordinate system and no preferred rest frame – a choice of global vs flat coordinates is an arbitrary labeling that contains no physics whatsoever. *However*, when we talk about the FRW universe, there is also matter, and the (average) rest frame of the matter sets a preferred frame. Therefore the closed slicing and flat slicing are inequivalent FRW universes. (The spacetimes are identical, but the initial conditions on the assume matter distribution are different. At late times, when the global sphere becomes large, the two become asymptotically the same.)

Horizon

Note that there is a horizon, and that comoving matter is escaping behind the horizon. Like in a black hole, this stuff clearly can't return to the observer sitting at the north pole (right edge of the Penrose diagram). To describe this horizon in more detail, it helps to consider yet another coordinate system that makes the horizon manifest...

2.2 Static patch

de Sitter has lots of timelike isometries, inherited from the boost isometries of the embedding space. One of them was written explicitly in (1.15). Yet the metrics we've considered so far — global and flat — are time-dependent, in the sense that t appears explicitly in the metric. Since there is a timelike Killing vector field it must be possible to find coordinates where 'time' does not appear in the metric. How do we construct such a coordinate system? Let's say we want to find coordinates such that the boost

$$K_1 = X_0 \partial_{X_1} + X_1 \partial_{X_0} \tag{2.8}$$

is manifest, ie this should be proportional to ∂_t for some time coordinate t. Well we know how to make this boost manifest in the embedding space: just go to Rindler coordinates in the (X_0, X_1) plane,

$$X_0 = \rho \sinh(t/\ell), \quad X_1 = \rho \cosh(t/\ell) . \tag{2.9}$$

This sends $-dX_0^2 + dX_1^2 \to -\frac{\rho^2}{\ell^2} dt^2 + d\rho^2$, and $\rho > 0$ covers just the right Rindler wedge of the embedding diagram in figure 1. The t-translation is the boost, as desired:

$$\partial_t = \frac{\partial X_0}{\partial t} \partial_{X_0} + \frac{\partial X_1}{\partial t} \partial_{X_1} \tag{2.10}$$

$$= \frac{1}{\ell} K_1 \tag{2.11}$$

This only works with other directions set to zero. More generally, we need to pick coordinates to solve the hyperbola equation $X^2 = \ell^2$. Inspired by the Rindler construction, we try

$$X_0 = \sqrt{\ell^2 - r^2} \sinh(t/\ell), \quad X_1 = \sqrt{\ell^2 - r^2} \cosh(t/\ell), \quad X_i = r z_i \tag{2.12}$$
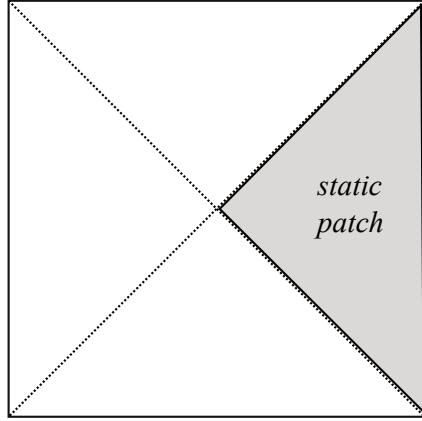
14

Figure 4: Static patch, on the Penrose diagram. This is the causal patch of an observer sitting at the north pole, ie $\theta = 0$ in global coordinates, ie $r = 0$ in static coordinates. The right edge of the diagram is $r_{static} = 0$; the bifurcate Killing horizon is $r_{static} = \ell$. The other three patches can also be covered by (independent) static coordinate systems, much like the four regions of the Penrose diagram for Schwarzschild black holes.

where $i = 2 \ldots D$ and $\vec{z}^2 = 1$ parameterize a $(D-2)$-sphere. I've replaced $\rho \to \sqrt{\ell^2 - r^2}$ to write things in a more standard way. The result is the **static patch metric**:

$$ds^2 = -(1 - r^2/\ell^2)dt^2 + \frac{dr^2}{1 - r^2/\ell^2} + r^2 d\Omega_{D-2}^2 \; , \tag{2.13}$$

with $0 < r < \ell$. The metric is manifestly static, since it's independent of t (and indeed $\partial_t = \frac{1}{\ell}K_1$). The horizon is at $r = \ell$. This horizon was drawn as a dashed line in the Penrose diagram, figure 2. The static coordinates cover only the right diamond of the Penrose diagram, see figure 4. The cosmological horizon for an observer sitting at the North pole is a bifurcate Killing horizon for ∂_t, and the bifurcation sphere is S^{D-2} at the middle of the Penrose diagram. This sphere is the equator of the $t_{global} = 0$ slice in global coordinates.

The diamond covered by static coordinates is called the causal patch, or static patch. This the region of de Sitter accessible to a single observer, in the sense that the observer and both send and receive signals to/from this entire region.