

Quantum theory of three-dimensional de Sitter space

Sidan A¹, Tom Banks,¹ and Willy Fischler²

¹*Department of Physics and NHETC, Rutgers University,
Piscataway, New Jersey 08854, USA*

²*Department of Physics and Texas Cosmology Center, University of Texas,
Austin, Texas 78712, USA*

 (Received 12 December 2023; accepted 2 January 2024; published 24 January 2024)

We sketch the construction of a quantum model of three-dimensional de Sitter space, based on the covariant entropy principle and the observation that semiclassical physics suggests the possibility of a consistent theory of a finite number of unstable massive particles with purely gravitational interactions. Our model is holographic, finite, unitary, causal, plausibly exhibits fast scrambling, and qualitatively reproduces features of semiclassical de Sitter physics. In an appendix we outline some calculations that might lead to further tests of the model.

DOI: [10.1103/PhysRevD.109.025011](https://doi.org/10.1103/PhysRevD.109.025011)

I. INTRODUCTION

In previous work on de Sitter (dS) space by two of the present authors [1] the underlying hypothesis has always been that dS space is an infrared regulator of a scattering theory of particles in Minkowski space. This point of view is problematic in three space-time dimensions because the work of [2] shows us that the conventional S -matrix does not exist. Locally flat space-time cannot support a total center-of-mass energy larger than the Planck mass, and there are other peculiar constraints on would-be Mandelstam invariants. Every different asymptotic boundary condition for particles changes the asymptotics of space-time.

In this paper we will study dS_3 , based on the covariant entropy principle (CEP) [3–5]. The maximal area causal diamond of dS_3 is finite, so according to the CEP its Hilbert space must be finite dimensional as well. The Schwarzschild–Sitter geometry suggests another important lesson about the theory; localized excitations in the bulk are constrained states of holographic degrees of freedom on the horizon, causing a reduction in the entropy [6–9]. Generic states of the holographic degrees of freedom, with respect to the density matrix of the diamond, carry energy (as measured along the geodesic in the diamond) that scales like $1/R$ as the diamond radius goes to infinity. Bulk localized energy must satisfy $ER \gtrsim 1$ and ER is proportional to the number of constraints on the boundary

qubits.¹ The fact that localized bulk excitations have low entropy suggests that they are all unstable.

In the first part of this paper we examine the instability of localized excitations from the bulk semiclassical point of view. Given a fixed static patch, instability is visible already in the classical limit, for any excitation whose center of mass follows a geodesic other than the one that connects the timelike tips of the diamond.² These states all have lifetimes roughly equal to the proper time it takes the geodesic to exit the diamond. Recall that this is independent of the mass of the particle. In the proper time along any trajectory connecting the timelike tips of the diamond, the excitation fades into the generic background equilibrium of the horizon. Only a particle on the geodesic connecting the tips of a given static patch is classically stable.

Quantum mechanics modifies this classical behavior in three ways:

¹In the $RM_p \rightarrow \infty$ limit, the boundary states might seem to become “zero-momentum massless particles”. This is probably an inaccurate description. In a generic causal diamond, massless particles with momentum of order the diamond radius would generally have wave functions spread throughout the bulk. The states that contribute the bulk of the entropy have wave functions localized infinitely close to the boundary and give rise to a UV divergence in the entropy.

²The theory of classical relativistic particles coupled to 3D gravity in de Sitter (dS) space was exactly solved by Deser and Jackiw [10], following their work with ’t Hooft on the analogous problem in locally asymptotically flat space-time [2]. A comprehensive study of the solutions in flat and AdS spaces can be found in the paper by Matschull [11]. The latter work makes it abundantly clear that the gravitational field has no independent degrees of freedom and merely modifies the structure of the particle phase space.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP³.

- (i) The uncertainty principle limits the localization of the object and its velocity fluctuations. The probability that it follows the static geodesic for an infinite time is zero.
- (ii) In $d \geq 4$ Gibbons-Hawking radiation will subject the object to momentum kicks, which would cause it to random walk toward the horizon. Once it gets sufficiently far, the further time to exit the horizon is aR , with a mass independent coefficient.
- (iii) If the localized object is a black hole, it will decay by Hawking radiation and the massless radiation will escape to the horizon. $d = 3$ pure gravity has no particle excitations, so it is not clear that every model must have Gibbons-Hawking radiation. Furthermore, the static solutions of classical Einstein gravity in $d = 3$ with non-negative cosmological constant do not have horizons apart from the cosmological one.

The first two sources of instability are minimal if the mass is large, and if the mass is large enough in $d \geq 4$ the object is surely a black hole, or will become one eventually. Since black holes can have masses of order R^{d-3} there are many long lived systems, gravitationally bound to the origin, which will have lifetimes $\gg R$. Furthermore, we will argue below that for a large gravitationally bound object with many semiclassical subsystems, the mathematical description of the system from the point of view of a geodesic in dS space is quite irrelevant. Different trajectories of the center of mass of the system decohere and the subsystems can use these to define a reference frame in which to do quantum mechanics. By its definition, this reference frame is stable until the entire object collapses into a black hole and begins to emit Hawking radiation.

In $d = 3$, the situation is quite different. The maximal mass of a localized object in dS₃ (or three-dimensional Minkowski space for that matter) is of order the Planck mass. Roughly speaking then, the wave function of a particle starting out at rest at global time $t = 0$ in dS₃, is spread over a distance of order at least 1 in Planck units. Then the nature of the instability in three dimensions is dominated by the first of the mechanisms above.

In Sec. II we explain and give a summary of previous results. In Sec. III we examine the evolution of particle wave functions in dS₃. By examining the transition amplitudes for single particle motion, we argue that a localized wave function becomes uniformly spread over the static patch *coordinates* in a time of order the dS Hubble scale, R . This means that the probability distribution is concentrated near the horizon in that time. We interpret this as an instability of the quantum state corresponding to a particle on the static geodesic to decay to a “typical” thermal state of the vacuum ensemble.³ Particles localized

³Typicality is meant in the sense of statistical mechanics and the eigenvalue thermalization hypothesis. The time for the state to become Haar random is much longer.

on geodesics other than the static trajectory connecting the two tips of the static patch are, of course, classically unstable and have lifetimes of order R as well. We exhibit the geometrical mechanism by which the “deficit angle” of such a moving particle disappears from measurements made by a detector localized on the static geodesic.

In Sec. IV, we discuss a set of finite-dimensional quantum mechanical models which have the qualitative properties of the semiclassical dynamics outlined above. The variables in these models are fermion fields $\psi(\sigma, \theta)$ on the holographic circle θ times a near-horizon interval σ . There are cutoffs on both the angular momentum in the θ direction and the linear momentum on the interval σ . The cutoffs are correlated in a way we will describe below. States with a deficit angle are constrained by

$$\psi(\theta_n, \sigma)|_S = 0, \quad (1)$$

if n is a multiple of an integer p . θ_n are a discrete set of angles such that the fields $\psi(\theta_n, \sigma)$ are independent canonical fermion fields. There are many choices of Hamiltonian that make all such states decay in a time of order R , and have other desirable gravitational attributes, like fast scrambling [12–14]. However, a natural mechanism for fast scrambling, the “fuzziness” of higher-dimensional geometry, does not work for the one-dimensional boundary of dS₃, and we have to insert explicit long range interactions into the boundary Hamiltonian to achieve fast scrambling of quantum information. The ideas of Carlip [15] and Solodukhin [16] provide a guide to finding the right description of these interactions.

II. SUMMARY OF PREVIOUS RESULTS

The basic principles on which this paper is based are the assertion [3,17] that the Bekenstein-Hawking area law for black holes holds for an arbitrary causal diamond in any model of quantum gravity

$$\langle K_\diamond \rangle = \frac{A_\diamond}{4G_N}, \quad (2)$$

and that states localized on a geodesic in de Sitter space have an entropy deficit

$$\Delta S = 2\pi RE. \quad (3)$$

Here R is the dS radius and E is the energy in static coordinates. In four and higher dimensions, this formula is only valid to the leading order when ER is small. The strongest argument for the first assertion is Jacobson’s demonstration [17] that the hydrodynamic equations of this entropy law are the double null projections of Einstein’s equations. It can also be demonstrated by a Euclidean path integral argument [18] or by the generalization to arbitrary diamonds [19] of the results of Carlip [15] and Solodukhin [16]. The latter authors argued that the effective modular

Hamiltonian⁴ on the horizon of a black hole was the L_0 generator of a (cutoff) $(1 + 1)$ -dimensional conformal field theory, with central charge proportional to the horizon area. The conformal field theory (CFT) lives on an interval, the stretched horizon, and its target space is related to the geometry of the holographic screen of the diamond.⁵

The second assertion follows from the first and from the formula for the Schwarzschild dS black hole metric. It gives a derivation of the Gibbons-Hawking temperature of dS space that depends only on the area law, rather than the details of quantum field theory. It tells us that static energy is a measure of the number of constrained qubits in a localized state. The localized state density matrix is the projection of e^{-L_0} for empty dS space onto a lower-dimensional subspace. This explains why dS space has a fixed temperature.

In a number of papers, two of the authors (T. B. and W. F.) have suggested that the proper variables for describing the fluctuating geometry of the holographic screen of any causal diamond were the expansion coefficients of a spinor field in eigenspinors of the Dirac equation on the background geometry. The idea is that the background geometry is a hydrodynamic variable. Connes [20] showed that Riemannian geometry was entirely encoded in the Dirac operator on the Riemannian manifold. The coefficients of the expansion of a solution of the fluctuating geometry's Dirac equation in eigenspinors of the background are a finite set of variables if we impose a UV cutoff on the Dirac eigenvalue. If they are quantized as (cutoff) $(1 + 1)$ -dimensional fermion fields, following the prescription of Carlip and Solodukhin, then we can correlate the UV cutoff on the transverse Dirac eigenvalue with the central charge of the CFT, and thus with the area of the holographic screen of the diamond. By contrast, the UV cutoff on the $(1 + 1)$ -dimensional Dirac momentum is associated with the very smallest diamonds. Our interpretation of this is that the Carlip-Solodukhin analysis is semiclassical/hydrodynamic and breaks down for very small systems. A theoretical lower bound on its use is that the single fermion CFT must have enough states for Cardy's formula to provide a good approximation to the spectrum. Probably of order 10–20 momentum eigenstates suffices for that. In the real world, one would have to determine the correct value of the $(1 + 1)$ -dimensional cutoff by experiment. In the imaginary world of $(2 + 1)$ -dimensional dS space we would have to realize this space-time as a subsystem of some rigorously defined asymptotically flat or anti-de Sitter (AdS) system, perhaps along the lines of [21,22], and hope that the larger system contains a sufficiently precise copy of the dS_3 subsystem to determine these delicate UV details. As appears to be the case for linear

dilaton gravity in $1 + 1$ dimensions, there might be many such embeddings, and no unique answer to the question.

The authors of [12–14] have argued that the quantum systems on stretched horizons are fast scramblers of quantum information. T. B. and W. F. have argued that the existing evidence for fast scrambling, for horizons of dimension 2 or higher, is accounted for if the systems are invariant under something like the group of volume preserving diffeomorphisms of the horizon, or some fuzzy deformation of it. dS_3 has a horizon of dimension 1 and length preserving diffeomorphisms preserve order. Thus any finite version of this group will be a Z_N translation group and the resulting system will not have fast scrambling unless it has long-range interactions. In the text, we have instead written random couplings between different points related by the Z_N symmetry. We imagine that different instances of these random couplings might be realized from different realizations of the model in AdS/CFT.

III. QUANTUM MECHANICS OF SINGLE PARTICLES IN dS_3

In quantum field theory, spin zero particles in dS_3 are described by the Klein-Gordon (KG) equation. We can view this as single-particle Hamiltonian evolution for particles with positive energy in some particular static patch. Generic local interactions between particles force us to give up the single particle picture, because a transition amplitude in one frame can be viewed as a particle production amplitude in another. This is not the case for gravitational interactions in $d = 3$. As shown by Matschull [11], the effect of gravitation in three dimensions is merely to modify the phase space of a fixed number of particles. While this argument is valid also for negative cosmological constant, no one has found a quantum theory dual to pure AdS gravity coupled to a finite number of particles. Note also that the static solutions of GR in three dimensions with asymptotically AdS boundary conditions have finite area horizons. The overwhelming evidence from the AdS/CFT correspondence is that the entropy/area law is valid and that those solutions correspond to complicated quantum systems with large Hilbert spaces. This is not true for dS_3 or Minkowski space.⁶ The static solutions have an entropy deficit, just like higher dimensions, but no entropy of their own. It is certainly impossible to put an infinite number of particles into dS_3 if each particle has any finite energy, and, at the classical level, no pair of particles remains in causal contact for more than a finite amount of proper time.

Our proposal, at the semiclassical level, is to study the quantum mechanics of individual particles. This model

⁴The modular Hamiltonian is minus the logarithm of the density matrix.

⁵The maximal area surface in a null foliation of the boundary of the diamond.

⁶The entire concept of asymptotically flat space is problematic in three dimensions. Classical “scattering” solutions exist for point particles [2], but traditional Mandelstam invariants are bounded and there is no obvious Hilbert space on which a traditional S -matrix could be defined.

does not explain the entropy of empty dS_3 or the entropy deficits of localized classical particles, but it does demonstrate the quantum mechanical decay back to the classical equilibrium state represented by empty dS_3 . We then propose a quantum model with a finite-dimensional Hilbert space that does explain the entropic properties of solutions. The localized solutions are realized as constrained states of the underlying variables, and the decay rates of those states back to equilibrium agree qualitatively with the semiclassical results.

A. Classical particles in dS_3

Although the physics of dS space can all be understood in a single static patch, it is convenient for calculations to work in the flat slicing metric,

$$ds^2 = -dt^2 + e^{2t/R} d\vec{y}^2, \quad (4)$$

through coordinate transformations from the static patch coordinates (t_s, r_s) to the flat slicing coordinates (t, r) ,

$$\begin{aligned} r_s &= r e^{t/R}, \\ e^{-2t_s/R} &= e^{-2t/R} - r^2/R^2, \end{aligned} \quad (5)$$

where $r^2 \equiv \vec{y}^2$. The geodesic equations can be solved by writing the conservation law resulting from spatial translation in these coordinates,

$$\left(1 - e^{2t/R} \frac{d\vec{y}^2}{dt^2}\right)^{-1/2} \frac{d\vec{y}}{dt} = \vec{v}_0. \quad (6)$$

The solution is

$$\vec{y}(t) = \vec{y}_0 - \frac{\vec{v}_0}{v_0^2} \left(\sqrt{1 - v_0^2} - \sqrt{1 - v_0^2 e^{-2t/R}} \right). \quad (7)$$

The classical geometry produced by a particle of mass m traveling on this geodesic is obtained from that of dS_3 by modding out by an $SO(1, 3)$ transformation that leaves the geodesic invariant. For geodesics with $\vec{v}_0 = 0$, this is just a rotation around the point \vec{y}_0 , producing a deficit angle $\delta = 8\pi G_N m$ [10]. Let us consider a particle at \vec{y}_0 , viewed from the static patch centered at $\vec{y} = 0$. Take a causal diamond of finite proper time centered on the $\vec{y} = 0$ geodesic with future tip t_2 and past tip t_1 . The particle enters the causal diamond at flat slicing time

$$t_i = -R \ln \left(e^{-\frac{t_1}{R}} - \frac{y_0}{R} \right), \quad (8)$$

and leaves it at

$$t_f = -R \ln \left(e^{-\frac{t_2}{R}} + \frac{y_0}{R} \right). \quad (9)$$

Consider past and future oriented nests of diamonds, all centered at the origin (Fig. 1). The entire effect of the particle is encoded in the lengths of circles at fixed flat time t . If a circle winds around the geodesic at \vec{y}_0 its circumference is smaller than it would have been in the absence of the particle. Note that the flat slicing time is equal to the proper time along both geodesics. If we look at future oriented diamonds with future tip T_+ such that

$$e^{T_+/R} \int_{-\infty}^{T_+} dt e^{-t/R} < t_i, \quad (10)$$

or past oriented diamonds with past tip T_- such that

$$e^{T_-/R} \int_{T_-}^{\infty} dt e^{-t/R} > t_f, \quad (11)$$

then there are no circles in those diamonds that circumnavigate the geodesic at \vec{y}_0 . In both cases, a detector in the diamond can receive signals from detectors that circumnavigate the particle trajectory, but cannot send signals back. The interpretation of this for the early, future-pointing nested diamonds is that the particle is part of the initial data on the past boundary of a larger diamond in the nesting. Since the particle is moving slower than light, the early detectors can receive signals of its imminent arrival,⁷ but cannot perform their own experiments to verify the shortened circles around the particle trajectory.

The question we want to answer is how this classical physics fits into the proposal of [5] that all of the physics of dS space is captured by a finite-dimensional system describing a single static patch. A classical relativist is apt to say that the static coordinate system is not good because it does not cover the manifold. A hypothetical quantum system defined on the full flat slices would certainly argue that on flat slices there are circles at all times, which exhibit a length deficit that is not redshifted. The association of an entropy deficit with a length deficit makes this a challenge for static patch holography. In the classical picture, there is a subsystem dual to the causal diamond in Fig. 1, in which the entropy deficits exist for the entire period that the particle ‘‘remains in the static patch’’.⁸ Unlike the length along spacelike circles, an entropy deficit

⁷This phrase uses the idea that there are things outside the static patch. From the framework of static patch holography, ‘‘imminent arrival’’ just means that the dynamics in the patch respects causality, in the sense that some of the initial conditions that are necessary to describe the full history of the diamond are imposed on degrees of freedom that are not in the algebra defining a smaller diamond in the nest.

⁸Here period is measured in terms of proper time along the particle’s geodesic.

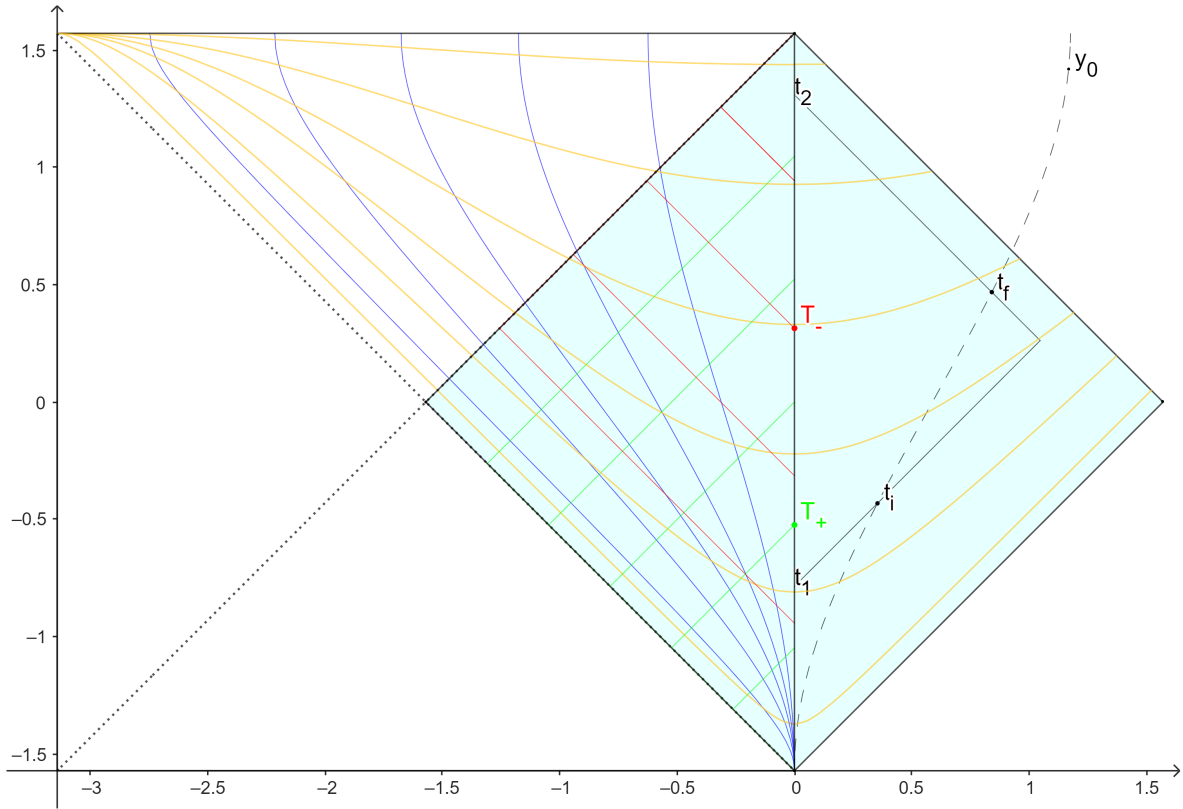


FIG. 1. The Penrose diagram of de Sitter space with flat slicing coordinates contours. Yellow curves are constant- t surfaces and blue curves are constant- y surfaces in flat coordinates. The static patch is highlighted in light blue. Green (red) lines are boundaries of future (past) oriented nests of diamonds centered at the origin. The dashed curve is the trajectory of a massive particle at \vec{y}_0 that enters and leaves the causal diamond at time t_i and t_f , respectively.

cannot be redshifted away by going to another reference frame. The quantum principle of relativity says that the density matrix corresponding to the diamond of Fig. 1 must have the same spectrum computed along the geodesic in the static patch as it does when computed along the geodesic along which the particle travels. In order for static patch holography to work, the low entropy particle state must decay back to the static patch equilibrium state.

B. Quantum particles in dS_3

Now we consider the evolution of the support of scalar particle wave functions in dS_3 . This provides a bulk semiclassical picture for how the low-entropy particle state decays back into the equilibrium state.

In flat polar coordinates the horizon is located at $r = Re^{-t/R}$ where R is the dS radius, and r the coordinate distance from the origin. The general massive scalar solutions are

$$\begin{aligned} \phi_{k\ell}(t, r, \theta) = & e^{i\ell\theta} e^{-t/R} J_\ell(kr) (A_{\ell k} J_n(kRe^{-t/R}) \\ & + B_{\ell k} J_{-n}(kRe^{-t/R})), \end{aligned} \quad (12)$$

where $n = \sqrt{1 - R^2 m^2} \approx imR$.

In the neighborhood of static time $t_s = 0$ and static radius $r_s = 0$, the flat t, r coordinates are also close to the origin. Therefore a lump, localized within a flat radius of about $1/m$ of the origin at $t = 0$, has most of its support in flat radial wave numbers $k < m$. The radial part of the solution above, $J_\ell(kr)$, has its first zero at $r \sim 1/k > 1/m$. So the horizon reaches the first Bessel zero for every relevant k at flat times $t < t_{\max} = R \log(mR)$. At later flat times, the solution inside the horizon is approximately homogeneous but exponentially damped like $e^{-t/R}$. (The time dependent Bessel functions $J_{\pm imR}(kRe^{-t/R})$ just oscillate with increasing t .)

Static time slices t_s intersect flat slices t if and only if $t_s > t$, and they do so at static coordinate radius $r_X(t, t_s)$ given by

$$r_X = R \sqrt{1 - e^{-2(t_s - t)/R}}. \quad (13)$$

For all smaller static radii $r_s < r_X$, the static time slice samples the solution at *later* flat times, where the solution is more exponentially suppressed. So for static times $t_s > t_{\max} + (\text{few}) \times R$, the solution in static coordinates is localized near r_X , and r_X is exponentially close to the horizon.

Note that there are two free mode coefficients $A_{\ell k}, B_{\ell k}$ in the solution above. One linear combination of them is determined by the initial condition at time $t_s = t = 0$, and we have analyzed the generic behavior for solutions localized within $1/m$ of the origin at that time. The other linear combination is determined by the requirement that the solution contains only positive frequencies with respect to static time. Since the considerations of the previous paragraph were independent of the choice of A and B for a given \vec{y} space initial condition, the positive frequency solutions of the KG equation in static coordinates, corresponding to the quantum behavior of a particle of mass m localized on the static geodesic, will predict overwhelming probability for the particle to be smeared over the horizon in a static time of order $R \ln mR$. Recognizing that R is the natural timescale for interactions on the horizon, this is consistent with the time for a fast scrambling system to equilibrate an entropy deficit of order mR .

Now let us consider a particle localized on the geodesic at \vec{y}_0 with $\vec{v}_0 = 0$. The momentum space solutions are exactly the same as the ones above. We just have to Fourier transform them with respect to $\vec{y} - \vec{y}_0$. Let us consider initial conditions with vanishing angular momentum around \vec{y}_0 . The resulting wave function remains circularly symmetric around \vec{y}_0 for all time. We can repeat the calculation above for a light particle initially localized near the static geodesic at \vec{y}_0 . The position space wave function is

$$\begin{aligned} \phi(\vec{y}) = e^{-t/R} \int d^2k e^{i\vec{k} \cdot (\vec{y} - \vec{y}_0)} [A(k)J_n(kR e^{-t/R}) \\ + B(k)J_{-n}(kR e^{-t/R})]. \end{aligned} \quad (14)$$

If $|y_0|$ is large then the value of the initial wave function in the static patch centered at the origin is exponentially small. As we bring \vec{y}_0 closer to the origin, it gets larger and is concentrated in the angular direction of \vec{y}_0 . However, the dominant effect we discussed for a particle sitting near the origin remains the same. The static patch at the origin shrinks exponentially in y coordinate size, but the wave function converges to

$$\phi(\vec{y}) = e^{-t/R} \int d^2k e^{i\vec{k} \cdot (\vec{y} - \vec{y}_0)} [A(k)J_n(0) + B(k)J_{-n}(0)]. \quad (15)$$

y_0 is now outside the static patch centered on the origin. Inside that patch, the solution in flat coordinates is again approximately homogeneous and exponentially damped in time, and so the dominant support in static coordinates is concentrated near the horizon, and the angular dependence disappears. The component of this solution with angular momentum ℓ around the origin is proportional to

$J_\ell(m|y_0|)J_\ell(m|y|)$. Since the horizon moves to exponentially smaller values of y all components with nonzero ℓ die off exponentially. Again we see return to the equilibrium state. For larger y_0 this happens more rapidly because the amplitude of the initial condition is small to begin with.

There is a subtle point in the analysis above. It is well known that the space of solutions of the Klein-Gordon equation is not a Hilbert space, but we have treated the KG solutions as quantum wave functions with the standard probability interpretation. This problem is actually resolved in dS space, if we remain within a single static patch. The positive frequency solutions of the equation *do* form a Hilbert space in the static patch. Positive static patch frequency wave functions are complicated linear combinations of the flat coordinate solutions that we have discussed. It is therefore important that our discussion did not depend on particular choices of the functions $A(k)$ and $B(k)$ but only their general support in momentum space.

Finally, we have to address the question of why this same discussion does not apply in higher-dimensional de Sitter space. The answer has two parts. The first is that the Nariai bound on the mass of localized objects in dS scales like R^{d-3} in d dimensions, so the wave functions of localized objects can be much more localized and it takes longer for them to dissipate. In addition, higher-dimensional gravity produces gravitational bound states so that even light objects like electrons can have wave functions that remain localized in a static patch for much longer than the dS time R , because they are bound to larger masses. In our own Universe, if the initial conditions are such that a galaxy can form, then the static patch in the rest frame of that galaxy will have localized objects in it until quantum fluctuations *of the galaxy center of mass position* predict a very high probability for the galaxy wave function to be spread uniformly over the horizon. For detectors bound to the galaxy, the lifetime is even longer. It is not until the galaxy collapses to form a black hole that these detectors “cease to exist.” The parts of the galaxy positional wave function that follow different timelike trajectories decohere from each other because the galaxy has so many quantum states for each position, and makes transitions between them on a time scale much shorter than the time it takes the center of mass to move. So the prediction of QM for a detector inside the galaxy is that it will find itself carried along on some timelike trajectory, after which observations of things bound to the galaxy are more or less unaffected by the fact that the galaxy is in dS space. Particles radiated by the galaxy will disappear into the Gibbons-Hawking haze in a time of order R , and it is only when that dissipation leads the galaxy to gravitationally collapse that the detector is affected by the dS background. Only cosmological observations of objects in the causal past of the galaxy, which are fading into the haze, can detect the global geometry of space-time. In three space time dimensions there are no large masses and no complex bound systems that can

exhibit this kind of interesting long term behavior. The existence of horizons for the static solutions of GR with non-negative cosmological constant, in dimension greater than three is an indication that this kind of complexity cannot be avoided in any model of quantum gravity. Although the internal dynamics of black holes does not provide us with many semiclassical observables, there are things like local rates of Hawking decay that are correlated with the center of mass position of the black hole and can be argued to have a decoherent, if evanescent, semiclassical existence.

The three-dimensional picture might be modified if we added electromagnetic interactions to the model since preliminary estimates suggest we could make a range of complex atoms and molecules with masses less than the Planck mass and sizes smaller than the dS radius. However, such a model raises a number of issues. Among these is the inevitability of particle production and how to make that compatible with a bound on the total energy in the system. In addition, the authors of [23] have suggested that in the presence of massless particles the static solutions of dS₃ gravity will develop genuine horizons and large entropy, so that our discussion would have to be modified drastically. We leave the exploration of models with other light fields to future work.

IV. TOWARDS A QUANTUM MODEL OF dS₃

The basic strategy of holographic space-time (HST) models of quantum gravity is to describe the variables on the holographic screen of a diamond as cutoff quantum fields on the screen, with a cutoff that preserves the symmetries of the diamond. The requirement that the diamond has finite entropy means that the basic fields are fermions.⁹ The spin-statistics connection and the observation of Connes [20] that Riemannian geometry is encoded in the Dirac operator, suggests that we view them as belonging to the spinor bundle over the screen. The work of Carlip [15] and Solodukhin [16] suggests that they be viewed as variables in a (1 + 1)-dimensional CFT, whose L_0 generator is the modular Hamiltonian of causal diamonds in dS₃ [24,25].

Let us elaborate a bit on this last point. Consider a causal diamond defined by a geodesic with proper time interval $[-T, T]$. There is a coordinate system (τ, x) on this diamond, which we will call diamond universe (DU) coordinates, where the metric takes the form

$$ds^2 = C^2(\tau, x)(-d\tau^2 + dx^2 + \sinh^2 x d\phi^2),$$

$$C = \frac{R \sinh(R_*/R)}{\cosh(\tau) + \cosh(x) \cosh(R_*/R)}, \quad (16)$$

where $\tau \in (-\infty, \infty)$ is a time coordinate and $x \in (0, \infty)$ is a radial coordinate. The relation between T and R_* is

$$T = \frac{\pi R}{2} \tanh(R_*/R). \quad (17)$$

There is a similar coordinate system for each of the subdiamonds corresponding to the intervals $[-T, \tau]$. In each of these systems, the lines of constant x are flow lines of a vector field $V_0(\tau)$, which leaves the diamond invariant and is a conformal Killing vector (CKV) of dS₃. The generalization of the conjecture of [15,16] made in [19] is that in models of quantum gravity, the quantum operator implementing the action of the CKV on the bifurcation surface of each of these diamonds is the L_0 generator of a (1 + 1)-dimensional CFT with central charge proportional to the length of the bifurcation surface. Connes' results, which motivated the HST formalism, suggest that this CFT be built from fermion fields $\psi_n(\sigma^0, \sigma)$ labeled by eigensections $\chi_n(y)$ of the one-dimensional Dirac operator on the bifurcation surface,

$$\gamma e^{-1}(y) i \partial_y \chi_n = \lambda_n \chi_n, \quad (18)$$

where $e(y)$ is the einbein on the boundary and γ is one of the Pauli matrices. y is the boundary coordinate and σ is a spatial coordinate on an interval where the model is defined.

The restriction to finite entropy implies that we must impose cutoffs on both the transverse Dirac operator and the spectrum of L_0 . We will require that the (1 + 1)-dimensional cutoff be independent of the size of the diamond, so that the entropy scaling is determined by the transverse cutoff. As we will see, the cutoffs are correlated, but depend on interactions that we must add to the free (1 + 1)-fermion Lagrangian in order to satisfy some important properties of a sensible model of quantum gravity. These properties have to do with the time evolution operator. Time evolution in the DU coordinates of necessity involves a time dependent evolution operator $U(t, -T)$. We can impose causality by insisting that at each discrete Planck time step τ , $U(\tau, -T)$ factorizes into an operator acting only in the Hilbert space of the fermionic variables in the diamond $[-T, \tau]$ and one which commutes with all those variables.

If we now consider two diamonds whose future tips are both many Planck times from $-T$, but differ by one Planck step along the geodesic we can define,

$$e^{-iL_P H(\tau)} = U(\tau + L_P, -T) U^{-1}(\tau, -T). \quad (19)$$

The geometric relation between the tangent vector we have chosen to represent by L_0 and the one that describes time

⁹Any finite-dimensional Hilbert space is a representation of a one-(discrete)parameter family of superalgebras since the Gell-Mann matrices are closed under both commutation and anti-commutation. Moreover, even if the dimension of the Hilbert space is not a power of 2, one can realize it as a constrained subspace of the representation space of a finite number of canonical fermions.

evolution along the geodesic between the diamond tips motivates the conjecture¹⁰ that $H(\tau) \propto L_0(\tau + L_P)$. Indeed $H(\tau)$ should be a Hermitian operator in the larger Hilbert space. Furthermore, since it should entangle the extant degrees of freedom with those that are being added, we conclude that the CFT in question cannot be free fermions. Note that any continuous deformation of the free fermion CFT along a conformal manifold will leave the value of the central charge unchanged, so the Carlip-Solodukhin calculation will be preserved by such a deformation.

We can get more insight into what kinds of interactions are needed by trying to understand the dissipation of localized information on the holographic screen. According to [12–14] this should occur by a nonlocal process of “fast scrambling”. For the one-dimensional metric $e^2(x)$ we can always introduce a new coordinate ϕ which measures length in Planck units,

$$\phi(y) = L_P^{-1} \int_0^y e(y') dy'. \quad (20)$$

We will normalize the length of *every* transverse circle to $2\pi L_P$ since the physical length is counted by the number of transverse fermion fields ψ_n . So ϕ goes from 0 to 2π . The variable ϕ is just the angle variable in DU coordinates on the diamond. A useful intuition from the work of Carlip and Solodukhin is to think of the spatial variable σ in the $(1+1)$ -dimensional CFT as a dimensionless interval in the DU coordinate x , near the boundary $x = \infty$. This is not strictly necessary, as we really only use the spectral properties of the Virasoro generator L_0 of the CFT to define our model.

In terms of ϕ , the Dirac eigenfunctions are just proportional to $e^{in\phi}$, and a cut-off keeps all modes between some wave numbers $\pm M$. Our prescription says that we have one $(1+1)$ D fermion field on a spatial interval for each $n \in [-M, M]$. Using the discrete Fourier transform on Z_{2M+1} we obtain linear combinations of these that represent fields labelled by $2M+1$ discrete angles $\phi(m)$. These latticized fields are canonical free fermion fields. The free Virasoro generators can be written as sums over either set of variables.

Beyond the total entropy, which we will return to below, another piece of physics we can try to realize in the model is the entropy deficit associated with localized particles. Let us try to find states of our system that could correspond to particles of mass μ sitting at the origin. They should

have entropy deficit $2\pi\mu R$. Candidate states are those satisfying

$$\psi_{\theta(m)(\sigma)}|s, \mu\rangle = 0, \quad (21)$$

for $m = 0 \bmod p$, with $p = MS_0/(\pi\mu R)$ and S_0 is the entropy of a single-fermion field theory which we discussed above. These states will have the proper entropy deficit. In the free fermion theory the constraints could be removed (but would not be if we had chosen to impose them on the Fourier transforms of the fields with respect to σ) by time evolution, but information would not be homogenized over the transverse dimension, as indicated by classical field theory computations in the bulk. This is a second indication that we need interactions.

A general class of interactions is suggested by the dual constraints of conformal invariance and transverse fast scrambling. From each of the free fields $\psi_{\theta(m)}(\sigma)$ we can construct a $U(1)$ current $J_{\theta(m)}^\alpha(\sigma)$ and the interaction

$$\delta\mathcal{L} = G_{mn} J_{\theta(m)}^\alpha(\sigma) J_{\theta(n)}^\beta(\sigma) \eta_{\alpha\beta}, \quad (22)$$

defines a well-defined conformal manifold of perturbations of the free fermion model for any symmetric positive definite matrix G_{mn} . The eigenvalues of L_0 of course depend on the choice of this matrix. For our problem we need a sequence of such choices, one for each sufficiently large sub-diamond of the diamond defined by $[-T, T]$. At each step in time we keep all the old matrix elements of G_{mn} and choose the new ones from a random distribution consistent with the constraints of symmetry and positivity. It is clear that once τ is large, the perturbation at each time step will be small. As far as the modular Hamiltonian is concerned, the main thing that will be affected is the probability for the previously nonexistent variables, and their correlations with those that existed in the prior diamond. It is also clear that the dynamics has no hint of locality in the transverse circle and so is likely to exhibit fast scrambling.

By causality, we only need to talk about is time evolution within a given causal diamond, but it is convenient to have a unitary operator on the entire Hilbert space at each instant of time. That operator should factor into a product of one that operates only on the degrees of freedom present in the diamond $[-T, \tau]$ and a unitary on its tensor complement in the full Hilbert space of the interval $[-T, T]$. We can implement this by choosing to make the coefficients G_{mn} time dependent. That is, the Hamiltonian is the sum of free fermion Hamiltonians for all of the fermion fields, with current-current interactions added gradually as time goes on and new degrees of freedom enter the expanding diamond. Note that this means that once τ is large in Planck units, the incremental change in the Hamiltonian in each Planck time step is small.

¹⁰This conjecture echoes an obscure result in algebraic QFT that T. B. learned from Nima Lashkari. Causal diamonds in QFT do not have density matrices, but they do have positive modular operators, which play a similar role. If we have two nested diamonds with infinitesimally close tips then the *modular flow* (the unitary transformation obtained by raising the modular operator to a continuous imaginary power) indeed generates Heisenberg evolution on the operator algebra in the wedge between the two diamonds.

The last remark implies that the natural time scale for any decay process, in geodesic proper time, is of order R , if we adjust the coefficient relating the dimensionless L_0 to the Hamiltonian appropriately. If, as we have claimed, the system is a fast scrambler, then the time scale for erasing an entropy deficit ΔS will be $CR \ln(\Delta S)$. The coefficient C will depend on the choice of the coefficients G_{mn} .

A. The cutoff procedure

The Cardy formula for the entropy of a CFT is given by

$$S = 2\pi \sqrt{\frac{c}{6} \left(L_0 - \frac{c}{24} \right)}. \quad (23)$$

Cardy and Solodukhin choose L_0 by finding an appropriate solution of the near horizon Liouville equation. Of course, in a typical CFT there will not be any degeneracies and L_0 certainly has a discrete spectrum. This formula represents the density of states in a particular band of eigenstates. We can calculate the fluctuations in L_0 for large c by integrating

$$Z \equiv \int dL_0 e^{A\sqrt{L_0} - \beta L_0}, \quad (24)$$

by saddle point for large A and calculating

$$\partial_\beta^2 (\ln Z)_{\beta=1} = A\sqrt{L_0^*}. \quad (25)$$

A is fixed by fitting to the Carlip-Solodukhin (CS) formula for diamond entropy.

This indicates a prescription for how we are to choose the two-dimensional cutoff in the fermion model. We first choose the transverse cutoff M to give a value of the central charge that scales like the diamond radius for all the causal diamonds $[-T, \tau]$. Then we fix the overall normalization by singling out a band of eigenstates of the CFTs that gives the CS result and the fluctuation formula $(\Delta K)^2 = K$ on the nose. (For more detailed discussion of the modular fluctuations, see Refs. [19,26].) We emphasize that the CS result is really the first term in an asymptotic expansion in L_p/L , where L is the size of the smallest causal diamond for which we trust the CS argument. In the language of the analogy with hydrodynamics, we are trying to bootstrap a quantum system from knowledge of its hydrodynamic equations. In condensed matter physics, the higher order corrections to hydrodynamic equations are replete with system dependent ambiguities, which also depend precisely on how one defines the hydrodynamic variables in terms of the microscopic degrees of freedom. The highly symmetrical structure of general relativity and quantum field theory has enabled us to make quite detailed guesses about the underlying quantum theory but one might imagine that one would need confrontation with an actual experimental system to get all of the microscopic details correct.

We want to emphasize that the role of the two-dimensional cutoff is to delineate the boundary where the hydrodynamic ansatz of Carlip and Solodukhin becomes an acceptable approximation to the dynamics of a large causal diamond in dS_3 . It is analogous to the choice of a block size L in the derivation of hydrodynamics from the Schrodinger equation for lattice quantum systems [27]. As emphasized in [27] there is a lot of ambiguity involved in that choice. One needs either a precise microscopic model or a comparison with experiment to tie down the details more accurately. At the moment, we have neither.

The details of the program we have outlined will of course depend on our choice of $G_{mn}(\tau)$. It is not clear whether this is a fundamental ambiguity or not. We are far from exhausting all of the constraints that a sensible quantum theory of three-dimensional de Sitter space must satisfy. On the other hand, it might be that one can obtain many different consistent quantum models of dS_3 , depending on the form of ‘‘compact dimensions’’. The latter situation seems to hold for two-dimensional models of quantum gravity with linear-dilaton asymptotics [28,29]. These arise as decoupling limits of linear-dilaton black holes in four-dimensional supersymmetric compactifications of Type II string theory. There seem to be many consistent nonperturbative models of linear dilaton black holes [30,31], which are plausibly related to the wide variety of consistent four-dimensional string models.

B. Constraints in two different dS static frames

Let us work in conformal coordinates. One geodesic at $\vec{y} = \vec{y}_0$, one at the origin. The proper times are synchronized. The cosmological horizons are just at the Euclidean distances R from each of these points. The proper time corresponding to conformal time η is

$$\tau = \int_{\eta^0}^{\eta} \frac{ds}{s - \eta^0}, \quad (26)$$

where η^0 is the place where we impose our initial condition. At conformal time s a detector on the geodesic at some point has a causal diamond that reaches out to $s\hat{n}$. Note that the range of s and η^0 could go as far back as $-\infty$. The timelike geodesics that are static in conformal or flat coordinates all begin within a single causal diamond as shown in Fig. 1. Thus, constraints on one geodesic have to appear as constraints on another, since constraints are by definition acausal. They are imposed globally on the Hilbert space of the holographic screen, and make sense as an initial condition but cannot be imposed at an intermediate time without violating causality.

The static geodesic at \vec{y}_0 leaves the causal diamond centered at the origin in a particular direction. In conformal coordinates this happens because the diamond for $\eta > \eta'$ shrinks in the \vec{y} coordinate as η' is made larger, so that the geodesic at \vec{y}_0 is no longer causally connected to a late time

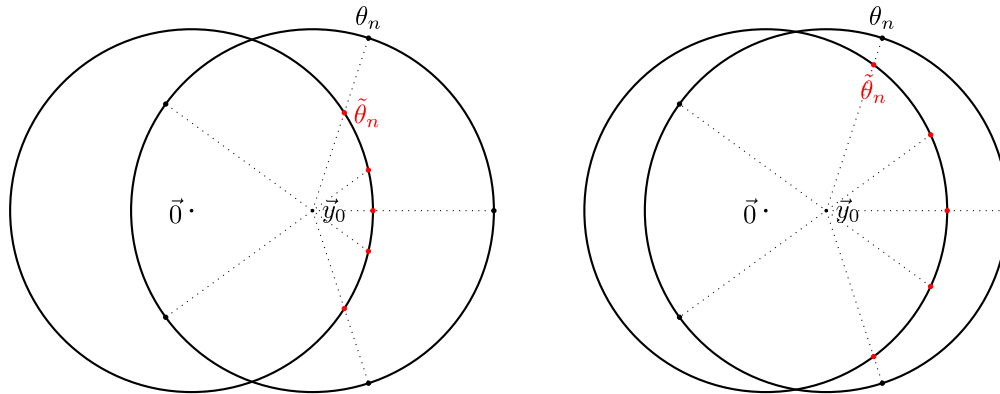


FIG. 2. Mapping constraints on the boundary of causal diamond D_1 centered at \vec{y}_0 onto the boundary of causal diamond D_0 centered at $\vec{0}$. Each state on D_1 boundary outside D_0 is mapped to the intersection of D_0 boundary and a line connecting this state with \vec{y}_0 (shown as a dotted line in the diagram). For states on D_1 boundary and inside D_0 , extend the dotted line so it intersects with the D_0 boundary as shown in the diagram, then the intersection gives the mapped state. The closer \vec{y}_0 is to the origin $\vec{0}$, the more spread out the mapped states are, and the locations of these states tell us where \vec{y}_0 is.

detector on the geodesic at the origin. The associated moving particle solution rapidly approaches the solution for an empty static patch, healing up a conical defect centered at \vec{y}_0 . We will model this behavior in the quantum theory by imposing initial conditions on the field $\psi(\theta, \sigma)$. We will impose $\psi(\tilde{\theta}_n, \sigma)|_S = 0$ on the initial state of the system. The angles $\tilde{\theta}_n$ are determined as follows. At very early times we consider the Z_N symmetric distribution of points θ_n on the circle of radius $|\eta^0|$ surrounding \vec{y}_0 . We take η^0 large and negative, but finite. Some of the θ_n are inside the circle of radius $|\eta^0|$ surrounding the origin and some are outside. We project each θ_n to the corresponding $\tilde{\theta}_n$ by drawing the line connecting θ_n to \vec{y}_0 . $\tilde{\theta}_n$ is the point where that line intersects the circle surrounding the origin as shown in Fig. 2.

Note that these constraints are imposed on the full set of microscopic degrees of freedom of the theory. The dynamical evolution of the system does not couple these degrees of freedom together immediately. At each proper Planck time step, a finite set of angular momentum modes interact via fast scrambling interactions of the maximally transversely localized $U(1)$ currents that we can form from them. At early times these currents do not have much overlap with the highly localized fields at individual θ_n , when the de Sitter radius is large in Planck units. The rest of the individual angular momentum modes evolve as free fermion fields, all obeying the same linear equation of motion. Thus, their contribution to the constraint equation is undisturbed by time evolution. As the proper time approaches the dS Hubble time and beyond, all the modes participate in the fast scrambling interaction, and the constraints are removed. It is also clear that the further \vec{y}_0 is from the origin, the constrained state is more of a distortion of the homogeneous dS equilibrium state. Since the fast scrambling interactions immediately transfer

information between points that are separated by arbitrary distances on the lattice, these will homogenize rapidly.

Much work remains to be done to find a more quantitative match between lifetimes and entropy deficits for particles in dS_3 as defined by the boundary quantum system, when compared to the semiclassical calculations based on the KG equation. The general qualitative behavior of our explicit quantum system clearly mirrors the semiclassical physics we have explored. In the Appendix we begin a more quantitative exploration by computing the overlap diamonds for pairs of intervals along different geodesics.

V. CONCLUSIONS

We have argued that the semiclassical physics of particles propagating in dS_3 , is compatible with a quantum theory based on the idea of static patch holography [5]. The appropriate quantum framework is a Hilbert bundle over the space of timelike trajectories in dS_3 with a finite-dimensional Hilbert space over each fiber. This is a more refined version of the HST formalism introduced by two of the authors. The dS group $SO(1, 3)$ acts on this bundle, with a typical transformation mapping one fiber into another. Within each fiber only the $SO(2) \times R$ subgroup leaving a particular geodesic invariant, acts. Dynamics within each fiber is provided by time dependent Hamiltonians adapted to either past or future oriented nests of causal diamonds. The future oriented nest provides us with a mathematical description of how the whole dS_3 universe evolves, given initial conditions on the past boundary of the static patch. The past oriented nest is useful for understanding what a detector traveling along the geodesic can probe after a given time.

The symmetry action guarantees that the dynamics along each geodesic is identical. Similarly, time reversal relates the Hamiltonians for the backward and forward oriented nests of diamonds. In both cases, the complication lies in

mapping a particular initial state along one sequence of nested proper time intervals on a particular geodesic to its corresponding state in another nesting or another geodesic. We have made a start on understanding this for the constrained states corresponding to single particles traveling on geodesics, whose relative velocity in flat coordinates is zero.

Perhaps the most disturbing consequence of our picture is the complete failure of quantum field theory in curved space-time to give any sort of approximation to our picture. One can view this as a drastic version of the Cohen-Kaplan-Nelson bound [32–35] on the validity of quantum field theory (QFT). Most states of QFT in dS_3 are simply not realizable once the classical backreaction of matter on geometry is taken into account. This is dramatically different from the situation in AdS_3 . Although the classical analysis of particles coupled to 3D gravity leads to similar conclusions for either sign of the cosmological constant, the Bekenstein-Hawking principle tells us that description is missing a huge number of quantum states compatible with asymptotically AdS_3 geometry. BTZ black holes have large entropy. In dS_3 the CEP identifies missing states on the cosmological horizon but not in the bulk interior. Thus, there seems to be a complete model of a small finite number of particles, with restricted phase space, interacting only with gravity. Note that this is a self consistent picture. In $2 + 1$ dimensions, gravitational interactions do not lead to processes with particle-antiparticle production.

Asymptotically flat three-dimensional space, to the extent such a concept exists in real models of quantum gravity, would appear to be in the same class as dS_3 . The argument that CFT correlators converge to S matrix elements fails in three dimensions, because the S -matrix does not exist, and any possible notion of scattering amplitudes for finite numbers of particles contains bizarre constraints on Mandelstam invariants. This is consistent with the empirical fact that every model of AdS_3 /CFT with an $R_{AdS}/l_S \rightarrow \infty$ limit has two or more compact dimensions whose radius (inverse mass of the lightest Kaluza-Klein state) goes to infinity linearly with R_{AdS} . Although it is conceivable that a model in asymptotically locally flat space could be defined by taking the $N \rightarrow \infty$ limit of our model of dS_3 , we make no such claim at this time.

The failure of quantum field theory in curved space time to adequately approximate our model deserves further study. A possible starting point is a model of charged particles coupled to electromagnetism and gravity. Such a model has local vertices, which lead to particle-antiparticle production. If the particles are sufficiently light, then the geometric back reaction on a single such process is minimal. It is clear however that multiparticle states with any finite energy per particle are forbidden by gravitational constraints whenever the particle number is too high. Thus, it would appear that neither the constructions of the present paper, nor conventional field theory descriptions can model this situation. It might lie in the class of low-energy field

theory models that do not arise as limits of models of quantum gravity.

Another feature of our analysis that we found surprising is that so far we do not have a time independent Hamiltonian description of our static patch model in static time. We have instead exploited a plausible connection between modular flow and Planck step geodesic evolution in the DU coordinates of [25]. In QFT, the Heisenberg evolution of operators in a small wedge between closely spaced nested diamonds can be related to modular flow, but is also generated by the action of the global Hamiltonian operator, restricted to the subalgebra of the diamond. One would have hoped that something similar happened in dS space, but so far we have not found the global description.

ACKNOWLEDGMENTS

We thank Patrick Draper for collaboration in the early stages of this work and for some ideas contributed to Sec. III B. The work of T. B. and S. A is supported in part by the DOE under Grant DE-SC0010008. The work of W. F. is supported by the NSF under Grant PHY-1914679.

APPENDIX: CAUSAL DIAMOND OVERLAPS

The entropy of the overlap of two causal diamonds is computed in this section. A causal diamond is a region accessible to a detector on an interval of a timelike trajectory. The intersection between two diamonds in $2 + 1$ dimensions has a boundary that ends, in general, in two lines rather than two points. One looks at all timelike geodesics connecting a point on the past line with a point on the future line and chooses the one whose diamond has the largest area. Figure 3 gives an example of an intersection of two diamonds.¹¹ In $2 + 1$ dimensions, the area reduces to the circumference, thus to determine the entropy we want to find the overlap radius. Start with the flat-slicing metric in de Sitter space Eq. (4), where R is de Sitter radius. A general solution for timelike geodesics is given by Eq. (7), and constant $\vec{y} = \vec{y}_0$ is the trial solution when $\vec{v}_0 = 0$. For null geodesics, vanishing space-time intervals give the following general solution:

$$\vec{y}_{\text{null}}(t) = \vec{u} + \vec{v}R e^{-t/R}, \tag{A1}$$

where \vec{u} and unit vector \vec{v} are parameters to be determined. Note that in a causal diamond, horizontal time slices correspond to constant conformal times η , which is related

¹¹We emphasize that time coordinates t_1, t_2, t'_1, t'_2 labeled in all figures in this appendix are flat-time coordinates, and constant flat time slices are curved in the conformal diagram Fig. 1. In the conformal diagram, a causal diamond along a timelike geodesic \vec{y}_0 has a rectangular side view, and it is stretched to a square diamond in Figs. 3–13 where the horizontal and vertical directions are in flat-space and conformal time coordinates, respectively.

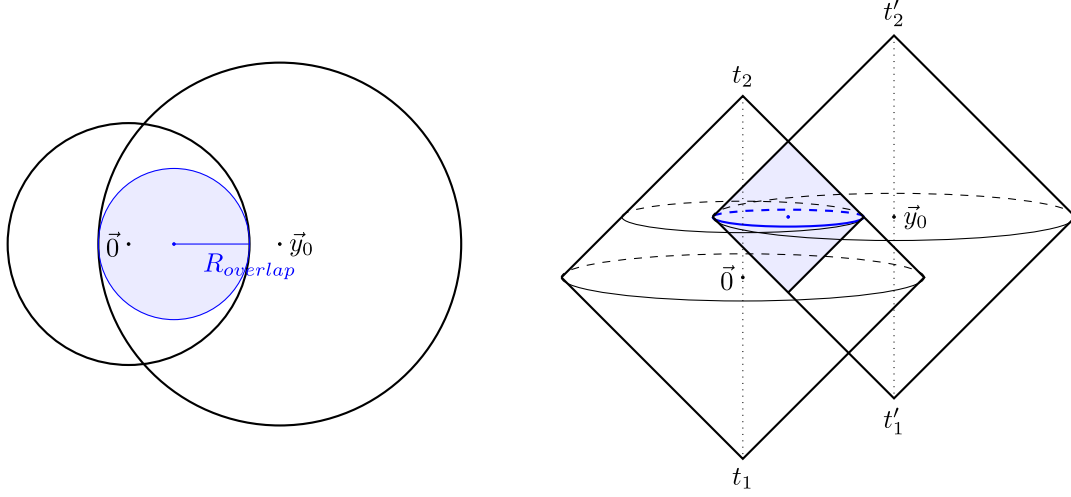


FIG. 3. An example top view (left) of intersecting spatial slices of two diamonds (right). The maximal overlap has radius R_{overlap} . This example corresponds to the case in Fig. 13, the particle at $\vec{0}$ is enclosed in the maximal overlap, leading to a deficit angle δ .

to the flat time t by $d\eta/dt = e^{-t/R}$. Constant-time flat slices are yellow curves in Fig. 1.

Now consider a causal diamond along timelike geodesic $\vec{y} = 0$. Let (t_1, t_2) be a time interval along $\vec{y} = 0$, and the past and future null boundaries of the causal diamond cross $\vec{y} = 0$ at t_1 and t_2 , respectively. Along with the intersection of the past and future boundaries of the causal diamond, we obtain three equations to determine parameters \vec{u} and \vec{v} ,

$$\begin{aligned} \vec{u}_1 + \vec{v}_1 R e^{-t_1/R} &= 0, \\ \vec{u}_2 + \vec{v}_2 R e^{-t_2/R} &= 0, \\ \vec{u}_1 + \vec{v}_1 R e^{-t_0/R} &= \vec{u}_2 + \vec{v}_2 R e^{-t_0/R}, \end{aligned} \quad (\text{A2})$$

where t_0 is the flat time of the bifurcation surface. In terms of conformal time $\eta = -R e^{-t/R}$, we have $\eta_0 = (\eta_1 + \eta_2)/2$. Solving Eq. (A2) leaves us one parameter, which is the direction of a free unit vector \vec{v}_1 . Express $\vec{v}_2, \vec{u}_1, \vec{u}_2$ in terms of unit vector \vec{v}_1 , we find the boundaries of the causal diamond to be

$$\begin{aligned} \vec{y}_1(t) &= \vec{v}_1 R (e^{-t/R} - e^{-t_1/R}), \\ \vec{y}_2(t) &= -\vec{v}_1 R (e^{-t/R} - e^{-t_2/R}). \end{aligned} \quad (\text{A3})$$

Let $e^{-t_0/R} = (e^{-t_1/R} + e^{-t_2/R})/2$ to solve for the intersection \vec{y}_{int} of the past and future null boundaries, and we find the radius of the causal diamond to be

$$y_{\text{int}} = \frac{R}{2} (e^{-t_1/R} - e^{-t_2/R}). \quad (\text{A4})$$

Then the area of a causal diamond is given by the circumference of the bifurcation surface, which is the two-dimensional surface centered at $\vec{y} = 0$ with boundary \vec{y}_{int} . The area of the causal diamond is

$$A = \Omega y_{\text{int}} = \frac{1}{2} \Omega R (e^{-t_1/R} - e^{-t_2/R}), \quad (\text{A5})$$

where $\Omega = 2\pi - \delta$ with $\delta = 8\pi G_N m$ [10] being the deficit angle due to the presence of a massive particle with mass m .

Now let us consider two maximal causal diamonds with dS radius R whose time runs from $-\infty$ to ∞ , centered along two timelike trajectories at $\vec{y} = 0$ and $\vec{y} = \vec{y}_0$. Two time intervals (t_1, t_2) and (t'_1, t'_2) along the two trajectories give two smaller causal diamonds. When the two diamonds overlap, there exists a maximal causal diamond within the overlapping region and we are interested in determining the entropy of this maximal overlap, the result depends on $y_0, t_1, t_2, t'_1, t'_2$ and we will discuss by cases.

1. When $y_0 > 2R$

The two maximal causal diamonds do not overlap, then no matter how large (t_1, t_2) and (t'_1, t'_2) are, there is no overlap. This case corresponds to when the causal diamond centered at \vec{y}_0 is completely unobservable viewed by an

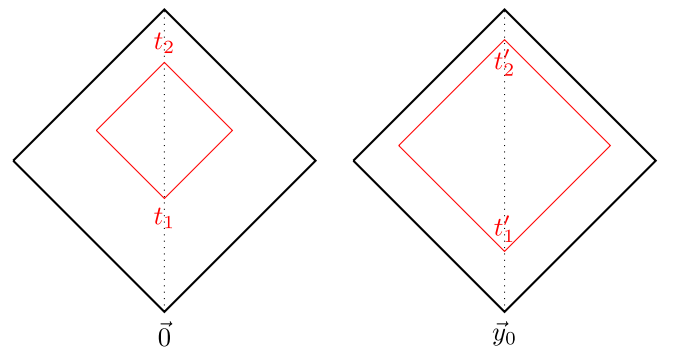


FIG. 4. When $y_0 > 2R$, there is no overlap between causal diamonds (t_1, t_2) and (t'_1, t'_2) .

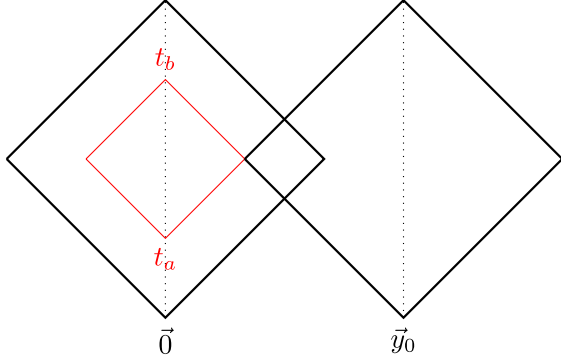


FIG. 5. When $R < y_0 < 2R$ and causal diamonds centered at \vec{y}_0 and (t_a, t_b) meet.

observer along $\vec{y} = 0$. See Fig. 4 for a front view of the case.

2. When $R < y_0 < 2R$

This case corresponds to when the particle along \vec{y}_0 trajectory is outside the observable Universe, but part of its causal diamond overlaps with the causal diamond centered at $\vec{y} = 0$. (t_1, t_2) and (t'_1, t'_2) need to satisfy conditions to allow overlap (which we will show below), then the entropy is given by $2\pi R_{\text{overlap}}$ since neither timelike trajectory passes through the overlapping causal diamond, thus the deficit angle is 0. The intersection of minor null surfaces of small causal diamond with (t_1, t_2) is given by Eq. (A4), which is the radius of the red diamond in Fig. 5. As shown in the Figure, any time interval with larger t_1 or smaller t_2 than such (t_1, t_2) in Fig. 5 gives no overlap, i.e. overlap exists when

$$t_1 < t_a = -R \ln \frac{y_0}{R} \quad \text{and} \quad t_2 > t_b = -R \ln \left(2 - \frac{y_0}{R} \right). \quad (\text{A6})$$

Consider when t_1, t_2 satisfy the above conditions, Fig. 6 gives the critical case when there is no overlap, which depends on

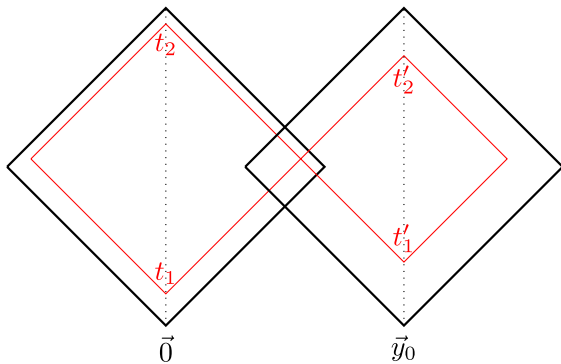


FIG. 6. When $R < y_0 < 2R$ and causal diamonds (t_1, t_2) and (t'_1, t'_2) meet.

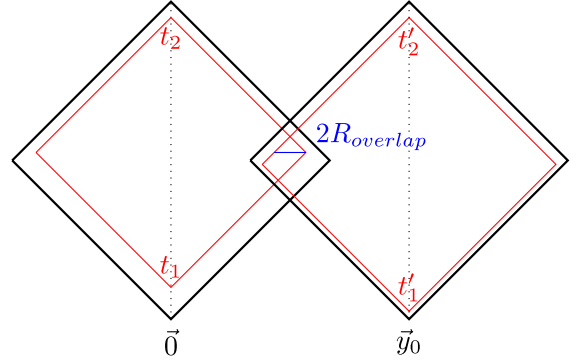


FIG. 7. When $R < y_0 < 2R$ and causal diamonds (t_1, t_2) and (t'_1, t'_2) overlap.

(t'_1, t'_2) . Similarly let y'_{int} be the radius of minor causal diamond with time (t'_1, t'_2) , then

$$y'_{\text{int}} = \frac{R}{2} (e^{-t'_1/R} - e^{-t'_2/R}). \quad (\text{A7})$$

Solve $y_{\text{int}} + y'_{\text{int}} = y_0$ for conditions on t'_1, t'_2 to allow overlap, where conformal time $\eta'_1 + \eta'_2 = \eta_1 + \eta_2$. We find that overlap exists only if

$$t'_1 < -R \ln \left(e^{-t_2/R} + \frac{y_0}{R} \right),$$

and

$$t'_2 > -R \ln \left(e^{-t_1/R} - \frac{y_0}{R} \right). \quad (\text{A8})$$

Now consider when all conditions are met, i.e., when Eqs. (A6) and (A8) are satisfied. Then overlap exists, as shown in Fig. 7, and R_{overlap} is a function of $y_0, t_1, t_2, t'_1, t'_2$ which is given by

$$R_{\text{overlap}} = \frac{1}{2} [y_{\text{int}} + y'_{\text{int}} ((\eta_1 + \eta_2)/2) - y_0]. \quad (\text{A9})$$

Here $y'_{\text{int}} ((\eta_1 + \eta_2)/2)$ is the radius of the (t'_1, t'_2) causal diamond evaluated at flat time t whose conformal time is $(\eta_1 + \eta_2)/2$,

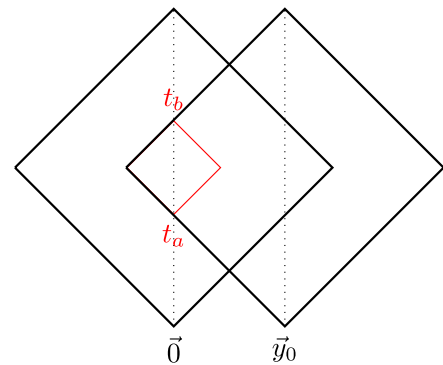


FIG. 8. When $y_0 < R$, there is no overlap if $t_1 > t_b$ or $t_2 < t_a$.

$$y'_{\text{int}}((\eta_1 + \eta_2)/2) = \begin{cases} \frac{R}{2}(e^{-t_1/R} + e^{-t_2/R} - 2e^{-t'_2/R}) & \text{if } \eta'_1 + \eta'_2 \geq \eta_1 + \eta_2. \\ \frac{R}{2}(2e^{-t'_1/R} - e^{-t_1/R} - e^{-t_2/R}) & \text{if } \eta'_1 + \eta'_2 < \eta_1 + \eta_2. \end{cases} \quad (\text{A10})$$

Note that when $\eta_1 + \eta_2 = \eta'_1 + \eta'_2$, the radius $y'_{\text{int}}((\eta_1 + \eta_2)/2) = y'_{\text{int}}$.

3. When $y_0 < R$

There are 3 cases:

- (1) when neither trajectory is enclosed by the overlap, the deficit angle is zero;
- (2) when overlapping causal diamond encloses one trajectory, one particle contributes to the deficit angle, either δ or δ' depending on which particle is enclosed;
- (3) when the overlap encloses both trajectories, both particles contribute, i.e., the deficit angle is $\delta + \delta'$.

Determine range for t_1, t_2 where $\eta_1 + \eta_2 = -R$ by symmetry about $\eta = -R$ as shown in Fig. 8, we find that overlap exists when

$$t_1 < t_b = -R \ln \frac{y_0}{R} \quad \text{and} \quad t_2 > t_a = -R \ln \left(2 - \frac{y_0}{R} \right). \quad (\text{A11})$$

Now consider conditions on t'_1, t'_2 to allow overlap.

4. When $y_{\text{int}} \leq y_0$

The conditions on t'_1, t'_2 to allow overlap is determined by solving $y_{\text{int}} + y'_{\text{int}} = y_0$ with $\eta_1 + \eta_2 = \eta'_1 + \eta'_2$ just like in case 2, and we get the same result, that overlap exists only if (t'_1, t'_2) satisfy Eq. (A8).

a. When $y'_{\text{int}} \leq y_0$

See Fig. 9, then results from case 2 hold, that the entropy is given by $2\pi R_{\text{overlap}}$.

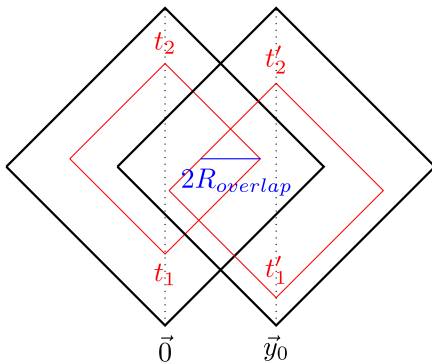


FIG. 9. When $y_0 < R$ and neither trajectory is enclosed in the overlap.

b. When $y'_{\text{int}} > y_0$

See Fig. 10, here $y'_{\text{int}} < R$ holds automatically because $t'_1 > -\infty$ and $t'_2 < \infty$, whether or not there is a deficit angle depends on the following:

- (a) If $\eta_1 + \eta_2 = \eta'_1 + \eta'_2$, the entropy of overlapping causal diamond is $(2\pi - \delta)R_{\text{overlap}}$ where δ is the deficit angle due to particle at $y = 0$.
- (b) If $\eta_1 + \eta_2 \neq \eta'_1 + \eta'_2$ and $y'_{\text{int}}((\eta_1 + \eta_2)/2) \leq y_0$, the entropy is $2\pi R_{\text{overlap}}$.
- (c) If $\eta_1 + \eta_2 \neq \eta'_1 + \eta'_2$ and $y'_{\text{int}}((\eta_1 + \eta_2)/2) > y_0$, the entropy is $(2\pi - \delta)R_{\text{overlap}}$ where δ is the deficit angle due to particle at $y = 0$.

5. When $y_{\text{int}} > y_0$

See Fig. 11 ($y_{\text{int}} < R$ holds because $t_1 > -\infty$ and $t_2 < \infty$), the conditions on t'_1, t'_2 are determined by solving

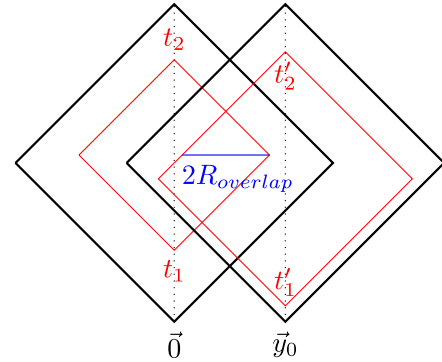


FIG. 10. Whether the maximal overlap encloses the trajectory is conditional.

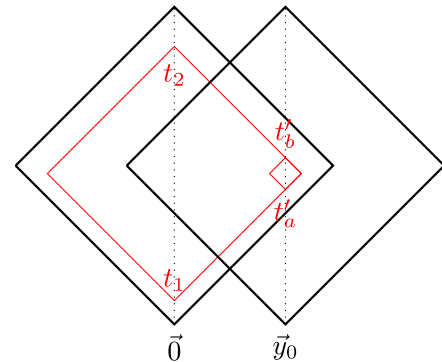


FIG. 11. When $y_{\text{int}} > y_0$, there is no overlap when $t'_1 > t'_b$ or $t'_2 < t'_a$.

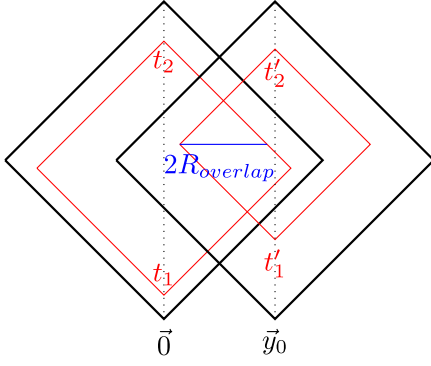


FIG. 12. Same results hold as in Fig. 10 by symmetry between $y = 0$ and $y = y_0$.

$y'_{\text{int}} = y_{\text{int}} - y_0$ with $\eta_1 + \eta_2 = \eta'_1 + \eta'_2$. We find the conditions for overlap to exist is

$$t_2' > t_a' = -R \ln \left(e^{-t_1/R} - \frac{y_0}{R} \right),$$

and $t_1' < t_b' = -R \ln \left(e^{-t_2/R} + \frac{y_0}{R} \right).$ (A12)

a. When $y'_{\text{int}} \leq y_0$

See Fig. 12. By symmetry of the geometry, the results, in this case, are exactly the same as in case when $y'_{\text{int}} > y_0$, except that the deficit angle is δ' instead of δ .

$$\{\delta, \delta'\}_{\min} = \begin{cases} \delta & \text{if } \min\{y'_{\text{int}}((\eta_1 + \eta_2)/2), y_{\text{int}}((\eta'_1 + \eta'_2)/2)\} = y_{\text{int}}((\eta'_1 + \eta'_2)/2), \\ \delta' & \text{if } \min\{y'_{\text{int}}((\eta_1 + \eta_2)/2), y_{\text{int}}((\eta'_1 + \eta'_2)/2)\} = y'_{\text{int}}((\eta_1 + \eta_2)/2). \end{cases} \quad (\text{A13})$$

(d) If $\eta_1 + \eta_2 \neq \eta'_1 + \eta'_2$ and $\min\{y'_{\text{int}}((\eta_1 + \eta_2)/2), y_{\text{int}}((\eta'_1 + \eta'_2)/2)\} > y_0$, the entropy is $(2\pi - \delta - \delta') \times R_{\text{overlap}}$.

To conclude, the radius of the overlapping causal diamond is always given by Eq. (A9) which is a function

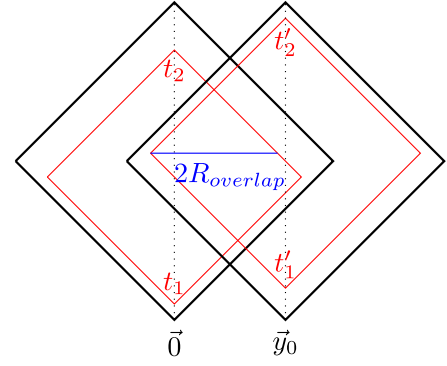


FIG. 13. The number of trajectories enclosed in the maximal overlap depends on conditions.

b. When $y'_{\text{int}} > y_0$

See Fig. 13. Whether there are one or two deficit angles depends on the situation. Here $y_{\text{int}}((\eta'_1 + \eta'_2)/2)$ is the radius of the (t_1, t_2) causal diamond evaluated at time $(\eta'_1 + \eta'_2)/2$, similar to the definition of $y'_{\text{int}}((\eta_1 + \eta_2)/2)$.

- (a) If $\eta_1 + \eta_2 = \eta'_1 + \eta'_2$, the entropy of overlapping causal diamond is $(2\pi - \delta - \delta')R_{\text{overlap}}$.
- (b) If $\eta_1 + \eta_2 \neq \eta'_1 + \eta'_2$ and $\max\{y'_{\text{int}}((\eta_1 + \eta_2)/2), y_{\text{int}}((\eta'_1 + \eta'_2)/2)\} \leq y_0$, the entropy is $2\pi R_{\text{overlap}}$.
- (c) If $\eta_1 + \eta_2 \neq \eta'_1 + \eta'_2$, $\max\{y'_{\text{int}}((\eta_1 + \eta_2)/2), y_{\text{int}}((\eta'_1 + \eta'_2)/2)\} > y_0$ and $\min\{y'_{\text{int}}((\eta_1 + \eta_2)/2), y_{\text{int}}((\eta'_1 + \eta'_2)/2)\} \leq y_0$, the entropy is $(2\pi - \{\delta, \delta'\}_{\min})R_{\text{overlap}}$ where the deficit angle given by

of $y_0, t_1, t_2, t_1', t_2'$. The deficit angle Δ takes value of $0, \delta, \delta', \delta + \delta'$, which value it takes depends on the relation among $y_0, t_1, t_2, t_1', t_2'$. The entropy is the area of the bifurcation surface of the overlapping diamond which is $(2\pi - \Delta)R_{\text{overlap}}$.

- [1] Tom Banks and W. Fischler, M theory observables for cosmological space-times, [arXiv:hep-th/0102077](#).
- [2] Stanley Deser, R. Jackiw, and Gerard 't Hooft, Three-dimensional Einstein gravity: Dynamics of flat space, *Ann. Phys. (N.Y.)* **152**, 220 (1984).
- [3] W. Fischler and L. Susskind, Holography and cosmology, [arXiv:hep-th/9806039](#).

- [4] Raphael Bousso, The holographic principle, *Rev. Mod. Phys.* **74**, 825 (2002).
- [5] Tom Banks, Cosmological breaking of supersymmetry?, *Int. J. Mod. Phys. A* **16**, 910 (2001).
- [6] Tom Banks, Bartomeu Fiol, and Alexander Morisse, Towards a quantum theory of de Sitter space, *J. High Energy Phys.* **12** (2006) 004.

- [7] Tom Banks and Willy Fischler, Holographic theory of accelerated observers, the S-matrix, and the emergence of effective field theory, [arXiv:1301.5924](#).
- [8] Patrick Draper and Szilard Farkas, de Sitter black holes as constrained states in the Euclidean path integral, *Phys. Rev. D* **105**, 126022 (2022).
- [9] Edward K. Morvan, Jan Pieter van der Schaar, and Manus R. Visser, On the Euclidean action of de Sitter black holes and constrained instantons, *SciPost Phys.* **14**, 022 (2023).
- [10] Stanley Deser and R. Jackiw, Three-dimensional cosmological gravity: Dynamics of constant curvature, *Ann. Phys. (N.Y.)* **153**, 405 (1984).
- [11] Hans-Jurgen Matuschull, The phase space structure of multi particle models in $2 + 1$ gravity, *Classical Quantum Gravity* **18**, 3497 (2001).
- [12] Leonard Susskind and James Lindesay, *An Introduction to Black Holes, Information and the String Theory Revolution* (World Scientific, Singapore, 2004), [10.1142/5689](#).
- [13] Patrick Hayden and John Preskill, Black holes as mirrors: Quantum information in random subsystems, *J. High Energy Phys.* **09** (2007) 120.
- [14] Yasuhiro Sekino and Leonard Susskind, Fast scramblers, *J. High Energy Phys.* **10** (2008) 065.
- [15] Steven Carlip, Black hole entropy from conformal field theory in any dimension, *Phys. Rev. Lett.* **82**, 2828 (1999).
- [16] Sergey N. Solodukhin, Conformal description of horizon's states, *Phys. Lett. B* **454**, 213 (1999).
- [17] Ted Jacobson, Thermodynamics of spacetime: The Einstein equation of state, *Phys. Rev. Lett.* **75**, 1260 (1995).
- [18] Tom Banks, Patrick Draper, and Szilard Farkas, Path integrals for causal diamonds and the covariant entropy principle, *Phys. Rev. D* **103**, 106022 (2021).
- [19] Thomas Banks and Kathryn M. Zurek, Conformal description of near-horizon vacuum states, *Phys. Rev. D* **104**, 126026 (2021).
- [20] A. Connes, *Noncommutative Geometry* (Academic Press, New York, 1994), ISBN: 9780080571751.
- [21] Evan Coleman, Edward A. Mazenc, Vasudev Shyam, Eva Silverstein, Ronak M. Soni, Gonzalo Torroba, and Sungyeon Yang, De Sitter microstates from $TT^- + \Lambda_2$ and the Hawking-Page transition, *J. High Energy Phys.* **07** (2022) 140.
- [22] Xi Dong, Eva Silverstein, and Gonzalo Torroba, De Sitter holography and entanglement entropy, *J. High Energy Phys.* **07** (2018) 050.
- [23] Roberto Emparan, Juan F. Pedraza, Andrew Svesko, Marija Tomašević, and Manus R. Visser, Black holes in dS_3 , *J. High Energy Phys.* **11** (2022) 073.
- [24] Horacio Casini, Marina Huerta, and Robert C. Myers, Towards a derivation of holographic entanglement entropy, *J. High Energy Phys.* **05** (2011) 036.
- [25] Ted Jacobson and Manus R. Visser, Entropy of causal diamond ensembles, *SciPost Phys.* **15**, 023 (2023).
- [26] Tom Banks and Patrick Draper, Comments on the entanglement spectrum of de Sitter space, *J. High Energy Phys.* **01** (2023) 135.
- [27] Tom Banks and Andrew Lucas, Emergent entropy production and hydrodynamics in quantum many-body systems, *Phys. Rev. E* **99**, 022105 (2019).
- [28] Curtis G. Callan, Jr., Steven B. Giddings, Jeffrey A. Harvey, and Andrew Strominger, Evanescent black holes, *Phys. Rev. D* **45**, R1005 (1992).
- [29] Jorge G. Russo, Leonard Susskind, and Larus Thorlacius, The endpoint of Hawking radiation, *Phys. Rev. D* **46**, 3444 (1992).
- [30] T. Banks, Holographic space-time models in $1 + 1$ dimensions, [arXiv:1506.05777](#).
- [31] T. Banks, Microscopic models of linear dilaton gravity and their semi-classical approximations, [arXiv:2005.09479](#).
- [32] Andrew G. Cohen, David B. Kaplan, and Ann E. Nelson, Effective field theory, black holes, and the cosmological constant, *Phys. Rev. Lett.* **82**, 4971 (1999).
- [33] Tom Banks and Patrick Draper, Remarks on the Cohen-Kaplan-Nelson bound, *Phys. Rev. D* **101**, 126010 (2020).
- [34] Nikita Blinov and Patrick Draper, Densities of states and the Cohen-Kaplan-Nelson bound, *Phys. Rev. D* **104**, 076024 (2021).
- [35] Tom Banks, Patrick Draper, and Szilard Farkas, Path integrals for causal diamonds and the covariant entropy principle, *Phys. Rev. D* **103**, 106022 (2021).